

Complexidade e organização do conhecimento: desafios de nosso século

Organizadores: Vera Dodebei e José Augusto Chaves Guimarães



ISKO-BRASIL



Organizadores

Vera Dodebei
José Augusto Chaves Guimarães

Complexidade e Organização do Conhecimento:
Desafios do nosso século

3

Rio de Janeiro, Brasil

Sociedade Brasileira de Organização do Conhecimento
(ISKO-Brasil)
Fundação para o Desenvolvimento do Ensino, Pesquisa e
Extensão (FUNDEPE)

2013

Organizadores

Vera Lúcia Doyle Louzada de Mattos Dodebei – *Universidade Federal do Estado do Rio de Janeiro*

José Augusto Chaves Guimarães – *Universidade Estadual Paulista*

Comitê Científico

Marisa Bräscher Basílio Medeiros – *Universidade Federal de Santa Catarina*

Maria Luiza de Almeida Campos – *Universidade Federal Fluminense*

Nanci Oddone – *Universidade Federal do Estado do Rio de Janeiro*

Editoração:

Editoração, capa e projeto gráfico: Maíra Fernandes Alencar - Universidade Estadual Paulista

© A reprodução deste livro na íntegra ou em parte é permitida, desde que citados os créditos. Proibida a venda.

C737

Complexidade e organização do conhecimento: desafios de nosso século.
[recurso eletrônico] Vera Dobedei, José Augusto Chaves Guimarães.
(organizadores). – Rio de Janeiro : ISKO-Brasil ; Marília : FUNDEPE,
2013.

310 p. il..

e-Book

ISBN: 978-85-98176-51-2

1. Organização do Conhecimento. 2 ISKO. I . Dobedei, Vera II. Guimarães, José Augusto Chaves. Título

CDD 025.4

SUMÁRIO

Prefácio.....13 - 15

A DIMENSÃO EPISTEMOLÓGICA DA ORGANIZAÇÃO E REPRESENTAÇÃO DO CONHECIMENTO

The epistemological dimension of knowledge organization.....17 - 25
Richard Smiraglia

Panorama de investigación en Organización del conocimiento en su dimensión epistemológica.....26 - 33
Rosa San Segundo Manuel

Aspectos teóricos, procedimentais, normativos e pragmáticos como categorias para uma epistemologia da organização da informação.....34 - 38
Cristina Dota Ortega

**Terminologia, linguística e filosofia: contribuições interdisciplinares em artigos sobre sistemas de organização do conhecimento publicados em revistas nacionais
Marisa Bräscher Basílio Medeiros.....39 - 44**

7

Temas e questões em análise documentária de charge.....45 - 51
Thulio Pereira Dias Gomes | Vânia Lisbôa da Silveira Guedes | Maria José V. da Costa Santos

**Tendências temáticas e opções metodológicas da produção científica de pesquisadores brasileiros na advances in Knowledge Organization: 2010 e 2012
.....52 - 59**
Leilah Santiago Bufrem | Murilo Artur Araújo da Silveira | Bruna S. do Nascimento

Visibilidade da produção científica sobre metateoria: análise bibliométrica no período entre 2000 a 201260 - 65
Bruno Henrique Alves | Ely Francina Tannuri de Oliveira

A produção científica na temática indexação: análise bibliométrica de teses e dissertações no período de 2002 a 201166 - 72
Paula Regina Dal'Evedove | Roberta C.D. Tartarottiro | Mariângela Spotti Lopes Fujita | Vera Regina Casari Boccato

Mapeamento conceitual sob o enfoque da Análise de Domínio: uma discussão de conceitos fundamentais.....73 - 78
Maria Luiza de Almeida Campos | Rosana P. Tavares de Moraes

Continuidades e rupturas em representação do conhecimento.....	79 - 85
<i>Vera Dodebei</i>	
Aproximación sistemática al concepto de “control de vocabulario”	86 - 92
<i>Mario Barité Juan Carlos Fernández-Molina</i>	
A autoria e a representação documental.....	93 - 96
<i>Naira Christofolletti Silveira Maria de Fátima G.M. Tálamo</i>	
Autopoiesis e formação discursiva na consolidação de linguagens de indexação para análise.....	97 - 102
<i>Maria Aparecida Moura.....</i>	
Representação de fotografias para publicidade em bancos de imagens: princípios para análise.....	103 - 108
<i>Joice C. Cardoso Ennes de Souza Rosali Fernandez de Souza</i>	

A DIMENSÃO APLICADA DA ORGANIZAÇÃO E REPRESENTAÇÃO DO CONHECIMENTO

Consideraciones para construir un tesouro sobre leyes político-electorales usando tecnología digital.....110 - 131

Catalina Naumis Peña

Ariel Antonio Morán Rey

Rosa Atzimba Morales Monroy

Ontologias e Definições: a explicitação do compromisso ontológico.....132 - 140

Maria Luiza de Almeida Campos

Modelagem conceitual e sistemas de organização de conhecimento; uma reflexão baseada em experiências acadêmicas e profissionais no Brasil no domínio da Biblioteconomia e Ciência da Informação.....141 - 146

Lídia Alvarenga

A importância teórica e prática da indexação na fundamentação científica da organização e representação do conhecimento.....147 - 159

Mariângela S. L. Fujita

UMLS como suporte ao processamento semântico de artigos em ciências biomédicas.....160 - 164

Carlos Henrique Marcondes

A norma ISO 25964 e a semântica latente das folksonomias.....165 - 170

Juliana de Assis / Maria Aparecida Moura

Folksonomias na Organização do Conhecimento: tags atribuídas a artigos sobre Medicina171 - 176

Luis Fernando Herbert Massoni / Regina Helena van der Laan / Rita do Carmo Ferreira Laipelt / Sônia Elisa Caregnato

Ontologias a partir de parâmetros da norma ANSI/NISO Z39.19.....177 - 182

Ana Cristina de Albuquerque / Brígida M. Nogueira Cervantes

Reflexões sobre processamento e representação automática de conhecimento

Viviane Neves dos Santos / Nair Yumiko Kobashi.....183 - 188

Análisis de Dominio como abordaje teórico en la construcción de ontologías: estudio de caso, desarrollo de ontologías para la IDE-Uruguay.....189 - 194

Diana Comesaña / Alicia Díaz

Uso de BISAC como sistema de organización del conocimiento en el catálogo.....195 - 199

Daniel Martínez-Ávila / Margaret E.I. Kipp / Hope A. Olson

O banco de dados como fonte de pesquisa e recuperação da informação – Projetos sobre a Região Platina/MERCOSUL.....200- 204
Gilberto F. Rodrigues Viana / Telma C. de Carvalho Madio

Elaboração de vocabulário controlado em formato SKOS usando Tematres: implicações metodológicas e web-semântica.....205 - 210
Cibele Araújo Camargo Marques dos Santos / Michely Jabala Mamede Vogel / Vânia Mara Lima / João Carlos Holland de Barcellos

O trabalho de descrição de acervo arquivístico no Brasil.....211 - 216
Eliezer Pires da Silva / Evelyn Goyannes Dill Orrico

A DIMENSÃO SOCIAL, CULTURAL E POLÍTICA DA ORGANIZAÇÃO E REPRESENTAÇÃO DO CONHECIMENTO

A presença brasileira no cenário de produção científica da ISKO.....218 - 227
José Augusto Chaves Guimarães

Facets, views and frames: enhancing connectedness in information organization.....228- 230
Kathryn La Barre

Imagens e filmes: o potencial informativo dos documentos visuais para o acesso em ambientes de informação.....231 - 233
Rosa Inês de Novaes Cordeiro

Comunicação da mesa : Panorama da pesquisa em OC em sua dimensão social, cultural e política (formação profissional, ética, cultura e identidade, contextos, sustentabilidade.....234 - 237
Maria José Lopez-Huertas

Problemas da organização do conhecimento na contemporaneidade.....238- 245
Marilda Ginez G. de Lara

Metáfora e ortofemismo na representação de assunto.....246 - 251
Fabio Assis Pinho | Suellen Oliveira Milani

Organização e representação do conhecimento na web: desafios para a construção colaborativa de uma ontologia do samba.....252 - 259
Marcos Miranda | Jair Miranda

Democracia digital e as suas implicações para a organização participativa do conhecimento.....260 - 265
Miriam Gontijo de Moraes

Documentação jurídica: reflexões sobre a função social do documento legislativo
Simone Torres | Maurício Barcellos de Almeida.....266 - 271

Breves reflexões sobre a ética na representação do conhecimento.....272 - 276
Dulce Maria Baptista

O caráter técnico, social e mediador da organização do conhecimento.....277 - 282
Aida Varela Varela | Marilene Lobo Abreu Barbosa

Modelagem, Organização e Representação do Conhecimento: proposta de elaboração da Biblioteca Virtual Temática em Vigilância Sanitária dos serviços de estética do município do Rio de Janeiro.....283 - 288
Jorge C. de Almeida Biolchini | Aline Vieira do Nascimento

Sistema de Organização do Conhecimento sobre Intensificação Agropecuária: agregando valor às redes de conhecimento da Embrapa.....289 - 295

Ivo Pierozzi Junior | Leandro H. M. de Oliveira | Maria de Cléofas F. Alencar | Gladis M. de Barcellos Almeida

A representação e a organização do conhecimento em uma instituição pública de pesquisa: a implementação do Repositório do Conhecimento do Instituto de Pesquisa Econômica Aplicada (RCIpea).....296 - 301

Fábio Ferreira Batista | Veruska da Silva Costa

A integração museu/biblioteca na organização e representação do conhecimento de documentos artísticos..... 302 - 307

Maria de Fátima G. M. Tálamo | Giovana Deliberali Maimone

Ethics and epistemology of classification.....308

Jens-Erik Mai | Royal School of Library and Information Science, Copenhagen, Denmark

Knowledge organization for its own sake: relationships between theory and applications.....309

Claudio Gnoli | Università di Pavia

Trabalhos premiados310

Introdução

Estudos Avançados em Organização e Representação do Conhecimento - v. 2 apresenta os textos completos da produção científica de pesquisadores em organização do conhecimento, brasileiros e estrangeiros, que participaram do II Congresso Brasileiro em Representação e Organização do Conhecimento, realizado no Rio de Janeiro durante os dias 28, 29 e 30 de maio de 2013. O evento, dando continuidade a uma trajetória de sedimentação e de interlocução científica na área, em âmbito brasileiro, iniciada em 2011, por ocasião do I Congresso de ISKO-Brasil, apontou para perspectivas bastante promissoras no que se refere a pesquisas conjuntas, diálogos interdisciplinares e visibilidade científica, sinalizando a necessidade de sua continuidade e verticalização em edições futuras.

Com o tema *Complexidade e Organização do Conhecimento: desafios de nosso século* registra-se o diálogo sobre a organização do conhecimento em um mundo pautado pela diversidade cultural, por um ideal ético de respeito às diferenças e por uma busca de liberdade de expressão e colaboração. Segundo Edgar Morin, o problema do conhecimento é um desafio porque só podemos conhecer as partes se conhecemos o todo em que se situam, e só podemos conhecer o todo se conhecemos as partes que o compõem. Esse desafio leva à reflexão sobre o caráter distinto do pensamento e, ao mesmo tempo, aquele que é conexo. Segundo o autor, *complexus* significa o que tece junto e, assim, o objetivo do pensamento é ao mesmo tempo unir (contextualizar e globalizar) e aceitar o desafio da incerteza. Isto demonstra o que Morin¹ denomina de pensamento complexo: “o pensamento complexo deve ligar a autonomia à dependência [...] Esta concepção de pensamento dá-nos uma lição de prudência, de método, de modéstia”.

Para os especialistas em organização do conhecimento visando sua localização e uso, o campo da representação vem se tornando um desafio à medida que as bases de registro se virtualizam e a informação é lançada na rede de computadores em múltiplos formatos e com múltiplas etiquetas. Se, por um lado, isto garante a diversidade cultural no uso dos recursos, por outro lado, o alto índice de dispersão temática desses recursos de informação vai obrigar os pesquisadores da organização do conhecimento a

¹ MORIN, Edgar. 2003. “Da necessidade de um pensamento complexo” In *Para navegar no século 21: tecnologias do imaginário e cibercultura*, Martins, F. M., Silva, Juremir M. org. Porto Alegre: Edipucrs/Sulina.

desenharem, em conjunto com profissionais de campos correlatos, sistemas e programas que auxiliem os processos de indexação e classificação compartilhados. A colaboração entre produtores/usuários do conhecimento no mundo contemporâneo parece ser a melhor estratégia para realizar tarefas altamente dispendiosas como aquelas típicas da área da OC nos exemplos da classificação bibliográfica, da indexação e da elaboração de linguagens documentais ou documentárias.

Nesse contexto, a realização do II ISKO-BRASIL propôs a discussão sobre tradição e inovação nos sistemas, processos, instrumentos e produtos, objetivando ampliar as trocas de experiências teóricas e práticas nos seus três eixos temáticos. O primeiro eixo ou subtema, *A dimensão epistemológica da organização do conhecimento*, coordenado por Marisa Bräscher Basílio Medeiros (UFSC) reuniu, entre conferências, mesas redondas e comunicações, 16 (dezesesseis) trabalhos a cargo de 28 (vinte e oito) autores. O segundo eixo, *A dimensão aplicada da organização do conhecimento*, coordenado por Maria Luiza de Almeida Campos (UFF) contou com 15 (quinze) trabalhos a cargo de 30 (trinta) autores. O terceiro eixo, coordenado por Nanci Oddone (UNIRIO), reuniu 15 (quinze) trabalhos e 25 (vinte e cinco) autores, totalizando um conjunto de 46 (quarenta e seis) trabalhos e 85 (oitenta e cinco) autores.

Contando com 111 (cento e onze) pesquisadores inscritos e 127 (cento e vinte sete) certificados emitidos, o evento pautou-se pela diversidade institucional de seus participantes. Nesse sentido, em âmbito internacional, registrou-se a participação de Richard Smiraglia (University of Wisconsin-Milwaukee - EUA), Jens-Eric Mai (Royal School of Library and Information Science - Dinamarca), Kathyne Anne La Barre, University of Illinois at Urbana-Champaign, Rosa San Segundo e Daniel Martínez Ávila (Universidad Carlos III de Madrid – Espanha), Maria José López Huertas e Juan Carlos Fernández Molina – Universidad de Granada – Espanha), Claudio Gnoli (Università de Pavia – Itália), Mario Barité (Universidad de la Republica - Uruguai), e Catalina Naumis Peña (Universidad Autonoma de Mexico).

Em relação à comunidade brasileira, observou-se uma significativa participação de docentes e pesquisadores provenientes de distintas universidades e instituições de pesquisa, a saber: Empresa Brasileira de Pesquisa Agropecuária, Instituto Brasileiro de Informação em Ciência e Tecnologia, Universidade de Brasília, Universidade de São Paulo, Universidade do Estado do Rio de Janeiro, Universidade Estadual Paulista, Universidade Estadual de Londrina, Universidade Federal da Bahia, Universidade Federal de Minas Gerais, Universidade Federal de Pernambuco, Universidade Federal

de Santa Catarina, Universidade Federal do Rio de Janeiro, Universidade Federal do Rio Grande do Sul e Universidade Federal Fluminense

À vista das atividades científicas do evento, foi possível observar um esforço dos grupos de pesquisa no sentido de incentivar a produção científica transdisciplinar no campo de estudos da organização e representação do conhecimento, seja no que tange às temáticas dos trabalhos seja com relação aos participantes, provenientes de áreas como ciência da computação, ciência da informação, ciências humanas e ciências sociais, artes entre outras.

Os trabalhos aqui apresentados dão testemunho efetivo do atual estágio de desenvolvimento científico da área de Organização e Representação do Conhecimento no Brasil enquanto um campo de estudos já consolidado nos domínios da Ciência da Informação e, em outros fronteiriços, nota-se uma forte aproximação com seus fundamentos teóricos-conceituais e metodológicos.

Vera Dodebei

José Augusto Chaves Guimarães

A dimensão epistemológica da ORC

The Epistemological Dimension of Knowledge Organization

Richard P. Smiraglia
University of Wisconsin, Milwaukee
smiragli@uwm.edu

What is epistemology? Epistemology is the study of that which is known. What is a dimension? A dimension is an expression of the extent of a space. What is knowledge organization? Knowledge organization is the science of the order of knowledge. The domain of knowledge organization is a discourse community in which rigorous, self-conscious inquiry takes place concerning that which is known, and its various orderings or sequences, both those that are natural or heuristic, and those that are imposed. The products of the domain, then, are ordered segments of that which is known, and the rules either for discovering their natural orders, or the rules for imposing a useful sequence. All applied knowledge organization then is a form of discourse, in which the structures and rules are objects of communication, and which takes place in a cultural milieu or among actors in various cultural milieus. Thus a very important component of the science of knowledge organization must be epistemology, which is the science of knowledge itself. So my task is to answer these three questions, and to describe how it is that epistemology *is* one dimension of the domain of knowledge organization.

17

Epistemology is the division of philosophy that investigates the nature and origin of knowledge. Epistemology is therefore the science of knowledge, and the self-conscious inquiry of knowledge. If knowledge organization is the science of the order of knowledge, then we—its scientists—must be concerned first and foremost with the question of what, exactly, knowledge is. Epistemology, through rigorous inquiry, tells us where knowledge comes from. The central problems for epistemology are the definition of knowledge, and the means of its acquisition. Historically there are just two main schools of epistemological thought: rationalism and empiricism, which arise from mathematics and logic, and the natural sciences, respectively. In other words, knowledge can encompass that which we can logically discern, and knowledge can encompass that which we can observe.

Although philosophers have identified many approaches to epistemology, in knowledge organization we have come to rely on a framework set forth by Birger

Hjørland. (Together with Jeppe Nicolaisen, Hjørland has constructed a web tool called a “lifeboat”— [http://www.iva.dk/jni/lifeboat/.](http://www.iva.dk/jni/lifeboat/)) Hjørland lists four basic epistemological stances (or positions) (Hjørland 1998, 608):

- Empiricism: derived from observation, perception, and experience;
- Rationalism: derived from the employment of reason over sensory experience;
- Historicism: derived from cultural hermeneutics; and,
- Pragmatism: derived from the consideration of goals and their consequences.

That which we know from our own experience of it, and in particular that which is known through the positivist sciences, is what we call empirical. We have solid evidence for the empirical, and we can point to the evidence as a means of prediction. That which we know from reasoning about it, and in particular that which is known through humanistic scholarship, is based in rationalism. There is no evidence, per se, for the rational; rather there are explanatory statements that seem to be logical when taken together. Historicist epistemology interprets evidence through a cultural lens, relying in particular on past experience. Pragmatism is knowledge derived from assumptions about the best means to an end. Pragmatic solutions work in the moment but do not necessarily rely on empirical evidence, and therefore do not necessarily pass the test of time. Rational solutions also often ignore empirical evidence and thus frequently yield unworkable schemas.

Epistemology is an essential tool of knowledge organization and a dimension is an expression of the extent of a space. Therefore, in knowledge organization, epistemology represents one dimension, because it is how we can measure or express the space within our domain ranging from the empirical to the rational, the two primary epistemological stances. The diagram in Figure 1 shows the three dimensions of a sphere. Each dimension travels along one of those arrows that bisect the space. They define height, width, and depth (in colloquial language) or colatitude, φ , longitude, θ , and radius, ρ , (in mathematical terms).

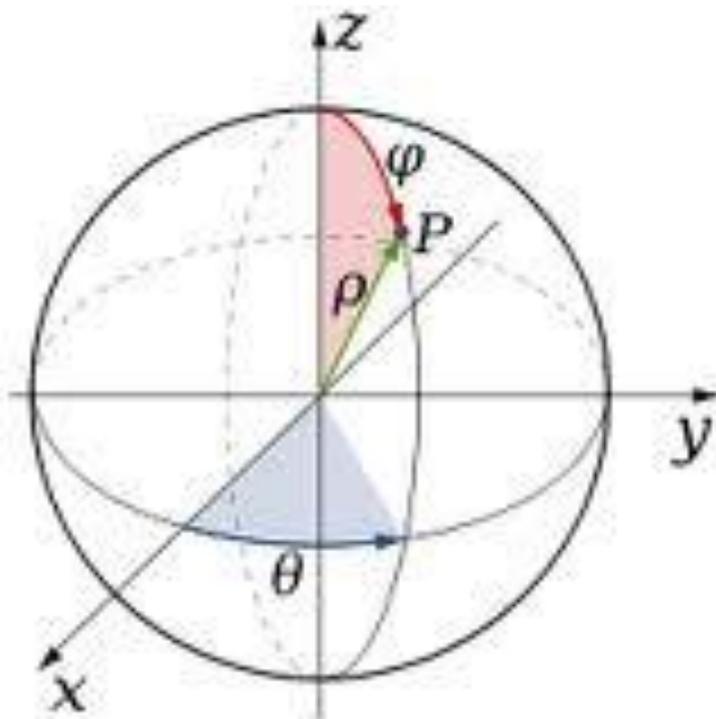


Figure 1. Three dimensions of a sphere (Wikipedia, “Dimension (Mathematics and Physics).”)

If we consider the intellectual space of the domain of knowledge organization to be something like this sphere, then we can see that each dimension both bisects the space and provides a trajectory to transit through or across it. In other words, the dimensions of our domain define what we call its intension (its depth) and its extension (its external parameters). An hypothesis is that in knowledge organization there always is tension along the epistemological dimension stretching from the empirical to the rational. We will see this demonstrated in the modes of thought used by knowledge organization researchers.

But first, a few fundamentals about knowledge organization as a science. In one way of thinking, there is a fundamental theoretical dimension in our domain represented by our chief research questions. Our main concern is with the identification of concepts that represent the totality of what is known. A second critical concern is with the order of these concepts. Therefore our main research questions always must be “what is knowledge?” and “what is its order?”

Meta-analysis of KO domain analyses demonstrates the evolution of the domain from its original goal of creating a universally-applicable knowledge organization system, to its current post-modern embrace of multiple goals. Among this multiplicity of goals we find first and foremost an embrace of domain-specific ontologies that can be combined as necessary using faceted techniques, leading to the perception of multiple,

but representable, universes of knowledge. In KO there always is work to solidify concept theory, which is at the core of the discipline. KO as a domain demonstrates coherence across time and across geo-political boundaries, particularly as it concerns its theoretical foundations. There are consistently marked dimensions within the domain— theoretical versus applied on one continuum, humanistic versus scientific on another, which is another way of stating the epistemological dimension. All of these dimensions serve to maintain constructive and dynamic tension within the domain, which in turn keeps the research front constantly in a state of renewal, constantly stretching the intension and the extension (the boundaries) of the domain.

Figure 2 shows internal author-cocitation from the Rome 2010 International ISKO Conference.

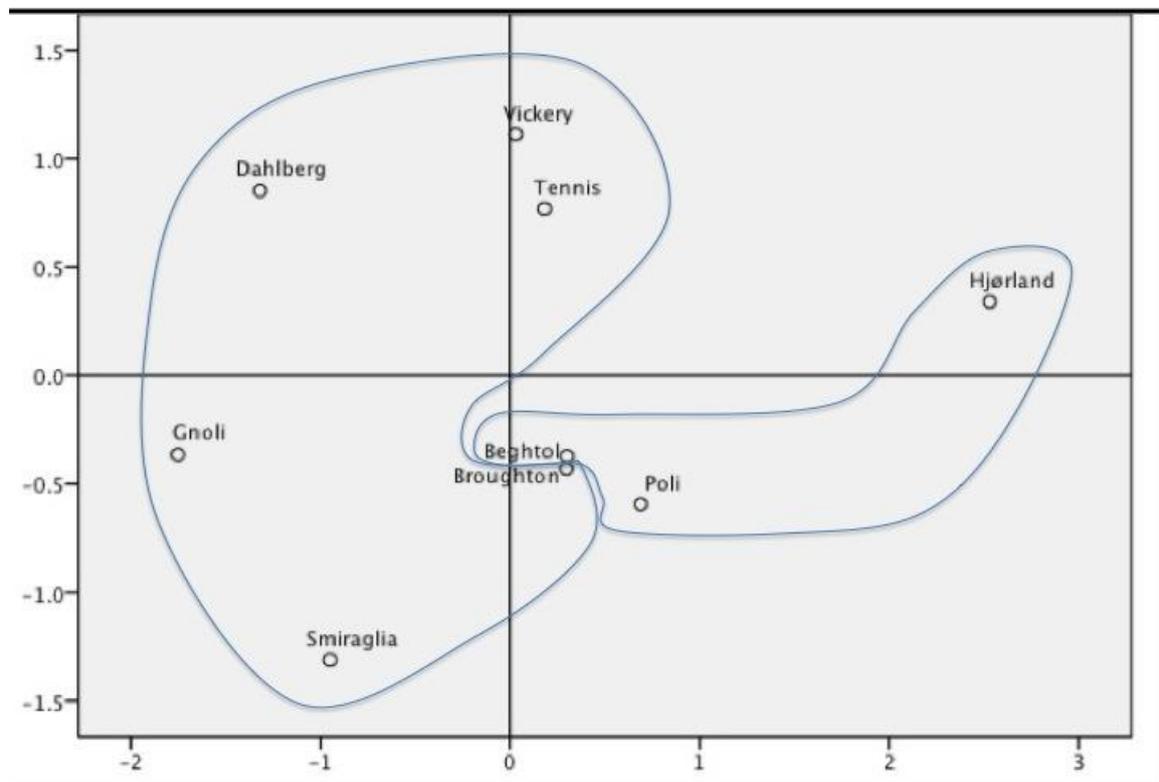


Figure 2. Internal Author Cocitation from ISKO 2010 Rome.

Like the illustration of the sphere, this map shows us some of the dimensions of knowledge organization as a domain. As it happens, although it is a map of only one conference, it is also a pretty good representation of the whole domain. The parameters, or dimensions, are given by domain analysis, which is the representation of a discourse community as can be determined from the formal products, or published research emanating from the community. From domain analytical studies emerge means for

visualizing interaction within the scholarly community, as well as points of shifting emphasis as theoretical paradigms are tested and hypotheses are generated.

According to Tennis (2003) the extension of a domain is the breadth of its topical universe, and the intension is the depth of the paradigms that comprise it. Bibliometric tools and meta-analysis can be used to visualize the extension and intension of KO as a domain. Analysis of the domain of knowledge organization is relatively easily generated from the formal publications in biennial ISKO conference proceedings, articles in our journal *Knowledge Organization*, and proceedings of biennial regional chapters, whose meeting years alternate with the ISKO international conferences.

Knowledge Organization is devoted to the conceptual order of knowledge. In the broadest sense KO is the arena in which the heuristics of ordering knowledge are studied. Specifically, KO is the research community devoted to classification and ontology, thesauri and controlled vocabulary, epistemology and warrant, as well as the development of applied systems for all of the preceding (often, especially in North America, resource description is also considered to be a part of KO). There is a long tradition of the activities and tools of KO; classification, taxonomy, and typology, for example, have always been key to the development of scholarship. Specific applications for information storage and retrieval, such as indexes, bibliographic classifications, etc., have been part of the practice of librarianship and the research agenda of information science since the late 19th century. The formal KO domain, represented by the International Society for Knowledge Organization (ISKO) and its chapters, councils and publications, dates from 1989 (Dahlberg, 2010).

The classical geopolitics of KO as a domain traditionally has consisted of anchors in six countries. Dr. Dahlberg, founder of the modern form of the domain is from Germany. Much of the scholarship in the first decades of the domain came from North America and the UK, especially with regard to the classification research group. Essential theoretical grounding found its base in the Royal School in Copenhagen, and the beginnings of epistemological and ontological thought in the domain came from Italy. New-ISKO chapters holding regional conferences or universities generating KO research are shifting the emphasis away from Northern hemispheric dominance to a more global perspective. To the classical mix, we now have added Slovenia, Morocco, Tunisia, Belgium, India, Brazil, France, and Spain. This is based on the formal publications of regional chapters. There also is a recent influx of papers from Singapore,

Taiwan, and China, the countries of affiliation of authors at the 2012 Mysore ISKO International Conference. The geopolitics of knowledge organization are shifting. Brazil now accounts for almost a third of the research at the most recent international conference. Is there a geopolitical dimension to KO? If all KO authors are working on the same set of problems are they doing so from different cultural perspectives.

A meta-analysis of knowledge organization based on 17 studies of KO literature incorporated 3494 source papers, of which 1100 appeared in journals such as *CCQ*, *LRTS*, or *LQ*, 444 appeared in KO, but 600 appeared in ISKO conference proceedings, and 1350 were papers in ISKO regional conference proceedings. The importance of conference proceedings as a venue for communication in the domain is clear. 56% of the research over time has been presented in conference papers. Only 13% of the total over time has appeared in the domain's formal journal.

The extension of the domain is consistently represented as including theoretical foundations, such as classification and ontology, and epistemology, which lies at the heart of both. The intension is represented by development and testing of applications. The international nature of the domain means that there are occasional overlapping emphases on multicultural and multilingual issues. Geographical diversity does not preclude domain coherence. In the co-citation map in Figure 2 there were two poles on the extension dimension: semantics and KO systems, or if you will, concepts and order. All of the rest is intension.

Figure 3 shows author-cocitation maps from ISKO Montréal 2008 and IKSO Mysore 2012 respectively.

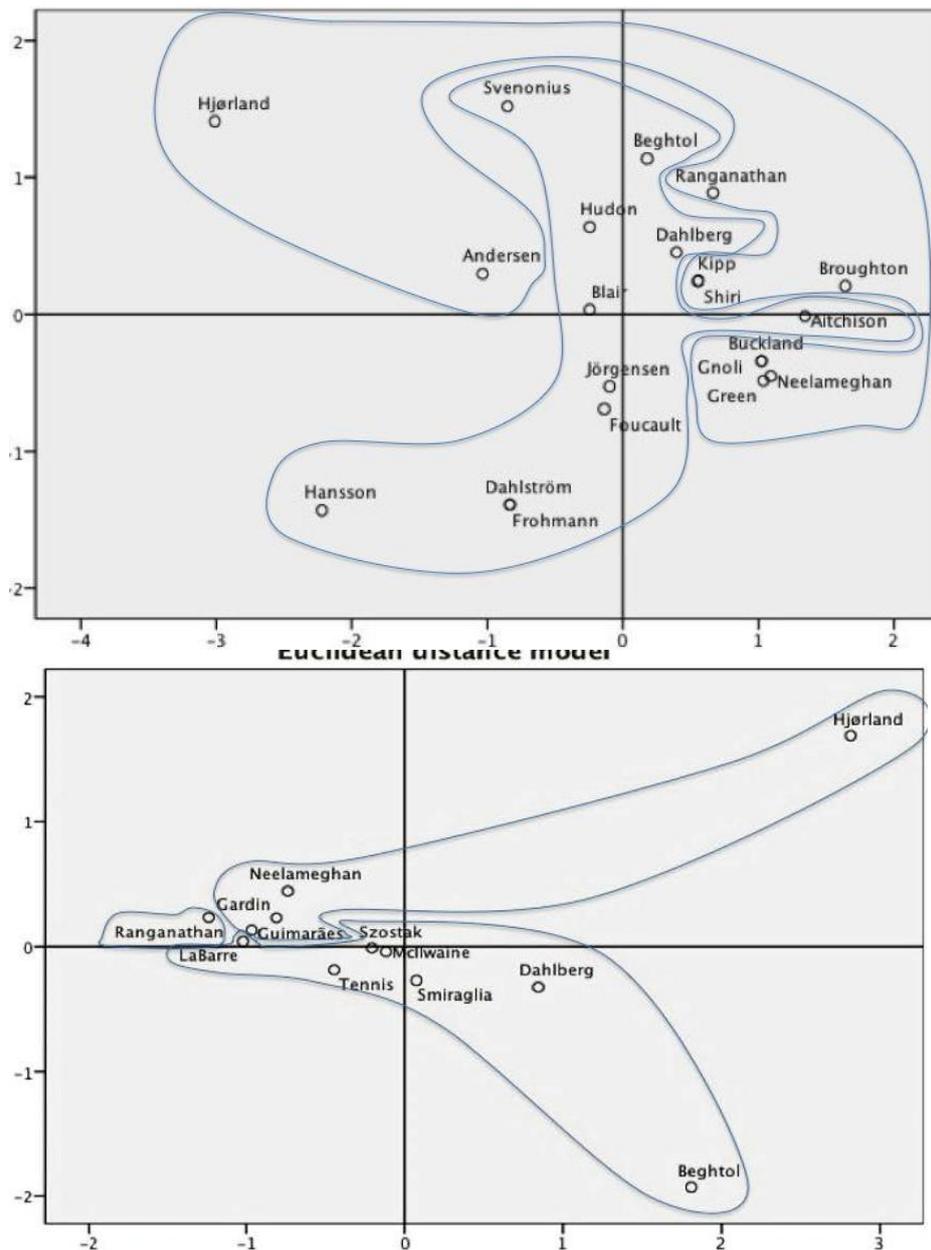


Figure 3. Author co-citation maps from ISKO Montréal 2008 and IKSO Mysore 2012.

ISKO 2008 in Montréal shows that “information retrieval” and “models and methods” represent the KO systems portion (interestingly enough associated this time with Hjørland and Andersen), or 36% of the conference, while “epistemology” and “multilingual multicultural” and “discourse communities” for good measure represent the semantics portion (Ranganathan, Broughton, Neelamegha), or 31% of the conference. The remaining roughly 33% is intension, largely applied research (led by Kipp and Shiri). ISKO 2012 in Mysore has “Ontology” “Categories” “Relationships” “Domain of KO” and “Domain specificity” or 47% of the conference as roughly constituting conceptology (the cluster with La Barre, Tennis, Dahlberg and Beghtol). “General classifications” and “Design and Development” give us 16% clearly in the KO

systems category (with Guimaraes, and Hjørland). That leaves 38% as intension, applied research, stretching the boundaries. It also shows how there is tension along the “concept – KO systems” dimension, as the weight of each shifts back and forth from conference to conference. We note the centrality of Ranganathan. We have to ask ourselves whether some of this is regional. We also can ask whether any of it is geopolitical.

Here is an epistemological map of the Mysore conference, in which we see themes arrayed by methodology. The dark blue is informetric, or domain analytical, the green is quantitative, which we see in ontology navigation relationships and information mining, the qualitative is found in user studies. This is the empirical part of the domain, or as I will show in a moment about half. The rest is rational, covering most of the KO systems and applied research.

Knowledge organization clearly is a coherent domain. It has an ontological base, its extension, which is a dimension stretching from concept theory (or semantics) to KO systems. It has an underlying teleology, which is the attempt to grasp the order of knowledge. It has, therefore, a set of common hypotheses, and epistemological consensus, although as we have seen the epistemological consensus forms another dimension bisecting the domain, stretching from empiricism to rationalism. And it has social semantics, which is a way of saying we have formal venues for our research, and the knowledge published there is clearly cumulative. There also is a geopolitical form of social semantics, which means we work globally and we bring our cultural characteristics with us.

Finally, we see that knowledge organization as a domain demonstrates coherence across time and across geopolitical boundaries, particularly as it concerns its theoretical foundations. Knowledge organization has robust and continuous formal publication venues that help to maintain domain coherence. Theoretical poles that anchor bisecting dimensions are both conceptual and methodological. The domain is scientific, but also has deep roots in humanistic rational methods and modes of thought. Differences that emerge in intension reflect shifting cultural approaches across regions and across time. There has been a shift over time in intension as the domain moved from emphasis on universal classifications to interoperability. But there are consistently marked dimensions within the domain:

- theoretical versus applied on a conceptual continuum,
- empirical versus rational on an epistemological continuum.

These dimensions serve to maintain constructive and dynamic tension within the domain, which in turn keeps the research front constantly in a state of renewal. Faced with different universes of knowledge, KO as a domain approaches the problems of analysis and concept designation from within the dynamic tension demonstrated here. Collins (1998) suggests no school of thought can comprise more than six points of view without either concretizing or splitting. KO itself has concretized, maintaining its status as a domain but using this tension constructively to evolve.

References

Smiraglia, Richard P. and Hur-Li Lee, eds. 2012. *Cultural frames of knowledge*. Würzburg: Ergon Verlag.

Smiraglia, Richard P. 2012. Universes, dimensions, domains, intensions and extensions: knowledge organization for the 21st Century. In A. Neelameghan and K.S. Raghavan eds. *Categories, contexts, and relations in knowledge organization: Proceedings of the Twelfth International ISKO Conference, 6-9 August 2012, Mysore, India*. Advances in knowledge organization 13. Würzburg: Ergon Verlag, pp. 1-7.

Hjørland, Birger and Jeppe Nicolaisen, EDS. 2005. Epistemological lifeboat.
<http://www.iva.dk/jni/lifeboat/>

Hjørland, Birger. 1998. Theory and metatheory of information science: a new interpretation. *Journal of documentation* 54: 606-21.

Dimensions (mathematics and physics). 2012. *Wikipedia*
[http://en.wikipedia.org/wiki/Dimension_\(mathematics_and_physics\)](http://en.wikipedia.org/wiki/Dimension_(mathematics_and_physics)).

Tennis, Joseph T. 2003. Two axes of domains for domain analysis. *Knowledge organization* 30: 191-95.

Smiraglia, Richard P. 2012. Epistemology of domain analysis. In Smiraglia, Richard P. and Hur-Li Lee, eds. *Cultural frames of knowledge*. Würzburg: Ergon-Verlag, p. 114.

Panorama de investigación en Organización del conocimiento en su dimensión epistemológica

Rosa San Segundo
Universidad Carlos III de Madrid
rsan@bib.uc3m.es

1. Paradigmas conceptuales en epistemología en OC

En Organización del conocimiento confluyen disciplinas con metodologías y categorías heterogéneas, intercambiando metodologías entre estas subdisciplinas. Y en este contexto científico la investigación está supeditada a la resolución colectiva o transdisciplinar de problemas, más que a la formulación de objetos y problemas aislados.

En Organización del conocimiento ha sido poco abordado su método desde sus fundamentos teóricos. Y así se han consolidado, en el último siglo, dos paradigmas fundamentales, primero el Paradigma positivista, y con posterioridad el Paradigma pragmatista. El paradigma positivista, y su correlación en nuestro contexto, el paradigma bibliográfico, ya se ha abandonado. El modelo teórico de recuperación de información basado en la mera equiparación invariable del paradigma bibliográfico está cayendo en desuso², ya no es prioritaria la fórmula entre lo buscado y lo recuperado con equiparación total, que generó los lenguajes documentales y las clasificaciones universales a finales del siglo XIX e inicios del XX. Está siendo sustituido por un proceso de búsqueda de información interactivo, evolutivo sin total equiparación denominado paradigma de dialogo persona-ordenador, basado en el Pragmatismo, en este paradigma pragmatista, subjetivo lo prioritario es que interactúa la persona y el ordenador.

Este cambio hacia la interacción, ha modificado las herramientas de organización y representación y se han realizado casi con la ausencia de metodologías de análisis a priori. Además, en la actualidad hay predominio de otro paradigma que genera las interfaces entre el usuario y el ordenador, que son las tecnologías de la información, estas han transformado todo el quehacer científico. De forma tal, que la nueva Organización del conocimiento se sitúa en el pragmatismo con la interacción persona ordenador y en entorno digital. La configuración epistemológica se ha abordar en el

² SAN SEGUNDO, Rosa. *A New Conception of Representation of Knowledge*, 2004

tratamiento de un volumen multiplicado de información, en un mundo que se ha vuelto digital.

2. Paradigma tecnológico de la información

Se ha generado la emergencia de un nuevo ‘paradigma tecnológico’ producido por las tecnologías de información y comunicación. Lo que caracteriza a la revolución tecnológica actual no es solo el carácter central del conocimiento y la información, sino la aplicación de ese conocimiento. La tecnología de la información es para esta revolución lo que las nuevas materias primas fueron para las sucesivas Revoluciones Industriales. La acción del conocimiento sobre el conocimiento, como la fuente fundamental de productividad. Este paradigma tecnológico posee además un efecto de retroalimentación de flujos de información que transforman y constituye las relaciones sociales y que, en última, estas generan y construyen a las mismas.

Las tecnologías de la información se han convertido, desde finales del siglo pasado, en la fuerza motriz de la sociedad de la información y conocimiento. Se trata, pues, de un nuevo paradigma tecnológico, de un modo de pensar, hacer y vivir en torno a las tecnologías de la información. Este paradigma de la información digital se basa la recopilación de información para luego proceder a la elaboración de un producto, se produce en serie pero se involucra en la producción a las innovaciones individuales. De igual forma que el positivismo generó las clasificaciones documentales, el actual pragmatismo del contexto digital genera folksonomías, filtros burbujas y herramientas de recuperación de contenidos basadas en las búsquedas anteriores de los usuarios.

3. Choque de paradigmas cognitivos

Este proceso es interactivo, a modo de un ciclo de retroalimentación mutua, en el que el aprendizaje y la creación se entrelazan de manera prácticamente indistinguible. Se aprende creando y se crea aprendiendo, se modifican las tecnologías para servir a las necesidades del usuario a la vez que el usuario se adapta a las posibilidades casi infinitas de tecnologías y programas. Esto se observa en la transformación de la web de un emisor de información a un lugar de creación dinámica, donde la “sabiduría de las masas” da forma a los contenidos.

La interacción persona ordenador, las aportaciones de información individual son constitutivas del nuevo entramado tecnológico informacional, articulado, en gran medida, por los nativos digitales, nuevos usuarios con nuevas relaciones, con nuevos entretenimientos, nuevos aprendizajes y nuevos juegos. Estos han construido sus conceptos de espacio, tiempos, identidad y memoria en un entorno de tecnología, tienen habilidades innatas con relación al lenguaje y todo ello revierte en que tienen ya nuevas habilidades cognitivas y una nueva estructura neuronal³.

E incluso, es posible que con las nuevas modalidades de conocimiento se lleguen a activar nuevos módulos o funciones de la mente y alterar nuestros procesos neuronales *el uso sostenido y reiterativo en el tiempo de las herramientas que brindan las Tecnologías de la Información y la Comunicación (TIC), tan celebradas en la denominada sociedad del conocimiento, pueden llegar a modificarnos físicamente. Y esto es debido a la especial naturaleza de nuestro cerebro. Las TIC influyen en nuestra vida diaria porque conforme más tiempo pasamos trabajando con herramientas derivadas del uso de Internet, a la vez que procesamos mayor cantidad de información y somos, teóricamente, más eficientes, nuestra capacidad de atención y concentración en actividades que no tienen relación con las nuevas tecnologías, así como nuestra capacidad para el planteamiento y la resolución de problemas, disminuyen*⁴. Leer modificó el cerebro, y las nuevas estructuras de lectura digital también lo están modificando, estamos en un nuevo periodo o era, el acceso a la información digital comporta nuevas formas cognitivas, miramos el correo, navegamos, aparecen mensajes parpadeantes que reclaman la atención, o sea internet distrae, ya que hay múltiples asuntos, pero se pierde capacidad de concentración o profundización. Algunas investigaciones hacen pensar que la inteligencia secuencial que nos proporcionó la linealidad de la escritura, es más evolucionada, ya que necesita ser educada, entrenada a la visión alfabética y a su soporte típico: el texto. Se está produciendo una renuncia a la conquista evolutiva de la visión alfabética y la inteligencia secuencial, para volver a unos medios más naturales, más primitivos, más complacientes.

Se ha iniciado con internet nuevas formas cognitivas, *con la aceleración del cambio, con la mutaciones cognitivas, con el zapping que gobierna hoy nuestra capacidad de atención y con las herramientas de aumento de la inteligencia que*

³ GARCÍA, Felipe; PORTILLO, Javier; ROMO, Jesús ; BENITO, Manuel. *Nativos digitales y modelos de aprendizaje*, 2007

⁴ CARR, Nicholas. *Superficiales. ¿Qué está haciendo internet con nuestras mentes?* 2012

*existen, lo que necesitamos son competencias complementarias, y potenciadores, lo que debe primar en nuestras epistemologías*⁵. Las consecuencias que esta mediamorfosis implica para el aprendizaje ha generado un choque de paradigmas cognitivos, basados en la la neuroplasticidad de nuestro cerebro, que se están constituyendo en torno a los neoalfabetismos que integrarán el conocimiento y la acción en el siglo XXI.

También se están generando nuevas formas de representación del mundo con primacía de lo virtual sobre lo real. Los sistemas de visualización virtual, que proporcionan la ilusión de una inmersión en la imagen, es un campo en rápida evolución, ello está modificando toda nuestra realidad, *la influencia de lo virtual en nuestra civilización de información de flujos irá creciendo y acabará, sin duda alguna, alterando para siempre nuestra visión del mundo*⁶, lo que genera problemas filosóficos y éticos nuevos provocados por el progreso de la imagería virtual, con consecuencias en nuestra nueva manera de representar e interpretar el mundo.

Los mundos generados por ordenador modifican nuestras vidas, en tanto que las tecnologías de la información, están dispuesta a abolir la frontera que separa lo real de lo irreal. La realidad virtual un campo de investigación tecnológica prioritaria. O sea *la realidad virtual es sin duda el campo de investigación mas importante del momento, el más prometedor y el mas inquietante, el mysterium tremendum , que el hombre tendrá que afrontar en su próximo salto iniciático*⁷.

4. Interacción colectiva con la tecnología

La especie humana siempre ha tenido gran fascinación por el cambio tecnológico. Se está constituyendo una inteligencia digital colectiva, nos encontramos ante multitudes inteligentes, con acceso a ingentes cantidades de información, se especula sobre las consecuencias de esta impresionante conexión mundial y sobre lo que podría ocurrir si se convirtiera en una red global de "inteligencias individuales en conexión". Podría dar lugar esta conexión a una inteligencia muy superior a la de cada uno de las

⁵ PISCITELLI. *Internet: la imprenta del siglo XXI*, 2005 PISCITELLI. *Internet: la imprenta del siglo XXI*. Barcelona, Gedisa, 2005

⁶ QUEAU, Philippe. *Lo virtual. Virtudes y vértigo*, 1995

⁷ RHEINGOLD. *Realidad virtual*, 1994

personas. La tecnología ha determinado nuestro proceso evolutivo, en la eclosión tecnológica actual se contempla la posibilidad de que en el presente estemos experimentando uno de los mayores saltos en la evolución de nuestra especie. La conexión de las inteligencias, bien podría ser el paso próximo en la evolución de la inteligencia humana.

Pero a la inversa también se produce que los buscadores, por ejemplo Google, no nos proporcionan información estandarizada de una misma búsqueda igual para todos los usuarios, sino que las proporciona según algoritmos de contenido personalizados y por tanto da resultados personalizados. Esta no es la web semántica que necesitamos⁸ los buscadores de la red se esfuerzan por adaptar sus servicios (incluyendo noticias y resultados de búsqueda) a las inquietudes y preferencias personales, surge así una consecuencia no deseada y peligrosa: quedar atrapados en una burbuja de filtros que obstaculiza el acceso a la información, que nos impide desafiar o ampliar nuestra visión del mundo, en última instancia, es nocivo para nosotros y para el ejercicio de ciudadanía. De esta forma se puede crear la memoria digital como significante despótico, la actual memoria se enfrentará a una memoria despótica y documental, *la cultura está dejando de comportarse como, principalmente, una memoria de archivo, para hacerlo en cambio como una memoria de procesamiento de interconexión de datos y de sujetos de conocimiento. Acaso habría que decir aquí que el uso de la metáfora, de los modos de la memoria propia del ordenador, es probablemente algo más que una mera metáfora... La irrupción histórica de los modos de la gestión de lo cultural por la eficacia de las nuevas herramientas tecnológicas, es seguramente mucho más que una mera metáfora*⁹.

5. Patrón exponencial de progreso tecnológico

Estamos también ante un progreso tecnológico de gran envergadura. Los avances tecnológicos en el mundo de los ordenadores darán lugar a máquinas cada vez más potentes, numerosas y baratas, ello puede generar un patrón exponencial de crecimiento

⁸ PARISER, Eli. *Cuidado con los Filtros Burbuja en línea*

http://www.ted.com/talks/lang/es/eli_paiser_beware_online_filter_bubbles.html

⁹ BREA, José Luis. *Cultura RAM. Mutaciones de la cultura en la era de su distribución electrónica*, 2007

tecnológico, este patrón es probable que culmine en un progreso tecnológico inimaginable en el Siglo XXI, y puede conducir a lo que se ha denominado a la Singularidad tecnológica. Este es un evento futuro en el que se predice que el progreso tecnológico se acelerará, según Ray Kurzweil, nos dice *Así es como gradualmente he tomado consciencia de un impactante acontecimiento que se cierne sobre la mitad del siglo XXI. Tal y como un agujero negro en el espacio altera la forma dramática los patrones de materia y de energía acelerándolos hacia su horizonte de sucesos...Entonces ¿qué es la Singularidad? Es un tiempo venidero en el que el ritmo del cambio tecnológico será tan rápido y su repercusión tan profunda que la vida humana se verá transformada de forma irreversible. Aunque ni utópica ni distópica, esta era transformará los conceptos de los que dependemos a la hora de dar significado a nuestras vidas*¹⁰.

En este crecimiento exponencial puede que se origine una inteligencia superior a la humana, y se entraría en una etapa post humana. Kurzweil previó que la primera inteligencia artificial será construida por una simulación por ordenador de un cerebro humano. Este momento futuro denominado singularidad tecnológica, se ha nombrado así por analogía con la singularidad gravitacional, observada en los agujeros negros, donde existe un punto en el que las reglas de la física dejan de ser válidas, y donde la convergencia hacia valores infinitos hace imposible el definir una función. Así el desarrollo de inteligencia superhumana, puede cambiar nuestro entorno de manera tal, que un ser humano anterior a la Singularidad es incapaz de predecir

6. El modelo totalizante de interacción con la tecnología determina las epistemologías

En la actualidad, se ha llegado a un modelo totalizante tecnológico de interacción que determina e inunda el sustrato de las epistemologías. Tenemos que nuestro cerebro se adapta tanto a lo digital que está generando un choque de paradigmas cognitivos. Además la interacción entre usuario y tecnología conforma contenidos de inteligencias en conexión, y también genera su efecto inverso donde la interacción entre usuario y tecnología conforma contenidos producto del aislamiento personalizado. La interacción con la tecnología en sus múltiples facetas va a ser el objetivo a estudiar. El

¹⁰ KURZWEIL, Ray. *La singularidad está cerca*, 2012

patrón exponencial de crecimiento tecnológico es tan grande que nos sumerge en un modelo tecnológico totalizante, por lo que el advenimiento de Internet es el nuevo horizonte de discusión epistemológica.

REFERENCIAS BIBLIOGRÁFICAS

BREA, José Luis. *Cultura RAM. Mutaciones de la cultura en la era de su distribución electrónica*. Barcelona, Gedisa, 2007.

CAPURRO, Rafael. *La hermenéutica frente al desafío de la cultura digital*. Liinc (Laboratório Interdisciplinar em Informação e Conhecimento) Vol. 6, No. 2 (2010), 235-249. http://www.capurro.de/hermeneutica_porto.html

CARR, Nicholas. *Superficiales. ¿Qué está haciendo internet con nuestras mentes?* Taurus, 2012

COBO ROMANÍ, Cristóbal. PARDO KUKLINSKY, Hugo. *Planeta 2.0. Inteligencia colectiva o medios fastfood*. Grup de Recerca d'Interaccions Digitals, Universitat de Vic. Flasco México. Barcelona / México DF, 2007. E-book de acceso gratuito. Versión 0.1 / Septiembre de 2007. <http://www.planetawebdospuntocero.net>

GARCÍA CANCLINI, Néstor. *Lectores, espectadores e internautas*. Barcelona, Gedisa Editorial, 2007.

GARCÍA, Felipe; PORTILLO, Javier; ROMO, Jesús ; BENITO, Manuel. *Nativos digitales y modelos de aprendizaje*. IV Simposio Pluridisciplinar sobre Diseño, Evaluación y Desarrollo de Contenidos Educativos Reutilizables (SPDECE07), 2007

GARCÍA GUTIÉRREZ, Antonio Luís. *Epistemología de la documentación*, Stonberg Editorial, 2011

HJORLAND, B. (1997). *Information Seeking and Subject Representation. An Activity Teorical Approach to Information Science*. London, Greenwood Press.

KARCKHOVE, Derrick de. *Inteligencias en conexión. Hacia una sociedad de la web*. Barcelona, Gedisa Editorial, 1999. QUEAU, Philippe. *Lo virtual. Virtudes y vértigo*. Barcelona, Paidós, 1995
<http://es.scribd.com/doc/51888568/Queau-Philippe-Lo-virtual-Virtudes-y-vertigos>

KURZWEIL, Ray. *La singularidad está cerca*. Berlin, Lola Books, 2012

QUEAU, Philippe. *Lo virtual. Virtudes y vértigo*. Barcelona, Paidós, 1995
<http://es.scribd.com/doc/51888568/Queau-Philippe-Lo-virtual-Virtudes-y-vertigos>

PADRÓN, José. *Interpretaciones históricas acerca del conocimiento científico*
http://padron.entretemas.com/interpretaciones_hist.htm

PISCITELLI. *Internet: la imprenta del siglo XXI*. Barcelona, Gedisa, 2005
<http://tpcomunicaciontresunlz.files.wordpress.com/2011/11/piscitelli-alejandro-internet-la-imprenta-del-siglo-xxi.pdf>

RHEINGOLD, Howard *Multitudes inteligentes. La próxima revolución social*. Barcelona, Gedisa, 2004.

RHEINGOLD, Howard. *Realidad virtual*. Barcelona, Gedisa, 1994

<http://es.scribd.com/doc/62836521/Rheingold-Howard-Realidad-Virtual>

SAN SEGUNDO, Rosa. *A New Conception of Representation of Knowledge*, En: Knowledge Organization. International Journal. , Vol 31, 2004, n.2, p. 106-111

SAN SEGUNDO, Rosa. *Epistemologías posmodernas en Información y Documentación*. En: FORMACIÓN, investigación y mercado laboral en información y documentación en España y Portugal. Formação, investigação e mercado de trabalho em informação e documentação em Espanha e Portugal. Editor: José Antonio Frías, Crispulo Travieso. Ediciones Universidad de Salamanca, universidad, 2008 p. 675-684. Colección Aquilafuente, n.135. Ediciones Universidad de Salamanca.

http://books.google.es/books?id=UJ_xnapYvBMC&pg=PA675&lpg=PA675&dq=Rosa+san+segundo&source=bl&ots=WkCBD0xbV-&sig=JVFzFCNmNZab-F54KtKn2nXqpQ&hl=es&ei=rE3-TKW2IZCp8QOEuLSMCw&sa=X&oi=book_result&ct=result&resnum=2&ved=0CCAQ6AEwATgU#v=onepage&q=Rosa%20san%20segundo&f=false

SCOLARI, Carlos. *Hacer clic. Hacia una sociosemiótica de las interacciones digitales*. Barcelona, Gedisa Editorial, 2004.

SIMONE, R. *La tercera fase: formas de saber que estamos perdiendo*. Madrid: Taurus, 2001

ZALLO Ramón. *Paradojas de la cultura digital*

<http://sociedadinformacion.fundacion.telefonica.com/url-direct/pdf-generator?tipoContenido=articuloTelos&idContenido=2011072809010001&idioma=es>

WOLF, Maryanne. *Cómo aprendemos a leer. Historia y ciencia del cerebro y de la lectura*. Ediciones B, 2008

Aspectos teóricos, procedimentais, normativos e pragmáticos como categorias para uma epistemologia da organização da informação

Cristina Dotta Ortega
Universidade Federal de Minas Gerais
ortega@eci.ufmg.br

1 Organização do Conhecimento e Organização da Informação

A exemplo da polêmica em torno do termo gestão do conhecimento, não tratamos aqui de organização do conhecimento como a operação cognitiva realizada pelo indivíduo, mas como organização do que foi externalizado pela fala ou pela escrita e, tornado, deste modo, tangível. Denominamos como informação a unidade operacionalizável do conhecimento em um processo de organização do mesmo. A organização do conhecimento científico é, então, aquela que representa uma aproximação a uma configuração ideal deste conhecimento, sempre condicionada por questões políticas, sociais e culturais.

É preciso distinguir a organização do conhecimento (científico ou não) e a organização que é realizada com fins de promoção do acesso a este conhecimento, visando seu uso e nova produção, ou seja, a atividade de elaborar representações que possam ser significadas e manipuladas a favor de um certo público. Adotaremos o termo Organização da Informação para tratar desta última que é, neste texto, objeto da organização do conhecimento que se buscará explorar brevemente em torno de questões epistemológicas.

34

2 Para uma epistemologia da Organização da Informação

Partindo de García Gutiérrez (1999, p. 24), cultura refere-se aos modos distintos de os seres humanos entenderem e incorporarem o mundo e de se organizarem frente aos desafios da natureza. É neste contexto que se dá a Organização da Informação como o conjunto de procedimentos e ferramentas que ajudam os seres humanos a compreender e ordenar o caos cognitivo. Ainda segundo o autor, trata-se de dar respostas à sociedade para o problema do crescimento dos documentos, por meio de critérios científicos, como uma extensão da capacidade de organização e transferência daquilo que o ser humano sabe.

Diversas vertentes teórico-práticas constituídas desde ao menos o século XIX até hoje – nunca estanques ou independentes uma da outra – permitem entender a Organização da Informação, quais sejam, a Bibliografia, a Biblioteconomia, a Documentação e a Ciência da Informação. A Organização da Informação seria a atividade-meio para o objetivo ou fim da área, atividade segundo a qual convergem outros elementos constitutivos, como os aportes gerenciais e os tecnológicos. Este quadro permite construir um olhar crítico sobre a estrutura *ad hoc* da teoria da área a que se refere Smit (2009, p. 64), ao questionar a validade de estudos que seriam opcionais, pois seria possível prescindir deles sem prejuízo de uma compreensão da área. O fato de definições de Organização da Informação serem similares a algumas de Ciência da Informação, embora se reconheça a primeira como sub-área da segunda, pode ter relação com a compreensão de que a Organização da Informação seja o núcleo duro da área, como afirma Smit (p. 62).

Como matriz fundamentadora da Organização da Informação, trabalhamos com a linguagem como princípio operatório do processo informacional, e por a mesma constituir-se como elemento fundante da cultura, oferecendo aproximação aos modos sociais de produção e uso do conhecimento.

Ao tratar de uma epistemologia da Organização da Informação, estamos, portanto, construindo as fronteiras que fornecem o escopo da área, distinguindo-a de qualquer outra, e elaborando parâmetros para a eleição de aportes de outras áreas, assim como, os modos efetivamente científicos que permitem esta apropriação. O processo de apropriação não se dá, por exemplo, por importação direta nem por alteração arbitrária dos conceitos das outras áreas.

Para uma epistemologia da Organização da Informação nos parece produtivo refletir sobre categorias que a configuram como tal. Propomos as categorias teórica, procedimental, normativa e pragmática, não de modo definitivo, mas como um possível cenário para o debate. A categoria teórica é observada por meio do conjunto de conceitos básicos que sustentam a Organização da Informação no que tange a seus objetos, processos, instrumentos e produtos, assim como a explicitação da função de socialização de conteúdos a públicos específicos. A categoria procedimental refere-se à identificação dos processos essenciais e instrumentos correspondentes, observando-se aquilo que fundamenta cada um destes instrumentos e as metodologias mais rigorosas para seu uso. A categoria normativa inclui o conjunto de normas adotadas para a efetivação dos processos e para a construção dos instrumentos, elaboradas por

instituições para aplicação aos seus sistemas ou como referência nacional, regional ou internacional aos gestores dos diversos sistemas de informação. No caso das normas referenciais, trata-se de prescrições fornecidas em certos momentos como soluções norteadoras para novas situações surgidas em função de uma determinada tecnologia, tipologia documental e/ou necessidade de uso, a serem contextualizadas para uso em cada sistema. A categoria pragmática refere-se aos parâmetros que permitem elaborar procedimentos e instrumentos que envolvem as características linguísticas das comunidades discursivas de produção e uso de conhecimento, o que envolve os fatores culturais correspondentes, sempre condicionados institucionalmente.

Cada uma das categorias apresentadas envolve as outras e só pode ser compreendida a partir de sua relação com as mesmas. A categoria teórica seria a mais abrangente e a categoria normativa a mais pontual e datada.

Para ilustrar, segundo García Gutiérrez (1999, p. 27), na proposta emblemática de Otlet em seu Tratado de Documentação, tem-se a Documentologia como a ciência documentária, a Documentotecnia para tratar das técnicas documentárias, e a Documentonomia que se ocupa da normativa documentária. Os aspectos pragmáticos, assim como a própria noção de usuário, embora considerados, apenas mais tarde foram explicitamente observados.

3 Problemas em torno de uma epistemologia da Organização da Informação

Para tratar da epistemologia de uma área, estão em questão sua singularidade e suas fronteiras. Fronteiras, não articuladas ao núcleo duro, poderiam ser produtivamente adotadas como lugar privilegiado de pesquisa? Seria o caso de entender que estas pesquisas estariam indicando outra área de conhecimento?

Constatamos a dificuldade em articular conteúdos da Organização da Informação aos conteúdos de áreas variadas que, de algum modo, perpassam a informação. Como esta dificuldade tende a reforçar a abordagem operacional-mecanicista da Organização da Informação, qual seria o papel dos procedimentos documentários para uma configuração epistemológica?

Haveria proposta epistemológica que contemplasse de modo relativamente abrangente as diversas perspectivas da pesquisa brasileira, sem perder de vista a pergunta relativa a que tipo de problema a área deve responder? Seguindo outro

caminho, como elaborar posições epistemológicas passíveis de serem compreendidas e analisadas por pesquisadores atuantes em perspectivas variadas?

A categoria teórica apresentada é muitas vezes entendida como a representante de uma epistemologia da Organização da Informação, em especial no que tange à agregação de elementos de outras áreas. Por outro lado, a categoria normativa tende a ser considerada o elemento norteador concreto da prática profissional, do ensino e até da pesquisa. Sob esta concepção, faltaria solidariedade e interdependência entre as ações de teorizar e de normatizar?

Quais as implicações para uma epistemologia da Organização da Informação frente à ampla presença de processos e estratégias de gestão que não se articulam aos processos que visam à circulação da informação em uma instituição (no sentido amplo de pessoa jurídica, grupo ou pessoa), mas antes às questões da área de gestão das organizações?

A tecnologia é elemento intrínseco à área, que molda os processos, mas não os determina. Que compreensão da área é possível ao restringir-se a questão para as contingências da tecnologia eletrônica, mais especificamente, para as possibilidades de contribuição da Ciência da Computação, portanto, segundo seus objetivos e história? Quais as implicações de um entendimento de tecnologia como a instância da aplicação e que, por este motivo, não compõe a configuração epistemológica da área?

A noção de usuário é imprescindível a uma epistemologia da Organização da Informação, como indicamos nas categorias propostas. Quais as decorrências, no desenvolvimento da área, de uma possível predominância de estudos sobre usuários centrados no uso social ou cognitivo da informação, prescindindo da sua relação com as ações que buscam responder à complexidade do acesso e uso de informação?

A configuração epistemológica de uma área é necessária à transposição didática efetivada nos diversos níveis de ensino. Como fazer uso dos consensos historicamente constituídos em Organização da Informação, mesmo que distantes de um nível mínimo de compartilhamento para além de certas escolas, assim como dos diversos dissensos, tão fracamente explicitados? Como superar a orientação fortemente normativa do ensino, baseado no modelo proposto pela Biblioteca do Congresso dos Estados Unidos, passando a contemplar os aspectos conceituais e a historicidade deste e de outros modelos?

Se a Organização da Informação evidencia processos, trata-se de considerar que a orientação destes é dependente do objetivo que se persegue. Deste modo, como

questão principal, apontamos o desafio colocado por uma epistemologia de uma ciência social aplicada que deve responder pela Organização da Informação, e sustentá-la, fornecendo compreensão da mesma e fundamentação às aplicações.

Estamos imersos em diversas (falsas) dicotomias, como: graduação (Biblioteconomia – formação profissional) *versus* pós-graduação (Ciência da Informação – formação para a pesquisa); processamento manual *versus* processamento automático, aspectos técnicos *versus* aspectos sociais e, principalmente, teoria *versus* prática. Identificamos como pano de fundo destas antigas mas ainda recorrentes dicotomias os resquícios de um pensamento fragmentado, não historicizado, nem elaborado sob base conceitual consistente.

A despeito do acúmulo de conhecimento que caracteriza a Organização da Informação e a distingue, vivenciamos muitos problemas extra-epistemológicos na quase ausência do enfrentamento de problemas efetivamente epistemológicos.

Referências:

GARCÍA GUTIÉRREZ, Antonio. Aproximación al concepto y al objeto de la Información/Documentación. In: _____(Ed.). **Introducción a la documentación informativa y periodística**. Alcalá de Guadaira (Sevilla): Ed. Mad, 1999. Cap. 1: p. 23-61.

SMIT, Johanna Wilhelmina. Novas abordagens na organização, no acesso e na transferência da informação. In: SILVA, Helen de Castro; BARROS, Maria Helena T. C. de. **Ciência da Informação: múltiplos diálogos**. Marília: Oficina Universitária UNESP, 2009. p. 57-66.

Terminologia, linguística e filosofia: contribuições interdisciplinares em artigos sobre sistemas de organização do conhecimento publicados em revistas nacionais

Marisa Bräscher
Universidade de Santa Catarina
marisa.brascher@ufsc.br

Filosofia, Linguística e Terminologia são áreas que, tradicionalmente, contribuem com a organização e representação do conhecimento, tanto do ponto de vista teórico quanto metodológico. Quando procuramos por fundamentação teórica acerca dos relacionamentos semânticos em sistemas de organização do conhecimento (SOC), nos deparamos com algumas questões: que correntes teóricas dessas três áreas efetivamente contribuem para os estudos sobre SOC? Como essas contribuições evoluem no tempo? Em que aspectos específicos dos SOC aplicam-se fundamentos teóricos dessas três áreas?¹¹ Partindo de tais indagações, efetuamos o levantamento e análise da literatura brasileira a fim de identificar como a Filosofia, a Linguística e a Terminologia tem contribuído com os estudos brasileiros sobre SOC. O objetivo, nesta etapa da pesquisa, é verificar a evolução temporal e temática dessas contribuições interdisciplinares no Brasil.

Utilizamos a Base de Dados Referencial de Artigos de Periódicos em Ciência da Informação (BRAPCI)¹², que indexa 37 periódicos nacionais impressos e eletrônicos da área de Ciência da Informação, com cobertura temporal de 1972 (data do primeiro fascículo de periódico indexado pela base de dados) até os dias atuais. Efetuamos a busca com as seguintes palavras-chave, pesquisadas individualmente: sistemas de organização do conhecimento, filosofia, linguística, terminologia, linguagem documentária, tesouro, classificação, taxonomia, ontologia e semântica. Foram eliminados artigos duplicados e selecionados apenas aqueles que continham informação sobre a base teórica utilizada em seus títulos, palavras-chave ou resumo. É necessário

¹¹ Lembramos que o artigo FRANCELIN, M.M; KOBASHI, N. Concepções sobre o conceito na organização da informação e do conhecimento. *Ciência da Informação*, v. 40, n. 2, p. 207-228, maio/ago. 2011 aponta abordagens adotadas nos estudos brasileiros sobre o conceito. Mesmo que os objetivos dos dois estudos sejam distintos, uma vez que nós procuramos identificar as contribuições das três áreas citadas nos estudos sobre SOC, os resultados das duas pesquisas poderão ser comparados futuramente.

¹² <http://www.brapci.ufpr.br/>

ressaltar que grande parte dos artigos recuperados não apresentou as bases teóricas definidas em seus resumos ou palavras-chave e, portanto, não foram considerados na pesquisa. As análises aqui apresentadas foram limitadas, ainda, ao conjunto de artigos que tratam de sistemas de organização do conhecimento, no sentido em que definimos: entendemos que os SOC são representações de domínios do conhecimento que delimitam o significado de termos no contexto desses domínios, estabelecem relações conceituais que auxiliam a posicionar um conceito no sistema conceitual e são utilizadas como instrumentos de organização e recuperação da informação.

Dentre os distintos tipos de SOC, nosso estudo restringe-se aos tesouros, classificações, taxonomias e ontologias. Após realizar os recortes definidos, chegamos ao total de 73 artigos que corresponderam aos critérios da pesquisa. Eliminamos do total de artigos recuperados na ABRAPCI, aqueles que não apresentaram de maneira evidente, em seus resumos, a base teórica em que se apóiam. A análise dos 73 artigos evidencia algumas tendências que passamos a apresentar.

A Filosofia apresentou predomínio sobre a Linguística e Terminologia ao ser utilizada como base teórica única em 23 trabalhos. Os trabalhos que se apoiam na Teoria da Classificação Facetada de Ranganathan foram considerados no âmbito desta vertente e totalizaram 10 artigos, sendo, portanto, o aporte teórico mais utilizado nos estudos com base na Filosofia.

Os filósofos e teóricos citados nos resumos são: Michel Foucault, Marcel Mauss, Claude Lévi-Strauss, Jean Piaget, Shiyali Ramamrita Ranganathan, Luciano Floridi, Jeremy Bentham, Aristóteles, Immanuel Kant, Charles Sanders Peirce, David Hume, Martin Heidegger, Hans-Georg Gadamer, Ludwig Wittgenstein. A diversidade de correntes filosóficas e linhas de pensamento de tais filósofos e teóricos nos leva a concluir que, com exceção da Teoria da Classificação Facetada de Ranganathan, não há corrente predominante nos estudos de SOC que se apoiam na Filosofia.

A distribuição por período de produção revela um crescimento do interesse pelos estudos de base filosófica. Encontramos 4 artigos publicados na década de 1970 e 1 na década de 1990. Já no século atual, temos 10 trabalhos na primeira década (2000 a 2009) e 8 nos 2 anos iniciais da segunda década (2010-2012).

Quanto à temática dos 23 trabalhos com base teórica da Filosofia, temos os seguintes números de artigos por assunto: classificação (15), aspectos gerais da

representação do ¹³conhecimento Os trabalhos foram incluídos nesta temática quando não abordam nenhum tipo específico de SOC (3), ontologia (2), tesouro (2) e taxonomia (1). A classificação é a temática mais estudada na abordagem filosófica, resultado que não nos surpreende, se consideramos que 10 trabalhos dessa vertente teórica apóiam-se nos estudos da classificação facetada.

A Terminologia é a segunda área em que maior número de trabalhos se apóiam, num total de 22 artigos que a mencionam como base teórica única. A evolução temporal apresenta-se com os seguintes números de trabalhos por décadas: 1970 (1), 1990 (4), 2000 (14) e 2010 (3).

No que diz respeito aos assuntos, os artigos que tratam de linguagem documentária e de aspectos gerais da representação do conhecimento são os mais frequentes, com 8 e 7 artigos, respectivamente. São abordados, ainda: ontologia e tesouros, ambos com 3 artigos e 1 artigos que trata dos dois sistemas conjuntamente. É interessante notar que não foram identificados artigos sobre classificação e que se apóiam na Terminologia.

A Linguística é mencionada como base teórica única em 15 artigos, distribuídos por décadas da seguinte maneira: década de 1970 (1), 1980 (4), 2000 (7) e 2010 (3). Os temas tratados nessa vertente são: aspectos gerais da representação do conhecimento (6) linguagens documentárias (5), ontologias (2), classificação (1) e tesouros (1). Observa-se que, assim como a Terminologia, a Linguística e a Terminologia tem as mesmas temáticas mais frequentes: aspectos gerais da representação do conhecimento e linguagens documentárias. De maneira geral, os estudos nessas temáticas discutem as possíveis contribuições dessas duas áreas para a elaboração de SOC.

Encontramos artigos que se apoiam em bases teóricas provenientes de mais de uma área. Na intersecção da Filosofia com a Linguística, temos 2 trabalhos, 1 na década de 1970 e outro na de 2000. Filosofia e Terminologia são abordadas conjuntamente em 3 trabalhos publicados nos anos 1996, 2010 e 2011 e todos discorrem sobre contribuições dessas áreas aos estudos das linguagens documentárias ou sobre aspectos gerais da representação do conhecimento. Estudos com bases teóricas da Terminologia e Linguística foram identificados apenas a partir de 2000, com 5 artigos e 3 na década de 2010. Tratam de linguagens documentárias (5), aspectos gerais da representação do conhecimento (1), tesouros (1) e SOC (1).

¹³ Os trabalhos foram incluídos nesta temática quando não abordam nenhum tipo específico de SOC e tratam de aspectos mais gerais da elaboração desses sistemas.

Somando-se as contribuições das três áreas e suas intersecções, identificamos que os temas mais tratados são: aspectos gerais da representação do conhecimento (23), linguagens documentárias (18), classificação (16), tesouros (7) e ontologias (7). Tesouros e ontologias tratados conjuntamente em um artigo, assim como taxonomias e SOC, também abordados apenas uma vez. Observa-se uma quase ausência de estudos sobre taxonomias com base teórica nas três áreas analisadas, não sabemos se é realmente um tema que não tem merecido a mesma atenção dada a outros tipos de SOC na área de ciência da informação ou se os estudos são mais orientados aos aspectos práticos da elaboração de taxonomias e não identificam de maneira evidente a base teórica que utilizam. Outro aspecto a ser destacado é o número igual de trabalhos sobre tesouros e ontologias, fato interessante, se levarmos em conta que tesouro é uma temática que vem sendo pesquisada há mais tempo na área, sobretudo a partir da década de 1970 e que as ontologias surgem já na década de 1990.

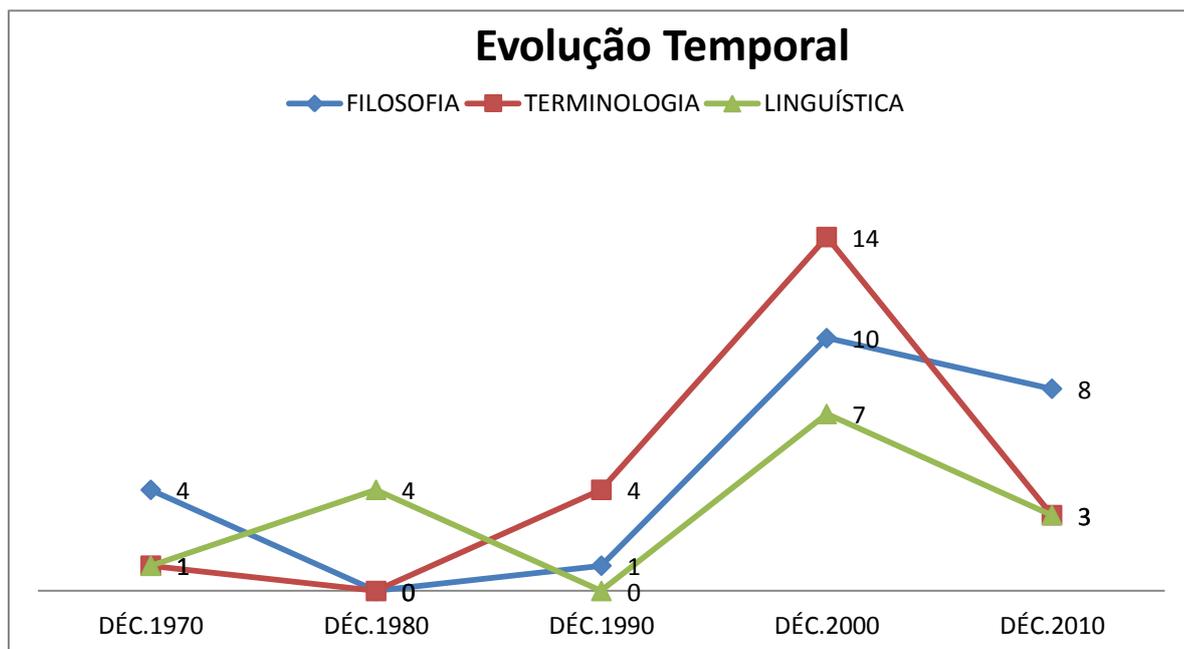


Fig.1 – Evolução do número de artigos em cada vertente teórica por década

Fonte: produção própria

A figura 1 apresenta a evolução temporal dos artigos nas três vertentes teóricas. Apesar dos resultados refletirem uma realidade parcial, uma vez que limitada aos artigos dos periódicos indexados pela BRAPCI, observa-se o crescimento dos estudos na década de 1990. Talvez alguns fatores contribuam para esse quadro: o aumento do número de revistas especializadas na área de Ciência da Informação, sobretudo as publicadas em formato eletrônico, o que abre espaço para a divulgação das pesquisas realizadas no país e o aumento do número de doutores que tem a área de organização do conhecimento como linha de pesquisa, muitos dos quais finalizaram seus estudos de doutorado no final da década de 1990, como muitos dos colegas que se encontram presentes neste evento. Ainda com base na figura 1, se tomarmos os dois anos iniciais da década atual como indicadores de tendências, nos chama a atenção ver que a Filosofia apresenta o maior número de trabalhos (8), em relação à Terminologia e a Linguística, ambas com 3 artigos. Análises futuras poderão corroborar ou não essa tendência.

Levando em conta as contribuições de cada área e as intersecções entre elas, observamos que a Terminologia passa a ser a que apresenta maior número de artigos (33), a Filosofia, com 28 trabalhos é a segunda e a Linguística, com 25 artigos, permanece como a que apresenta menor contribuição na literatura brasileira. Esse resultado consolidado reflete nosso sentimento no início da pesquisa, pois, em nossa

experiência na área, identificávamos uma presença marcante da Terminologia nos estudos brasileiros de SOC. Por outro lado, o número de artigos na vertente da Filosofia nos surpreendeu, sobretudo pela diversidade de caminhos possíveis, em função das correntes filosóficas presentes na literatura brasileira.

Os resultados aqui apresentados nos permitem concluir que a Filosofia, a Linguística e a Terminologia tem contribuído efetivamente com os estudos sobre SOC no Brasil. Nossa pesquisa segue no sentido de aprofundar as análises acerca das contribuições das três áreas nos estudos de SOC e identificar as teorias e métodos empregados nos estudos brasileiros. Pretende-se com os futuros resultados, colaborar com a definição de arcabouço teórico para a elaboração de SOC.¹⁴

¹⁴ O presente trabalho contou com a participação da bolsista Pibic Elisabete Lima Machado na fase de coleta e sistematização dos dados, a quem manifestamos nosso agradecimento.

Temas e questões em análise documentária de charge

Thulio Pereira Dias Gomes
Instituto Brasileiro de Informações em Ciência e Tecnologia
thuliogomes@gmail.com

Vânia Lisbôa da Silveira Guedes
Universidade Federal do Rio de Janeiro
vanialisboa@facc.ufrj.br

Maria José Veloso da Costa Santos
Universidade Federal do Rio de Janeiro
msantos1402@gmail.com

1 Introdução

Este trabalho desenvolve um estudo exploratório sobre análise documentária (AD) de charge, a partir dos pressupostos teóricos da Linguística Documentária. Oferece uma contribuição para a Organização do Conhecimento (OC), vez que procura suscitar algumas reflexões acerca da contribuição da teoria dos gêneros discursivos para a AD da charge.

Para a fundamentação teórica foi realizada pesquisa bibliográfica sobre a AD de charges em periódicos selecionados na página da Associação Nacional de Pesquisa e Pós-Graduação em Ciência da Informação. A busca revelou escassez de estudos sobre o assunto, fato que, entre outros, revela uma lacuna nas pesquisas sobre o tema e sugere a necessidade de intensificação de estudos sobre esse gênero discursivo como objeto informativo na OC, sobretudo, discussões voltadas para abordagens teóricas e metodológicas de AD da charge. .

2 Objetivos

O estudo apresenta como objetivos:

2.1 *Objetivo geral*

- Contribuir para o avanço teórico-metodológico de tratamento temático da informação verbal-imagética contida em charges.

2.2 *Objetivos específicos*

- Levantar questões para pesquisa sobre especificidades do processo de AD de charge.

- Incentivar estudos sobre AD de charge no domínio da OC.
- Intensificar a interface entre a Ciência da Informação (CI) e a Linguística.

3 Pressupostos Teóricos

A AD consiste em “[...] atividade metodológica específica no interior da Documentação, que trata da análise, síntese e representação da informação, com o objetivo de recuperá-la e disseminá-la” (Cintra et al., 1994: 24). García Gutierrez (1990) esclarece que a AD consiste em extrair ideias centrais de um documento com o fim de disponibilizá-lo para sua recuperação mediante representações sintéticas. Araújo Júnior (2007) aponta a AD como principal processo pelo qual passa um documento em um sistema informativo documentário (SID) que, segundo Rendón Rojas (2008), é formado pela interação entre informação, documento, usuário e instituição informativa documental.

O conceito de "registro textual" tem sido crescentemente substituído pela ênfase na análise de gênero discursivo, na Sociolinguística e Análise do Domínio na OC, objetivando investigar as linguagens especializadas de comunidades discursivas. Nesse contexto, os documentos apresentam aspectos linguísticos que os caracterizam como gêneros discursivos, porque desempenham função comunicativa pelo uso de um sistema de linguagem. Em Swales (1990), a noção de gênero discursivo é fundamentada nos conceitos de propósito comunicativo e de comunidade discursiva. Para ele, a análise de gêneros refere-se à análise de qualquer tipo particular de discurso e, por sua vez, as comunidades discursivas são categorizadas pelos gêneros discursivos que utilizam. Nessa perspectiva, cada comunidade de discurso define os gêneros mais relevantes para atingir seus objetivos (Guedes, 2010). Swales acrescenta que o trabalho que os membros de uma comunidade de discurso estão engajados envolve o processamento de tarefas que refletem uma linguística específica, habilidades retóricas e discursivas.

O conceito de charge adotado neste trabalho segue o conceito operacional, sugerido por Gomes (2013: 26-27). Assim, a charge é “um gênero discursivo de uso híbrido das linguagens verbal e imagética, caracterizado pela temporalidade marcada pela sátira e pela crítica referentes a determinado evento, em geral de natureza política”.

3 Análise Documentária De Charge

A charge é considerada objeto informativo em diferentes domínios do conhecimento (Gomes, 2013), o que implica diferentes usos do documento. Portanto, o

tratamento da charge em um SID deve observar a utilização específica que se faz do gênero diretamente associados ao contexto de uso. Cada domínio interpreta, representa e utiliza a charge, segundo seus próprios critérios de relevância (Capurro & Hjørland, 2007). De acordo com Hjørland (2001), a melhor análise temática é aquela que faz melhor prognóstico quanto ao uso do documento. Assim, a qualidade da AD de charge exige o conhecimento sobre usuários que atuam em comunidades discursivas, porque responde aos parâmetros de relevância do domínio atendido pelo SID, facilitando o desenvolvimento de linguagens documentárias compatíveis com a linguagem natural dessa comunidade.

A literatura apresenta pesquisas sobre a AD em outros tipos de documentos de linguagem imagética (cf. Smit, 1996; Manini, 2002; Moreiro González; Robledano Arillo, 2003; Shatford, 1986). A aplicabilidade dessas metodologias à charge deve ser verificada sob a perspectiva da eficácia e da eficiência da representação e recuperação de informação contida em charges e pode contribuir para o desenvolvimento de metodologias que considerem os aspectos informativos e linguísticos comuns entre a charge e os outros tipos de documento investigados.

Gawryszewski (2008) e Peláez Malagón (2003) sugerem a necessidade de estabelecimento de novas categorias para a caricatura e a charge. Guedes (2010) e Mollica e Guedes (2011) falam da viabilidade com fins didáticos e da pesquisa do agrupamento de gêneros discursivos. Dessa forma, a categorização de charge pode contribuir para a AD do gênero, sobretudo aquela que envolve a representação descritiva. O estabelecimento de diferentes categorias poderá apontar a necessidade de desenvolvimento de metodologias distintas, consoantes às possíveis tipologias de charge.

Outra problemática relevante para AD da charge são os níveis de informação a ser descritos. Por exemplo, Panofsky (1979) propõe níveis de significação para a análise de imagens pictóricas, os quais são utilizados por Shatford (1986) na OC. A possibilidade de transposição desses níveis para a AD de charge deve ser verificada. Acredita-se que a linguagem híbrida pode reconfigurar esses níveis de significação, o que tornaria necessário o estabelecimento de outros níveis para a charge.

Smit (1989) aponta problemas no uso em análise de imagens de linguagens documentárias criadas para tratamento de textos verbais, tendo em vista os diferentes códigos linguísticos desses gêneros documentais. Argumenta que os níveis de descrição

devem ser outros. Há necessidade de verificar se na charge o híbrido uso da linguagem e sua relação com o contexto de enunciação modificam a relação apontada.

Os conhecimentos da estrutura textual e das condições de produção discursiva favorecem a compreensão do fluxo de informação durante a leitura do discurso (Mollica & Guedes, 2011). Em termos de AD, o conhecimento da estrutura do discurso e de como se dá sua produção sófica permitem a caracterização de processos cognitivos ocorrentes durante a leitura documentária (Fujita, 2004; Kobashi, 1996). Parafraseando Fujita (2004), é possível desenvolver estratégias metacognitivas de leitura documentária, permitindo o alcance de maiores índices de eficiência e de eficácia na AD, tornando necessário desenvolver pesquisas sobre a estrutura e o fluxo de informação em charges e suas implicações na leitura documentária.

O contexto de produção chargístico se coloca como desafio para a AD, vez que a compreensão da mensagem depende do conhecimento das condições de enunciação da mesma (Flôres, 2002). Deve-se também atentar para o tempo necessário para recuperar o contexto de produção da charge, o que pode tornar a AD morosa. Lara (2011) reconhece que é grande a quantidade de documentos com que lida a AD. Assim, uma metodologia para AD de charge deve considerar os recursos disponíveis para execução, considerando o tempo médio necessário para cada documento. A identificação de estratégias metacognitivas que permitem a recuperação do contexto se torna, destarte, condição de sucesso para a AD de charge.

Para haver informação, é necessária uma pré-estrutura de conhecimento capaz de contextualizar e de interpretar a mensagem, gerando uma nova estrutura de conhecimento (Capurro, 2007). O conhecimento prévio do leitor documentário é fundamental para a compreensão da charge (Flôres, 2002). A leitura da charge, desse modo, consiste em processo informativo de contextualização do conhecimento objetivado no documento, a partir de um pressuposto já conhecido. Na AD, quanto maior o conhecimento pressuposto, maior a chance de o processo informativo obter êxito.

Gawryszewski (2008: 16) afirma que o riso ou o prazer na leitura da charge só existirá “se houver certa cumplicidade, uma possibilidade de acordo entre o desenhista e o leitor”. Cunha (1990: 71) traz à discussão o problema da ideologia do analista documentário na execução de suas atividades profissionais. Para a autora, a negação do componente ideologia em análise de texto para fins documentais “é não colocar em questão o papel do indivíduo no processo de produção, consumo e tradução do conteúdo

para o todo/conceito/palavra-chave”. Alerta, ainda, que essa postura também consiste em negar que a linguagem do discurso ou da linguagem documentária é aquisição, construção e instrumento “elaborado respectivamente pela sociedade por e para um grupo social, refletindo, em ambos os casos, o pensamento e os interesses dominantes”. Estas questões provocam discussões a respeito da postura ética para a AD da charge, e das limitações de tal processo considerando as diferenças entre os contextos ideológicos de produção e de uso.

4 Considerações Finais

As especificidades textuais da charge devem ser observadas durante o processo de AD. Esse procedimento se justifica tendo em vista as especificidades informativas, diretamente relacionadas às suas características discursivas.

A AD se apresenta como abordagem teórico-metodológica profícua para a organização da informação de charge, considerando suas interlocuções com a Linguística Textual. Nessa perspectiva, é importante ressaltar o potencial de contribuição das pesquisas no âmbito da análise crítica do discurso em gêneros discursivos e a interação discurso e contexto para a AD da charge na OC.

Finalmente, provavelmente, o uso híbrido da linguagem na charge e sua relação com o contexto de enunciação, podem ser representados a partir das abordagens teóricas- metodológicas da Análise de Gêneros em interação com a CI.

Referências

- Araújo Júnior, Rogério Henrique de. (2007). *Precisão no processo de busca e recuperação da informação*. Brasília, DF: Thesaurus.
- Capurro, Rafael. (2007). Epistemología y ciencia de la informacion. *Enl@ace*, 4 (1): p. 11-29.
- Capurro, Rafael & Hjørland, Birger. (2007). O conceito de informação. *Perspectivas em Ciência da Informação*, 12 (1): 148-207.
- Cintra, Anna Maria Marques et al. (1994). *Para entender as linguagens documentárias*. São Paulo: Editora Polis.
- Cunha, Isabel Maria Ribeiro Ferin. (1990). *Do mito à análise documentária*. São Paulo: EDUSP.
- Flôres, Onici. (2002). *A leitura da charge*. Canoas, RS: Ed. ULBRA.
- Fujita, Mariângela Spotti Lopes. (2004). A leitura documentária na perspectiva de suas variáveis: leitor-texto-contexto. *Datagramazero – Revista de Ciência da Informação*, 5 (4). Disponível em: <http://www.dgz.org.br/ago04/F_I_art.htm> Acesso em: 27 mar. 2013.

- García Gutiérrez, Antonio Luis. (1990). *Estructura lingüística de la documentación: teoría y método*. Murcia: Universidade de Murcia, Secretariado de Publicaciones.
- Gawryszewski, Alberto. (2008). Conceito de caricatura: não tem graça nenhuma. *Domínios da Imagem*, 1 (2): 7-26.
- Gomes, Thulio Pereira Dias. (2013). *Temas e questões em análise documentária de charges*. Trabalho de Conclusão de Curso—(Bacharelado em Biblioteconomia) - Universidade Federal do Rio de Janeiro, Rio de Janeiro.
- Guedes, Vânia Lisbôa da Silveira. (2010). *Nominalizações deverbais em artigos científicos: uma contribuição para a análise e a indexação temática da informação*. Tese (Doutorado em Linguística)—Universidade Federal do Rio de Janeiro, Rio de Janeiro.
- Hjørland, Birger. (2001). Toward a theory of aboutness, subject topicality, theme, domain, field, content... and relevance. *Journal of the American Society for Information Science and Technology*, 52: 774-78.
- Lara, Marilda Lopes Ginez de. (2011). Conceitos de organização e representação do conhecimento na ótica das reflexões do Grupo Temma. *Informação e Informação*, 16 (3): 92-121.
- Kobashi, Nair Yumiko. (1996). Análise documentária e representação da informação. *Informare*, Rio de Janeiro, 2 (2): 5-27.
- Manini, Miriam Paula. (2002). *Análise documentária de fotografias: um referencial de leitura de imagens fotográficas para fins documentários*. Tese (Doutorado em Ciências da Comunicação) —Universidade de São Paulo, São Paulo.
- Mollica, Maria Cecilia de Magalhães & Guedes, Vânia Lisbôa da Silveira. (2011). *Gêneros discursivos e Ciência da Informação*. IBICT/CNPq. No prelo. Disponível em: <<http://www.ppgci.ufrj.br/index.php/teste?download=213%3Ageneros-discursivos-e-ciencia-da-informacao>>. Acesso em: 27 set. 2012.
- Moreiro González, José Antonio & Robledano Arillo, Jesus. (2003). *O conteúdo da imagem*. Curitiba, PR: Editora da UFPR.
- Panofsky, Erwin. (1979). *Significado das artes visuais*. 2. ed. São Paulo: Perspectiva.
- Peláez Malágon, José Enrique. (2002). El concepto de caricatura como arte en el siglo XIX. *Sincronía*, Primavera. Disponível em: <<http://sincronia.cucsh.udg.mx/caricatur.htm>>. Acesso em 14 jan. 2013.
- Rendón Rojas, Miguel Ángel. (2008). La ciencia de la información en el contexto de las ciencias sociales y humanas. Ontología, epistemología, metodología e interdisciplina. *Datagramazero - Revista de Ciência da Informação*, 8 (4). Disponível: <http://www.dgz.org.br/ago08/Art_06.htm>. Acesso em: 30 abr. 2013.
- Shatford, Sara. (1986). Analyzing the subject of a picture: a theoretical approach. *Cataloging & Classification Quarterly*, 6 (3): 39-62.
- Smit, Johanna W. (1989). A análise da imagem: um primeiro plano. In: Grupo Temma. *Análise documentária: a análise da síntese*. 2. ed. Brasília, DF: IBICT: 101-13.
- _____. (1996). A representação da imagem. *Informare*, Rio de Janeiro, 2 (2)]: 28-36.

Swales, John M. (1990). *Genre analysis: English in academic and research settings*. Cambridge: Cambridge University Press.

Tendências temáticas e opções metodológicas da produção científica de pesquisadores brasileiros na *Advances in Knowledge Organization* (2010 e 2012)

Leilah Santiago Bufrem
Universidade Federal de Pernambuco/
Universidade Estadual Paulista-FFC
santiagobufrem@gmail.com

Murilo Artur Araújo da Silveira
Universidade Federal de Pernambuco
muriloas@gmail.com

Bruna S. do Nascimento
Universidade Federal do Rio Grande do Sul.
brusnascimento@gmail.com

1 Introdução

Um dos elementos substantivos para a institucionalização científica, do ponto de vista de sua estrutura social é, segundo Whitley (1974), a organização do campo, expressa na existência de periódicos, concretização de eventos e associações, como evolução natural de uma especialização científica. Contudo, a estrutura social é interdependente da estrutura cognitiva, que se orienta para as questões teóricas e metodológicas de um campo científico (WHITLEY, 1974). Como essas estruturas diversificam-se conforme os campos, apresentando modos de expressão e características particulares, para que se possa compreendê-las é importante acompanhar simultaneamente as manifestações do processo de institucionalização (social e cognitiva) que, segundo o autor, diz respeito à concordância sobre certas questões relativas à demarcação da área.

Entre essas questões, reserva-se aqui um lugar especial aos modos de tratamento ou metodologias para a concretização dos estudos, alternativas reveladas na literatura especializada, pois o estudo sobre as práticas institucionalizadas de pesquisa revela um dinamismo gerador e renovador de domínios e métodos científicos, embora sem perder sua vertente reprodutora.

Considera-se para este estudo que as comunicações em eventos, entre as expressões da produção científica, são cronologicamente as mais próximas do ato de criação. Sendo assim, ao representarem um momento dialógico entre pesquisador e

pares, podem ser consideradas como elementos originais constitutivos da institucionalização cognitiva e social do domínio. Oferecem-se, portanto, quando de sua apresentação, à análise e crítica dos pares pesquisadores, como alvo dos olhares e observações em prol de seu aperfeiçoamento, gerando discussões profícuas e repercutindo a pesquisa científica.

Essa compreensão motivou-nos a escolher como constituintes do *corpus* específico deste estudo um conjunto de 27 trabalhos de autores brasileiros sobre organização do conhecimento publicados na *Advances in Knowledge Organization* nos dois últimos volumes (2010 e 2012). Buscou-se identificar e analisar essas comunicações do ponto de vista de sua trajetória metodológica e temática e em que se aproximam essas trajetórias selecionadas dos objetos de estudo.

A importância dos anais de eventos para a compreensão desse discurso é enfatizada por Smiraglia (2012) uma vez que as limitações de publicações em periódicos são tantas que apenas pequena proporção da pesquisa no domínio neles aparece. A questão que surge, então, considerando-se o tema organização do conhecimento, é sobre a possibilidade de se compor um mapeamento das tendências temáticas e das trajetórias metodológicas selecionadas pelos autores dos artigos para a resolução de seus problemas científicos, partindo-se da análise de suas comunicações.

O objetivo deste trabalho – identificar e analisar modalidades de pesquisas como opções metodológicas dos autores brasileiros responsáveis por 27 artigos publicados nos anais da ISKO - implica a distinção de procedimentos, a partir da análise e reflexão sobre um campo do saber cujos componentes disciplinares realizam-se nas interfaces com outras disciplinas. Destacam-se como objetivos específicos a identificação dos trabalhos com autoria brasileira e a caracterização das pesquisas segundo suas opções metodológicas e temáticas.

Com apoio no referencial existente sobre o campo empírico pesquisado, busca-se contribuir para a compreensão do processo de investigação e de raciocínio empregado pelos autores dos artigos, não somente como recurso para o aperfeiçoamento do fazer científico, como também, para o reconhecimento dos modos de investigação. Observam-se, assim, meios de produção do conhecimento, como forma de contribuição à ampliação das visões sobre a pesquisa e à divulgação dos saberes na área.

2 Procedimentos Metodológicos

O *corpus* analisado foi extraído dos dois últimos volumes da *Advances in Knowledge Organization*, publicação composta pelos trabalhos apresentados nos eventos promovidos pela ISKO nos anos de 2010 e 2012. Escolheu-se a aplicação de técnicas diversas objetivando contemplar a análise do objeto em sua complexidade. Para tanto, empregou-se a análise de conteúdo (BARDIN, 1977) que permite não só o reconhecimento das opções metodológicas aplicadas, mas também que oferece um retrato acerca de como as pesquisas estão sendo estruturadas e apresentadas à comunidade científica. Para verificar as temáticas mais recorrentes utilizaram-se técnicas bibliométricas que possibilitam “[...] conhecer, pelo estudo das publicações, o desenvolvimento de área determinada da ciência, em âmbito geográfico específico, com recortes relativos a temáticas, materiais ou períodos específicos.” (BUFREM; PRATES, 2005: 13).

As etapas de pesquisa foram:

Etapa 1 - Levantamento dos Trabalhos Constantes nos Anais da ISKO (2010 e 2012):

- >seleção dos números das duas últimas edições da publicação;
- >identificação e arquivamento dos trabalhos dos pesquisadores brasileiros.

Etapa 2 - Formatação da Base de Dados e Inclusão dos Dados Bibliográficos:

- a) estabelecimento de categorias sobre os elementos pertinentes à pesquisa dos trabalhos dos autores brasileiros;
- b) formatação da base (Excel) tendo como parâmetro as categorias eleitas para alcance dos objetivos da pesquisa;
- c) seleção e inclusão dos dados presentes nas contribuições identificadas.

Etapa 3 - Análise dos Dados e Discussão dos Resultados:

- a) edição e cruzamento dos dados e elementos dos trabalhos por meio das categorias estabelecidas: os enfoques metodológicos, os procedimentos metodológicos, as técnicas e os instrumentos de análise de resultados, a aproximação com a realidade temática e as tendências temáticas;
- b) definição e elaboração de tabelas e representações gráficas resultantes da análise dos dados;
- c) discussão dos resultados com base nos gráficos e tabelas geradas em consonância com os objetivos traçados pela pesquisa.

3 Resultados

Os trabalhos foram categorizados preliminarmente quanto à aproximação com a realidade, com distribuição de ocorrências conforme a Tabela 1.

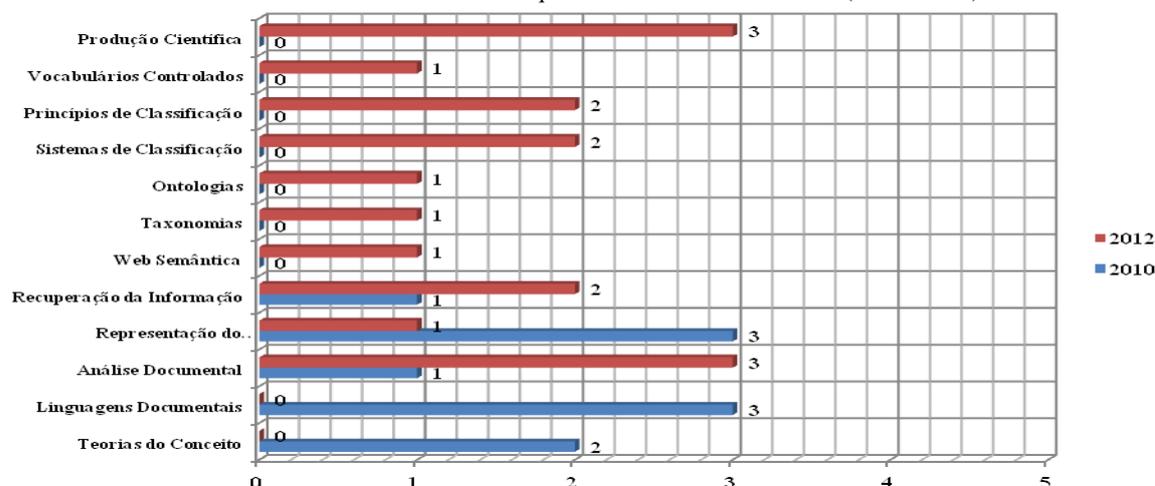
Tabela 1 - Distribuição dos Trabalhos quanto à Aproximação com a Realidade

Aproximação	2010		2012		Acumulado (2010 e 2012)	
	Freq. Abs.	Freq. Rel. (%)	Freq. Abs.	Freq. Rel. (%)	Freq. Abs.	Freq. Rel. (%)
Fundamentos Teóricos	6	22,22%	8	29,63%	14	51,85%
Fundamentos Metodológicos	4	14,81%	6	22,22%	10	37,04%
Aplicações	-	-	3	11,11%	3	11,11%
TOTAL	10	37,04%	17	62,96%	27	100,00%

Fonte: dados da pesquisa

Os trabalhos comunicados pelos pesquisadores brasileiros, em relação às temáticas mais recorrentes nas duas edições do *Advances in Knowledge Organization*, são visualizados no Gráfico 1.

Gráfico 1 - Tendências Temáticas dos Trabalhos de Pesquisadores Brasileiros no ISKO (2010 e 2012)



Fonte: dados da pesquisa

Em relação à distribuição temática dos trabalhos sobre organização do conhecimento, o gráfico sugere, por um lado, que as comunicações distribuem-se em dois blocos temáticos, com proeminência de temas que se relacionam entre si. O primeiro, abrangendo trabalhos de caráter mais tradicional, em 2010, e o segundo, em 2012, apresentando temas inovadores como *web semântica*, ontologias e taxonomias. A análise de conteúdo permitiu que se verificasse relação mais íntima entre o enfoque teórico e os temas Produção Científica, Representação do Conhecimento, Teorias do Conceito e Análise Documental, enquanto que o enfoque metodológico relacionou-se fortemente com as temáticas Princípios de Classificação, Sistemas de Classificação, Recuperação da Informação e Vocabulários Controlados. Já os temas Web Semântica, Taxonomias e Ontologias relacionam-se com os trabalhos que abordam as aplicações da organização do conhecimento.

Os resultados obtidos sobre as opções metodológicas dos pesquisadores que apresentaram comunicações nos dois eventos analisados são apresentados na Tabela 2.

Tabela 2 - Distribuição dos Trabalhos quanto aos Fins

Quanto aos Fins	2010		2012		Acumulado (2010 e 2012)	
	Freq. Abs.	Freq. Rel. (%)	Freq. Abs.	Freq. Rel. (%)	Freq. Abs.	Freq. Rel. (%)
Descritiva	2	7,41%	8	29,63%	10	37,04%
Explicativa	1	3,70%	-	-	1	3,70%
Exploratória	4	14,81%	4	14,81%	8	29,63%
Metodológica	3	11,11%	5	18,52%	8	29,63%
TOTAL	10	37,04%	17	62,96%	27	100,00%

Fonte: dados da pesquisa

A tabela ilustra o equilíbrio de trabalhos exploratórios e metodológicos e considerável crescimento dos estudos descritivos no período analisado (2010 e 2012). Pode-se inferir que, no primeiro ano analisado, parte significativa dos estudos é exploratória (14,81%), com objetos de estudo analisados de modo mais incipiente, enquanto que no segundo ano os trabalhos indicam estar em estágios mais avançados, com destaque aos estudos descritivos (29,63%).

Tabela 3 - Distribuição dos Trabalhos quanto aos Meios

Quanto aos Meios	2010		2012		Acumulado (2010 e 2012)	
	Freq. Abs.	Freq. Rel. (%)	Freq. Abs.	Freq. Rel. (%)	Freq. Abs.	Freq. Rel. (%)
Bibliográfica	6	22,22%	3	11,11%	9	33,33%
Documental	3	11,11%	-	-	3	11,11%
Estudo comparativo	-	-	1	3,70%	1	3,70%
Estudo Experimental	1	3,70%	2	7,41%	3	11,11%
Estudo de Caso	-	-	7	25,93%	7	25,93%
Estudos Métricos	-	-	3	11,11%	3	11,11%
Operacional	-	-	1	3,70%	1	3,70%
Total	10	37,04%	17	62,96%	27	100,00%

Fonte: dados da pesquisa

Em relação aos meios, o panorama das comunicações sinaliza a predominância de pesquisas bibliográficas nos dois eventos, como de estudos de caso no ano de 2012, período em que também se percebe a diversidade maior de percursos metodológicos dos pesquisadores na condução de seus trabalhos.

Sobre os enfoques dos trabalhos comunicados, bem como as técnicas e análises, as tabelas que seguem apresentam o panorama encontrado.

Tabela 4 - Distribuição dos Trabalhos quanto ao Enfoque*

Enfoque	2010		2012		Acumulado (2010 e 2012)	
	Freq. Abs.	Freq. Rel. (%)	Freq. Abs.	Freq. Rel. (%)	Freq. Abs.	Freq. Rel. (%)

* O total de trabalhos analisados no acumulado (2010 e 2012) foi artificialmente aumentado, pois enquadraram-se em mais de uma categoria.

Bibliometria	-	-	1	3,45%	1	3,45%
Cientometria	-	-	2	6,90%	2	6,90%
Crítico	-	-	1	3,45%	1	3,45%
Histórico	-	-	1	3,45%	1	3,45%
Linguístico	1	3,45%	-	-	1	3,45%
Reflexivo	2	6,90%	6	20,69%	8	27,59%
Teórico	7	24,14%	6	20,69%	13	44,83%
Não se Aplica	-	-	2	6,90%	2	6,90%
Total	10	34,48%	17	58,62%	29	100,00%

Fonte: dados da pesquisa

Constata-se uma variedade de enfoques nos trabalhos analisados, prevalecendo o teórico, evidenciando a preocupação com questões estruturais e constituintes das configurações epistemológicas da organização do conhecimento, que se forjam na medida em que se institucionaliza uma área de conhecimento. Por sua vez, o enfoque reflexivo sugere uma atenção ao movimento de tensionamento e retroalimentação constante entre teoria e prática na área.

Na tabela 5, verifica-se a técnica análise de políticas e a observação como as mais utilizadas pelos pesquisadores no período, indicando o cenário de mudanças trazidas pela complexidade dos conhecimentos disponíveis, relativos à organização do conhecimento. Nota-se ainda a significativa presença da análise de conteúdo enquanto instrumento para apresentação de resultados, revelando uma tradição na pesquisa brasileira, especialmente voltada aos estudos sobre *corpora* delimitados em função de objetos de pesquisa.

Tabela 5 - Distribuição dos Trabalhos quanto às Técnicas*

Técnicas	2010		2012		Acumulado (2010 e 2012)	
	Freq. Abs.	Freq. Rel. (%)	Freq. Abs.	Freq. Rel. (%)	Freq. Abs.	Freq. Rel. (%)
Análise de Políticas	6	21,43%	1	3,57%	7	25,00%
Análise de Projeto	-	-	2	7,14%	2	7,14%
Comparação	3	10,71%	2	7,14%	5	17,86%
Grupo Focal	-	-	1	3,57%	1	3,57%
Mineração	-	-	1	3,57%	1	3,57%
Observação	-	-	7	25,00%	7	25,00%
Observação Direta	-	-	1	3,57%	1	3,57%
Não se Aplica	1	3,57%	3	10,71%	4	14,29%
Total	10	35,71%	17	60,71%	28	100,00%

Fonte: dados da pesquisa

* O total de trabalhos analisados no acumulado (2010 e 2012) foi artificialmente aumentado, pois enquadraram-se em mais de uma categoria.

Na tabela 6, é possível observar que a análise de conteúdo foi a mais representativa no período de 2012, fato que pode estar relacionado não só ao aparecimento dos trabalhos sobre produção científica, mas também, como pode ser observado na Tabela 4, com a presença dos enfoques bibliométrico e cientométrico. Os pesquisadores se utilizam da análise de conteúdo, tanto em trabalhos sobre o tema quanto sob o enfoque, para discutir e contextualizar seus estudos.

Tabela 6 - Distribuição dos Trabalhos quanto às Análises*

Análises	2010		2012		Acumulado (2010 e 2012)	
	Freq. Abs.	Freq. Rel. (%)	Freq. Abs.	Freq. Rel. (%)	Freq. Abs.	Freq. Rel. (%)
Bibliográfica	-	-	1	3,57%	1	3,57%
Citação	-	-	2	7,14%	2	7,14%
Conteúdo	2	7,14%	9	32,14%	11	39,29%
Contexto	-	-	1	3,57%	1	3,57%
Discurso	1	3,57%	1	3,57%	2	7,14%
Domínio	-	-	2	7,14%	2	7,14%
Para-análise	-	-	1	3,57%	1	3,57%
Tarefas	-	-	1	3,57%	1	3,57%
Não se Aplica	7	25,00%	-	-	7	25,00%
Total	10	35,71%	17	60,71%	28	100,00%

Fonte: dados da pesquisa

Do conjunto apresentado nas tabelas 4, 5 e 6, destacam-se os seguintes aspectos: os enfoques bibliométrico e cientométrico na segunda edição do encontro, para apresentação do processo de produção do conhecimento no recorte estudado; a técnica de observação como a mais recorrente no ano de 2012, indicando que as atividades de pesquisa estão orientadas para análise de determinados fenômenos, sujeitos e realidades.

4 Considerações Finais

Enquanto subárea da Ciência da Informação, a Organização do Conhecimento configura-se como um dos segmentos mais consolidados. Essa sedimentação resulta dos esforços de pesquisa de comunidades que se organizam para fins de avaliação do estágio do conhecimento alcançado, baseada na crítica científica. Entre os elementos mais significativos de institucionalização científica, as associações e os eventos, no caso específico a International Society for Knowledge Organization e os Advances in Knowledge Organization.

* Idem.

Com base nos anais analisados, verifica-se que os estudos apresentam equilíbrio nos níveis temático e metodológico. Constata-se também o predomínio dos estudos descritivos, revelando um nível de aprofundamento que transcende ao dos estudos exploratórios. Nota-se ainda a pluralidade de técnicas e análises para apresentação de resultados, demonstrando as possibilidades reais frente aos dados e às informações coletadas, bem como à natureza e aos objetivos das investigações.

Referências

BARDIN, L. **Análise de conteúdo**. São Paulo: Edições 70, 1977.

BUFREM, L. S.; PRATES, Y. O saber científico registrado e as práticas de mensuração da informação. **Ciência da informação**, Brasília, v. 34, n. 2, p. 9-25, maio/ago, 2005.

SMIRAGLIA, R. P. Universes, Dimensions, Domains, Intensions and Extensions: knowledge organization for the 21st century. In: NEELAMENGHAN, A.; RAGHAVAN, K. S. (Ed.). **Categories, contexts and relations in Knowledge Organization**. Würzburg: Ergon Verlag, 2012. p. 1-7.

WHITLEY, R. Cognitive and social institutionalization of scientific specialities and research areas. In: WHITLEY, R. (Ed.). **Social processes of scientific development**. London: Routledge and Kegan, 1974. p. 69-95.

Visibilidade da produção científica sobre metateoria: Análise bibliométrica no período entre 2000 a 2012.

Ely Francina Tannuri de Oliveira
Universidade Estadual Paulista
etannuri@marilia.unesp.br

Bruno Henrique Alves
Universidade Estadual Paulista
bruninkmkt@hotmail.com

1 Introdução

A metateoria pode ser considerada um processo de reflexão da teoria sobre si própria, de forma a teorizar sobre a natureza da própria teoria em uma determinada ciência. Considera-se que, o conhecer a produção científica neste tema, de fundamental importância para entendimento do *domínio* na temática metateoria.

Nesse sentido, esta pesquisa objetiva analisar a produção científica, na tipologia artigo, sobre “Metateoria”, no período de 2000 a 2012. Os objetivos específicos são: identificar os pesquisadores mais produtivos, as temáticas que mais têm contribuído para desenvolvimento do saber no tema e analisar a coautoria entre os pesquisadores.

60

2 Metateoria, Organização Do Conhecimento: Uma Análise Bibliométrica.

A metateoria esta relacionada com a epistemologia que trata da teoria do conhecimento. Para (Gomez, 2001: 6), a “epistemologia seria o estudo das possibilidades, origem, natureza e extensão do conhecimento humano”. Para Tennis (2008, tradução nossa), a epistemologia é a forma que se conhecer algum objeto. Em Organização do conhecimento (KO), preparam-se declarações implícitas epistêmicas sobre o conhecimento de conceitos, atos (como as representações), entidades e sistemas. Algumas questões centrais são: a origem do conhecimento; o lugar da experiência e da razão na gênese do conhecimento; a relação entre o conhecimento e a certeza, e entre o conhecimento e a impossibilidade do erro; a possibilidade do ceticismo universal; e as formas de conhecimento que emergem das novas conceitualizações do mundo (Blackburn, 1997).

Hjørland (2008) considera a perspectiva do ambiente institucional e a concepção do conhecimento registrado. Distingue a Organização do Conhecimento (OC) em sentido específico, como a organização cognitiva do conhecimento e, em sentido amplo, como a organização social do conhecimento. Nessa distinção, considera que, em sentido específico, a Biblioteconomia e a Ciência da Informação (CI) são as disciplinas centrais da OC e compreende atividades, tais como “[...] descrição documentária, indexação e classificação”. (Hjørland, 2008: 86).

Tennis (2008) destaca que as estruturas de pesquisa em Organização da Informação (OI) são resultados da investigação metateórica. Ainda, de acordo com o autor, a metateoria subsidia quatro propósitos apontados pelo autor: inicialmente uma compreensão mais aprofundada do trabalho teórico; em seguida fornece uma perspectiva mais abrangente e importante deste trabalho; em terceiro lugar serve como mecanismo para a avaliação e por último serve como um prenúncio ao trabalho teórico futuro (Ritzer, 1991a, 1991b).

Processo paralelo à socialização do conhecimento, a avaliação constante da atividade científica, torna-se necessária nas diferentes áreas do conhecimento. A “avaliação da atividade científica representa um processo fundamental em países onde a ciência é financiada majoritariamente por investimentos públicos” (Vanz, 2009: 17), criando-se a necessidade da construção de procedimentos metodológicos para avaliação da atividade científica.

Para melhor contextualizar essa questão, Oliveira e Grácio (2011: 18) destacam que “as análises bibliométricas têm se mostrado procedimentos tangíveis e confiáveis, utilizam indicadores de produção, ligação e citação, explicitando, além da produtividade, a relevância e impacto de autores”.

A colaboração científica entre autores ou instituições supõe uma consociação de hipóteses e objetivos centrais de um projeto. (Olmeda Gómez; Perianes-Rodriguez; Ovalle-Perandones, 2008).

Wasserman e Faust (1994) afirmam que o termo “rede social” se refere ao conjunto de “atores” e suas ligações. A análise de rede tem por objetivo explicitar as relações entre os atores, a fim de descrever e visualizar a estrutura de um determinado grupo.

3 Metodologia

Como procedimento de pesquisa, realizou-se um levantamento na base *Scopus*, no dia 5 de fevereiro de 2012, utilizando-se como termo de busca a expressão “*Metatheory*”, no subcampo “*Articles title, Abstract and Keywords*”, no período de 2000 a 2012. Aplicou-se o filtro “*Articles*” na subárea do conhecimento “*Social Sciences & Humanities*”.

Encontraram-se 109 artigos científicos que foram analisados um a um, por meio do título, resumo e palavras-chave, a fim de se ratificar a pertinência dos mesmos na temática “*Metatheory*”, resultando 106 artigos, constituindo-se este o universo de pesquisa, a partir do qual foram identificados os autores mais produtivos e as temáticas.

Para análise de rede de colaboração entre os autores, foram considerados todos os autores, dos 142 presentes, que trabalharam em coautoria, a fim de se verificar a abrangência da temática no contexto mundial. Utilizou o *Software Ucinet* para gerar a rede de colaboração entre os autores.

4 Apresentação E Análise Dos Dados

Em relação ao total dos 142 pesquisadores, encontrou-se 9 autores mais produtivos, com pelo menos 2 artigos publicados, representados na Tabela 1.

Tabela 1 – Autores mais produtivos encontrados na base *Scopus* sobre “Metateoria”.

Autores mais produtivos e sua afiliação	Nº de artigos
Gergen, K.J. (Swarthmore College - United States)	3
Anchin, J.C. (University of New York - United States)	2
Conrey, F.R. (Macro International - United States)	2
Dulany, D.E. (University of Illinois - United States)	2
Hibberd, F.J. (University of Sydney - Australia)	2
Käpylä, J. (University of Tampere - Finland)	2
Mikkola, H. (Ulkopoliittinen instituutti - Finland)	2
Smith, E.R. (Indiana University - United States)	2
Weiner, J. (Indiana University - United States)	2

Fonte: Elaborado pelos autores.

Em destaque, o pesquisador Gergen, K. J., com três artigos, ocupa a primeira posição entre os pesquisadores mais produtivos.

A dominância de afiliação dos autores mais produtivos é estrangeira, ocorrendo maior concentração, nos seguintes países: *United States, Australia e Finland*. Nesse

sentido, destaca-se que os *United States* é um dos principais países que desenvolvem pesquisas mais rigorosas na temática em estudo, considerando seu contexto histórico.

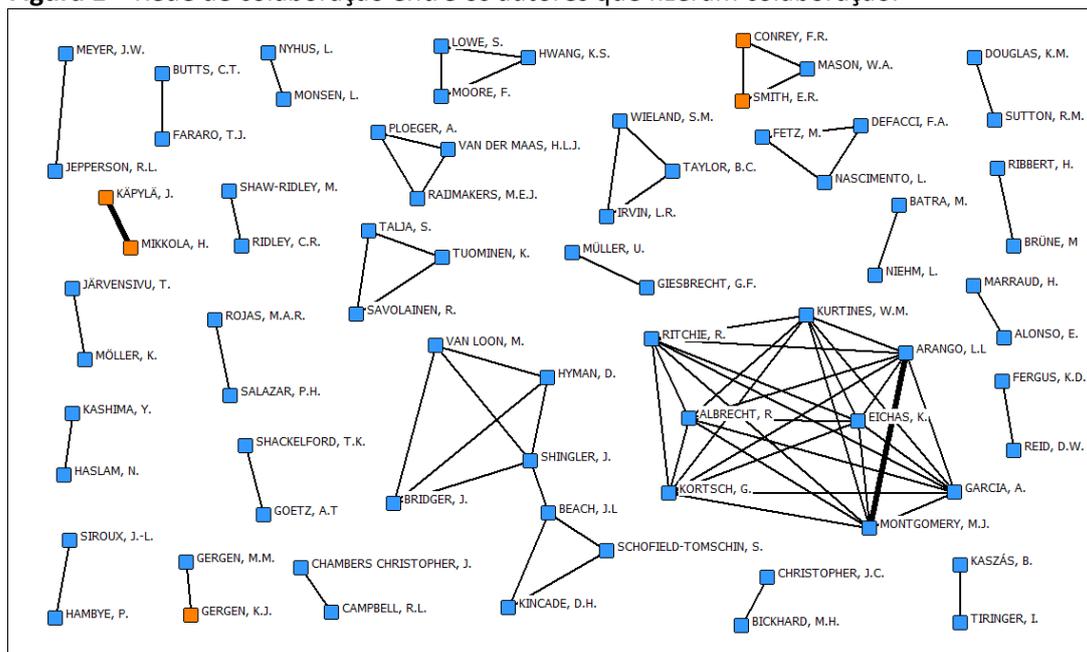
Quanto às temáticas mais presentes levantadas a partir das palavras-chave dos artigos destacam-se em ordem decrescente de frequência: *Metatheory*, *Critical realism*, *Philosophy of Science*, *Psychotherapy Integration*, *Constructivism*, *Economic Growth Theory* e *Epistemology*, entre outras.

As temáticas destacadas podem ser analisadas partindo-se das relações que existem em comum entre si. Cite-se como exemplo, as temáticas *Critical realismo* e a *Philosophy of Science*, temas que são desenvolvidos nos contextos cientificistas, materialistas e positivistas que contribuíram para o desenvolvimento teórico da ciência.

Nesse sentido, pode-se também relacionar o *Constructivism e Epistemology*, partindo da construção e contextualização do conhecimento considerando sua origem.

A seguir, apresenta-se a rede de colaboração entre os autores.

Figura 1 – Rede de colaboração entre os autores que fizeram colaboração.



Fonte: Elaborado pelos autores.

Observa-se na rede de colaboração científica apresentada na Figura 2, algumas “Comunidades”. Segundo Capra (2002: 119) “cada comunidade gera pensamentos e um significado, os quais dão origem a novas comunicações”.

Observa-se que os autores mais produtivos representados na rede estão na cor laranja, e seus respectivos coautores, em azul. A densidade da rede, isto é, se ela tem em

sua totalidade, maior ou menor número de conexões, é expressa pelo valor 2,9%, significando uma rede de baixa densidade.

Destaca-se uma “comunidade”, com presença de dois dos autores mais produtivos, que são Conrey, F.R., com a seguinte afiliação: *Macro International - United States* e Smith, E.R., com a seguinte afiliação: *Indiana University - United States*.

Na rede estão apresentados cinco dos autores mais produtivos, pois os demais quatro autores não apresentaram coautorias no período em estudo, que sinaliza a pouca dialogicidade dos autores mais produtivos. Dos 142 autores encontrados, 73 fizeram colaboração.

5 CONCLUSÕES

Destaca-se a significativa quantidade de autores estrangeiros que trabalharam temas de pesquisas voltadas a “Metateoria”, advindos majoritariamente dos *United States*, especialmente com temáticas que se relacionam entre si, no contexto da origem e desenvolvimento teórico da ciência.

A rede de colaboração entre autores, embora frágil em sua densidade, já mostra um pequeno grupo de pesquisadores dialogando na referida temática.

Considerando-se a relevância do tema, recomenda-se a extensão desta pesquisa para outras bases de dados relevantes da área e demais tipologias documentais para que se possa ter um panorama mais completo e amplo da realidade do tema em estudo.

REFERÊNCIAS

Blackburn, S. 1997. *Dicionário Oxford de Filosofia*. Tradução de Desidério Murcho et al. Rio de Janeiro: Zahar.

Capra, F. 2002. *As conexões ocultas: ciência para uma vida sustentável*. 2 ed. São Paulo: Cultrix.

Gomez, M. N. G. 2001. *Para uma reflexão epistemológica acerca da ciência da informação*. Perspectivas em Ciência da Informação, Belo Horizonte, 6: 5-18. Disponível em: <http://portaldeperiodicos.eci.ufmg.br/index.php/pci/article/view/433> Acesso em: 14 fev. 2013.

Hjørland, B. 2008. *What is knowledge organization (KO)?* Knowledge Organization, 35: 86-101.

Oliveira, E. F. T.; Grácio, M. C. C. 2011. *Indicadores bibliométricos em ciência da informação: análise dos pesquisadores mais produtivos no tema estudos métricos na base Scopus*. Perspectivas em Ciência da Informação, 16: 16-28. Disponível em: <http://portaldeperiodicos.eci.ufmg.br/index.php/pci/article/view/1299> Acesso em: 14 fev. 2013.

Olmeda Gómez, C.; Perianes-Rodríguez, A.; Ovalle-Perandones, M. A. 2008. *Estructura de las redes de colaboración científica entre las universidades españolas*. Ibersid: revista de sistemas de información e comunicación, p.129-140. Disponível em: http://eprints.rclis.org/bitstream/10760/3955/1/ibersid2008_ok.pdf Acesso em: 14 fev. 2013.

Ritzer, G. 1991a. *Metatheorizing in sociology*. Lexington, MA: Lexington Books.

Ritzer, G. 1991b. *Reflections on the rise of metatheorizing in sociology*. *Sociological perspectives* 34 : 237-248.

Tennis, J. T. 2008. *Epistemology, theory and methodology in Knowledge Organization: toward a classification, metatheory, and research framework*. *Knowledge Organization*, 35: 102-112.

Vanz, S.A.S. 2009. *As redes de colaboração no Brasil (2004-2006)*. Tese (Doutorado em Comunicação e Informação) – Faculdade de Biblioteconomia e Comunicação, Universidade Federal do Rio Grande do Sul, Porto Alegre, p. 204.

Wasserman, S.; Faust, K. 1994. *Social network analysis: methods and applications*. Cambridge: Cambridge University Press.

**A produção científica na temática indexação:
Análise bibliométrica de teses e dissertações no período de 2002 a 2011**

Paula Regina Dal'Evedove
Universidade Estadual Paulista
p.dallevedove@gmail.com

Roberta Cristina Dal' Evedove Tartarotti
Universidade Federal de São Carlos
roberta_tartarotti@yahoo.com.br

Mariângela Spotti Lopes Fujita
Universidade Estadual Paulista
mariangelaslf57@gmail.com

Vera Regina Casari Boccato
Universidade Federal de São Carlos
vboccato@ufscar.br

1 Introdução

Estudos nacionais e internacionais têm se pautado na abordagem bibliométrica no intuito de relatar as pesquisas nas diversas áreas do conhecimento e permitir a avaliação da produção científica de modo objetivo e confiável, que, associada às análises contextuais, constituem um dos instrumentos metodológicos que contribuem para a visualização do comportamento da ciência em um dado campo (Danuello e Oliveira, 2012:67; Tsay, 2011:2).

Visando uma compreensão qualitativa da produção científica em Organização e Representação do Conhecimento, diversas são as pesquisas que se debruçam na análise e interpretação de dados quantitativos, utilizando os resultados para projetar o cenário atual, propor novas conjecturas e apoiar tomadas de decisão no campo da ciência.

Considerando-se a relevância da observação da área de Organização e Representação do Conhecimento em todo o seu entorno científico, a presente pesquisa teve como objetivo realizar uma análise bibliométrica de teses e dissertações provenientes de Programas de Pós-Graduação em Ciência da Informação do Brasil na temática indexação a partir do Banco de Teses da CAPES - Coordenação de Aperfeiçoamento de Pessoal de Nível Superior.

Embora as três abordagens teórico-metodológicas do Tratamento Temático da Informação (*subject cataloguing*, *indexing* e *analyse documentaire*) serem, por vezes,

interligadas no que tange ao arcabouço teórico-conceitual e prático-aplicado (Guimarães, 2008), optou-se por investigar apenas a indexação com o propósito de tornar a pesquisa possível operacionalmente. Ademais, este recorte deve-se ao número limitado de estudos destinados à análise da produção científica desta temática e pela relevância da mesma na área de Organização e Representação do Conhecimento.

Conceitualmente, a indexação consiste em um processo formado por subprocessos ou etapas que tem como objetivo identificar o conteúdo de um documento, por meio de uma metalinguagem construída – a linguagem documental - no intuito de promover a efetiva recuperação da informação. Em termos teóricos, qualquer objeto é passível de ser indexado ou representado por meio de conceitos em busca da satisfação de necessidades específicas de informação. Embora não consensuais na literatura, tem-se como principais etapas da indexação: leitura documental; análise de assunto ou identificação de conceitos; seleção de conceitos e tradução de conceitos (Gil Leiva, 2008:64; Tartarotti e Boccato, 2013:40).

A definição do *corpus* de análise – exclusivamente teses e dissertações –, decorre da representatividade deste tipo de documento científico, pois demonstra a preocupação dos pesquisadores sobre a temática e, sobretudo, pelas diferentes perspectivas de observação que podem ser trabalhadas neste suporte de informação, fato que amplia as análises e inferências sobre as configurações de seus campos de estudo.

2 Procedimentos Metodológicos

A metodologia adotada é de natureza exploratória e descritiva com abordagem bibliométrica para o levantamento dos indicadores da produção científica na temática indexação. O percurso investigativo desenrolou-se a partir da seguinte trilha metodológica: a) construção do referencial teórico por meio da leitura especializada na área de Organização e Representação do Conhecimento; b) escolha da base de dados como fonte de pesquisa para a seleção do material bibliográfico: Banco de Teses da CAPES, no qual se utilizou do resumo como ferramenta de busca e consulta às informações sobre teses e dissertações defendidas junto aos Programas de Pós-Graduação em Ciência da Informação do país, sendo estes responsáveis pela veracidade dos dados, pois fornecem as informações diretamente à CAPES; c) definição do período de busca: 2002 a 2011; d) seleção do material bibliográfico.

As expressões de busca utilizadas para a coleta dos registros foram: indexação, indexador(es), linguagem(ns) de indexação e política(s) de indexação. Para a análise da

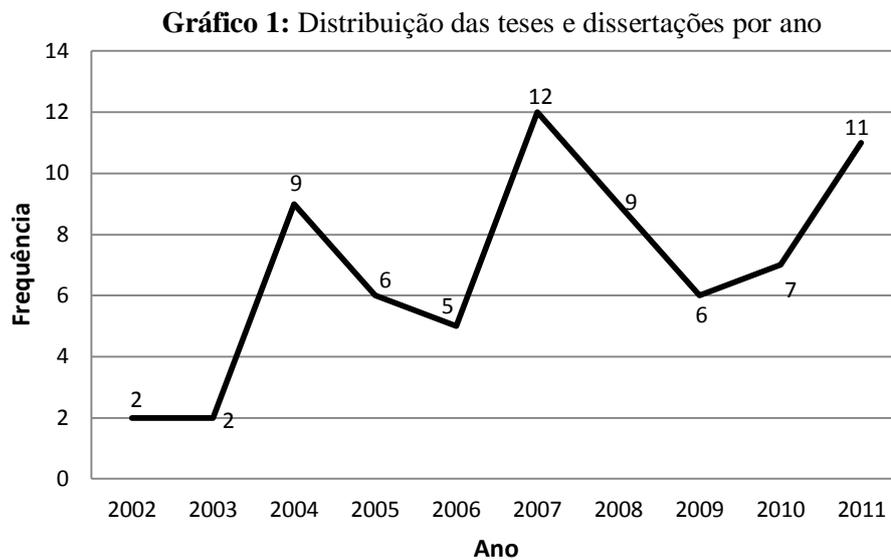
produção científica sobre indexação oriunda de teses e dissertações, foram definidos os seguintes parâmetros bibliométricos: autoria, ano de defesa do documento, origem geográfica e institucional, nível acadêmico (doutorado, mestrado e mestrado profissional) e palavras-chave atribuídas pelos autores das teses e dissertações para verificar os principais focos de investigação.

Os dados obtidos foram sistematizados com o objetivo de gerar indicadores da produção científica em indexação, visando agregar valor à informação. Para o tratamento estatístico e representação gráfica dos dados, utilizou-se o *software Microsoft Office Excel 2007*.

3 Apresentação e análise dos dados

Foram recuperados 103 registros no Banco de Teses da CAPES na temática indexação, referentes a um período de 10 anos (2002 a 2011). Para cada registro, foram efetuadas leituras dos títulos, palavras-chave e resumo, com o propósito de verificar se de fato a tese ou dissertação tratava do tema indexação no campo científico da Ciência da Informação. Como resultado, obteve-se 69 registros, distribuídos em 47 (68%) dissertações de mestrado, 20 (29%) teses de doutorado e 2 (3%) dissertações de mestrado profissional, com predominância de autoria do sexo feminino, perfazendo um total de 45 documentos ou 65% dos documentos recuperados. As autorias do sexo masculino correspondem a 24 registros ou 35%.

De acordo com o portal CAPES, o Sistema de Avaliação dos Programas de Pós-graduação, implantado em 1976, vem cumprindo papel fundamental para o desenvolvimento da pós-graduação e da pesquisa científica e tecnológica no Brasil ao estabelecer, dentre outros objetivos, o padrão de qualidade exigido dos cursos de mestrado e de doutorado, identificando os cursos que atendem a tal padrão (CAPES, 2013). Neste cenário, pode-se inferir que a presença de um número maior de dissertações decorre do fato da existência de mais vagas para os cursos de mestrado do que de doutorado em Ciência da informação no país, gerando um maior número de dissertações em comparação com o número de teses defendidas nos programas de pós-graduação. Observa-se, a seguir, a distribuição anual destes 69 registros:

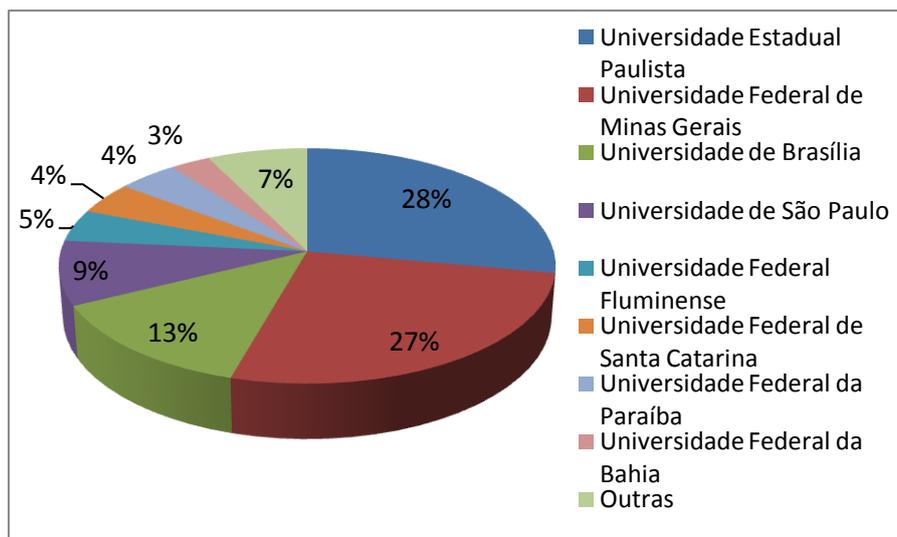


Fonte: elaboração própria.

O Gráfico 1 representa um aumento acentuado na produção científica sobre indexação a partir do ano de 2004, no qual foram identificados 9 (13%) registros, número expressivo quando comparado aos dois anos anteriores (2002 e 2003), com 2 registros cada ou 6% do total. Nota-se uma média de 6,9 teses e dissertações desenvolvidas por ano no Brasil na temática, com destaque para o ano de 2007, com 12 (17%) registros no período de dez anos (2002 a 2011).

Foram identificadas 14 diferentes instituições que produziram conhecimento na temática no período investigado (Gráfico 2). A maior concentração dos registros ocorreu na Universidade Estadual Paulista (19), seguida pela Universidade Federal de Minas Gerais (18), Universidade de Brasília (9) e Universidade de São Paulo (6), que juntas reúnem 52 ou 75% do total dos registros identificados. Por sua vez, a Universidade Federal Fluminense, a Universidade Federal de Santa Catarina e a Universidade Federal da Paraíba aparecem com três registros cada, seguidas pela Universidade Federal da Bahia com dois registros. Outras cinco instituições aparecem com um registro cada e representam 7% do total:

Gráfico 2: Distribuição das teses e dissertações por Instituição



Fonte: elaboração própria.

Quanto à distribuição das regiões geográficas das instituições de ensino superior pelo Brasil, a Região Sudeste se destacou com sete instituições, seguida pelas regiões Centro-Oeste, Nordeste e Sul com duas instituições cada, enquanto que a Região Norte possui apenas uma instituição. Este resultado indica uma concentração da produção de teses e dissertações em indexação na região Sudeste do país, decorrente de uma maior representatividade de Programas de Pós-graduação em Ciência da Informação nesta região.

Com relação aos temas mais trabalhados nas teses e dissertações envolvendo o contexto científico da indexação, foram identificadas 169 palavras-chave atribuídas pelos autores, com frequência de aparecimento de uma a 17 ocorrências. Desse modo, as 24 palavras-chave mais abordadas nos estudos, com ocorrência de no mínimo 3 vezes, são apresentadas a seguir:

Tabela 1: Frequência das palavras-chave das teses e dissertações na temática indexação

PALAVRAS-CHAVE	FREQUÊNCIA
Indexação	17
Linguagem documental	9
Indexação automática	8
Organização e Representação do Conhecimento	6
Processamento de Linguagem Natural	6
Leitura documental	6
Necessidades de informação	5
Organização da Informação	5
Representação da Informação	5
Taxonomia	5
Tesouro	5
<i>Folksonomia</i>	4

Fotografia	4
Modelos de busca	4
Ontologia	4
Protocolo Verbal	4
Recuperação da informação	4
Análise de assunto	3
Avaliação de linguagem documental	3
Bibliotecas universitárias	3
Classificação facetada	3
Indexação de imagens	3
Indexação manual	3
Representação temática	3

Fonte: elaboração própria.

A Tabela 1 apresenta uma frequência considerável de palavras-chave relativas aos processos (leitura documental, análise de assunto e representação temática) e instrumentos (linguagem documental, taxonomia, tesouro e ontologia) da indexação no universo científico do Tratamento Temático da Informação. Outro fator que merece destaque é a presença do termo “bibliotecas universitárias”, fato que demonstra preocupação da área com o processo desenvolvido em contextos especializados. Pode-se inferir, também, que a frequência das palavras-chave “indexação automática”, “*folksonomia*” e “taxonomia” indica uma tendência nas pesquisas em indexação sob a ótica da Organização e Recuperação do Conhecimento na web.

4 Considerações Finais

Devido à relevância de investigações bibliométricas no âmbito da Organização e Representação do Conhecimento, recomenda-se a extensão deste tipo de pesquisa, abarcando outras tipologias documentais como livros, artigos de periódicos e demais formas de comunicação científica que visem divulgar e dar visibilidade à produção científica da área, especificamente sobre indexação no campo da informação, a fim de traçar um panorama mais amplo e completo da realidade desta temática.

REFERÊNCIAS

Avaliação da Pós-graduação. 2013. *Portal CAPES*. Disponível em: <<http://www.capes.gov.br/avaliacao/avaliacao-da-pos-graduacao>>. Acesso em: 21 fev. 2013.

Danuello, J. C.; Oliveira, E. F. T. de. 2012. Análise cientométrica: produção científica e redes colaborativas a partir das publicações dos docentes dos programas de pós-graduação em Fonoaudiologia no Brasil. *Em Questão*, Porto Alegre, 18:65-79. Disponível em: <<http://seer.ufrgs.br/EmQuestao/article/view/33178>>. Acesso em: 04 mar. 2013.

Gil Leiva, I. 2008. Manual de indización: teoría y práctica. Gijón: Trea.

Guimarães, J. A. C. 2008. A dimensão teórica do tratamento temático da informação e suas interlocuções com o universo científico da International Society for Knowledge Organization (ISKO). *Revista Ibero-Americana de Ciência da Informação*, 1:77-99. Disponível em: <<http://seer.bce.unb.br/index.php/RICI/article/viewArticle/2761>>. Acesso em: 03 mar. 2013.

Tartarotti, R. C. D.; Boccato, V. R. C. 2013. A abordagem teórica bakhtiniana e o processo de indexação: diálogos. *Versão Beta: sob o signo da palavra*, 9:33-47.

Tsay, M. Y. 2011. A bibliometric analysis on the Journal of Information Science. *Journal of Library and Information Science*, 5:2:1-28. Disponível em: <<http://lac3.glis.ntnu.edu.tw/vj-attachment/2011/07/attach69.pdf>>. Acesso em: 05 mar. 2013.

Mapeamento conceitual sob o enfoque da análise de domínio: Uma discussão de conceitos fundamentais

Rosana Portugal Tavares de Moraes
Universidade Federal Rural do Rio de Janeiro
rosanabiblio@gmail.com

Maria Luiza de Almeida Campos
Universidade Federal Fluminense
marialuizalmeida@gmail.com

Introdução

A área de Organização e Representação do Conhecimento representada pelo GT2 do ENANCIB (Encontro Nacional de Pesquisa em Ciência da Informação e Biblioteconomia) tem se destacado pelo número de comunicações apresentadas, são 282 no período compreendido entre 1994 e 2012. As temáticas produzidas nessas comunicações têm muito a informar sobre como as vertentes conceituais se desenvolveram, quais são mais representativas e quais não são consideradas relevantes em determinados períodos, dentre outros.

Nesse sentido, o estudo se desenvolve como uma proposta de investigação nas bases teóricas e metodológicas da Análise de Domínio visando à elaboração de um mapeamento temático para o campo de Organização e Representação do Conhecimento. A Análise de Domínio fornece apoio metodológico ao mapeamento temático, pois identifica o domínio como uma comunidade discursiva com especial atenção ao contexto onde as trocas informacionais acontecem. Como aporte teórico buscou-se na literatura embasamento conceitual para Contexto, Domínio e Comunidade Discursiva; uma vez que, entende-se que a definição desses termos é essencial para a pesquisa.

73

Análise de Domínio

Na Ciência da Informação os estudos de Análise de Domínio foram sistematizados por Birger Hjørland e Hanne Albrechtsen, em 1995, como um paradigma na área, afirmando que: “[...] a melhor forma de entender a informação é através do estudo de domínios do conhecimento através das comunidades discursivas, que são partes da divisão de trabalho na sociedade” (1995, p. 400, tradução nossa).

Segundo os autores a Análise de Domínio é uma nova abordagem na Ciência da Informação, denominada como sócio cognitivista, em contraposição ao paradigma

cognitivo com suas preocupações que se colocam no espaço de um só indivíduo, onde a possibilidade de existência se coloca a partir de uma mente que interpreta. Estes espaços cognitivos ou mentais são discutidos por Popper (1999) como a Teoria da Mente Objetiva. A perspectiva da Análise de Domínio é diferente, pois percebe a informação como algo construído socialmente, através da interação do usuário com o seu específico contexto de informação, em um espaço discursivo formado pelas principais temáticas de assunto, em seus menores ou mais detalhados ambientes de uso.

A ênfase da visão de Análise de Domínio está no tratamento do assunto do documento visando sua recuperação. Hjørland comenta que conceitos devem ser estudados em domínios e disciplinas, relacionando, entre outros, a teorias e paradigmas desses domínios. (2009, p. 1520) Isso se justifica, diante da defesa que o autor faz, declarando que será uma maior contribuição à pesquisa do usuário, quando este é informado sobre qual perspectiva teórica trata determinado documento.

Hjørland (2009, p. 1527) evidencia que o entendimento do conceito de um documento vem sendo tratado pela Ciência da Informação de acordo com quatro principais visões epistemológicas: *empiricista*, - define conceitos agrupando assuntos similares; *racionalista* - define conceitos de um conjunto primitivo de conceitos; *historicista* - define conceitos genealogicamente e explicita suas relações entre teorias e discursos; *pragmatista* - define conceitos decidindo que classes de coisas melhor servem a um dado propósito, e então fixa essa classe em um signo.

Segundo o mesmo autor predomina a teoria empírica e racionalista na forma de tratar assuntos na Ciência da Informação. A abordagem da Análise de Domínio está apoiada na visão histórica e pragmática por considerar os aspectos culturais, sociais e históricos tanto do contexto de produção do documento como do usuário deste documento.

Para o autor (Id., 2009, p. 1527), “[...] a definição de termos e suas relações semânticas são relativas ao contexto.” Ressalta que a Organização do Conhecimento deveria fornecer métodos e diretrizes para definição do contexto. Nesse sentido, justifica-se a busca por essa definição, como também dos conceitos de domínio e comunidade discursiva, apresentados aqui de forma bem resumida.

Contexto

Ao introduzir o tema contexto, Dervin (1997, p. 13) começa afirmando que existe uma boa e uma má notícia, ela identifica o assunto como “quente” e crescente na

literatura de ciências sociais e humanas; o que é uma boa notícia. A parte ruim consiste em que, a própria questão acaba por ser embaraçosa e leva a uma busca que exige tolerância extraordinária do caos. Nesse sentido, afirma que diante do esforço em revisar o tratamento dado na literatura, conclui que - a única possível conclusão reside no fato de que - não existe: termo mais frequentemente usado, menos frequentemente definido, e quando definido, definido de variadas formas.

Diante desse quadro, buscou-se o tratamento conceitual para contexto em artigo de revisão de Courtright (2007), sob o ponto de vista de necessidade, busca e uso de informação e a teoria de contexto de Van Dijk (2001, 2012). Diante das abordagens dos dois autores, nota-se, como era de se esperar, diferenças no que é observado como contexto. Os estudos de Courtright (2007) com ênfase nas práticas informacionais, relatam que contexto está ligado ao espaço físico ou a um conjunto de fatores e as interações e influências trocadas entre o usuário e esse espaço. O enfoque está na análise do comportamento do usuário, suas motivações e necessidades envolvidas com o ambiente informacional.

A proposta de Van Dijk é diferenciada, pois para ele contexto em uma perspectiva social comunicativa “[...] não é algo externo ou visível, ou está fora dos participantes, mas algo que os participantes constroem como representação mental.” (2001, p. 72) O contexto é construído subjetivamente pelo participante, através de suas interpretações, como experiência única, com linguagem e discurso também únicos. Cada pessoa tem um modelo de contexto com base nas influências e interações com o espaço social, político e cultural, ou também conhecido como, experiências cotidianas. “O modelo de contexto não representa todos os aspectos pessoais ou sociais da situação comunicativa, mas somente os aspectos, que em um dado momento, são relevantes para cada participante.” (VAN DIJK, 2001, p. 72, tradução nossa)

O estudo de contexto é essencial para pesquisa em Análise de Domínio, o contexto fornece informações sobre as relações semânticas entre os termos, a palavra e seus múltiplos sentidos, as quais estão imersas em práticas sociais e culturais.

Domínio

Hjorland define domínio como: “[...] uma disciplina científica ou um campo acadêmico, também pode ser uma comunidade de discurso ligada a um partido político, uma religião, um comércio ou um hobby.” (HJORLAND, 2005, p. 131, tradução nossa)

Mai (2005, p. 606) considera que um domínio fora do ambiente acadêmico e científico pode ser definido através das metas e objetivos de uma organização onde as estratégias de indexação são formuladas, como por exemplo, uma empresa farmacêutica ou um arquivo de filmes.

Hjorland e Hartel (2003, p. 239) explicam que os domínios são basicamente constituídos por três tipos de teorias e conceitos: a dimensão ontológica com destaque para explicações dos seguintes termos: áreas, domínios, elementos, campos, tipos, objetos, problemas, assunto (substância/diferença substancial) e tópicos; a dimensão epistemológica, sendo melhor identificada pelos termos: abordagens, metateorias, movimentos, paradigmas, filosofias, escolas de pensamento, entre outros; e a dimensão sociológica com expressões nos seguintes conceitos: disciplinas, subdisciplinas, comunidades discursivas, comunidades epistêmicas, profissões, especialidades, entre outros. Para os autores a principal forma de definir um domínio é atentando para os diferentes paradigmas atuantes no campo. Eles asseguram que: “[...] uma definição será sempre mais relacionada para uma visão ou paradigma, e relativamente insatisfatória para outros paradigmas.” A análise das vertentes conceituais atuantes no domínio pode ser considerada uma das formas de delimitá-lo.

Comunidade discursiva

Comunidade discursiva é um termo relatado nos textos sobre Análise de Domínio, contudo não se encontra uma definição clara do que vem a ser uma comunidade discursiva, como é formada, como se organiza, quais suas características e quais seus limites. Na busca desse entendimento, Swales (1990 apud BIASI-RODRIGUES; HEMAIS; ARAÚJO, 2009, p. 23) sob o enfoque dos estudos de gêneros textuais, considera comunidades discursivas como: “[...] relacionada à produção de textos como uma atividade social que se realiza de acordo com convenções discursivas específicas e revela o comportamento social e o conhecimento dos membros do grupo.” O entendimento de um domínio acontece através do estudo da comunidade discursiva, como é o seu processo de comunicação, sob quais critérios se constituiu, como se desenvolveu e quais fatores a influenciam ou chegam a alterá-la. Para estudar uma comunidade discursiva, Saldanha (2011, p. 63), comenta da necessidade de conhecer e conviver com seus jogos de linguagem, os critérios utilizados para construir e compartilhar seus saberes. Devem ser priorizados estudos dos canais de comunicação utilizados pela comunidade, com ênfase tanto em sua forma como em seu conteúdo.

Percebe-se que, qualquer estrutura informacional que careça de tratamento com objetivo de recuperação deve ser concebida e analisada sob o ponto de vista de uma comunidade discursiva. Esta pode ser científica, acadêmica ou profissional, com comunicações formais ou informais, mas que devem ter as formas de representar o conhecimento analisadas sob o aspecto social, cultural e histórico, inerente aos seus produtos de informação.

Considerações finais

O estudo da área de Organização e Representação do Conhecimento, sob a perspectiva da Análise de Domínio, contribui para que se estabeleça um olhar sobre a construção teórica da área respeitando suas particularidades históricas, sociais e culturais.

A possibilidade de conhecer o teor do discurso produzido nesse período tem informações valiosas para Ciência da Informação, proporciona uma visão e uma possível avaliação retrospectiva nos fundamentos da área, o que também permite um direcionamento mais consciente de suas temáticas de pesquisa.

Referências

Biasi-Rodrigues, B.; Araújo, J. C.; Sousa, S. C. T. de. 2009. Análise de gêneros na abordagem de Swales: princípios teóricos e metodológicos. In _____. (Org.) *Gêneros textuais e comunidades discursivas: um diálogo com John Swales*. Belo Horizonte: Autêntica, 17-32

Courtright, Christina. 2007. Context in information behavior research. *Annual Review of Information Science and Technology*, 41(1), 273-306.

Dervin, Brenda. 1997. Given a context by any other name: Methodological tools for taming the unruly beast. In *Information seeking in context*. Ed. by Pertti Vakkari, R. Savolainen, Brenda Dervin. London: Taylor Graham, 13-38.

Hjørland, B.; Albrechtsen, H. 1995. Toward a new horizon in information science: domain-analysis. *Journal of the American Society for Information Science*, 46(2), 400-425.

Hjørland, B.; Hartel, J. 2003. Ontological, Epistemological and Sociological dimensions of domains. *Knowledge Organization*, 30(3-4), 239-245.

Hjørland, B.; Pedersen, K. N. 2005. A substantive theory of classification for information retrieval. *Journal of Documentation*, 61(5), 582-597.

Hjørland, B. Concept theory. 2009. *Journal of the American Society for Information Science and Technology*, 60(8), 1519-1536.

Saldanha, Gustavo S. 2011. Entre retórica e a filologia: do pragmatismo ao humanismo na epistemologia da Ciência da Informação. *InCID: Revista de Ciência da Informação e Documentação*, 2(1), 47-67.

Mai, Jens-Erik. 2005. Analysis in indexing: document and domain centered approaches. *Information processing and management*, 41, 599-611.

Popper, Karl R. 1999. *Conhecimento objetivo: uma abordagem evolucionária*. Tradução de Milton Amado. Belo Horizonte: Itatiaia.

Van Dijk, Teun A. 2012. *Discurso e contexto: uma abordagem sociocognitiva*. São Paulo: Contexto.

Van Dijk, Teun A. 2001. Algunos principios de una teoría del contexto. *ALED, Revista Latinoamericana de estudios del discurso*, 1(1), 69-81.

Continuidades e rupturas em representação do conhecimento

Vera Dodebei
Universidade Federal do Estado do Rio de Janeiro
dodebei@gmail.com

1 Introdução

Ao acompanhar os contextos de produção e uso dos sistemas de informação ao longo das últimas décadas do século vinte e das primeiras do vinte e um, temos verificado que a história dos modelos lógico-semânticos que organizam o conhecimento produzido pelas sociedades mostra continuidade teórica em relação ao conceito como unidade epistêmica para a construção de linguagens de representação e, ao mesmo tempo, sugere descontinuidades que se referem aos modos de apresentar redes conceituais.

Recuperando anotações de aulas ministradas por Manoel Adolpho Wanderley entre os anos 1960-1980 aos cursos de Biblioteconomia no Rio de Janeiro, encontramos alguns resumos ainda datilografados que procuram sintetizar as discussões sobre conceitos, funções e problemas relativos à complexidade do campo da organização do conhecimento. Muito do que é discutido nesses roteiros de aulas, envelhecidos apenas no que toca ao suporte amarelado, pode ser recuperado de sua dissertação de mestrado, uma das primeiras do curso pioneiro do IBICT no Brasil (Wanderley: 1973).

Com o objetivo de socializar essas notas de aula, discutiremos nesta comunicação as continuidades e rupturas das relações conceituais na história do desenvolvimento das redes que objetivam representar campos de conhecimento científico. Utilizamos como *corpus* para guiar a discussão o resumo da aula ‘Limitações dos sistemas tradicionais e inovações neles introduzida’ (Wanderley:1980), procurando analisar as limitações ou restrições que os princípios lógicos e semânticos impõem às redes conceituais contemporâneas e as propostas sugeridas para minimizá-los. Partimos do pressuposto de que uma possível distinção ontológica entre os instrumentos de representação do conhecimento se constitui, não na lógica de construção das relações hierárquicas, mas no nível semântico ocasionado por uma redução a favor do controle terminológico no caso dos tesouros, e por sua ampliação, no caso das ontologias de domínio e aplicadas.

2 Divisão lógica e outras limitações dos sistemas gerais de organização do conhecimento – OC

A partir da segunda metade do século vinte, os sistemas tradicionais de classificação e o modelo lógico-semântico que as inspirou não tardaram a revelarem-se insatisfatórios. Com o aparecimento das novas condições de pesquisa científica, agravaram-se de tal sorte, em certas áreas, os problemas da informação especializada que a classificação e suas regras e métodos se viram diante de um verdadeiro impasse. Segundo Wanderley, era importante se atentar para certas limitações que a literatura crítica vinha assinalando nas classificações ditas convencionais. Inspirando-se em Mills, “A modern outline of library classification” (1960), Wanderley enumera, quanto à divisão lógica, sete restrições:

1ª restrição: a *relação de gênero para espécie*, de uma coisa para suas variedades, é apenas uma das relações de fato utilizadas nas classificações bibliográficas: não se alegue tratar-se de ‘classes de documentos caracterizados por assuntos’ e não de assuntos mesmos, porque sempre subsistiriam relações finais entre as próprias matérias, por exemplo: adaptação é processo em Biologia; dureza é propriedade em Metalurgia; destilação é operação executada em Química.

2ª restrição: é inadequado limitar-se a divisão às *características* que constituem qualidades *genéricas essenciais*: pode-se optar, dentro da mais estrita lógica, entre o que há de ser tido como ‘definição’ ou como ‘propriedade’. Por muito que a ‘diferença’ e a ‘propriedade’ interessem à correlação de propriedades- fundamental para a questão das classificações naturais e artificiais – não há razão em colocar o problema, antes metafísico do que lógico, de uma ‘essência’ física e imutável. A questão aqui se refere às regras da divisão lógica em que a essência, ou características essenciais dos conceitos se elegeriam como mais apropriadas àquelas ditas acidentais, ou pragmáticas.

3ª restrição: nem sempre ocorre rejeitarem os esquemas tradicionais as características relativas a acidentes e aterem-se tão só às que participem da natureza essencial do gênero. Em princípio, o gênero ‘soldado’ será dividido por modo de combate (infantaria, artilharia etc.); ‘condições de serviço’ (ativo, reserva, convocado etc.); ‘postos’ e assim por diante. Mas não o seria por fatores não relevantes, ou por características arbitrárias, como ‘classe social de origem’(operários, agricultores etc). Características semelhantes a essas são de emprego comum, do ponto de vista dos documentos, como, por exemplo, determinações de natureza espacial e temporal, ainda que os conceitos de lugar e tempo (Universidade inglesa) não sejam específicos do

gênero (Universidade). *A observação e a reflexão é que admitem, ou descartam, o accidental.* Assim, em relação à classe ‘Indústria’ ou ‘Negócios’ pode requerer-se, por significativa, a distinção entre ‘grandes’ e ‘pequenas’.

4ª restrição: da mesma forma, não é sempre exequível, ao dividir-se um gênero por características sucessivas, ater-se a uma regra de todo coerente. Adotado, na divisão, ‘modo de combate’ como critério, seria mister prosseguir-la, na mesma ordem de noções, por subdivisões como, no caso da ‘artilharia’, as que indicam variedades de ação bélica, definidas pelas expressões ‘artilharia de costa’, ‘artilharia de montanha’ e assim por diante. Não se aplicaria, antes de esgotada uma série de discriminações, outras que não lhes fossem condizentes, por exemplo, para o critério citado, a subdivisão por ‘postos’. Contudo, nos sistemas tradicionais, a formação de classes pela *combinação de características de todo distintas é inevitável, mormente quando se tratar de entidades complexas.* É de notar-se que a classificação facetada comporta o uso exaustivo de todos os princípios relativos a uma das facetas (i. é, todas as suas subfacetes) antes do recurso a outra, em perfeita conformidade, pois, com a exigência aludida.

5ª restrição: a divisão lógica não impõe uma *ordem privilegiada de aplicação de características.* A classe ‘pintura’ será dividida por ‘processo’ (óleo, aquarela, guache etc.) antes de o ser por ‘temas’ (paisagens, retrato, marinha). Isto se dá, porém, graças a considerações extra-lógicas, tais como a finalidade da própria classificação.

6ª restrição: o mesmo ocorre quanto ao *arranjo mesmo das classes, o qual, teoricamente, não obrigaria à anteposição do geral para o particular, do abstrato ao concreto,* nem privilegiaria qualquer ordem dentro de uma série de subclasses coordenadas (*array*);

7ª restrição: há frequentes superposições quando se classificam e arranjam fisicamente, quanto ao conteúdo, os documentos, i. é, os objetos a que os termos das classes em última análise se referem. Verifica-se a *impossibilidade de uma correspondência biunívoca entre documentos e termos de classes,* o que se soma à natural complexidade das relações entre as próprias ciências e ao caráter interdisciplinar da pesquisa moderna. Na prática, o que fazem os sistemas é afrouxar a conotação: ‘impostos’ equivale a ‘documentos relativos principalmente a impostos’.

3 Redes conceituais: para além da lógica

Como vimos nos exemplos anteriores, a prática da “desobediência” ao cumprimento estrito dos preceitos lógicos por parte das classificações bibliográficas para a organização de conceitos de um determinado domínio do conhecimento abre uma possibilidade para pensar que a hospitalidade dos sistemas conceituais, de fato, deve ser considerada primordial quando da sua construção. Hierarquias, essência e acidente, ordem e sucessividade na derivação conceitual, complexidade e especificidade são princípios lógicos que ainda cabem à construção de linguagens artificiais. Mas, estas precisam flexibilizar o princípio da completude em razão de que se constituem como processo e não como produto.

Paul Thagarde (2007) em “A estrutura conceitual da revolução química¹⁵”, utilizando técnicas extraídas da inteligência artificial, investiga as mudanças conceituais revolucionárias que ocorreram quando a teoria do flogisto de Stahl foi substituída pela teoria do oxigênio de Lavoisier. O autor fornece um pequeno fragmento de rede conceitual referente a animais para explicar que ligações de tipo e ligações de regra podem ser codificadas por generalizações universais tais como (x) (*canário* (x) → *pássaro* (x)), e ligações de instância e ligações de propriedade podem ser codificadas por sentenças atômicas tais como *pássaro* (*Piu-Piu*). Segundo o autor, de uma perspectiva computacional, as ligações têm uma importância que transcende a informação que elas expressam e podem ir além da lógica.

¹⁵ Este artigo é uma tradução de “The Conceptual Structure of The Chemical Revolution”, publicado originalmente em *Philosophy of Science*, número 57, p. 183-209, 1990.

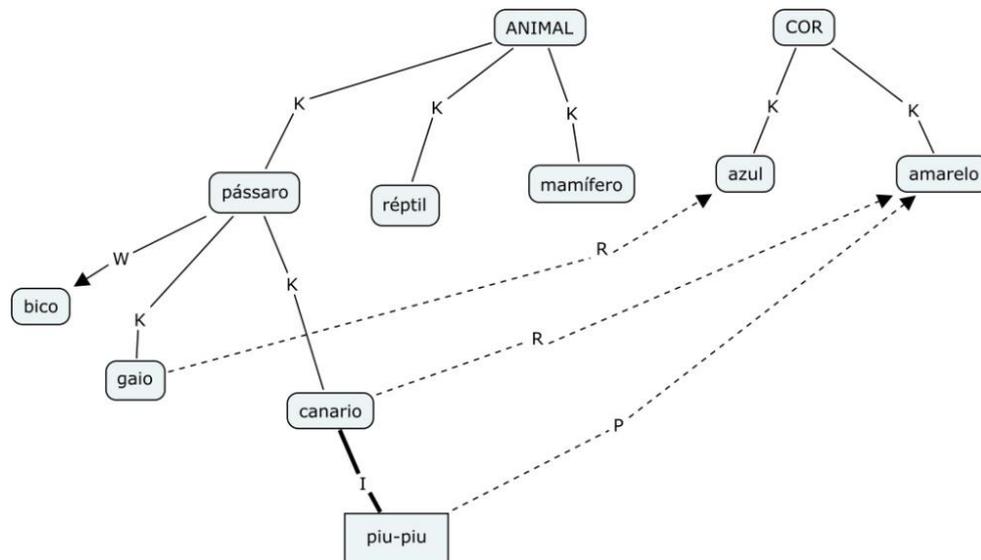


Figura 1: Parte de um framework conceitual para animais. Thagarde (2007:268)¹⁶

A pesquisa de Thagarde, além de demonstrar a possibilidade de representar uma revolução científica, mostra que as redes conceituais concebidas para representar o conhecimento devem ser abertas, sempre incompletas ou provisórias, pois ao surgir um novo esquema conceitual fundamental para o desenvolvimento científico ele é um novo ponto de vista a ser incorporado nas representações históricas do conhecimento. Ao mesmo tempo, o exemplo na figura 1 confirma o uso de hierarquias e suas derivações para representar ‘frames’ de uma rede conceitual e sugere a hospitalidade do sistema a relações entre os conceitos.

4 Algumas considerações

As linguagens de representação do conhecimento baseiam-se em seu contexto de produção e no acesso às informações. Esse contexto varia em função das dimensões espaciais e temporais em que os objetos são representados em redes de relação de acordo com o momento tecnológico em que essas são construídas. As limitações

¹⁶ Ligações de *tipos*, marcadas por linhas retas rotuladas como ‘K’. Estas ligações indicam que um conceito é um tipo de outro: por exemplo, canário é um tipo de pássaro, e pássaro é um tipo de animal. Ligações de *instâncias*, marcadas por linhas retas rotuladas como ‘I’. Estas indicam que algum objeto particular, marcado não por uma elipse mas por uma caixa, é uma instância de um conceito: Piu-Piu é um canário. A rede deixa claro que Piu-Piu é também um animal. Ligações de *regra*, marcadas por linhas pontilhadas terminadas com flechas e rotuladas como ‘R’. Elas expressam relações gerais (ainda que nem sempre universais) entre conceitos; por exemplo, que os canários têm cor amarela. Ligações de *propriedade*, marcadas por linhas pontilhadas terminadas com flechas mas rotuladas como ‘P’. Elas indicam que um objeto tem uma propriedade tal como: Piu-Piu é amarelo. Ligações de *parte-todo*, marcadas por linhas retas terminadas com flechas e marcadas como ‘W’. Elas indicam que um todo tem uma dada parte: um bico é uma parte de um pássaro.

apresentadas por Wanderley são um indício de que o contexto pode apresentar focos de interesses distintos do modo como as informações foram analisadas. Esse descompasso é natural em razão das rápidas transformações tecnológicas que mediam as operações de redução semântica na representação da linguagem natural. Assim, é lícito afirmar que recursos de informação possuem um viés multidimensional e a lógica da representação por vezes se restringe a uma unidimensionalidade.

Ontologias gerais, de domínio ou de aplicação construídas em ambiente digital colaborativo demonstram uma gradação de abertura à incompletude, e buscam romper as barreiras do domínio que representam em função de que a complexidade da organização do conhecimento é sempre da ordem do incompleto e, portanto, as operações de associação de conceitos devem prever novas relações, sob o risco de implodirem ao simples contato com outras possibilidades.

Quanto aos tesouros, eles foram concebidos no campo da documentação, como instrumentos facilitadores do processo de tradução da língua natural para as linguagens artificiais. Para cumprir com este objetivo os tesouros, apesar de fazerem uso da lógica (não formalizada) e de categorias para a organização de classes e subclasses, necessitavam reduzir a polissemia dos códigos comunicacionais em prol de uma precisa recuperação da informação. Até que ponto as ontologias podem substituir os tesouros, não podemos ainda responder. Convivemos com todos estes instrumentos de organização e representação do conhecimento de forma cumulativa incluindo, no conjunto, as classificações bibliográficas.

Por fim, resta reforçar que o ambiente tecnológico da comunicação na atualidade favorece a modelização cooperativa de redes conceituais em direção à extensibilidade categorial sem contornos definidos, portanto, inclusivas e com alto grau de hospitalidade. Diríamos então que representar conhecimento requer instrumentos que combinem razão, sensibilidade e formalização para que haja o diálogo entre humanos, mediados pela máquina e, quiçá, diretamente das pessoas com as inteligências artificiais por elas construídas.

Referências

- Mills, J. 1960. *A modern outline of library classification*. London: Chapman and Hall.
- Thagarde, Paul. 2007. A estrutura conceitual da revolução química. *Princípios*, Natal, 14/22: 265-303.
- Wanderley, M. A. 1980. *Limitações dos sistemas tradicionais e inovações neles introduzidos*. Rio de Janeiro: Unirio (Notas de aula)
- Wanderley, M. A. 1973. Linguagem documentária: acesso à informação. *Ci. Inf.*, Rio de Janeiro, 2/2 p. 175-217.

Aproximación sistemática al concepto de “control de vocabulario”.

Mario Barité
Universidad de la República
mabarite@gmail.com

J. Carlos Fernández-Molina
Universidad de Granada
jcfernan@ugr.es

1. Introducción.

El control de vocabulario es un conjunto de técnicas y procedimientos que se ejercen sobre expresiones lingüísticas denominadas 'términos', que representan conceptos en un determinado contexto temático. En Organización del Conocimiento permite formalizar estructuras conceptuales destinadas a representar el saber (tesauros, listas, taxonomías), y normalizar terminología utilizada en sistemas de información, para favorecer los procesos de búsqueda, acceso y recuperación temática de información. Subyace la idea de que tales operaciones jerarquizan y mejoran la calidad de los términos de indización y las búsquedas temáticas, en contraposición al uso del lenguaje natural, no controlado ni estructurado; idea interpelada cotidianamente por la mejora continua de las aplicaciones informáticas basadas en el lenguaje.

Existe una estrecha relación entre los conceptos de control de vocabulario y vocabulario controlado, basada en el elemento común de modelización de las expresiones lingüísticas y las nociones de las que son referencia, para fines de representación de conocimiento y de recuperación de información. Un vocabulario controlado solo puede construirse a partir de sucesivas operaciones de control de vocabulario. Inversamente, el control de vocabulario adquiere sentido pleno si se realiza en el contexto del desarrollo de vocabularios controlados, sin perjuicio de su aplicación en otros ámbitos donde la mencionada modelización pueda ser útil.

El control de vocabulario supone la preexistencia de un vocabulario, seleccionado previamente a partir de alguna de las garantías: literaria, de usuario, cultural, etc. (Barité, 2011). Así, el control de vocabulario no incide sobre la calidad de la selección previa, pero necesita disponer de una caja terminológica suficiente,

pertinente y calificado.

Los estándares desarrollados a escala internacional (British Standards Institution, 2005; National Information Standards Organization, 2005; International Standard Organization, 2011), juegan un papel trascendente: proponen reglas de aplicación universal y buscan unificar criterios metodológicos para desarrollar vocabularios controlados, clasificar e indizar recursos de información convencionales o electrónicos.

A partir de la revisión de literatura y del estudio de los estándares, puede afirmarse que hay al menos cuatro perspectivas de análisis del control de vocabulario. La perspectiva más general refiere al rigor con que se aplica en sistemas de organización del conocimiento (SOC). Así, el estándar norteamericano establece una graduación de los vocabularios controlados, desde aquellos con control más riguroso y sistemático a los que tienen control bajo, en este orden decreciente: tesauros, taxonomías, anillos de sinónimos y listas (National Information Standards Organization, 2005, p. 16-17). Con menor precisión, esa graduación es recogida por el estándar británico (British Standards Institution, 2005).

Las otras tres perspectivas relevantes de análisis están basadas respectivamente en los objetivos del control de vocabulario, su objeto (términos individuales y relaciones entre términos), y los niveles de aplicación (lingüístico, semántico, funcional). Estas perspectivas no son necesariamente excluyentes; más bien brindan diferentes ópticas para entender, interpretar y mejorar los procesos de control de vocabulario.

El esquema previo pretende aportar las bases para la sistematización de las cuestiones que atañen y están implicadas en el control de vocabulario, con la finalidad de utilizarla en futuras investigaciones.

2. Análisis Basado En Objetivos.

Este trabajo se centra específicamente en la perspectiva de análisis basada en los objetivos del control de vocabulario. Desde esta aproximación, los procesos se discriminan así:

1. Restricción. El control de vocabulario establece límites en cuatro direcciones, respecto a: i) la cantidad de términos a ser incluidos en un vocabulario controlado, ya

que persigue la organización de universos restringidos del conocimiento; ii) la extensión de cada término (un término de indización no debe sobrepasar las tres palabras, salvo excepciones); iii) el alcance conceptual de cada término, pues se tiende a asignar a cada uno un solo significado, en cumplimiento del principio de univocidad (un término – un significado, y viceversa); y, iv) la cantidad de relaciones que se establece entre un término y otros afines. Así visto, el control de vocabulario es una expresión de economía lingüística.

2. *Desambiguación.* Homonimia, polisemia, variantes, sinonimias y/o cuasi-sinonimias son manifestaciones habituales de la ambigüedad del lenguaje, provocando dudas, solapamientos y confusiones en torno al significado de los términos propios de una disciplina. Según Bakhtin, "*no plano ideológico, a palavra é uma unidade 'neutra', apta a se adequar a diferentes padrões culturais. E isso ocorre, porque ela é portadora de uma gama de significação que a torna capaz de assumir sentidos ou valores diferentes, dependendo do contexto*" (Bakhtin apud Cintra et al, 2002, p. 67-68). Si como afirma Moreiro (1994, p. 207), "*los términos controlados se aplican para caracterizar contenidos en los documentos*", las técnicas de desambiguación procuran la mayor precisión respecto al alcance conceptual y al sistema de relaciones de cada término. En este sentido, surgen regularmente en la literatura opiniones favorables a la inclusión sistemática de definiciones en los SOC (Buchan, 1989; Hudon, 1996; Jorna & Davies, 2001), especialmente en las áreas sociales y humanas, debido a la imprecisión o reformulación de su terminología (Aitchinson, Gilchrist & Bawden, 2000). Desde la óptica de la desambiguación, el control de vocabulario se constituye en un proceso de determinación conceptual y terminológica.

3. *Normalización.* Por sus características y objetivos, el control de vocabulario se aboca al establecimiento de formas padronizadas de representación del conocimiento. En esta medida se vincula con procedimientos de control de autoridades (Esteban Navarro, 1997). La normalización comprende la fijación de forma y de sentido, y la situación ponderada de cada término en la estructura conceptual de pertenencia, tendiendo a una mayor coincidencia entre la indización y las búsquedas de los usuarios (Aitchison, Gilchrist & Bawden, 2000). La normalización se alcanza mediante reglas formales para la elección de singulares y plurales, ortografías correctas, formas gramaticales preferidas y cuestiones similares.

4. *Jerarquización.* Si bien los sinónimos y cuasi-sinónimos son intercambiables

en un discurso o en la comunicación habitual, en la mayoría de los SOC solo uno de ellos es reconocido como término autorizado o preferido, mientras los demás conservan solamente valor de referencia, y aún otros son dejados de lado. Así, la distinción entre términos autorizados, no autorizados y excluidos, es constante en el desarrollo de un SOC. Esta jerarquización debería realizarse a partir de la garantía utilizada (literaria, de uso, académica, etc.)

5. *Relación.* El control de vocabulario trabaja sobre dos vínculos: las relaciones intra-término, donde se dilucida la situación de sinónimos, cuasi-sinónimos y variantes; y las relaciones entre términos, que tradicionalmente se establecen bajo distintas modalidades de relación jerárquica o asociativa (Cintra *et al*, 2002, Aitchison & Clarke, 2004, Naumis, 2007). Estructuras nuevas como las ontologías permiten diversificar y enriquecer el sistema de relaciones. Si bien las tipologías tradicionales de relaciones intra y entre términos han dado firmeza al proceso de control de vocabulario, deben considerarse estrategias diferentes, pues las disciplinas tienen distintas consistencias terminológicas, conforme con la estabilidad de las relaciones internas de sus conceptos (Mote, 1962; Svenonius, 1986).

Los estándares en curso colocan diferentes énfasis respecto a los objetivos que persigue el control de vocabulario. El estándar norteamericano pone el acento en el proceso de organización de una lista de términos, a partir de operaciones de desambiguación, mencionando también la jerarquización (ANSI/NISO 2005: 10). El estándar británico prioriza la restricción, pues ella puede “aumentar la posibilidad de que los indizadores y buscadores encuentren el mismo término representando al mismo concepto” (BSI, 2005, p. 6). Por último, la norma ISO 25.964-1 establece un equilibrio entre la desambiguación y la restricción, y menciona lateralmente la tarea de jerarquización entre términos (International Standard Organization, 2011, p. 12).

3. Conclusiones.

El control de vocabulario está vigente como herramienta intelectual para organizar y estructurar listas de términos destinados a la representación del conocimiento a través de formas o códigos de clasificación padronizados. Reafirma su importancia comprobar que las listas normalizadas de términos sugeridos aparecen habitualmente en ventanas de ayuda en Internet -reino del lenguaje natural-

para orientar la búsqueda, como se da en Google o Youtube.

Permite una aproximación racional y estructurada al conocimiento a través de los SOC, al tiempo que ofrece una perspectiva de usuario, favoreciendo la recuperación temática en sistemas de información, y reduciendo la ambigüedad propia del lenguaje natural. Proporciona consistencia en la descripción temática de recursos de información, aumenta la eficiencia de las herramientas de búsqueda, y la efectividad de los sistemas de almacenamiento y recuperación de información, los sistemas de navegación en Internet y otros entornos.

La investigación en esta materia se ha dado como apoyo al desarrollo de estándares. Por ello se hace sentir en la literatura una mayor preocupación de los investigadores por aspectos teóricos relacionados con el control de vocabulario, pues un mayor conocimiento de sus potencialidades puede incidir decisivamente en el debate entre defensores del lenguaje natural por oposición a enfoques normalizadores, de larga tradición en nuestra disciplina. Internet es testigo de primera línea en este debate, del que aprovechará las enseñanzas que brinde la investigación.

En ese marco, se ha propuesto una primera sistematización, para interpretar y consolidar las técnicas y procedimientos del control de vocabulario. Asimismo, se puede contribuir a una superación conceptual de los estándares actuales, proveyendo nuevos elementos para la elaboración, revisión y evaluación de sistemas de organización del conocimiento. Se observa que los cinco procesos analizados – restricción, desambiguación, normalización, jerarquización y relación – se presentan especialmente útiles para la evaluación de lenguajes de representación del conocimiento, sean ellos naturales o de indización.

Referencias bibliográficas

- Aitchison, J. & Clarke, D. 2004. The thesaurus: a historical viewpoint, with a look to the future. *Cataloging & Classification Quarterly*, vol. 37, nº 3/4: 5-21.
- Aitchison, J., Gilchrist, A. & Bawden, D. 2000. *Thesaurus construction and use: a practical manual*. Londres: ASLIB.
- Barité, M. 2011. La garantía cultural como justificación en sistemas de organización del conocimiento: aproximación crítica. *Palabra Clave*, vol. 1, nº 1: 2-11.
- British Standards Institution, 2005. *Structured vocabularies for information retrieval: BS 8723-1*. S.l.: BSI.

- Buchan, R. L. 1989. Intertwining thesauri and dictionaries. *Information Services & Use*, vol. 9, nº 3: 171-175.
- Cavalcanti, C. 1978. *Indexação & Tesouro: metodologia & técnicas*. Ed. preliminar. Brasília: ABDE.
- Cintra, A. M. M....[et al]. 2002. *Para entender as linguagens documentárias*. 2a. ed. São Paulo: Polis; APB.
- Dodebei, V.L.D. 2002. *Tesouro : Linguagem de representação da memória documentária*. Niterói: Intertexto; Rio de Janeiro: Interciência.
- Esteban Navarro, M.A., 1997. *Construcción y mantenimiento de clasificaciones documentales*. In Maria Pinto, ed. *Manual de clasificación documental*. Madrid: Síntesis, p. 131-174.
- Gil Urdiciain, B. 1996. *Manual de lenguajes documentales*. Madrid: Noesis.
- Green, R. 2001. *Overview of relationships in Knowledge Organization*. In Carol Bean, ed. *Relationships in the Organization of Knowledge*. Dordrecht: Kluwer.
- Hudon, M. 1996. *Preparing terminological definitions for indexing and retrieval thesauri: a model*. In *Advances in Knowledge Organization*. V. 5. Würzburg: Ergon Verlag, p. 363-369.
- International Standard Organization. 2011. *International Standard ISO 25.964-1: Information and Documentation : Thesauri and interoperability with other vocabularies: Part 1: Thesauri for information retrieval*. Geneva: ISO.
- Jorna, K. & Davies, S. 2001. *Multilingual thesauri for the modern world: no ideal solution?* *Journal of Documentation*, vol. 57, nº 2: 284-295.
- Martínez, A.M. & Valdez, J.C. 2008. *Indización y clasificación en bibliotecas*. Buenos Aires: Alfagrama.
- Mitchell, J.N. 2007. *Should be control vocabulary?* NELINET Annual Conference Worcester, Massachusetts November 16, 2007. Disponible en: www.nelinet.net/edserv/conf/cataloging/2007/ohnmitchell.pdf [consultado el 10 de febrero de 2013]
- Moreiro, J.A. 1994. *Documentación y lingüística: conceptos de relación esenciales*. *Ciencias de la Información*, vol. 25, nº 4: 202-211.,
- Mote, L.J. 1962. *Reasons for the variations in the information needs of scientists*. *Journal of Documentation*, vol. 18, nº 4: 169-175.
- National Information Standards Organization. 2005. *Guidelines for the construction, format and management of monolingual controlled vocabularies: an American National Standard developed by the National Information Standards Organization*. Bethesda: NISO Press. Disponible en: http://www.niso.org/kst/reports/standards?step=2&gid=None&project_key%3Austring%3Aiso-8859-1=7cc9b583cb5a62e8c15d3099e0bb46bbae9cf38a [consultado el 10 de febrero de 2013]
- Naumis, C. 2007. *Los tesauros documentales y su aplicación en la información impresa, digital y multimedia*. México: UNAM; Buenos Aires: Alfagrama.

- Svenonius, E. 1986. Unanswered questions in the design of controlled vocabularies. *Journal of the American Society for Information Science*, vol. 37, n° 5: 331-340.

A autoria e a representação documental

Naira Christofolletti Silveira
Universidade de São Paulo
naira_csilveira@yahoo.com.br

Maria de Fátima G.M Tálamo
Universidade de São Paulo
mfgmtala@usp.br

Introdução

É possível observar que, inicialmente, a representação documental era simplificada, com poucos elementos descritivos e os catálogos eram mais próximos de inventários do que de um equipamento de busca e recuperação. Entre outros fatores, isso também se justifica devido a pouca produção bibliográfica daquela época e as características dos documentos.

Na medida em que a produção bibliográfica aumentou, os catálogos passaram a ser mais elaborados e a identificação da autoria passou a ser relevante para identificar um documento. Desde então, tornou-se um elemento constante na representação documental, em especial na cultura ocidental.

Neste trabalho, tem-se como objetivo discutir a questão autoral no contexto da representação documental, revelando como a autoria está refletida nos registros bibliográficos e a sua relação com a sociedade. Acredita-se que a limitação a este enfoque contribui com uma pequena parte para o estabelecimento da comunicação efetiva entre a informação e o usuário.

Algumas observações sobre a autoria e o catálogo

A função autor é abordada em *O que é um autor?*, comunicação de Foucault à *Société Française de Philosophie*, em 1969. Para ele, alguns “[...] discursos são providos da função ‘autor’, ao passo que outros são dela desprovidos” (FOUCAULT, 2006: 46). Há quatro características reconhecidas da função autor, assim resumidas:

[...] a função autor está ligada ao sistema jurídico e institucional que encerra, determina, articula o universo dos discursos; não se exerce uniformemente e da mesma maneira sobre todos os discursos, em todas as épocas e em todas

as formas de civilização; não se define pela atribuição espontânea de um discurso ao seu produtor, mas através de uma série de operações específicas e complexas; não reenvia pura e simplesmente para um indivíduo real, podendo dar lugar a vários “eus” em, simultâneo, a várias posições-sujeitos que classes diferentes de indivíduos podem ocupar (FOUCAULT, 2006: 56-57).

Do mesmo modo que se observam alterações na noção de autor ao longo do tempo na sociedade, observa-se que sua representação também sofreu alterações, aos poucos a autoria ganhou um espaço de destaque na representação documental.

“Os textos, os livros, os discursos começaram efetivamente a ter autores [...] na medida em que o autor se tornou passível de ser punido, isto é, na medida em que os discursos se tornaram transgressores” (FOUCAULT, 2006: 47). Para condenar aqueles que transgrediam a ortodoxia política ou religiosa era necessário identificá-los, por isso a figura do autor aparece durante a Inquisição (CHARTIER, 1999). É importante destacar que o surgimento do autor na Idade Média coincide com o surgimento de catálogos mais estruturados.

Antes da Idade Média, existiam tabletes de argilas e listas com dados bibliográficos, mas se desconhece se tais listas eram mecanismos de busca e recuperação da informação ou se eram apenas listas de inventários (GARRIDO ARILLA, 1996). O processo de elaboração de catálogos mais estruturados é decorrência de ações realizadas séculos antes. No século II, por exemplo, a organização bibliográfica tornou-se mais específica e individualizada, “[...] quando Galeno, médico grego, fez a relação de seus trabalhos para que não fossem confundidos com os de outros autores” (CALDEIRA, 1984: 261).

Nesse caso, a introdução da autoria foi um mecanismo de organização que permitiu a personificação do conhecimento, além de um importante elemento de individualização. Desde então, as listas com dados bibliográficos tornaram-se mais elaboradas e iniciou-se um processo de identificação dos documentos, no qual são incluídos elementos descritivos com a finalidade de fornecer uma representação minuciosa e fidedigna, distinguindo um documento dos demais presentes no acervo.

Somente a partir da Idade Média é que a identificação dos autores tornou-se recorrente. Porém, essa constatação não significa a inexistência de autores antes desse período. Ela revela a alteração da relação entre o autor e o texto, ocorrida durante a Idade Média, e a evolução dos catálogos que começaram a incorporar em seus elementos descritivos o nome do autor para identificar uma obra.

Uma vez que o autor passou a ser reconhecido como responsável pelo seu discurso, tornou-se um elemento muito importante na representação documental. O surgimento da função “autor” e o olhar da sociedade sobre ele alteraram a identificação da autoria nos catálogos. Após a Idade Média, a presença da autoria fica cada vez mais evidente. Existem duas justificativas para o acréscimo da autoria como pontos de acesso nos catálogos:

- a) para encontrar um documento específico quando o autor é conhecido;
- b) para encontrar quais documentos de um determinado autor há na biblioteca (NEEDHAM, 1964).

Existem outras justificativas para a inclusão da autoria nos catálogos: a identificação e a seleção de documentos semelhantes, com autorias diferentes. Com a explosão documental é cada vez mais necessário elencar elementos que possam ajudar na organização, identificação e seleção de documentos. Os autores e outros responsáveis, como tradutores, editores, entre outros, são elementos fundamentais para a busca, recuperação e seleção de documentos.

A representação da autoria está presente nos três elementos que compõem o registro bibliográfico tradicional: dados de localização, descrição bibliográfica e pontos de acesso.

Quando a responsabilidade é registrada na descrição bibliográfica, ela tem o papel de individualizar um documento, permitindo que o usuário o identifique. Nos dados de localização, a notação de autor, seja pelo uso da Tabela de Cutter, pela Tabela PHA ou pela notação nominal, tem como função reunir os autores dentro da mesma classe temática quando o arranjo é relativo. A relação entre autoria e o documento aparece no ponto de acesso quando se estabelece o ponto de acesso de responsabilidade, na qual há controle para a forma do nome do autor para que todas suas *obras* e *expressões* estejam reunidas e possam ser buscadas.

Portanto, a autoria é um elemento constante em todos os processos que envolvem a organização, representação, busca, recuperação e obtenção de um documento. A ressalva está na forma como a autoria é representada no registro bibliográfico que, na maioria das vezes, adota padrões internacionais de representação (códigos de catalogação entre outros instrumentos) sem refletir na construção e no uso do documento representado.

Um documento ao ser representado em um sistema documental, segundo regras internacionais, pode vir a dificultar sua representação adequada às comunidades

específicas de usuários e não encontra ressonância na sociedade em que o equipamento cultural (biblioteca, centro de documentação entre outros) está inserido. O conhecimento e a informação são gerados e utilizados dentro de contextos socioculturais específicos, os padrões de representação documental em âmbito internacional devem ser estudados e adaptados aos modos de produção e uso da informação.

Considerações Finais

A discussão sobre a questão autoral na representação documental inicia-se com o percurso histórico do autor na sociedade e na produção do conhecimento. Como a autoria nem sempre é atribuída do mesmo modo entre as sociedades, é necessário estudar suas características gerais e específicas para se estabelecer formas de representação que contemplem as particularidades da autoria em determinados documentos e sociedades.

A preocupação em estabelecer canais de comunicação efetivos entre a informação e o usuário está presente na construção de instrumentos para a representação documental. Porém, observa-se uma lacuna em pesquisas que contextualizem o usuário em seu universo cultural na Representação Descritiva. Neste sentido, este trabalho buscou incentivar pesquisas sociais e teóricas, contribuindo para o desenvolvimento e consolidação desta área como uma prática social.

Referências

- CALDEIRA, P. da T. 1984. A situação do Brasil em relação ao controle bibliográfico universal. *R. Esc. Bibliotecon. UFMG*, Belo Horizonte, v. 13, n. 2, p. 260-283.
- CHARTIER, R. 1999. *A aventura do livro: do leitor ao navegador: conversações com Jean Lebrun*. São Paulo: Ed. da UNESP.
- FOUCAULT, M. 2006. *O que é um autor?* 6. ed. Lisboa: Vega Passagem.
- GARRIDO ARILLA, M. R. 1996. *Teoría e historia de la catalogación de documentos*. Madrid: Síntesis.
- NEEDHAM, C. D. 1964. *Organizing knowledge in libraries: an introduction to classification and cataloguing*. London: Andre Deutsch.

Autopoiesis e formação discursiva na consolidação de linguagens de indexação: A elaboração do tesauro brasileiro de estudos semióticos (Semavoc)

Maria Aparecida Moura
Universidade Federal de Minas Gerais
mamoura@eci.ufmg.br

A ampliação dos espaços sociais de informação semântica (SSIS¹⁷) incorporada aos produtos e serviços de informação contemporâneos tornou possível a intensa manifestação dos usuários desde a descrição intelectual da informação. Os espaços sociais de informação semântica são ambientes virtuais fortalecidos pelo contexto digital em que coexistem comunidades especializadas e organizadas em torno do compartilhamento de informações através de temas e conceitos. Estes espaços semânticos são criados por grupos autônomos ou usuários individuais que partilham diferentes padrões de qualidade da informação a partir de distintas visões de mundo.

A digitalização da informação, associada ao fenômeno da globalização cultural e às tecnologias de telecomunicações incidiram radicalmente sobre nossa “fé perceptiva” exigindo novas posturas diante dos fenômenos informacionais contemporâneos. Atualmente, uma perspectiva semiótica do documento é reivindicada. Tal reivindicação incorpora a ideia de desmaterialização do documento, demanda a sua relativização enquanto artefato portador de informação e acentua a sua inscrição em um tempo histórico o que requer e uma abordagem interpretativa no campo da Ciência da informação.

De acordo com Pinto (1995) é preciso sempre considerar a vagueza e a dinamicidade da informação, sobretudo no atual contexto de produção e disseminação da informação. A composição dos instrumentos de representação da informação foi, ao longo de sua história, fortemente determinada pela tríade de garantias (literária, uso e estrutural) (Svenonius, 2000). No entanto, nos últimos anos, a categorização da informação em ambientes digitais motivada pelos processos colaborativos online e pela auto-referência dos usuários tornou possível a hipótese de incorporação de novas garantias que possam dar visibilidade a inevitável movimentação dos usuários nos cenários de produção, descrição e recuperação da informação.

¹⁷ Social Semantic Information Spaces.

Beghtol (1986, 2001) abordou a perspectiva de formalização das garantias epistemológica e cultural no qual considerou o contexto social de produção, descrição e uso da informação. As garantias assinaladas por Beghtol ressaltam, sobretudo, o caráter mediador das ferramentas e dos profissionais que atuam nesse contexto.

Mai (2011), em análise do contexto de produção dos sistemas de classificação, destacou a possibilidade de consolidação de uma garantia autopoietica. A garantia autopoietica (auto "próprio", poiesis "criação") refere-se à autoridade classificatória que emana de um sistema auto-organizado e regulado por seus usuários, no qual as classes incluídas e a autoridade do sistema emergem de sua utilização. Dessa perspectiva, Mai (2011) assinalou que o sistema de representação da informação autopoietico pode contribuir para facilitar a ordenação do universo do conhecimento, de forma "democrática, responsável e significativa" devido à habilidade colaborativa de comportar as diferentes interpretações.

Para Andrade (2012),

“a preocupação central da autopoiesis não é exclusivamente a realidade e a existência do mundo, mas a forma como interpretamos o mundo e compreendemos a realidade, partindo do princípio de que é preciso compreender como compreendemos, ou seja, compreender o modo pelo qual fazemos uso de nossa própria compreensão, enquanto seres humanos e observadores.”

98

Desse ponto de vista, a garantia autopoietica incide sobre os elementos semânticos e semióticos através dos quais os usuários buscam, recuperam, compartilham, produzem e nomeiam as informações em um dado contexto.

De acordo com Foucault (2010) o discurso é conjunto de enunciados que pode pertencer a campos diferentes, mas que estão submetidos a regras de funcionamento comuns (discurso clínico, médico, psiquiátrico). Para o autor os discursos devem ser tratados no jogo de sua instância, visto que não representam um lugar tranquilo a partir do qual outras questões podem ser levantadas (estrutura, coerência, sistematicidade e transformações). O autor assinala que o caráter desconcertante do discurso está na descontinuidade, na necessidade de assumir recortes provisórios e de escolher domínios de análise em que as relações correm o risco de ser numerosas e densas.

Para Foucault a formação discursiva (2010) é “um conjunto de regras anônimas, históricas, sempre determinadas no tempo e no espaço, que definiram em uma época dada, para uma área social, econômica, geográfica ou linguística dada as condições de exercício da função enunciativa.” A formação discursiva se manifesta em três níveis:

superfícies de emergência, instâncias de delimitação e grades de especificação. As superfícies de emergência assinalam as circunstâncias do surgimento, as diferenças individuais, os graus de racionalidade, os códigos conceituais e os tipos de teoria. A instância de delimitação refere-se à identificação das instituições reguladoras do domínio responsáveis por distinguir, designar, nomear, instaurar os processos de significação como objeto. Finalmente, as grades de especificação referem-se aos regimes de agrupamentos e classificações adotadas em um determinado contexto de formação discursiva.

Todavia,

“uma formação discursiva se define (pelo menos quanto a seus objetos) se se puder estabelecer um conjunto semelhante, se se puder mostrar como qualquer objeto do discurso em questão aí encontra seu lugar e sua lei de aparecimento; se se puder mostrar que ele pode dar origem, simultânea ou sucessivamente a objetos que se excluem, sem que ele próprio tenha que se modificar. (Foucault, 2010: p. 49-50)

Vale lembrar também que a formação conceitual que compõe o discurso não funciona como blocos de imobilidade e que são as relações estabelecidas entre as superfícies que podem auxiliar na delimitação, análise e especificação em um dado domínio.

Inevitavelmente as formações discursivas contemporâneas têm se consolidado sob a égide das atuais formas de difusão e compartilhamento da informação e do conhecimento.

Para a área de organização da informação tornou-se ainda mais complexo propor instrumentos de recuperação da informação em contexto sem compreender a dinâmica da formação discursiva em dado domínio do conhecimento. Sobretudo, porque os conceitos referem-se à descrição do campo onde os enunciados aparecem. Eles são formados a partir de um feixe de relações no qual se considera contexto, as regularidades e coações discursivas, as escolhas teóricas, a historicidade e articulam-se acontecimentos, transformações, mutações e processos.

Analisar os processos de construção de linguagens de indexação do ponto de vista da intensa mobilização dos usuários da informação implica em compreender o caráter autopoietico das trocas informacionais em espaços sociais semânticos e, ao mesmo tempo, compreender a dinâmica de formação discursiva nas diferentes áreas do conhecimento. De acordo com Foucault (2010: p.62) “antes de querer repor os conceitos

em um edifício dedutivo virtual, seria necessário descrever a organização do campo de enunciados em que aparecem e circulam.”.

À luz da reflexão de Mai (2011) acerca da hipótese do estabelecimento da garantia autopoiética no processo consolidação de linguagens de indexação e da regulação estabelecida pelos processos de formação discursiva sugerida por Foucault (2010) buscou-se compreender as implicações dessa movimentação aplicada aos estudos semióticos no contexto brasileiro.

A opção por formalizar em linguagem de indexação o debate em relação aos estudos semióticos se deve a notável centralidade alcançada pela semiótica nos mapas de conhecimento do campo da Ciência da Informação em nível nacional e internacional.

De acordo com o estudo realizado por Zins (2007), a semiótica apareceu como fundamento teórico, conceito, processo, campo relacionado e disciplina estruturante da Ciência da Informação. Dos 28 mapas de conhecimento apresentados pelos teóricos do campo notam-se, com distinção das escolas de pensamento e aproximações, dez mapas em que a semiótica está em evidência teórica. Ademais, a semiótica tem como substrato os processos de significação e compreende todos os domínios significantes, o que, *per se*, representa um território clivado por multiplicidades conceituais que oscilam entre o horizonte da idealidade e o curso empírico das ideias consolidadas.

O instrumento criado foi denominado de SEMAVOC – Tesouro Brasileiro de Estudos Semióticos. Para a identificação dos descritores incorporados ao tesouro buscou-se compreender a formação dos estudos semióticos em sua superfície de emergência, instâncias de delimitação e grades de especificação. Assim, procurou-se descrever os estudos semióticos de uma perspectiva funcional e organizar os conceitos a partir dos enunciados teóricos compartilhados na literatura e nos espaços sociais semânticos nos quais aparecem, circulam e são referendados seus objetos e discursos.

Para tanto, compreendeu-se como superfície de emergência, instância de delimitação e grades de especificação no qual os objetos surgem, são identificados e/ou transformados, os eventos culturais e científicos, recursos online, publicações culturais e científicas, projetos de pesquisa e fóruns virtuais e presenciais dedicados ao tema.

O SEMAVOC - Tesouro Brasileiro de Estudos Semióticos, fruto desse percurso investigativo, foi composto por 1430 descritores organizados nas seguintes facetas: Análise do discurso (Análise do discurso – teóricos); Comunicação social; Estruturalismo (Estruturalismo – teóricos); Filosofia da linguagem; Hermenêutica (Hermenêutica – Teóricos); Informação; Linguagem; Representação social; Semântica

(Semântica discursiva); Semiologia (Semiologia – teóricos, Semiologia da imagem, Semiologia do cinema, Semiologia interpretativa); Semiótica (Antroposemiótica, Biosemiótica, Ethosemiótica, Metodologia semiótica, Semiótica – teóricos, Semiótica aplicada, Semiótica cognitiva, Semiótica da ação, Semiótica da cultura, Semiótica da Escola de Paris, Semiótica das mídias, Semiótica das paixões, Semiótica discursiva, Semiótica do sujeito, Semiótica literária, Semiótica peirciana, Semiótica peirciana – categorias, Semiótica social (sócio-semiótica), Semiótica tensiva e fenomenológica, Semiótica textual, Terminologia (Socioterminologia, terminografia, terminótica); Centros de estudo, formação e investigação semiótica; Pesquisadores em Estudos semióticos; Especificadores de suporte de informação.

O percurso investigativo realizado permitiu a incorporação metodológica sugerida por Mai (2011) e Foucault (2010) na estruturação de linguagens de indexação. O percurso realizado sugeriu que contemporaneamente há limites na adoção linear das garantias tradicionais (literária, estrutural e de uso) para a composição de linguagens de indexação, sobretudo no que concerne à semiose¹⁸ presumida na concepção de tais ferramentas.

Ao considerar as trocas informacionais em contexto como um processo autopoético mediado por práticas discursivas, foi possível ressaltar nesse percurso, não apenas o vocabulário sistemático gerado, mas compreender o regime dos objetos que ata as palavras e as coisas aos quais nos referimos nos complexos processos de produção e representação do conhecimento.

AGRADECIMENTOS

Esse trabalho contou com a valorosa contribuição do bolsista de iniciação científica Rubeniki Fernandes de Limas e o apoio do CNPq e da FAPEMIG aos quais dirijo os meus sinceros agradecimentos.

REFERÊNCIAS

ANDRADE, Cláudia Castro de. A fenomenologia da percepção a partir da autopoiesis de Humberto Maturana e Francisco Varela. *Griot – Revista de filosofia*, v.6, n.2, dezembro/2012. Disponível em <http://www.ufrb.edu.br>

¹⁸ Por semiose entende-se, estritamente, a produção de sentido, processo infinito pelo qual, através de sua relação com o objeto, o signo produz um interpretante que, por sua vez, é um signo que produz um interpretante e assim por diante. PINTO, Júlio. 1,2,3 da semiótica. Belo Horizonte: Ed. UFMG, 1995. p. 49.

ARTIÈRES, Philippe et al. *Les mots et les choses de Michel Foucault: regards critiques 1966-1968*. Paris: presses Universitaires de Caen, 2009.

BEGHTOL, Clare. Relationships in classificatory structure and meaning. In: Bean, Carol A. and Green, Rebecca, eds. *Relationships in the organization of knowledge*. Dordrecht: Kluwer Academic Publishers, 2001. p. 99-113.

BEGHTOL, Clare. Semantic validity: concepts of warrant in bibliographic classification systems. *Library resources & technical services* 30: p.109-25. 1986

FOUCAULT, Michel. *A arqueologia do saber*. Rio de Janeiro: Forense Universitária, 2010.

MAI, J.E. Folksonomies and the new order: authority in digital disorder. *Knowledge Organization*. 38, n. 2. 2011.p. 114-122.

NÖTH, Winfried. *A semiótica no século XX*. São Paulo: Annablume, 1996.

PINTO, Júlio. Semiótica e informação. Belo Horizonte, *Perspectiva em Ciência da Informação*, v.1, n.1, p.87-92, jan./jun.1996.

SANTAELLA, Lúcia. *O que é semiótica*. São Paulo: Brasiliense, 1986.114 p. (Coleção Primeiros Passos, 103)

SVENONIUS, Elaine. *The intellectual foundation of information organization*. Cambridge MA: MIT Press, 2000.

102

ZINS, Chaim. Classification Schemes of Information Science: Twenty-Eight Scholars Map the Field. *Journal of the American Society for Information Science and Technology*, New York, v. 58, n. 5, p. 645–672, Mar. 2007.

Representação de fotografias para publicidade em bancos de imagens: Princípios para análise

Joice Cleide Cardoso Ennes de Souza
joyce.cardoso@gmail.com
Instituto Brasileiro de Informação
em Ciência e Tecnologia

Rosali Fernandez de Souza
rosali@ibict.br
Instituto Brasileiro de Informação
em Ciência e Tecnologia

1 Introdução

A proliferação de imagens fotográficas na atual sociedade da informação é constatada pela grande produção de imagens digitais, gerando o questionamento de como representar esse acervo para futura recuperação. Insere-se nessa conjuntura a divulgação rápida de fotografias, que é observada pelo crescimento do número de comunidades virtuais e *sites* de bancos de imagens na *web*.

A disseminação de bancos de imagens *online* permitiu que o usuário pesquisasse remotamente, sem a mediação do profissional de informação. Na *web*, a localização de itens relevantes é de responsabilidade do usuário uma vez que ele faz a busca e avalia o resultado conforme suas necessidades informacionais. No âmbito da publicidade, há o aumento de buscas pelos aspectos subjetivos e/ou abstratos das imagens, consoante com a mensagem simbólica transmitida pelo anúncio publicitário. Para isso, a definição dos atributos para representação da fotografia é decisiva para que determinada imagem seja recuperada ou não pelo usuário. Em bancos de imagens comerciais, a não recuperação da fotografia desejada significa a não realização da transação comercial. Neste contexto, questiona-se: como as fotografias são indexadas nos bancos de imagens? Quais atributos devem ser adotados na representação de fotografias, para que as mesmas sejam recuperadas em bancos de imagem comerciais para uso na publicidade, permitindo sua reutilização em diferentes contextos?

Com base nos questionamentos, objetivou-se investigar procedimentos para a representação e recuperação de fotografias arquivadas em bancos de imagens, visando obter subsídios para reflexão e desenvolvimento de modelos para indexação de fotografia em arquivos digitais, para publicidade.

A análise dos pontos de acesso atribuídos às fotografias pelos analistas indexadores do banco de imagens *Latinstock* Brasil, selecionado para o estudo, foi

fundamentada na literatura que foca os atributos aplicados na representação de imagens (Shatford, 1986; Smit, 1996; Shatford Layne, 1994; Armitage; Enser, 1997; Greisdorf; O'Connor, 2002; Jorgensen, 1998; Markkula; Sormunem, 1998) e Ornager (1997), e na perspectiva subjetiva e interpretativa da imagem (Barthes, 2009; Jorgensen, 1996).

Partiu-se da proposição que a função persuasiva da linguagem publicitária oferece recursos a serem aplicados na análise conceitual de fotografias arquivadas em bancos de imagens na *web*. Uma vez que a persuasão publicitária se dá pelo destaque de valores, qualidades e emoções que são atribuídos aos produtos/serviços, de modo a estimular a compra pelo consumidor, entende-se que o publicitário é um pesquisador de imagens que integra uma comunidade de usuários cuja característica é a busca por fotografias que atendam ao objetivo persuasivo da mensagem publicitária.

2 Metodologia

Tomando como princípio os estudos sobre publicidade e linguagem da propaganda, foram determinados conceitos associados às fotografias usadas na publicidade, visando obter embasamento para as estratégias de busca efetuadas na parte empírica. Como o público-alvo da mensagem publicitária é diversificado, focou-se na publicidade direcionada ao público feminino por apresentar diferentes estudos na literatura (Swain, 2001; Kronbauer, 2011; Ribeiro, 2011). A figura feminina é explorada na imagem publicitária pelo seu poder de consumo pessoal e da família, tornando-se foco de anúncios publicitários que exaltam as qualidades e valores dos produtos.

Dentro do contexto da publicidade direcionada à mulher, foram identificados os elementos de análise: **beleza**, **juventude** e **moda** para a elaboração das estratégias de busca realizadas no *Latinstock* Brasil. Para a formulação das **estratégias de busca** usou-se a técnica da fatoração semântica, uma vez que se considerou o uso de conceitos/características que definem o termo pesquisado. Para o levantamento dos dados, foram realizadas buscas pelas combinações: mulheres+beleza, mulher+beleza, mulheres+juventude, mulher+juventude, mulheres+moda e mulher+moda, no singular e no plural, pois foi constatado que a pesquisa no *site Latinstock* era sensível a flexão de número.

A seleção do *Latinstock Brasil* considerou os aspectos: indexação atributiva na língua portuguesa, divulgação dos atributos aplicados na representação das imagens, publicação dos procedimentos usados na indexação, acesso aos dados descritivos e temáticos por parte do usuário e oferta de agrupamento de fotografias destinadas ao

mercado publicitário. A língua portuguesa foi um dos critérios de escolha adotado por facilitar a análise linguística realizada. Com a explicitação dos atributos e procedimentos adotados na indexação das fotografias pelo banco de imagens foi possível identificar os pontos de acesso considerados relevantes para o usuário. Com a consulta à descrição da imagem e às palavras-chave atribuídas observou-se o uso de conceitos subjetivos, abstratos e simbólicos, de acordo com os pressupostos da mensagem publicitária.

3 Banco De Imagens *Latinstock* Brasil

Através dos bancos de imagens na internet, as agências de imagens oferecem seu conteúdo imagético para usuários variados e, para atender as necessidades informacionais, suas coleções reúnem diferentes áreas do conhecimento, além de dar tratamento técnico na análise das imagens fotográficas. Diante das diferentes atividades realizadas no âmbito de bancos de imagens, adotou-se a definição de bancos de imagens como sendo “serviço técnico de uma instituição, que seleciona, adquire, organiza, armazena e permite a recuperação de imagens fotográficas de acordo com políticas e princípios pré-estabelecidos” (Rodrigues, 2011, p. 199).

O *Latinstock* Brasil faz parte do grupo *Latinstock*, e tem com objetivo administrar a distribuição de conteúdos de diferentes agências internacionais. A empresa possui representações em onze países da América Latina e está estabelecida no Brasil desde 1992. Dispõe de aproximadamente quarenta coleções com cerca de sete milhões de imagens fotográficas, imagens em movimento e trilhas sonoras, além de representar diferentes agências fotográficas internacionais, entre elas *Corbis*, *Magnun* e *Reuters*.

Uma vez que o foco do trabalho foi imagens usadas na publicidade, foram analisadas as coleções *Platinum* e *Publicidade*, por serem organizadas especificamente para o mercado publicitário.

No *link* Dicas para busca os administradores do *Latinstock* explicam como pesquisar e esclarecem o sistema usado na representação do acervo fotográfico. As fotografias são indexadas por: **Objeto** (quem? objeto fotografado? e assunto principal composto por substantivo e adjetivos), **Ação** (como? o que?), **Contexto/Conceito abstrato** (significado global da imagem) e **Forma expressiva** (análise técnica da imagem: enquadramento, exposição, foco, lente, luminosidade, posição da câmera). O *Latinstock* tem como política analisar genericamente a imagem, evitando o excesso de

subjetividade, e incentiva em sua página que o usuário empregue o uso de palavras-chave concretas para representar conceitos.

4 Indexação De Fotografias No Banco De Imagens *Latinstock Brasil*

Foram recuperadas 4072 fotografias na coleção *Platinun* e 24.840 na coleção *Publicidade*. Para a análise dos dados, selecionou-se 36 fotografias da coleção *Platinun* e 42 da coleção *Publicidade*, totalizando 78 fotografias. As fotografias foram organizadas em 12 amostras, que correspondem a cada estratégia de busca realizada com base nas combinações descritas acima. As amostras foram organizadas em quadros e colunas com informações sobre Descrição da imagem, Palavras-chave, Número de palavras-chave, Classificação das palavras-chave segundo atributos do site e Quantidade de termos por atributo.

Com base nas informações sobre a estrutura de indexação obtidas no *site* do *Latinstock*, classificou-se as palavras-chave segundo os atributos usados na indexação (Objeto, Ação, Contexto/Conceito abstrato, Forma expressiva), sendo realizado um levantamento quantitativo da incidência de palavras-chave por atributo, a fim de identificar qual atributo foi mais observado na indexação das fotografias.

Analisando estatisticamente as 78 fotografias recuperadas nas coleções estudadas, observou-se que foram atribuídas 1770 palavras-chave. Pela tabela 1 é possível inferir sobre a quantidade de palavras-chave atribuídas por atributo, pelos analistas indexadores do *Latinstock*.

Tabela 1 - Total de palavras-chave por atributo nas coleções *Platinum* e *Publicidade*

ATRIBUTOS	OBJETO	AÇÃO	CONTEXTO/CONCEITO ABSTRATO	FORMA EXPRESSIVA	TOTAL
NÚMERO DE PALAVRAS-CHAVE	1052	233	339	146	1770

Fonte: as autoras

Examinando a tabela 1 é possível verificar a grande oscilação numérica entre os atributos. **Objeto**, que representa Quem? ou O que? exibido na fotografia, se destaca com 59,4%, seguido de **Contexto/Conceito abstrato** com 19,1%. **Ação** vem em terceiro lugar com 13,1% e por último, **Forma expressiva** representando 8,2%. A partir da crítica dos dados, constatou-se que o reconhecimento visual do que é apresentado na fotografia é o item predominante na análise da imagem. A descrição da **Ação** desse

objeto seria um prosseguimento natural do processo, uma vez que a pergunta Como? segue ao questionamento Quem? ou O que?. Contudo, nas amostras examinadas, o atributo **Contexto/Conceito abstrato** surgiu como o segundo aspecto na análise das fotografias do banco de imagens *Latinstock*. Tal fato não é surpreendente uma vez que as coleções estudadas são direcionadas ao mercado publicitário, que se caracteriza pela transmissão de valores, sentimentos e qualidades, que são conceitos subjetivos, em suas mensagens.

Outro aspecto identificado foi o número de 21 palavras-chave atribuídas por fotografia, com variação de 9 (valor menor) a 66 (valor maior) palavras-chave atribuídas às fotografias.

5 Considerações Finais

Da análise realizada concluiu-se que a indexação de imagens para publicidade compreende: **Atributos técnicos** - recursos técnicos da imagem, como composição, ponto focal, orientação, perspectiva, foco, textura, cor, formato; **Atributos figurativos** - elementos percebidos visualmente: Objetos (pessoa/coisa); lugar/localização; ação (atividade, evento); e **Atributos subjetivos** - aspecto simbólico e temático da fotografia: valor simbólico (representação, abstração, tema); ambiente (sentimento geral, atmosfera) e estímulos emocionais (reação individual, emoções).

Constatou-se que a atribuição de elementos figurativos e/ou simbólicos na indexação de fotografias em publicidade caracteriza o tipo de recuperação por parte do usuário. Os usuários de publicidade buscam fotografias em bancos de imagens para atenderem uma necessidade que usa a imagem como objeto, por exemplo, em produtos dos usuários como capas de livro.

O número de palavras-chave atribuído por fotografia indica que a política de exaustividade na indexação por parte do *Latinstock*. Contudo, a precisão mercadológica compreende a localização da fotografia desejada e sua compra, exigindo precisão na recuperação, de modo a evitar que documentos inúteis sejam recuperados.

A proposta deste trabalho, ao aproximar Publicidade e Ciência da Informação, foi lançar novas formas de pensar a indexação de fotografias arquivadas em bancos de imagens comerciais, proporcionando uma reflexão sobre a atuação do profissional de informação, conscientizando-o de sua importância na sociedade atual, onde a produção da imagem fotográfica é significativa, exigindo representação adequada para sua efetiva recuperação.

Referências

- Armitage, L.; Enser, P. Analysis of user need in image archives. **Journal of Information Science**, v. 23, n. 4, 287-299, 1997.
- Barthes, R. **O óbvio e o obtuso**. Lisboa: Edições 70, 2009.
- Greisdorf, H.; O' Connor, B. Modelling what users see when they look at images: a cognitive viewpoint. **Journal of Documentation**, v. 58, n. 1, p. 6-29, 2002.
- Jørgensen, C. Attributes of images in describing tasks. **Information Processing & Management**, 34 (2/3), 161-174. 1998.
- Jørgensen, C. Indexing images: testing an image description template. In: Asis Annual Conference, 1996, Baltimore. **Proceedings...** [S.l.: s.n.], 1996.
- Kronbauer, Ana Pauline Schäfer. **Publicidade impressa direcionada ao público feminino: valores e conceitos**. 2011. Trabalho de Conclusão de Curso (Graduação em Comunicação Social) – Universidade Regional do Noroeste do Estado do Rio Grande do Sul, Ijuí, 2011. Disponível em: <<http://bibliodigital.unijui.edu.br:8080/xmlui/bitstream/handle/123456789/789/Monografia%20PDF.pdf?sequence=1>>. Acesso em: 12 set 2012.
- Markkula, M.; Sormunen, E. Searching for photos: journalists' practices in pictorial IR. In: **The Challenge of Image Retrieval Research Workshop**, 1998, Newcastle. Disponível em: <http://www.bcs.org/upload/pdf/ewic_im98_paper8.pdf>. Acesso em: 12 jul 2012.
- Ørnager, S. Image Retrieval: theoretical analysis and empirical user studies on accessing information images. In: Asis Annual Meeting, 60., 1997, Washington. **Proceedings...** Medford: ASIS, 1997.
- Ribeiro, Marislei da Silveira. **Beleza feminina e publicidade: um estudo sobre as campanhas da marca Dove**. 2011. Tese (Doutorado em Comunicação Social) - Pontifícia Universidade Católica do Rio Grande do Sul, Porto Alegre, 2011.
- Rodrigues, R. C. **Análise e tematização da imagem fotográfica: determinação, delimitação e direcionamento dos discursos da imagem**. 2011. Tese (Doutorado em Ciência da Informação) - UnB, Brasília, 2011.
- Shatford, S. Analyzing the subject of a picture: a theoretical approach. **Cataloging and Classification Quarterly**, v. 6, n. 3, p. 39-62, 1986.
- Shatford Layne, S. Some issues in the indexing of images. **Journal of the American Society of Information Science**, v. 45, n. 8, 583-588, 1994.
- Smit, J. W. A representação da imagem. **Informare: Cadernos do programa de Pós-Graduação em Ciência da Informação**, v.2, n.2, p.28-36, jul./dez. 1996.
- Swain, Tania Navarro. Feminismo e recortes do tempo presente: mulheres em revistas femininas. **São Paulo em Perspectiva**, v.15, n.3, p. 67-81, 2001.

A dimensão aplicada da ORC

Consideraciones para construir un tesoro sobre leyes político-electorales usando tecnología digital

Catalina Naumis Peña
Universidad Nacional Autónoma de México
naumis@unam.mx

Ariel Antonio Morán Reyes
Universidad Nacional Autónoma de México
a.moran@comunidad.unam.mx

Rosa Atzimba Morales Monroy
Universidad Nacional Autónoma de México
rosa.morales@ife.org.mx

Introducción

El tesoro ha ido evolucionando como una herramienta de recuperación de información inestimable en los procesos de organización del conocimiento. Las designaciones terminológicas usadas para definir esa evolución no tienen un significado clarificador. En este caso la informática como disciplina que ha otorgado una funcionalidad diferente al tesoro ha impuesto su terminología, creando bastante confusión. En general, la difusión mayor para distinguir las dos grandes etapas han sido para los siguientes términos: tesauros terminológicos y tesauros conceptuales. Planteado así parecería que con la primera designación los tesauros no trabajarían con conceptos. Nada más alejado de la realidad porque en la organización del conocimiento se ha buscado desde un principio, la representación del contenido documental mediante conceptos y no los términos presentes en los documentos. A pesar de la opinión de muchos de los autores citados a lo largo de este trabajo se utilizará el término tesauros documentales, en lugar de tesauros terminológicos. Cesteurs, Smith y Goldberg critican la denominación de “entidades conceptuales” en el uso de los tesauros y, aún más, definirlos como “entidades abstractas que representan conceptos” (2005: 503).

No se deja de reconocer el beneficio que ha otorgado a los sistemas de información el manejo tecnológico del tesoro documental. Su desarrollo se ha ido adecuando paulatinamente a los cambios que las tecnologías de la información suponen, aunque su aplicación en la Web ha resultado hasta ahora poco fructífera. Muchos

trabajos están estudiando la relación de la semántica y las tecnologías, principalmente las representaciones del espacio semántico.

El espacio semántico es el medio en el que se involucran mínimo dos personas, las cuales pueden tener diferentes percepciones ya sea de lengua, cultura, posición política, relación familiar, nivel educativo, entre otros, por lo que es un espacio de intercambio de valores y significados, que inciden en las conductas informacionales. Por lo tanto es muy difícil transmitir el mensaje con claridad y precisión debido a que en el proceso del emisor al receptor pueden existir muchas interferencias que hagan más complicada y un poco distorsionada la comunicación. Desde una perspectiva pragmática, en la actualidad encontramos múltiples modelos y propuestas en que se aplican estos principios. Como muestra está la recuperación de información en Internet a partir de motores de búsqueda. En este caso, la inferencia en un sistema y un espacio semántico consistirían en vincular la información que se encuentra latente en la red y en las bases de datos con las necesidades de información del usuario de manera que se logre establecer la mayor precisión y relevancia en la búsqueda realizada. A esto se le denomina indización semántica latente (Erk, 2009: 106).

Los tesauros se pueden utilizar en diferentes situaciones o escenarios de aplicación, aunque en general cumplen la función de homogeneizar terminologías, controlar vocabularios o auxiliar al usuario para navegar por un espacio de conceptos. A pesar de la larga trayectoria de investigación acerca del mantenimiento de la calidad de los tesauros, hasta este momento se ha prestado poca atención a la interacción entre las especificidades estructurales de un tesoro y la calidad de los resultados en relación con los diferentes escenarios de aplicación de un tesoro. Si bien hay varias iniciativas que se centran en la calidad de los tesauros, de los metadatos (en términos de expresividad) y en la solidez estructural.

La Web semántica supone una oportunidad para utilizar los tesauros en los procesos de descripción y recuperación de información. Para muestra de ello existen múltiples iniciativas de representación de tesauros mediante RDF (Resource Description Framework)/XML (Extensible Markup Language), que han dado lugar al desarrollo de SKOS (Simple Knowledge Information System), un modelo para esquemas conceptuales aplicables en la Web semántica y que actualmente se encuentra en desarrollo por el World Wide Web Consortium (W3C). La organización del conocimiento más novedosa tiene por objeto combinar diferentes sistemas de organización del conocimiento en Internet mediante metadatos compartidos y

ontologías formales que funcionan a partir de los principios de la lógica deductiva. Inicialmente, se diseñaron modelos que seguían la lógica conceptual clásica que seguía los planteamientos neopositivistas, por lo que buscó utilizar un metalenguaje, o sea, un conjunto de instrumentos simbólicos y conceptuales independientes de las lenguas naturales para conseguir una comunicación sin ambigüedad en temas especializados.

Es necesario puntualizar que el desarrollo de un vocabulario RDF/XML para representar tesauros supone explotar de manera innovadora el modelo canónico que nos proponen las normas sobre construcción y mantenimiento de tesauros: internacionales (ISO 2788 y 5964), norteamericana (ANSI/NISO Z39.19), británica (BS 8723) y la institucional de la IFLA, vertida en la *Guidelines for Multilingual Thesauri*. Cabe mencionar que la norteamericana sólo contempla vocabularios controlados monolingües y que la británica incluye estándares sobre control de autoridades y una norma sobre mapas temáticos. Estas normas describen un tesoro como un conjunto de términos de diverso tipo entre los que se establecen relaciones semánticas, en un contexto que le otorga sentido. No obstante, la Web precisa de una visión en la que se comparta información, como por ejemplo que las jerarquías incluyan instancias, atributos y meronimias distinguidos entre sí, otro de los servicios a destacar es una mayor riqueza en las asociaciones asociativas. Existen, por lo tanto, múltiples iniciativas dirigidas a la definición de vocabularios bajo esquemas de RDF para sustentar tesauros.

Pese a que algunos autores hablan de los tesauros basados en la gestión de datos enlazados, abiertos y cerrados, lo anterior se traduce, como ya se explicó en el manejo de dos modelos básicos para la estructuración de tesauros: uno documental de relaciones en un espacio semántico (el clásico) y otros denominados tesauros conceptuales o informatizados porque existe:

[...] una navegación directamente en el documento, en una red semántica conceptual por la que se ingresa de términos más genéricos de una faceta hacia los más específicos, e inversamente (navegación vertical), que a su vez permite la transición de una clase hacia otra y de un campo de la ciencia hacia otro mediante las relaciones asociativas (navegación horizontal) por nudos polijerárquicos basados en conceptos (Deweze, 1998. p. 48).

En este segundo tipo de tesauros que son de carácter automatizado en la construcción y en la recuperación de información se aprecia un incremento pragmático de las relaciones en especial las de asociación, y se reflejan también ideas combinadas que se vinculan entre sí con una red de relaciones semánticas de distinto tipo, principalmente jerárquicas (término), asociativas (termino relacionado) y de equivalencia (término equivalente, término preferente) (Tudhope, Harith & Jones,

2001). En esta segunda alternativa se concibe al tesauro como un conjunto de conceptos entre los que se establecen relaciones semánticas similares a las que se establecen en los tesauros documentales. Los conceptos se vinculan con términos, a modo de etiquetas léxicas, entre los que pueden o no establecer relaciones léxicas que es la propuesta del modelo SKOS.

La red semántica establecida por los tesauros conceptuales no se agota en la explicitación de las relaciones conceptuales anteriormente mencionadas, sino que incluye las denominadas notas de alcance donde se acota el significado o alcance conceptual que es representado por un término en particular. Sacco y Bothorel arguyen al respecto que:

El objetivo usual de un tesauro es guiar al indizador y al especialista en recuperación de información para escoger un mismo término para un mismo concepto. Los conceptos son representados por términos, y para cada concepto es seleccionada una de las representaciones posibles como el término preferido (2010: 8).

Los tesauros documentales posibilitan la representación de esquemas más compactos y cercanos al modelo tradicional, pero los tesauros basados en conceptos relacionados automáticamente, que podríamos denominar, como se propuso anteriormente, tesauros informatizados permiten navegar en las estructuras de información, aunque en un principio su comprensión no resulte tan intuitiva como los primeros. (Cfr. Jones, 1993). La estructuración de los tesauros documentales se limita a los tipos de relaciones que se encuentran en los mismos términos. Además, en los tesauros informatizados está presente la elección de los descriptores (o sea, los términos que representan a los conceptos), entendiéndose que es una elección de forma, es decir, que los descriptores sean compuestos (formados por varios conceptos) o simples (formados por un único concepto). No obstante, un primer problema que se presenta con los tesauros informatizados es la necesidad de dotar a los usuarios de estructuras de conocimiento que le permitan un acceso rápido y amigable a la información. Los usuarios requieren de herramientas conceptuales y semánticas destinadas a una efectiva organización de la información digital, y los tesauros son una de las estructuras que pueden tener una participación activa en este sentido.

El cambio de paradigma en la modelización de tesauros supone un gran esfuerzo aunque dicho cambio se ve compensado con creces al optimizarse los procesos de elaboración, mantenimiento y productos derivados de la aplicación del tesauro en la indización de documentos (Pastor Sánchez, Martínez Méndez & Rodríguez Muñoz,

2009). La principal razón para este cambio se debe a que los tesauros documentales conllevan problemas de aplicación en procesos de indización cuando se realizan operaciones de cambio de términos preferentes y términos no preferentes. De esta manera, cuando un descriptor con el que se ha indizado un documento pasa a ser un no descriptor hay que realizar una reorganización de los índices que asocian dicho descriptor con los documentos. Hoy día es más que necesario contar con una visión amplia en la elaboración de herramientas conceptuales, sin circunscribirnos al modelo clásico de construcción de tesauros, sino considerando otros tipos de estructuraciones conceptuales como las taxonomías o las ontologías. Una ontología como recurso lingüístico tiene claros elementos de contacto con la construcción de herramientas documentales de indización y recuperación de información, ya que ofrece una estructuración formal y más eficaz de la información. Pastor Sánchez define el tesoro actual como una especie del género ontologías.

1. Los Sistemas de Información Jurídica

La información jurídica por su naturaleza posee peculiaridades determinantes para influir en el planeamiento de su organización. El hecho de que la información mencionada tenga un papel relevante radica en que el mensaje legal está integrado de un conjunto de expresiones normativas, ordenadas para indicar acciones humanas en un contexto social (global, general e inmediato) y formuladas por medio de signos que, generalmente, no van de forma definida del remitente (que son los legisladores) al receptor (que es la mayoría de las veces un individuo desconocido específicamente). La naturaleza de la documentación jurídica no hace fácil la tarea de elaborar un tesoro o cualquier tarea de control de vocabulario, ya que los descriptores evocan ideas categóricas, bajo las cuales se repiten las mismas clases en diferentes categorías y no un contexto lingüístico determinado, es decir su naturaleza está enmarcada en un ámbito socio-cultural-político con el lenguaje jurídico.

A diferencia de la mayoría de las otras disciplinas, el derecho no utiliza un lenguaje puramente descriptivo o preciso. El lenguaje jurídico no describe al derecho como objeto, sino en gran medida al derecho mismo; por otra parte, cada uno de los términos empleados no corresponde biunívocamente a una realidad o a un objeto: la ambivalencia estriba en la naturaleza profunda del derecho, a lo cual difícilmente se presenta objeción. Las palabras que integran los documentos jurídicos no pueden ser

consideradas unidades fundamentales con un sentido cada una. Así, cada documento se caracteriza de manera única por las palabras utilizadas (a causa del estilo, de la sintaxis y del vocabulario jurídico) y se emplean esos términos en su sentido más general. El estilo jurídico recurre con frecuencia a comparaciones o efectos que pueden hacer intervenir una noción análoga en el curso de un texto; pero también puede sugerir una noción o situación sin que el término que señala más habitualmente aparezca en realidad.

Ante el proyecto de desarrollo de un tesoro político-electoral, es necesario precisar que la salvedad descrita anteriormente vuelve sumamente compleja la creación de un sistema jurídico documental, y mucho más para un lenguaje controlado especializado en política electoral del sistema jurídico, ya que estaría compuesto de mensajes normativos vigentes y en constante cambio, que tienen que ser recuperados cuando sea pertinente a la necesidad del usuario, cuya situación se encuentra controlada por la norma (Rosales Salinas, 2012: 113).

Otra peculiaridad de la información jurídica radica en el hecho de que un concepto legal posee un contenido prescriptivo, de tal suerte que la función social de una ley o decreto emanado de la actividad parlamentaria requiere que tal mensaje sea reconocido por el receptor para que la pueda aplicar. Es decir, uno de los mayores problemas de los lenguajes jurídicos no es de carácter técnico sino de índole social (Perales Ojeda, 1981: 37). En toda búsqueda documental, incluso la más elemental, encontramos documentos que no son relevantes para el tema y, por el otro lado, documentos que siendo de una gran significación no son descubiertos, para lo cual se necesita un esquema de representación adecuado que combata “accidentes lingüísticos” como la polisemia (Matute Cifuentes, 1989: 111). Por tal razón, el control del lenguaje jurídico constituye una dura tarea, debido a la naturaleza híbrida de este lenguaje, cuyo contenido, como ya se dijo, es expresamente prescriptivo.

Xavier Linant explica que la elaboración de un tesoro para un sistema de información jurídica se alimenta por procesos tales como la investigación (servicios profesionales y el ámbito académico), las estadísticas para la coordinación política, los productos de la industria editorial especializados en la rama, y la colaboración entre la jurisprudencia (papel de los abogados) y la gestión de la información (tareas del bibliotecólogo). Nosotros agregaríamos a esto un elemento decisivo y determinante: la tecnología emergente. Sin embargo, estas relaciones son intrincadas; la única fuente que permea directamente en el tesoro son los trabajos de los legisladores, o sea, las leyes.

Por lo que la revisión de éstas es imperiosa. Un legislador puede innovar o revivir un concepto jurídico y este tendrá que tener su representatividad en el vocabulario irremediabilmente. De lo discurrido anteriormente, se puede observar en la FIGURA 1 los diversos escenarios de competencia informativa en donde interviene el legislador, quien interactúa con diferentes sistemas.

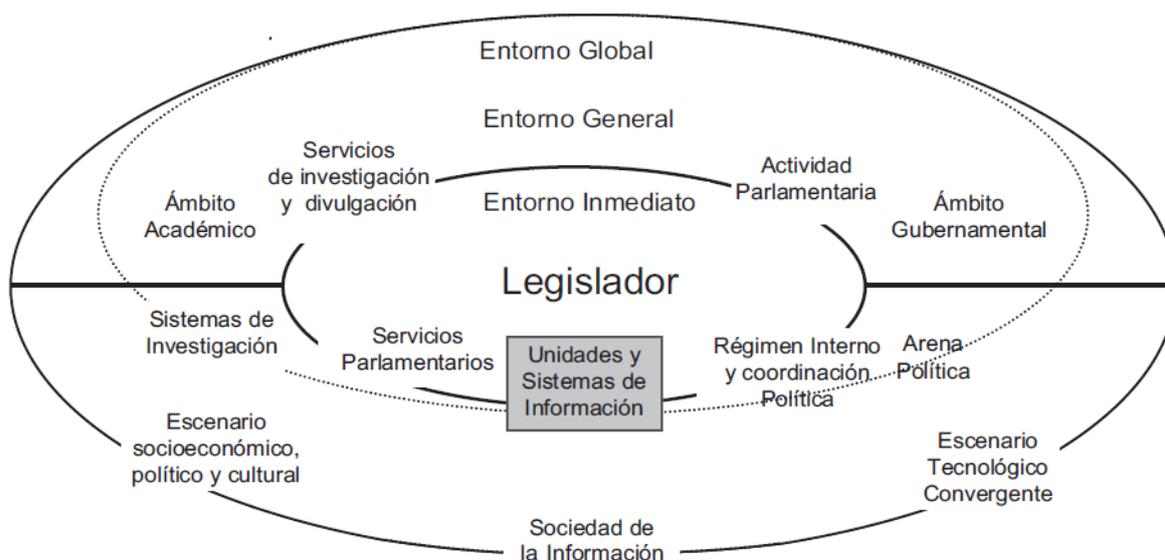


FIGURA 1

Del gráfico previo podemos aducir que la sociedad de la información, con el escenario *tecnológico convergente* y el *socioeconómico político y cultural*, engloban y determinan, a su vez, al escenario descrito anteriormente, así como a los entornos global, general e inmediato, es decir, es un macro entorno. Gran parte de este entorno se encuentra inmerso en el medio digital. Por ende, es posible la propuesta de aplicación de tesauros orientada hacia representar información usando la Web semántica, planteando la constitución de un modelo dirigido al desarrollo de aplicaciones web para la gestión de lenguajes controlados.

Las tecnologías actualmente utilizadas para la elaboración de la mayor parte de los contenidos Web carecen de estructuras que permitan una representación adecuada de los aspectos semánticos. Los motores de búsqueda utilizan técnicas basadas en la frecuencia de aparición de palabras en los procesos de recuperación, por lo que las nuevas herramientas asociadas a la Web 2.0 han hecho patentes ciertos problemas inherentes a este modelo de recuperación (Touminen, 2008). Generalmente, en los sistemas jurídicos el análisis de contenido se realiza mediante la medición del grado de

similitud semántico-sintáctica entre los diversos documentos fuente para lograr mayor confiabilidad en las respuestas en la búsqueda de información.

La Web semántica parece constituir una solución viable, ya que realiza una búsqueda similar, fundamentada en el uso de tecnologías como el lenguaje XML, los esquemas RDF y el lenguaje de marcado OWL (en su vertiente Full), en donde los metadatos y las ontologías constituyen el núcleo conceptual de desarrollo. No obstante, la siguiente declaración de Floridi debe servir como punto de inflexión para los profesionales de la información y para tomar con una postura crítica la propuesta de los tesauros informatizados para recuperar información en el entorno digital:

Demasiada retórica y muy poco de detalle hacen de la Web semántica un proyecto “conceptualmente” embrollado. Conceptos clave como “semántica”, “significado”, “comprensión”, “información”, “conocimiento” e “inteligencia”, son generosamente esparcidos en la literatura concerniente a la Web semántica, en donde parece que este calificativo es usado de manera holgada o simplemente de forma metafórica. La realidad es que los lenguajes, los protocolos y las ontologías para metadatos y metasintaxis sí pueden producir integración, sindicación y la capacidad de agregar, compartir y discernir entre datos de temas heterogéneos, pero esto por la interacción de diferentes bases de datos. Incluso en lo virtual, no parece haber nada de “semántico”. En 1997, por ejemplo, el W3C definió la primera especificación del RDF, lo cual se convirtió en una recomendación para 1999. RDF provee un lenguaje basado en tripletes de representación para identificadores Universales del Recurso (URI). Pero ninguna significación o inteligencia juega algún papel en esto (Floridi, 2009: 29).

La descripción de la semántica de los recursos de información resulta bastante compleja debido al estado de desarrollo actual de la Web, pero es posible dar un paso previo al realizar una descripción y caracterización de dichos recursos empleando metadatos y esquemas conceptuales, tales como los tesauros (Jones, 1993: 168). Un tesoro, de entrada, maneja una semántica simple tan sólo manejar una clasificación de sus descriptores, relacionándolos de forma jerárquica por generalización-especialización.

Para ello, se hace necesario definir las ventajas de los tesauros documentales y los informatizados. Ante lo cual, existen múltiples iniciativas de representación de tesauros que han dado lugar al desarrollo de SKOS, un modelo para esquemas conceptuales aplicables en la Web semántica y que actualmente se encuentra en desarrollo por el W3C.

2. El modelo de SKOS (Simple Knowledge Information System)

SKOS es una iniciativa de W3C en forma de aplicación de RDF que proporciona un modelo para representar la estructura básica y el contenido de esquemas conceptuales como listas de encabezamientos de materia, taxonomías, tablas de clasificación, tesauros o cualquier tipo de vocabulario controlado o lenguaje especializado. SKOS es un modelo que permite representar la estructura básica y el contenido de estos esquemas conceptuales, diseñado para crear nuevos sistemas de organización adaptándolos al uso en la Web semántica, ya que permite estructurar los datos en forma de tríadas que pueden ser codificadas en cualquier sintaxis válida para esquemas de RDF. En SKOS cada concepto se identifica con una cadena de caracteres de tipo URI (Uniform Resource Identifier). Estos conceptos pueden etiquetarse en cadenas de texto en uno o varios idiomas y estructurarse a través de relaciones de diversa tipología. Al respecto, Floridi nos recuerda que el idioma humano prospera al utilizar un mismo término para querer decir cosas diferentes, sin embargo esto es difícil en los procesos de automatización. Es por ello que usar a un URI diferente para cada concepto específico soluciona ese problema. El autor italiano nos indica que utilizar estos identificadores es como usar una dirección postal que puede ser distinguida de alguna que sea una dirección callejera, y ambas pueden ser distinguidas de una dirección que sea un discurso (Floridi, 2009: 28).

El RDF es un marco de trabajo que define a un conjunto estandarizado de conceptos y metadatos para representar la información en la web. Con RDF es posible caracterizar ciertas propiedades de conceptos de un dominio concreto al tiempo que es posible jerarquizar dichos conceptos. Aun así, es necesario alcanzar mayor grado de expresividad y abstracción para representar funciones y axiomas de una ontología. Y es que el modelo de datos SKOS es en realidad una ontología definida con OWL, que es un lenguaje de marcado para publicar y compartir datos usando ontologías en internet. Está diseñado expresamente para su uso en aplicaciones que requieran realizar funciones de procesamiento del contenido informativo y no limitarse únicamente a la presentación de la misma para su consulta por parte de personas.

SKOS puede ser utilizado conjuntamente con OWL para expresar formalmente estructuras de conocimiento sobre un dominio concreto, ya que no puede realizar esta función al no tratarse de un lenguaje para la representación de conocimiento formal (Pastor Sánchez, Martínez Méndez & Rodríguez Muñoz, 2009). La estructuración del

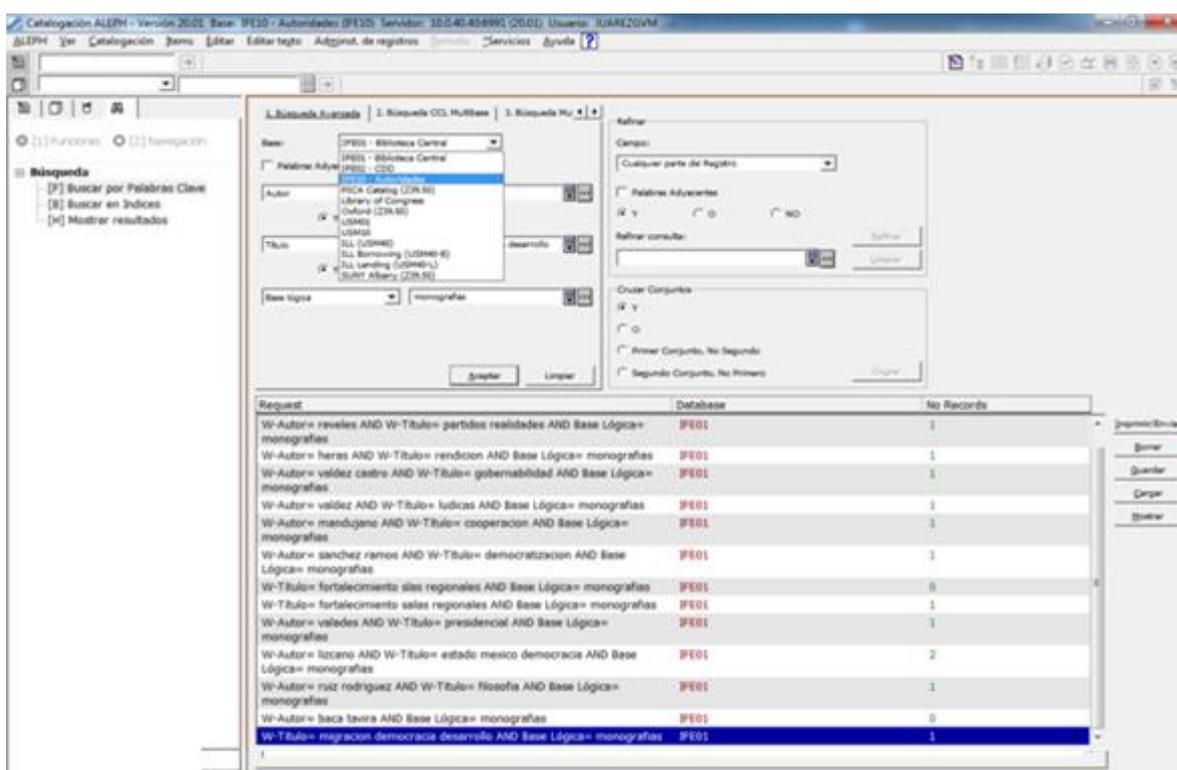
vocabulario SKOS se da en lenguaje XML, lo que permite el almacenamiento de estructuras de información en sistemas de gestión de información basados en aplicaciones Web, para el intercambio de datos entre aplicaciones y bases de datos. El vocabulario SKOS se divide en siete elementos o aspectos (TABLA 1), cada uno subdividido en etiquetas, mismas que identifican un concepto con una expresión del lenguaje natural. Un concepto se puede denotar mediante diversas etiquetas (dando expresión a los sinónimos), mientras que una etiqueta puede significar varios conceptos (dando expresión a los homónimos).

TABLA 1

Conceptos y esquemas de conceptos	skos:Concept skos:ConceptScheme skos:inScheme skos:hasTopConcept
Etiquetas léxicas	skos:prefLabel skos:altLabel skos:hiddenLabel
Relaciones semánticas	skos:semanticRelation skos:broaderTransitive skos:narrowerTransitive
La documentación	skos:note skos:ScopeNote skos:historyNote skos:changeNote skos:definition skos:editorialNote skos:example
Colecciones de conceptos	skos:Collection skos:OrderedCollection skos:member skos:memberList
Notaciones	skos:notation
Propiedades de mapeado	skos:mappingRelation skos:exactMatch skos:broadMatch skos:narrowerMatch skos:relatedMatch

3. Evaluación del software

Luego de una breve descripción del modelo SKOS es necesario evaluar varios software o paquetes que destaquen ciertas ventajas fundamentales para gestionar las relaciones y los formatos de visualización, como la capacidad de exportación de ficheros a otros sistemas. Dicha evaluación se centró en las necesidades del Instituto Federal Electoral, es decir, no se tenía contemplado un costo muy alto para la compra de un software y se necesitaba que fuera multilingüe y con una interfaz monousuario. Actualmente, el tesoro del IFE disponible en línea está estructurado de forma peculiar bajo el módulo de Aleph (FIGURA 2). Los términos de indización fueron asentados en las Autoridades del módulo, tendente a encabezamientos de materia. Por lo tanto, más que de un tesoro en términos propios, se cuenta con una lista de encabezamientos de materia “tesaurizados”.



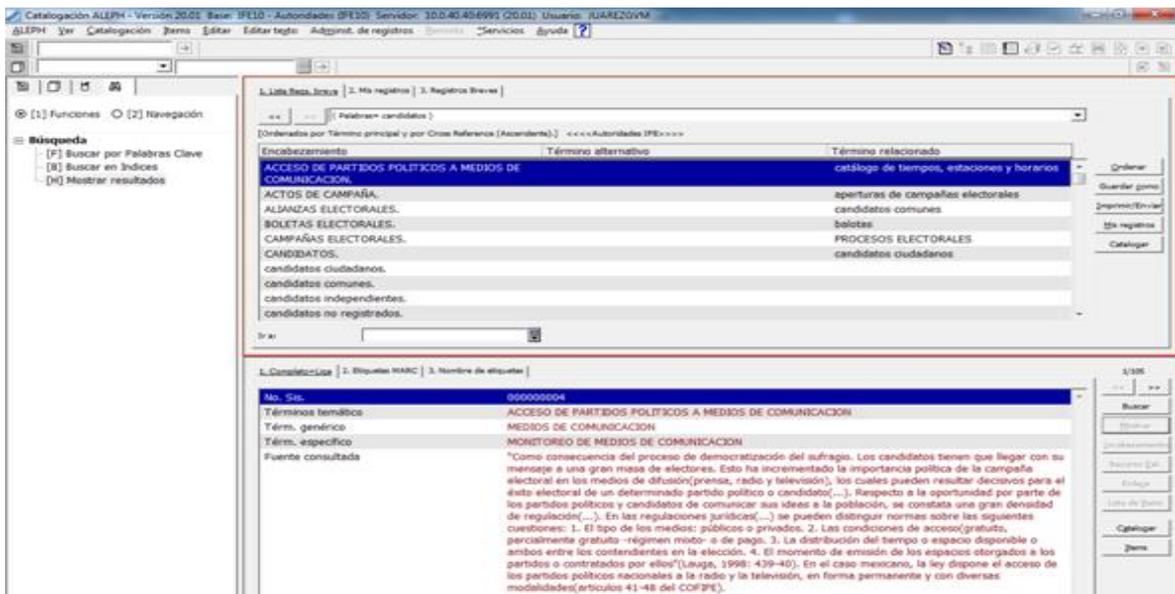


FIGURA 2

En la evaluación se consideró:

- ❖ Libre distribución
- ❖ Criterios
 - Funcionalidad
 - Herramientas conceptuales soportadas
 - Integridad
- ❖ Formatos de representación de tesauros (SKOS/RDF, XML)
- ❖ Distribución como software reusable (paquetes, servicios)

El aspecto que tiene que ver con los formatos de representación es de suma importancia, ya que la vinculación con bases de datos e integración de información depende de esto. Tanto el lenguaje XML como el modelo SKOS fueron creados por el W3C. Básicamente, se evaluaron cuatro softwares: TemaTres, Term Tree, PoolParty y ThManager. Los primeros están basados en SQL y en las normas de la American National Standards Institute, la NISO (National Information Standards Organization), y de la ISO (International Organization for Standardization), es decir son softwares para estructurar tesauros basados en términos. PoolParty, por su parte, es un sistema de gestión para tesauros diseñado para el uso y representación semántica en la web, principalmente en las funciones de análisis de texto. Este sistema ayuda a construir y

mantener tesauros multilingües ya que proporciona una interfaz de usuario sencilla y un par de servicios semánticos (buscadores semánticos y la navegación facetada), además de que hace el uso de las especificaciones de RDF/SKOS. El principal problema con PoolParty es que es sumamente inflexible en términos de una falta de escenarios de integración cambiantes en los actuales sistemas de información jurídica. De hecho no encontramos algún tesoro de tipo jurídico estructurado con PoolParty. No obstante, sí encontramos algunos con ThManager, como el *Tesoro de la UNESCO* y *La Nomenclatura Internacional de la UNESCO para los campos de Ciencia y Tecnología*.

Pero, específicamente, en cuanto a sistemas de información jurídica, es de destacar el tesoro multidisciplinar de la Unión Europea, es decir el *Tesoro multilingüe EuroVoc*, el cual sirve al Parlamento Europeo y la Oficina de Publicaciones Oficiales de las Comunidades Europeas. Está estructurado en 21 campos temáticos y 127 microtesauros. Algunos descriptores de los campos temáticos 72 (los geográficos) y 76 (las Organizaciones internacionales) son multi-jerárquicos, tienen varios términos genéricos de primer nivel diferentes. Eurovoc contiene 6645 descriptores, 519 de los cuales son *top terms*, 7756 no descriptores, 6669 relaciones jerárquicas recíprocas y 891 notas (Alvite Díez *et al.*, 2010: 48).

Al analizar la arquitectura de ThManager nos percatamos que diseñado para sistemas operativos Windows de 32 bits, es ligero y sencillo en cuanto al uso de sus herramientas, además que la descarga es gratuita. Es una aplicación de escritorio que requiere 100 megabytes (MB) de espacio libre en el disco duro. Con respecto a los requisitos de CPU y de memoria de acceso aleatorio (RAM), estos dependen en gran medida del tamaño y el número de tesauros cargados en la herramienta. De hecho, el software trae instalados 12 tesauros reconocidos elaborados con SKOS, entre ellos el de la UNESCO o el de AGROVOC (FIGURA 3). Este software permite reutilizar esos tesauros y modificarlos para elaborar uno nuevo o se puede comenzar uno desde un proceso en blanco. Teniendo en cuenta el número y el tamaño de tesauros utilizados como banco de pruebas, los rangos de consumo de memoria RAM pueden oscilar de 256 a 512 MB cada uno, y con un CPU de 3 Gigahertz (por ejemplo un procesador Pentium IV), los tiempos de carga para los tesauros son más que aceptables. Sin embargo, si el tamaño de los tesauros es más pequeño, se puede operar con un CPU de 1 GHz (por ejemplo un procesador Pentium III) y con 128 MB de memoria RAM.

Estándar	Id único	Idioma	Título	Tipo	Tema
DC: Thesaurus	AGROVOC	Inglés	AGROVOC	TEXT.REFEREN...	POLITICS, LAW A...
DC: Thesaurus	DCType	Inglés	[SYSTEM CONTROLLED LIST, DO NOT DELETE]. Types of r...	TEXT.REFEREN...	INFORMATION A...
DC: Thesaurus	DataLangCode	Inglés	[SYSTEM CONTROLLED LIST, DO NOT DELETE]. Data langu...	TEXT.REFEREN...	CULTURE.LINGUI...
DC: Thesaurus	EuropeanTerrit...	Inglés	Thesaurus of territorial units of Spain and France	TEXT.REFEREN...	COUNTRIES AND ...
DC: Thesaurus	GEMET	Inglés	General Multilingual Environmental Thesaurus	TEXT.REFEREN...	SCIENCE.NATUR...
DC: Thesaurus	ISO639	Inglés	Codes for the representation of names of languages	TEXT.REFEREN...	CULTURE.LINGUI...
DC: Thesaurus	MD_Restriction...	Inglés	[SYSTEM CONTROLLED LIST, DO NOT DELETE]. Metadata r...	TEXT.REFEREN...	POLITICS, LAW A...
DC: Thesaurus	MdLangCode	Inglés	[SYSTEM CONTROLLED LIST, DO NOT DELETE]. Metadata l...	TEXT.REFEREN...	CULTURE.LINGUI...
DC: Thesaurus	THES_NOTATI...	Inglés	[SYSTEM CONTROLLED LIST, DO NOT DELETE]. Values of h...	TEXT.REFEREN...	
DC: Thesaurus	THES_REL_LAN...	Inglés	[SYSTEM CONTROLLED LIST, DO NOT DELETE]. SKOS Relat...	TEXT.REFEREN...	
DC: Thesaurus	THES_REL_TYPE	Inglés	[SYSTEM CONTROLLED LIST, DO NOT DELETE]. SKOS Relat...	TEXT.REFEREN...	
DC: Thesaurus	UNESCO	Inglés	UNESCO thesaurus	TEXT.REFEREN...	EDUCATION

FIGURA 3

Inicialmente, SKOS trabaja en lenguaje JavaScript, pero a diferencia de los anteriores programas, permite transformar el tesauo a formato SKOS, por lo que se exporta el tesauo a lenguaje XML. Ciertamente, ThManager es un software de uso simple, empero, es imperioso y necesario conocer ciertas herramientas del software para diseñar un tesauo y poder exportar el contenido. Este programa se compone básicamente de tres capas: la primera, una capa de depósito donde se almacenan los tesauos y son identificados por medio de metadatos; en segundo lugar, una capa de persistencia que proporciona una interfaz de programación de aplicaciones para acceder a los tesauos almacenados en el repositorio, y, en tercer lugar, una interfaz gráfica de usuario que ofrece diferentes componentes gráficos para visualizar los tesauos, para buscarlos por sus propiedades, y para modificarlos de diferentes maneras (Lacasta, *et al.*, 2007: 43).

A estas alturas, se hace necesario explicar cómo se estructura ese tesauo, es decir cuáles son los principios y relaciones con los que se constituye antes de ser almacenado. Es importante aclarar que, antes de introducir términos y conceptos al tesauo, el programa nos insta a elegir uno o más idiomas, es decir es multilingüe (Figura 4). Esto es importante ya que si elaboráramos el tesauo en un software en inglés aquellos términos con signos especiales (á, ñ, ü), al ser ordenados alfabéticamente, serían mandados hasta el final de la lista de términos. Varios software —aún bajo el modelo SKOS—, a pesar de la norma de los Comités Técnicos de Normalización, UNE 50125 (acrónimo de Una Norma Española), para la construcción de un tesauo multilingüe, no manejan esta opción.

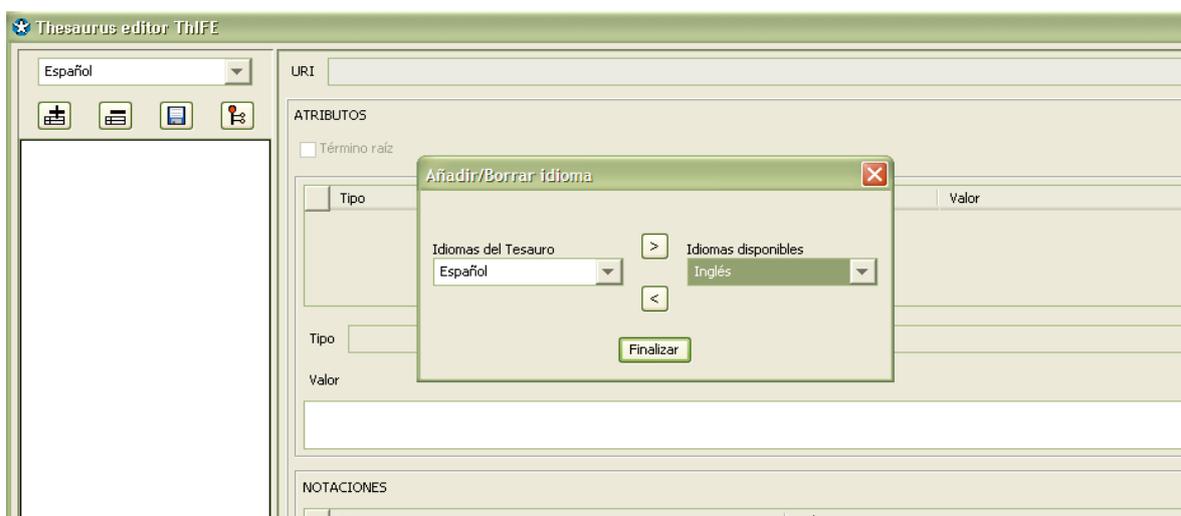


FIGURA 4

En cuanto a las interrelaciones semánticas, ThManager sólo permite relacionar términos con base en tres formas: abuelo-padre, hijo y relacionado. Esto indica, en primera instancia, que el término que se agrega está subordinado jerárquicamente a otro término, y también acepta que un término agregado tiene un término que se subordina a este. La relación del término “relacionado”, permite enlazar términos del mismo nivel, por ejemplo si un término padre tiene varios términos hijo, estos serán del mismo nivel, es decir su vínculo sería relacionado. Los términos ingresados al tesauro deberán ser descritos y relacionados en base a una serie de pasos del **Editor de tesauros** de ThManager. Los URI permiten ubicar el término en base a una línea de caracteres que se definen para el término en particular, lo que permite relacionar un término con otro a través de la asociación de los URI. Por ejemplo, para un término como Sistema jurídico su URI sería http://sistema_jurídico

. Bajo esta tónica, expondremos un ejemplo. El tesauro que se actualizará será el Tesauro Electoral del IFE, el cual maneja en su árbol conceptual once temas generales:

1. Conocimiento electoral: Producción teórica y su sistematización
2. Sistema político y elecciones
3. Sistemas electorales
4. Derecho y legislación electorales
5. Instituciones electorales

6. Organizaciones político-electorales
7. Sistema Jurídico
8. Geografía electoral
9. Comunicación político electoral
10. Cultura electoral
11. Historia electoral

Se elige para ejemplificar el concepto de SISTEMA JURÍDICO, que es el sétimo de la lista. Entre la ramificaciones de su primera clase se encuentra el descriptor Constituciones:

- 7 Sistema jurídico
 - 7.1 Constituciones
 - 7.1.1 Absolutismo
 - 7.1.2 Cambio constitucional
 - 7.1.3 Derecho (principios generales)
 - 7.1.4 Derecho electoral
 - 7.1.5 Derecho electoral (fuentes)

El tesauro nos insta a establecer las relaciones entre estos términos. SISTEMA JURÍDICO es el término abuelo, CONSTITUCIONES es el término padre y los puntos subsecuentes son términos hijo, y estos mantienen una relación ordenada entre sí. Esto expone el carácter de transitividad de algunas etiquetas de SKOS, principalmente las de las relaciones semánticas (*Vid.* TABLA 1), ya que no se puede vincular directamente SISTEMA JURÍDICO a ABSOLUTISMO, es decir de abuelo a hijo, sino que debe ser de abuelo a padre y de este a hijo.

En el caso anterior tendríamos un término abuelo, un padre y cuatro hijos (FIGURA 5) las cuales podemos revisar con el menú desplegable del **Visor del tesauro** (FIGURA 6), donde aparecen todas las relaciones establecidas en el **Editor de tesauros**: CONCEPTOS ELECTORALES tiene como término padre a SISTEMAS ELECTORALES y como término hijo a DERECHOS POLÍTICO-ELECTORALES:

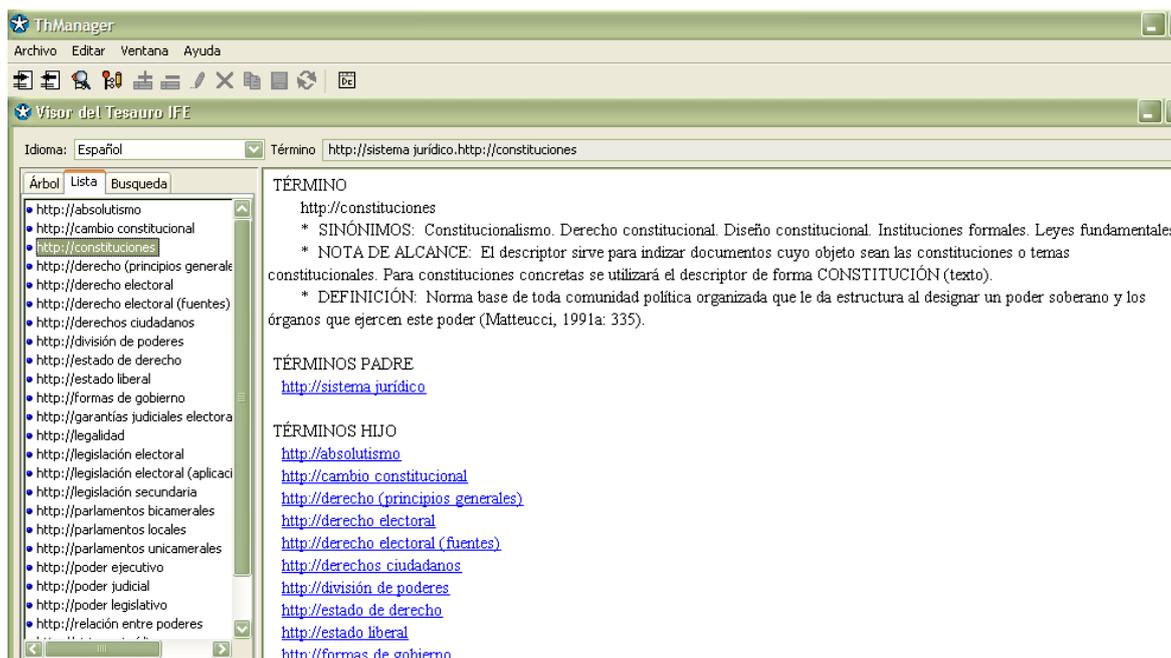


FIGURA 5

Para que aparezca la interfaz anterior, es necesario precisar cómo se asientan los UP, los TG, los TE, los TR, las NA y la definición para el término CONSTITUCIONES. En primer lugar, en ThManager todos estos aspectos se asientan en el Editor, cuyo símbolo en la barra de herramientas del software es: . La definición se asienta como se observa en la FIGURA 6:



FIGURA 6

Para la nota de alcance, véase la FIGURA 7:

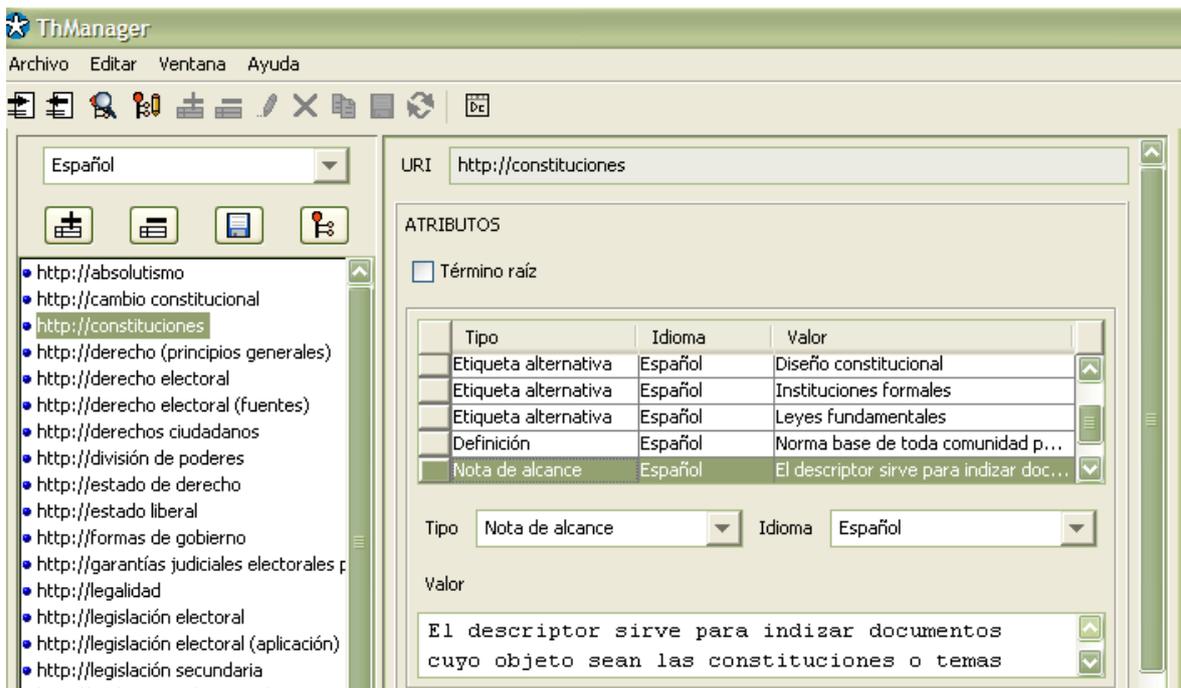


FIGURA 7

Para mirar cómo se asientan los ÚSESE POR, véase la FIGURA 8, donde se observará que se asientan como términos equivalentes y, luego, el *software* los toma como sinónimos (Vid. FIGURA 5):

FIGURA 8

Para los TG (abuelo o “padre del padre”), véase la FIGURA 9:

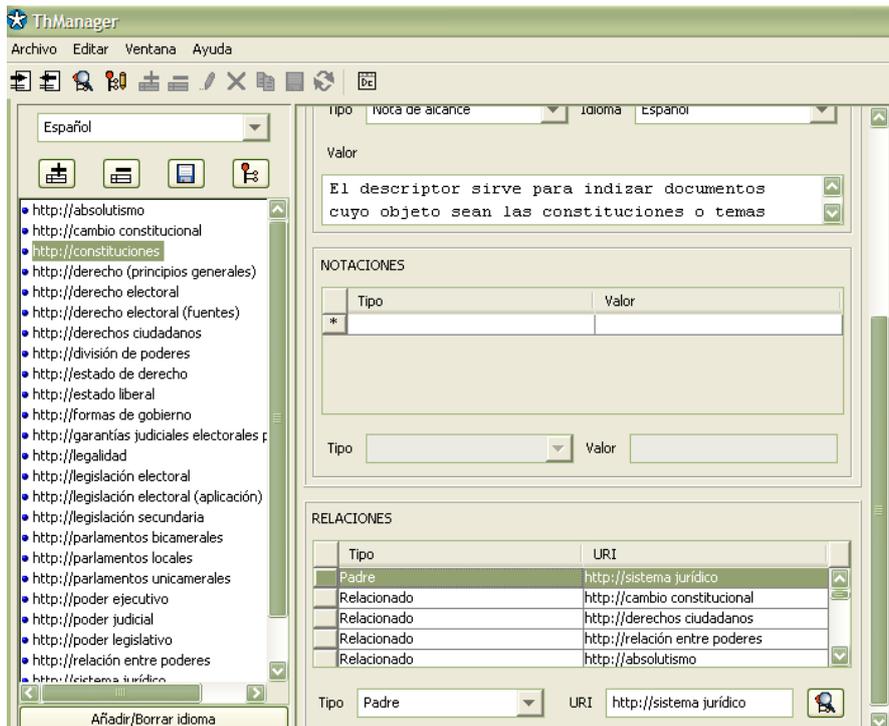


FIGURA 9

Para la gran cantidad de TE (los hijos), véase la FIGURA 10:

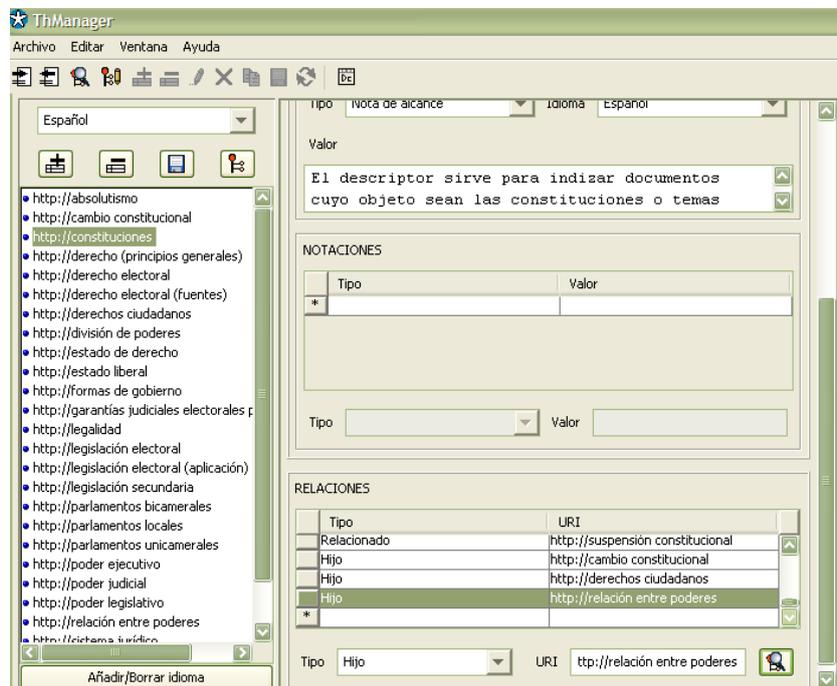


FIGURA 10

Para los TR, véase la siguiente figura:

:

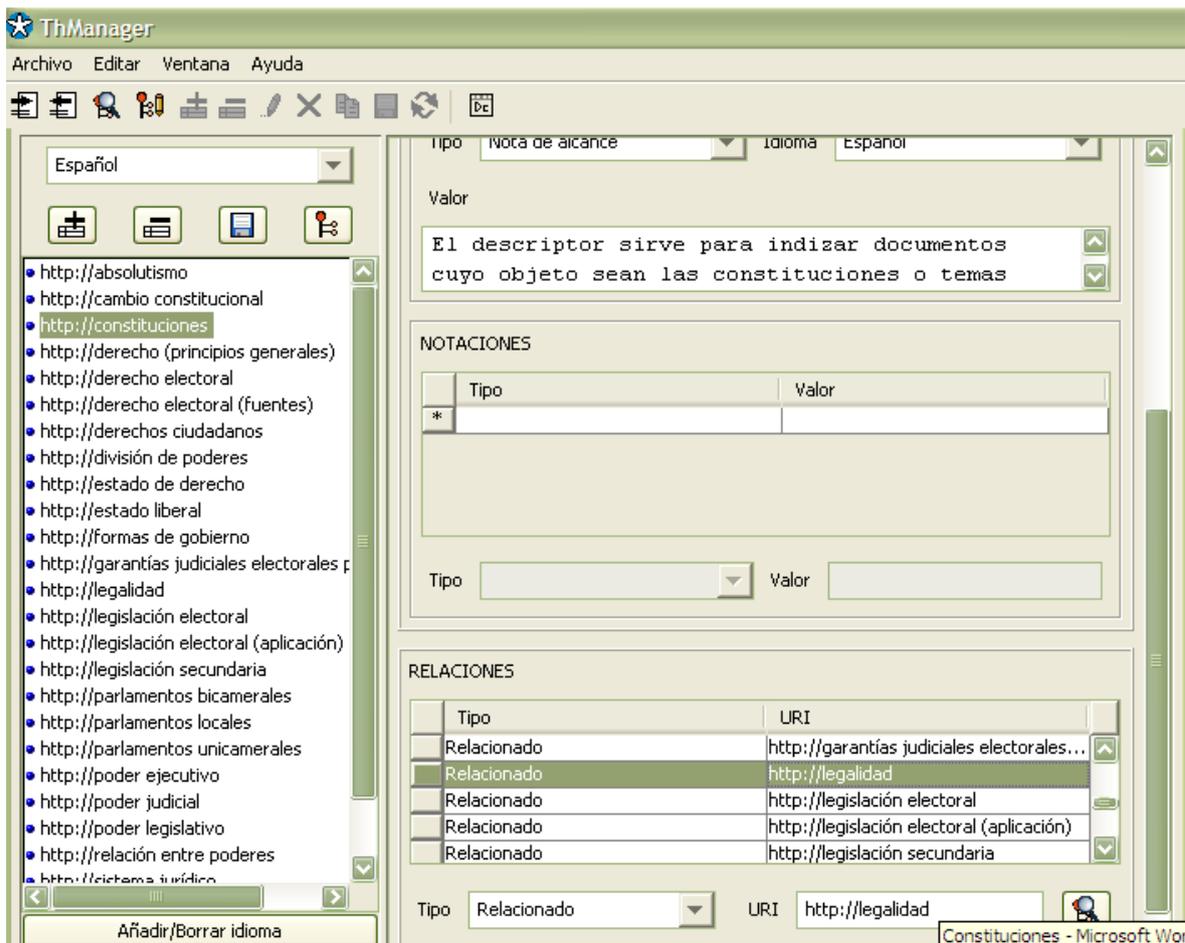


FIGURA 11

ThManager no sólo se ha creado como una herramienta independiente para facilitar la gestión de tesauros, sino también para permitir una fácil integración en las herramientas que necesitan utilizar tesauros, por lo que vincula diferentes bases de datos y repositorios con el tesauro. Por ejemplo, en una herramienta de creación de metadatos, se puede utilizar para proporcionar el componente gráfico para seleccionar los valores controlados de tesauros y automáticamente insertar en los metadatos. También se puede utilizar para proporcionar la lista de valores posibles para utilizar en un sistema de búsqueda Web, o para proporcionar una navegación basada tesauro de una colección de recursos en un sistema de búsqueda exploratoria.

Conclusiones

En este artículo se ha presentado una evaluación y descripción de las necesidades para la estructuración del Tesauro del IFE a partir de la herramienta para la gestión de los tesauros ThManager, bajo el modelo SKOS. La herramienta ThManager se ofrece como la solución a estos problemas. Es una herramienta de código abierto que puede gestionar tesauros almacenada en SKOS, lo que permite su visualización y edición. Gracias a la arquitectura de capas, sus componentes se pueden integrar fácilmente en otras aplicaciones que necesitan utilizar tesauros u otros vocabularios controlados. Adicionalmente, los componentes pueden ser utilizados para controlar los posibles valores utilizados en un servicio de búsqueda de Internet con la finalidad de facilitar búsquedas tradicionales o exploratorias basadas en un vocabulario controlado.

El modelo SKOS, y específicamente el software ThManager, son idóneos para crear tesauros de sistemas de información jurídica, en este caso, uno que atañe al derecho constitucional y a temas electorales. Debido a la naturaleza híbrida de los conceptos y mensajes legales, es necesario poder navegar en un tesauro estructurado para ser utilizado con la versatilidad que ofrece una herramienta lingüística integrada de manera computacional. Dicho en otras palabras, en un tesauro documental las relaciones semánticas se establecen entre los propios términos. A diferencia del nombrado tesauro conceptual, donde las relaciones entre conceptos suelen incluir las que desarrollan las estructuras jerárquica y asociativa, mientras por otro lado se denotan las relaciones existentes entre términos asociados a un mismo concepto (abreviaturas, uso común, tecnicismos, entre otros) e incluye una gama amplia de equivalencias entre términos, siempre vinculados con los propios documentos del sistema.

Obras consultadas

ALVITE DÍEZ, María Luisa *et al.* 2010. Propuesta de representación del tesauro EuroVoc en SKOS para su integración en sistemas de información jurídica. *Scire: Representación y Organización del Conocimiento*, 16 (2): 47-51.

CESTEURS, Werner, Smith, Barry & Goldberg, Louis. 2005. A terminological and ontological analysis of the NCI Thesaurus. *Methods of Information in Medicine* 44: 498-507.

DEWEZE, André. 1998. *Informatique documentaire*. París: Masson.

ERK, Katrin. 2009. Supporting inferences in semantic space: Representing words as regions. En Harry Bunt, Volha Petukhova y Sander Wubben (Eds.), *Proceedings of the 8th International Conference on Computational Semantics*, 1 (104-115). Tilburg, Países Bajos: Tilburg University, Department of Communication and Information Sciences.

FLORIDI, Luciano. 2009. Web 2.0 vs. the Semantic Web: A philosophical assessment. *Episteme*, 6 (1): 25-37.

JONES, Susan. 1993. A thesaurus data model for an intelligent retrieval system. *Journal of Information Science*, 19 (3): 167-178.

LACASTA, Javier *et al.* 2007. ThManager: An open source tool for creating and visualizing SKOS. *Information Technology and Management*, 26 (3): 39-51.

LINANT DE BELLEFONDS, Xavier. 1981. *L'informatique et le droit*. París: Presses Universitaires de France.

MATUTE CIFUENTES, Sergio Luis. 1989. Los sistemas de información, la informática jurídica y el sistema UNAM-JURE. En Jean Louis Bilon *et al.* (Coords), *Diálogo sobre la informática jurídica*. México: UNAM, Instituto de Investigaciones Jurídicas.

PASTOR SÁNCHEZ, Juan Antonio, Martínez Méndez, Francisco Javier & Rodríguez Muñoz, José Vicente. 2009. Advantages of thesaurus representation using the Simple Knowledge Organization System (SKOS) compared with proposed alternatives. *Information Research*, 14 (4). Documento en línea: <http://informationr.net/ir/14-4/paper422.html>.

131

PERALES OJEDA, Alicia. 1981. Traza de un sistema de información jurídica. *Anuario de Bibliotecología*, 4 (2): 7-78.

ROSALES SALINAS, Fabiola Elena. 2010. The parliamentary library as a bridge between Congress and citizenship. The Mexican Chamber of Deputies reference service. *Library Trends*, 58 (4): 510-516.

SACCO, O. & BOTHOREL, C. (2010). Exploiting semantic web techniques for representing and utilising folksonomies. En *Proceedings of the international workshop on modeling social media*. Nueva York: ACM.

TÉLLEZ VALDÉS, Julio Alejandro. 2008. *Derecho informático*. México: McGraw-Hill

TUOMINEN, Jouni *et al.* 2008. ONKI-SKOS - Publishing and utilizing thesauri in the semantic web. En *Proceedings of the 13th Finnish Artificial Intelligence Conference (STeP 2008): Finnish Artificial Intelligence Society*. Espoo, Finlandia: J. V. Editors.

TUDHOPE, Douglas, HARITH, Alanis & JONES, Christopher. 2001. Augmenting thesaurus relationships: Possibilities for retrieval. *Journal of Digital Information*, 1 (8): 1-20. Documento en línea: http://eprints.soton.ac.uk/254484/1/Tudhope_JoDI.pdf.

Ontologias e Definições: a explicitação do compromisso ontológico

Maria Luiza de Almeida Campos
Universidade Federal Fluminense
marialuizalmeida@gmail.com

1. Algumas Questões Iniciais

Esta pesquisa é parte dos estudos que estamos desenvolvendo no âmbito do Grupo de Pesquisa em “Ontologia e Taxonomia: aspectos teóricos e metodológicos” e que vem sendo financiada pelo CNPq através de bolsa de Produtividade em Pesquisa.

Em nossa proposta atual de projeto investigamos a questão relacionada ao compromisso ontológico, ou dito de outra forma, em que bases epistemológicas e ontológicas se dão a representação de domínios de conhecimento, enquanto um modelo classificatório para tratamento e recuperação de informação/conhecimento. Por bases epistemológicas, consideramos, neste estudo, discutir a possibilidade de representação em dados domínios, ou seja, discutir a própria natureza dos saberes consubstanciados nos modelos de representação. Estabelecer assim, um diálogo com a possibilidade de divisão dos saberes apresentados nestes modelos. Por bases ontológicas, consideramos, neste estudo, discutir quais elementos são passíveis de serem identificados em domínios tendo em vista a modelização com fins classificatórios, ou, dito de outra forma: Quais seriam as categorias formais que deveriam ter os campos diferenciados em Ciência e Humanidades? Que relações semânticas são passíveis de representação em tais campos?

De certa forma pretendemos demonstrar que o potencial semântico de uma dada ontologia de fundamentação, aqui entendida como uma teoria de representação sobre o domínio, possibilita testar e validar o seu modelo conceitual, onde definições operativas auxiliam na explicitação e entendimento do acordo firmado para entendimento do domínio em questão. (CAMPOS, 2012), sendo este o foco deste artigo.

2. Compromisso Ontológico em Ontologias de Fundamentação

As ontologias de fundamentação tem por objetivo identificar categorias gerais de certos aspectos da realidade que não são específicos a um campo científico, descrevendo conhecimento independentemente de linguagem, de um estado particular

das coisas ou ainda do estado de agentes (GUIZZARDI, 2005). Borgo e Masolo (2008) definem ontologias de fundamentação a partir de quatro itens: (i) têm grande alcance; (ii) podem ser altamente reutilizáveis em cenários de modelagem diferentes; (iii) são filosófica e conceitualmente bem fundamentadas; e (iv) são semanticamente transparentes e, portanto, ricamente axiomatizadas. Estas Ontologias permitem a elaboração de modelos de domínios, sendo altamente reutilizável. É também caracterizada por ser filosoficamente bem fundamentada, recebendo aportes da linguística, ontologia formal e ciências cognitivas, permitindo a explicitação de uma visão da realidade, ou seja, do acordo ontológico estabelecido, com determinação de regras de restrição, bem como conceitos, categorias e meta-propriedades.

Com efeito, nota-se a importância das ontologias de fundamentação no que tange a seus princípios filosóficos bem fundamentados para modelagem domínios. Esta importância é explicitada por Guizzardi (2005) e Guizzardi e outros (2008) ao relatarem que a ontologia de fundamentação permite a explicitação dos compromissos ontológicos da ontologia que representa um domínio, agregando fidelidade, consistência e clareza na representação.

Compromisso ontológico pode ser definido como um acordo firmado por uma comunidade sobre o significado que esta estabelece e é expressa, neste contexto, nas representações de conceituações, tanto do ponto de vista da compreensão pelo homem quanto do tratamento pela máquina, através dos agentes de software. Isso implica em definir o vocabulário de uma forma que venha a minimizar ambigüidades, de modo que seu uso possa ser partilhado para representar e recuperar conhecimento entre comunidades afins, que se proponham a estar comprometidas com o acordo adotado. (NODINE; FOWLER, 2002).

O estabelecimento de compromisso ontológico revela uma tomada de posição em relação aos objetos que se reconhecem como fundamentais em uma visão de mundo e que possibilitam o diálogo sobre esse mundo, e ainda em relação à maneira pela qual esses objetos são caracterizados e definidos. Vale ressaltar que existem diversos modos possíveis de um acordo sobre esta visão de mundo, e deve-se ter em mente, entretanto, que nas ontologias, principalmente nas chamadas ontologias de fundamentação, devido às restrições dos mecanismos de representação do conhecimento para as máquinas, a representação dos conceitos e de suas relações que constituem o seu modelo conceitual é sempre limitada e, portanto, aproximada.

Guarino situa o papel do compromisso ontológico como o de um elemento fomentador da precisão entre a *conceituação* e a representação de uma visão de mundo, esta última um *modelo* desse mundo, impreciso em algum grau em relação ao *significado pretendido* pelo homem. Essa imprecisão se dá devido ao fato de que as conceituações são entidades abstratas, que existem na mente de pessoas ou grupo de pessoas de uma comunidade (GUIZZARDI, 2007). Naturalmente, para que uma conceituação possa ser comunicada entre pessoas, esta deve ser expressa em termos de um artefato concreto que deve ser representado em uma linguagem, a qual não consegue ser suficientemente expressiva para representar a riqueza semântica da conceituação presente na mente humana.

Uma representação de conhecimento é um mecanismo usado para se raciocinar sobre o mundo, em vez de agir diretamente sobre ele. Nesse sentido, ela é, fundamentalmente, um substituto para aquilo que representa. (...). Deve haver alguma forma de correspondência especificada entre o substituto e seu referente planejado no mundo. Esta correspondência é a semântica da representação. (...). Fidelidade perfeita é, em geral, impossível, tanto na prática quanto em princípio. A única representação completamente precisa de um objeto é o objeto em si. Qualquer outra representação é imprecisa e, inevitavelmente, contém simplificações. (CAMPOS, 2004, p.24).

134

O compromisso ontológico surge então para preencher a lacuna semântica entre o que o homem identifica nas suas conceituações de um mundo e o que é possível representar através de uma linguagem de representação e recuperação de informação. O compromisso se desdobra em dois: as coisas cuja existência se admite e o seu significado pretendido.

Desta forma, o compromisso ontológico vem ocupar o espaço de mediação entre a conceituação, que está no plano da semântica do entendimento humano e a representação, que está no plano do processamento pela máquina. É formado por um conjunto de regras que objetiva minimizar ambigüidades ou entendimentos errôneos dos conceitos do mundo sendo representado de acordo com a conceituação do homem, e embora seja voltado para o entendimento humano, é adequado para a conversão formal em uma linguagem que possa ser entendida pela máquina. Sendo assim, embora possa ser definido através de uma linguagem textual ou gráfica, o uso de uma linguagem formal auxilia na sua passagem para a linguagem que pode ser tratada computacionalmente.

3. Definição e Compromisso Ontológico

O compromisso ontológico pode ser representado através de definições, voltados para o entendimento humano. No caso das ontologias, as definições propiciam a possibilidade de entendimento do objeto e de compatibilização semântica, pois descrevem o conteúdo significativo de um termo. Esta descrição, quando formalizada, possibilita que agentes inteligentes possam entender o significado de um termo e estabelecer inferências sobre esses significados, pois a definição é composta de características de conceitos, que são também conceitos que se relacionam formando o entendimento semântico dos termos em questão.

A definição em ontologias é um elemento essencial, pois irá garantir a explicitação do compromisso ontológico determinado. Entre outros aspectos as ontologias necessitam de um padrão para seus enunciados definitórios que possam apoiar o entendimento pelos agentes inteligentes. (CAMPOS, 2010)

Os problemas de definição tem sido objeto de estudos diversos na língua geral. Na língua especializada, nos últimos 50 anos, a definição passou a ser objeto de estudo sistemático. Até a década de 60 as definições tinham um caráter mais conceitual e filosófico e evidenciavam o que pensar acerca de um conceito. Entretanto, para atender às necessidades da Ciência foi necessário um esforço teórico visando à elaboração de definições operacionais – aquelas onde os elementos da definição não identificam somente o gênero próximo e a diferença específica -, mas também, elementos que procuravam definir certas operações onde o conceito seria aplicado. (HEGENBERG, 1974). Temos então o movimento chamado operacionismo, fruto direto do positivismo lógico defendido pelo Círculo de Viena, parente próximo do pragmatismo e do instrumentalismo. Assim, segundo Hegenberg (1974), as definições procedentes deste movimento denominam-se "operativas".

A idéia fundamental que norteia o operacionismo é simples, a saber, a de que não conhecemos os significados de um conceito a menos que possamos especificar as operações que foram realizadas, por nós ou por outrem, ao aplicar o conceito a qualquer situação concreta. Desta forma, as definições operativas revelam a maneira de aplicação dos conceitos dentro de um dado contexto, ou como se dá o uso de tais conceitos o que, mais tarde, Dahlberg denominará de relação funcional.

Dahlberg (1981) afirma que para se definir conceitos é necessário o conhecimento de: a) os referentes dos conceitos; b) a(s) categoria à(s) qual (is) pertence um conceito; e c) a expressão verbal adequada a um certo grupo de usuários. Defende a autora, portanto, que os conceitos se relacionam em um sistema de conceitos por suas características, que são também conceitos. Segundo a Teoria do Conceito, as características relevantes do conceito são, então, os elementos constitutivos da definição. O ponto principal no estabelecimento das definições dos conceitos está, portanto, na identificação das características. Ela fornece um padrão para definição, classificando-as conforme a categoria do conceito: definição genérica, partitiva, funcional (DAHLBERG, 1983). Dahlberg, na verdade, é integrante de um grupo de estudiosos que já há algum tempo vinham dando à definição um objetivo mais prático, relacionado com as demandas da terminologia e à necessidade de tornar claras as definições técnicas e científicas, e não mais um objetivo de interesse estritamente filosófico.

Observa-se, então, que os diferentes modos de definir podem levar aos diferentes tipos de definição. Não é possível privilegiar-se somente um tipo de definição, porque os conceitos de uma área de assunto são de níveis diversos (conceitos gerais e específicos) e de natureza categorial diversa (objeto, propriedade, processo).

É preciso, portanto, chegar a um modelo de enunciado capaz de atender às especificidades da área. Para tanto, consideramos que a forma definitória deve ser ditada pelo próprio nível do conceito, como, principalmente, por sua natureza categorial, para que se obtenham os elementos necessários à caracterização do conceito e, quando pertinente, sua função, numa dada área de assunto. Em cada caso, deve-se tentar identificar os modelos mais apropriados de definição, ou seja, os padrões de enunciados definitórios em cada contexto.

3.1 Um exemplo de Modo de Definir em Ontologias

Como apresentado, para a elaboração de um enunciado definitório é necessário reunir os elementos constitutivos que explicitam as características de um conceito. Onde algumas características possibilitam identificar o posicionamento do conceito em um sistema de conceitos, e outras características possibilitam ter uma visão mais pragmática do uso do conceito dentro de uma dada área. Teremos assim, reunidas em um mesmo enunciado definitório, definições do tipo analítico/conceituais e operativas, as primeiras então com a função de posicionamento do conceito e a segunda com a função do uso que

é atribuído ao conceito em questão; esta última característica, ou seja, a de uso/aplicação, é de suma importância para as ontologias formais de domínios.

Em Ontologias a unidade de informação é o conceito, que pode ser classificado como um conceito geral ou individual¹⁹ (DAHLBERG, 1978), onde a seqüência de conceitos representa um tipo de hierarquia, e esta hierarquia é partilhada por grupos de conceitos. A consistência desta hierarquia depende da definição explícita das propriedades (atributos, características) que descrevem os conceitos, fornecendo a base sobre a qual eles podem ser agrupados em conjunto, diferenciando um conceito do outro. Diferentemente dos dicionários onde a unidade de informação é uma palavra e que, neste sentido, as definições especificam os diferentes significados de uma palavra. A definição em dicionários é realizada de forma isolada, sem qualquer relação com as outras palavras.

Assim, a coerência nas definições e, por conseguinte, na classificação, exige uma perspectiva unificadora requerida pelo contexto a ser especificado para a representação dos conceitos. Desta forma, se em um primeiro momento do entendimento do conceito, as explicações sobre os conceitos são recolhidas a partir do Universo de Discurso pré-estabelecido, em um segundo momento será necessário que os elementos da definição sejam determinados a partir do acordo firmado, do compromisso ontológico estabelecido com a comunidade e explicitado a partir de um enunciado definitório.

O enunciado definitório é elaborado tendo em vista primeiramente a categoria conceitual do conceito, pois esta é também determinante na escolha dos conceitos que farão parte da definição.

Neste sentido, o exemplo que apresentamos é considerado uma entidade, ou seja, um objeto concreto, segundo a Teoria do Conceito, no domínio da Agroecologia, e é denominado “Cobertura Morta”.

No campo dos estudos das Ontologias de Fundamentação, Guizzardi (2005) apresenta a UFO(A), que têm sido usada para apoiar a modelagem de conceitos. Nos princípios apresentados pela UFO “Cobertura Morta” seria classificada na categoria de endurantes (também conhecidos como continuantes). Endurantes são entidades que não mudam sua essência ao longo do tempo, preservando sua identidade durante as mudanças que ocorrem durante o período de sua existência: “Você é hoje a mesma pessoa que você

¹⁹ No caso, o que Dahlberg define por conceito individual em sua Teoria do Conceito, pode ser denominado no campo de estudo das Ontologias como instâncias.

foi ontem” (GRENON; SMITH, 2004). Organismos, células e moléculas são exemplos de entidades cuja natureza é *endurante*.

Considera-se então que um Padrão Definitório para a definição de objetos, no caso “Cobertura Morta”, poderia ser composto dos seguintes elementos, a saber:

1. Definição analítica, onde as características que possibilitam a identidade e componentes do objeto, naquele contexto, são explicitadas, ou seja, O que ele é – é sempre um tipo de objeto (gênero próximo e diferença específica); e Quais são seus elementos constitutivos – partes componentes (relação todo-parte). Estes elementos pertencem sempre a mesma categoria do conceito definido (são de mesma natureza);

2. Definição Funcional ou Operativa, que evidencia a função/aplicação naquele contexto. Estes elementos são, na maioria das vezes, pertencentes a categorias conceituais diferentes.

Desta forma, a definição final, a partir da adoção deste princípio para a constituição de um enunciado definitório, para o conceito de “Cobertura Morta”, poderia ter como proposta, o seguinte enunciado: Capa de material morto de plantas ou outros materiais como papel ou plástico sobre o solo, para criar microclima na superfície do solo independente do tempo. Onde “Capa de Material” é também um objeto concreto, respeitando o princípio de identidade no âmbito de um contexto, e “material morto de plantas e outros materiais como papel” seriam os elementos constitutivos, esses elementos, ou características do conceito, são os que formam uma Definição Analítica. A Definição Funcional seria representada pelos elementos “para criar microclima na superfície do solo independente do tempo”, onde criação de microclima representa a função que exerce o material morto na superfície do solo.

Até aqui elaboramos um enunciado para a compreensão do humano, mas para possibilitar o raciocínio automático em Ontologias, pois estas requerem mecanismos que possibilitem que se extraia conhecimento, será necessário que estas definições possam ser apresentadas em linguagem formal para que as máquinas de inferências possam funcionar. Isto irá requerer também que profissionais de informação se capacitem nestas linguagens formais.

4. Considerações Finais

Como vimos um enunciado definitório pode ser explicitado como o conjunto de características que descrevem um conceito em um dado contexto; sua função não é de

somente permitir o posicionamento do conceito em um sistema de conceitos como as definições conceituais se propõem, mas de ampliar toda a complexidade de entendimento daquele conceito em um dado domínio, ou seja, quais elementos ontológicos (funcionais) podem ser agregados aos elementos lógicos (relações de gênero-espécie). Neste sentido, se coloca os tipos de definições analíticas e operativas.

Em Ontologias de Fundamentação a elaboração de enunciados definitórios, para cada contexto de aplicação, a partir do estabelecimento de um acordo firmado, ou seja, de seu compromisso ontológico, se faz necessário devido a função das Ontologias de possibilitarem a representação de uma dada realidade, que é sempre contextual.

5. Referências

BORGO, S.; MASOLO, C. **Foundational choices in DOLCE**. In: R. Poli, Healy M., and Kameas A., editors, *Theory and Applications of Ontology*, v. 2, C.12. Springer Verlag, 2008.

CAMPOS, M. L. A. **Ontologia de Fundamentação como uma teoria de representação para a Web semântica: estratégias interdisciplinares na pesquisa e no ensino**. In: Murguia, Eduardo Ismael; Rodrigues, M.E.F.. (Org.). *Arquivologia, Biblioteconomia e Ciência da Informação: identidades, contrastes e perspectivas de interlocução*. 1ed.Niterói: Editora UFF, 2012, v. 2, p. 101-116.

CAMPOS, M. L. A. Modelização de Domínios de Conhecimento: uma investigação de princípios fundamentais. **Ciência da Informação**, Brasília, v. 33, n. 1, p. 22-32, 2004.

CAMPOS, M. L. A. . O papel das definições na pesquisa em ontologia. **Perspectivas em Ciência da Informação**, v. 15, p. 220-238, 2010.

DAHLBERG, I. A referent-oriented analytical concept theory of interconcept. **International Classification**, Frankfurt, v.5, n.3, p.142-150, 1978.

DAHLBERG, I. **Terminological definitions: characteristics and demands**. In: *Problèmes de la définition et de la synonymie en terminologie*. Québec, GIRSTERM, 1983. p. 13-51.

DAHLBERG, Ingetraut. Conceptual definitions for interconcept. **International Classification**, v. 8, n. 1, 1981.

GRENON, P.; SMITH, B. SNAP and SPAN: **Prolegomenon to geodynamic ontology**. In: *Spatial Cognition and Computation*, 2004.

GUIZZARDI, G. **Ontological foundations for structural conceptual models**. 416 f. Tese (PhD em Computer Science) – Twente University of Technology, Twente, Holanda, 2005.

GUIZZARDI, G. **On Ontology, ontologies, Conceptualizations, Modeling Languages, and (Meta) Models**. In: Olegas Vasilecas; Johan Edler;Albertas Caplinskas. (Org.). *Frontiers in*

Artificial Intelligence and Applications, Databases and Information Systems IV. Amsterdã: IOS Press, 2007.

GUIZZARDI, G.; FALBO, R.; GUIZZARDI, R. S. S . A importância de Ontologias de Fundamentação para a Engenharia de Ontologias de Domínio: o caso do domínio de Processos de Software. **Revista IEEE América Latina**, v. 6, n.3, p. 244-251, 2008.

HEGENBERG, Leonidas. **Definições: termos teóricos e significado**. São Paulo: Cultrix, 1974.

NODINE, M.; FOWLER, J. On the Impact of Ontological Commitment. In: **Proceedings of The Workshop on Ontologies In Agent Systems**, Bologna, Italy, 2002.

Modelagem conceitual e sistemas de organização de conhecimento; uma reflexão baseada em experiências acadêmicas e profissionais no Brasil no domínio da Biblioteconomia e Ciência da Informação

Lidia Alvarenga
Universidade Federal de Minas Gerais
lidiaalvarenga@eci.ufmg.br

Aos profissionais e pesquisadores do domínio da Biblioteconomia e Ciência da Informação, BCI, competem os exercícios de: conceber, desenvolver, gerir, disseminar e avaliar sistemas de recuperação de informações documentais; apreender e compreender produtos e processos de várias naturezas deles constantes; e estabelecer conexões entre esses sistemas e os interesses e necessidades de usuários reais e potenciais.

Assim como os profissionais e pesquisadores vinculados aos campos acadêmicos e organizacionais do universo da produção conhecimentos, os profissionais do BCI, visando ao intercâmbio amplo de informações documentais, orientam-se usando modelos, normas e códigos; destacam-se dentre esses, propostas de estruturas conceituais e terminológicas que contribuem para os processos de produção, organização e recuperação de conhecimento.

Na área BCI, além de práticas e ferramentas compreendendo modelos conceituais dos campos científicos e profissionais, vale ressaltar a existência do modelo de descrição e relacionamento de entidades documentais, denominado Requisitos Funcionais para Registros Bibliográficos, conhecido na Língua Inglesa pela sigla FRBR. Concebido pela IFLA tal modelo trouxe avanços na compreensão desse importante objeto de estudo, o documento, assim como para a construção de códigos de catalogação e formatos de intercâmbio entre sistemas de recuperação de informações documentais.

Também como instrumento para a organização dos acervos para permitir uma comunicação ampla entre usuários e documentação produzida cita-se outro metamodelo, o desenvolvido e proposto por Ranganathan para a construção de classificações facetadas nas diversas áreas de conhecimento. Ressalta-se que, a partir deste podem ser criadas classificações ou modelagens conceituais específicas, relativas aos domínios de conhecimento.

Nos dias atuais constata-se que o trabalho de concepção e aplicação de modelos conceituais tem evoluído de forma significativa, graças à teoria e prática acumuladas na BCI, enriquecidas pela disponibilização de tecnologias para apresentação de tais aplicações. Por outro lado sabe-se que no Brasil, cursos, projetos e grupos de pesquisas, no âmbito da BCI, envolvendo a construção de linguagens documentárias, têm enriquecido há décadas o ensino, a pesquisa e a extensão em instituições acadêmicas e de pesquisa.

Com a evolução desses estudos e as necessidades provenientes das tecnologias de comunicação e informação, verifica-se que os esforços voltados à construção de linguagens de indexação extrapolam os domínios da classificação facetada e dos tesouros, tendo seu uso ampliado para além deste espaço. Hoje a estruturação semântica de termos (conceitos), relacionadas aos diversificados domínios vem sendo demandada em outros contextos, criando-se ferramentas que aqui denominaremos de sistemas de organização de conhecimento, SOC. Muitos dentre tais sistemas requerem no todo ou em parte o conhecimento da modelagem conceitual que vinha sendo estudada, aperfeiçoada e produzida no contexto da BCI.

O conhecimento proveniente das ações humanas e organizacionais deve ser organizado, modelado, estruturado, sistematizado, para ser conhecido e usado, seja por pessoas ou máquinas. Se a *web* semântica está avançando, a criação de sistemas de organização de conhecimento se constitui em uma dentre as chaves desse importante processo.

Segundo Dagobert Soergel, *Visiting Scholar* (2012), junto programa de Pós-Graduação em CI da UFMG, em *syllabus* de disciplina ofertada, um sistema de organização de conhecimento, SOC, pode ser caracterizado por: ser usado por pessoas, visando encontrar informações que façam sentido para elas; sustentar as pessoas, dando sentido às suas questões; apresentar estruturas significantes de conceitos em determinado domínio; ser usado por programas de computador, com a finalidade de tomada de decisões, a partir de dados; representar conhecimento formalizado sobre conceitos.

Atualmente tais sistemas também são conhecidos por nomes tais como: ontologias, esquemas de metadados, taxonomias, classificações, terminologias, estruturas de diretório da *web*, tesouros, folksonomias, e outros. Cada um desses tipos é usado em diferentes contextos e por diferentes comunidades, significando coisas diferentes, com significados nem sempre precisos. Embora considerando-se que do uso

dessas ferramentas pode advir grandes benefícios, não se pode ignorar o grande investimento requerido no desenvolvimento de bons sistemas de organização de conhecimento.

Enfatizando, afirma-se que um SOC provê uma estrutura ou esquema para armazenar e organizar dados, informação ou conhecimento sobre o mundo e sobre pensamentos, visando à compreensão, recuperação ou descoberta, ao raciocínio e a outros propósitos.

Ainda dentre as múltiplas funções de um SOC destacam-se: suportar pensamentos, decisões, integração de conhecimento; descobrir novos conhecimentos e de *gaps* entre pessoas e programas de computadores; suportar e planejar tomadas de decisões; dar suporte a indivíduos no processo de pensar e decidir, provendo uma estrutura conceitual que permita a organização e integração de dados; identificar *gaps* (defasagem, desníveis) no conhecimento individual; dar suporte ao planejamento de qualquer tipo de produto, processo ou serviço, incluindo o planejamento de estratégias e planos de *software* e projetos de bases de dados.

Muitas são as aplicações potenciais para produtos advindos desse processo e que podem servir de implementos para sistemas de organização de conhecimento, (*Knowledge Organization Systems*), como afirma o pesquisador Dagobert Soergel: *design* de domínios disciplinares; construção, desenvolvimento e avaliação de sistemas de informação documentais; construção de linguagens de indexação; construção de produtos que podem contribuir para a construção de ontologias de domínios; apoio ao desenvolvimento de outros métodos de modelagem, usados na análise de sistemas, como o Modelo Entidade-Relacionamento – MER; organização de *sites*, páginas na *web*, bibliotecas digitais e portais corporativos; organização de sistemas de conhecimento, seja para máquinas de inferência (sistemas especialistas); gestão empresarial, ou aquisição de conhecimento; delineamento de mapas conceituais; análise, delineamento e escopo de um domínio; organização do conceito de uma entidade. Nesses sentidos, um SOC lida com temas bastante familiares à BCI, tais como conceitos, categorias, classes, relacionamentos entre estes.

Tendo como objetivo refletir sobre a vocação e capacitação da BCI para o desenvolvimento de sistemas de organização de conhecimento, esta apresentação se fundamenta em experiências prévias de criação e desenvolvimento de tesouros para sistemas de recuperação de informações mantidos por órgãos brasileiros, assim como no

trabalho acadêmico e de pesquisa, realizado na Universidade Federal de Minas Gerais, UFMG, com alunos de graduação, mestrado e doutorado.

A presente reflexão parte de alguns pressupostos: de que uma modelagem conceitual, fundada nas teorias, métodos e modelos da BCI, se constitui em ferramenta muito potentes para a organização, não somente para gestão de informações documentais, mas também para a gestão do conhecimento, para a construção de SOC; e de que qualquer tentativa de modelar uma realidade conceitual compreende um trabalho subjetivo e sujeito às limitações de visões diferenciadas de mundo, mesmo quando praticada por especialistas dos campos de conhecimento.

Torna-se, portanto, desnecessário advogar a proficiência do profissional da BCI nesse tipo de trabalho, como consequência de sua formação. Os manuais de construção de tesouros, traduzidos para a Língua Portuguesa, os textos de Ranganathan e I. Dahlberg, em seus originais, ou traduzidos e interpretados por mestres que sobre eles se debruçaram, serviram de base para a aquisição desse tipo de conhecimento acumulado. As diversas aplicações de modelagem para a construção de SOC, a partir do material estudado por profissionais da BCI, viabilizam hoje sólidos fundamentos para o trabalho de modelização conceitual, tornando-o mais racional e objetivo.

Esse arcabouço teórico envolve outras competências tais como: análise de assuntos e de conteúdo; o conhecimento ou assessoria em campos específicos a serem modelizados; o rastreamento das entidades do domínio a partir do princípio da garantia literária. Em relação a esse último processo, merece destaque o fato de ser o mesmo facilitado quando desempenhado por profissionais da BCI, considerando-se que as bibliografias ou os repertórios bibliográficos se constituem como objetos centrais em torno dos quais gravitam os fazeres deste domínio.

Sem exaustividade e sem minimizar a contribuição de outros autores clássicos e contemporâneos, seguem-se referências alusivas à contribuição de S.R. Ranganathan para a modelagem de sistemas de organização de conhecimento, extrapolando o contexto para o qual ela foi concebida. LaBarre (2004), em tese defendida na *School of Library and Information Science* da *Indiana University, Bloomington, US*, trata da influência do pensamento de Ranganathan e apresenta uma revisão ampla de estudos teóricos e aplicados. Segundo a autora, a partir de 2000, assiste-se a um crescimento no número de estudos sobre as propostas de Ranganathan. Ainda com base em sua pesquisa, ela afirma que a tendência dos *web designers* não é construir *faceted classifications*, FC, completas, com notações e esquemas gerais; por outro lado, a autora

considera ser mais possível que os componentes dos *websites*, tais como os que promovem a navegação e acesso a informações do *site*, por exemplo, possam se beneficiar com o uso de algumas formas de análise de categorias e criação facetas, fundamentadas por essa teoria.

Neelameghan (1992), discute e apresenta alguns exemplos de aplicações de postulados, princípios e técnicas, formulados por Dr. S. R. Ranganathan, constantes do esquema de sua teoria geral de classificação de conhecimento e aplicados a vários estágios, no desenvolvimento de bases de dados especializadas: na conceituação, estruturação e organização da informação como percebida pelos usuários especialistas e no preparo de bases de dados a partir delas; na preparação de esquema de definição do domínio; e na análise orientada a objetos e modelização.

Devido à sua ampla potencialidade, a influência do pensamento de S.R. Ranganathan extrapola-se da criação de esquemas facetados de classificação e indexação, visando-se Sistemas de Recuperação de Informações (SRIs), até as novas aplicações de análise conceitual, produção, organização, representação e modelização de conhecimento, nos mais diversificados campos de interesse.

Restringindo-se ao campo acadêmico recente (últimos 15 anos) foram inúmeras as experiências de modelagem de domínios desenvolvidas na Escola de Ciência da Informação da UFMG (sem desconhecer certamente, como comprova a literatura, a produção de conhecimento sobre o tema em outras instituições de ensino e pesquisa no Brasil). Destaca-se que na UFMG esse trabalho se reverteu em conteúdos de disciplinas na graduação, voltadas prioritariamente para a construção de linguagens documentárias. Nelas os alunos trabalharam com as mais variadas áreas de conhecimento, desde as inseridas no escopo científico-acadêmico institucionalizado (Direito, Geologia, Psicologia), como em subáreas sob o prisma do senso comum, circunscritas a interesses aplicados ou culturais, tendo como alguns exemplos: moda, futebol, hipismo, cinema, fotografia. Também foram desenvolvidos alguns projetos mais detalhados de modelagem conceitual na pós-graduação da ECI/UFMG, citando-se como exemplos: cultura do sorgo; gestão Arquivística; biomonitoramento de águas; análise de assuntos, dentre outros. Em decorrência da presente reflexão é possível que as disciplinas até então oferecidas mudem suas denominações, tirando seu foco da criação específica de linguagens documentárias.

A BCI como campo específico e institucionalizado de ensino e pesquisa detém em seu cabedal teórico-metodológico um corpo consistente de conhecimento, já

bastante aplicado, capaz de subsidiar a construção de sistemas de organização de conhecimento. Entretanto, como já dito, deve-se ter sempre em mente as peculiaridades das realidades conceituais submetidas à modelagem, devido à multiplicidade de visões a que esse trabalho está sujeito.

Referências

La BARRE, Kathryn .*The Use of Faceted Analytico-Synthetic Theory as Revealed in the Practice of Website Construction and Design*; Thesis. School of Library and Information Science, Indiana University, Bloomington, US, 2004.

NEELAMENGHAN, A. Application of Ranganathan's General Theory of Knowledge Classification in Designing Specialized Databases, *Libri*, Vol 42, 1992. p. 202-226.

SOERGEL, Dagobert. *Knowledge organization systems in digital libraries*; tutorial. New York, Department of Library and Information Studies, Graduate School of Education University at Buffalo, NY, 2012. [Apresentado no PPGCI/UFMG]

SOERGEL, Dagobert. *Illuminating Chaos*; using semantics to harness the web; slides. New York, Department of Library and Information Studies, Graduate School of Education University at Buffalo, NY, 2012. [Apresentado no PPGCI/UFMG]

A importância teórica e prática da indexação na fundamentação científica da organização e representação do conhecimento

Mariângela Spotti Lopes Fujita
Universidade Estadual Paulista
fujita@marilia.unesp.br

A possibilidade que se abre nesta Segunda Reunião da ISKO-BRASIL de refletirmos sobre os avanços e perspectivas em Organização e Representação do conhecimento permite que façamos uma divisão necessária entre a fundamentação e a aplicabilidade científicas de modo a verificar quais articulações favorecem a intervenção na realidade com trocas substantivas entre uma e outra de modo a, de um lado, buscar a consolidação científica da área e de outro lado, buscar a legitimação social. Toda investigação em qualquer área do conhecimento necessita ter relevância científica e relevância social para justificar seu desenvolvimento. Passa por essa situação a área de Indexação no contexto da Organização e Representação do Conhecimento. O objetivo deste trabalho é apresentar uma reflexão sobre as contribuições teóricas e práticas para o desenvolvimento científico e social da Indexação cujos efeitos revelam a fundamentação científica da Organização e Representação do Conhecimento.

A Indexação, enquanto área de estudo e investigações, é compreendida por diferentes visões, entretanto, pela perspectiva da área de Organização e Representação do Conhecimento, representada pela sistematização de Dahlberg (1993) é vista como processo em Organização do Conhecimento e o resultado de sua aplicação como Representação do Conhecimento.

Dahlberg (2006:12) inclui na Organização do conhecimento “[...] os objetos e atividades da teoria do conceito, classificação e indexação e representação do conhecimento [...]”. Esta é, sem dúvida, a visão da área teórica em Ciência da Informação que fundamenta a Indexação.

O esquema de classificação concebido por Dahlberg, denominado “*Systematifier*”, contém as subdivisões da área de Organização do conhecimento em 10 grupos de 0 a 9. A explicação do *Systematifier*, conforme Fujita (2008), é fornecida por Dahlberg (1993) a partir da seqüência de sistematização dos grupos de 1 a 9, sem considerar o Grupo 0, em três subdivisões:

- Grupos 1-3: representam as divisões componentes/constituintes da área, caracterizadas por 1) Fundamentos teóricos, 2) Estrutura e construção de sistemas de classificação e tesouros, e 3) Classificação e Indexação;
- Grupos 4-6: representam as aplicações das divisões componentes/constituintes da área em: 4) Sistemas Universais, 5) Sistemas de classificação orientados ao objeto e tesouros, 6) Sistemas orientados à assuntos específicos;
- Grupos 7-9: representam a influência, aplicação e ambientes da área, 7) influências externas: problemas da representação do conhecimento pela linguagem e terminologia, 8) aplicação da classificação e indexação para diferentes tipos de dados, enunciados, documentos, e 9) levar da área de organização do conhecimento para fora ou externo, a organização da área em nível nacional e internacional, sua educação e treinamento, seus aspectos legais e econômicos, estudos de usuários e normalização.

O Grupo 3, em específico, considera a indexação no âmbito da Organização do Conhecimento a partir de seus aspectos teóricos e metodológicos relativos à fundamentos conceituais, processo e técnica, Indexação automática e avaliação da indexação.

Grupo 3 – Metodologia de classificação e indexação. Conteúdos: Teoria de Classificação e Indexação, Análise de assunto, Técnicas de classificação e indexação, Classificação e Indexação automática, Ordenação manual e automática, Geração de índices e programas, reclassificação, avaliação de classificação e indexação, Codificação e Decodificação (Fujita, 2008:9)

Com a perspectiva do ambiente institucional e a concepção do conhecimento registrado, Hjørland (2008) considera que, em sentido específico, a Biblioteconomia e a Ciência da Informação são as disciplinas centrais da Organização do Conhecimento e compreende atividades tais como “[...] descrição documentária, indexação e classificação realizadas em bibliotecas, bases de dados bibliográficas, arquivos e outros tipos de “instituições da memória” por bibliotecários, arquivistas, especialistas da informação, especialistas de assunto” (Hjørland, 2008:86).

Com história recente, a Organização do Conhecimento, em sentido específico, teve origem na institucionalização da ISKO, entretanto, tanto a indexação quanto a classificação vistas como áreas de atividades da Biblioteconomia e da Ciência da Informação, tem origens mais distantes e estão atreladas, na concepção de Guimarães (2008), ao desenvolvimento do Tratamento Temático da Informação (TTI). Nesse âmbito, Guimarães (2008:79) considera que o TTI

[...] ao longo da história foi, pouco a pouco, sendo permeado por distintas concepções, que propiciaram o delineamento de três correntes teóricas – *catalogação de assunto*, *indexação* e *análise documental* – que hoje encontram importante espaço de interlocução no âmbito da *International Society for Knowledge Organization* (ISKO).

O Tratamento Temático da Informação é parte das atividades e operações do tratamento da informação que envolve o conhecimento teórico e metodológico disponível quanto ao tratamento descritivo do suporte material da informação e ao tratamento temático de conteúdo da informação. Esta dicotomia que se apresenta no tratamento da informação é explicada, de um lado, pelo desenvolvimento teórico e metodológico distinto alcançado pelas duas áreas e de outro, pela diferença existente entre os aspectos da informação – o material e o conteúdo, que exigem tratamento diferenciado.

O tratamento temático do documento realiza uma descrição do conteúdo mediante representação condensada com vistas à sua acessibilidade temática e, nesse sentido, as operações de elaboração de resumos, classificação, catalogação de assuntos e indexação possuem graus de condensação nas representações documentais (verbais, não verbais ou codificadas), de modo que um resumo tem menor grau de condensação que palavras-chaves, cabeçalhos de assunto ou um número de classificação.

149

A indexação como processo de análise de assunto tem raízes teóricas e metodológicas ligadas à linha inglesa e a catalogação de assunto à linha norte-americana. Ambas tiveram desenvolvimentos próprios em ambientes institucionais e tipologias documentárias diferentes além de áreas de assunto mais especializadas no caso da indexação. Assim, a indexação é realizada em serviços de indexação e resumos com artigos de periódicos e documentação científica em geral, como por exemplo, na Biblioteca Virtual em Saúde [1], e a catalogação de assuntos em bibliotecas com livros e documentação publicada convencionalmente.

Com o crescimento dos serviços de indexação e resumo em todo mundo e a importância das bases de dados bibliográficas para o desenvolvimento científico e tecnológico, cada vez mais se buscava o aprimoramento teórico e metodológico da indexação com o objetivo de melhorar a recuperação por assuntos.

Em decorrência da necessidade de recuperação cada vez mais rápida, precisa e especializada por parte de Instituições informacionais, a indexação, enquanto operação do tratamento temático da informação, passou a contar com um aparato metodológico e

instrumental mais diversificado e muito mais voltado para o contexto de cada documento, domínio de assunto e ambiente.

Reflexo desse contexto, a indexação é ainda entendida por profissionais da informação, bibliotecários de modo geral, como operação realizada somente em serviços de informação que produzem bases de dados, porém a evolução científica e tecnológica que ocorreu em todas as áreas de conhecimento e atividades profissionais e sociais, alterou de forma irreversível o modo como se armazena, trata e recupera informação e conhecimento, atingindo significativamente os serviços de informação e as bibliotecas. Por um lado, os serviços de informação que antes produziam bases de dados referenciais de artigos de periódicos com base em indexação e elaboração de resumos, atualmente abrigam, também, bases de dados de periódicos eletrônicos de texto completo que não necessitam de indexação e elaboração de resumos. Por outro lado, as bibliotecas que construíram seus catálogos durante séculos para a comunidade local e frequentadora, passaram a disponibilizá-lo na web em formato on-line. Além disso, individualmente podemos ter acesso a uma imensa quantidade de documentos e outros recursos de informação na própria web sem que acessemos bases de dados ou catálogos de biblioteca.

Essas diferenças residem principalmente no fato de que a catalogação na biblioteca apresenta um conjunto de princípios firmados e reconhecidos mundialmente que fornecem padrões para a elaboração de registros bibliográficos contribuindo para a construção da área da catalogação e para o intercâmbio de informações. Os serviços de indexação e resumo, por sua vez, não têm, e apresentam diferenças entre seus métodos para descrição e pontos de acesso e diferentes padrões.

Além disso, devem ser considerados os objetivos e funções de um catálogo e de uma base de dados; a estrutura e conteúdo de um registro bibliográfico do catálogo e da base de dados e a escolha e as formas de pontos de acesso.

Esta dualidade da indexação entre serviços de indexação e resumos que produzem bases de dados e bibliotecas que produzem catálogos foi exposta por Lancaster (1993) quando refere ao tema de indexação abordado em seu livro que, embora esteja voltado principalmente para a indexação e redação de resumos, da forma como são praticadas por serviços de indexação e resumos em forma de publicação (seja em papel, seja em suporte eletrônico), e não em catálogos de biblioteca, os princípios são os mesmos, e alimenta a expectativa de que seu livro poderá interessar e ser útil

àqueles que se preocupam com o melhoramento do acesso temático nos catálogos em linha.

Por outro lado, a catalogação de assuntos, restrita ao ambiente institucional de bibliotecas, teve sua evolução atrelada às perspectivas técnicas da prática profissional do catalogador a partir das regras de Charles Ammi Cutter. Segundo Guimarães (2008), a catalogação de assuntos, assim como a indexação, são linhas teóricas do tratamento temático da informação (TTI), entretanto, considera que a primeira abordagem da catalogação de assuntos foi voltada diretamente para a atividade profissional em bibliotecas e sob forte influência da Escola de Chicago que decorreu dos princípios de catalogação alfabética de Cutter e da tradição de cabeçalhos de assunto da *Library of Congress*, cuja ênfase reside no catálogo enquanto produto do tratamento da informação em bibliotecas. O avanço do catálogo manual como instrumento de recuperação da informação, conforme Guimarães (2008) teve contribuições teóricas importantes de Cutter (1904), Coates (1960) e Kaiser (1911) na construção e organização de enunciados de assuntos para a formação de cabeçalhos de assunto. Atualmente contribuem para os estudos teóricos e metodológicos dessa linha autores como Olson & Boll (2001) e Šaupperl (2002) que se destacam no estudo da articulação entre a catalogação de assunto e a indexação.

A indexação na catalogação representa uma estratégia inteligente para que os catálogos possibilitem a recuperação por assuntos mais precisa e específica e, além disso, é possível, de antemão, prever que com os conhecimentos teóricos e práticos da área de indexação as bibliotecas terem linguagem mais especializada e abrangente, métodos de avaliação da indexação para melhoria da recuperação da informação, método de indexação, manual de política de indexação que assegure a qualidade da recuperação por assuntos na atual conjuntura de catalogação cooperativa, software de indexação automatizada.

É inegável o desenvolvimento teórico, prático, instrumental e pedagógico da Indexação que se traduz na construção de linguagens documentais específicos de cada domínio, considerados instrumentos de representação da informação documental, tais como tesouros, vocabulários controlados, listas de cabeçalhos de assuntos, ou no aprimoramento do processo de indexar de indexadores aprendizes e proficientes com propostas pedagógicas que visam a abrangência de conhecimentos específicos e gerais, na evolução teórica da indexação pontuada por estudos cuja fundamentação tem obtido consolidação mediante abordagens interdisciplinares com a Lingüística, Semiótica,

Ciência Cognitiva e até da Ciência da Computação no caso da indexação automatizada, dos métodos de avaliação da indexação que sempre acompanharam a prática de indexação visando verificar os efeitos produzidos na recuperação da informação e, ainda, a perspectiva da atuação do indexador frente à uma política de indexação. Tal abrangência tem consistência teórica, metodológica, instrumental e pedagógica completa que torna a indexação, sob ponto de vista da área de Organização e Representação do Conhecimento, uma área importante e estratégica com conhecimento teórico, metodológico e de inovação para a pesquisa e para a formação e atuação profissional.

Portanto, a sistematização teórica sobre indexação tem como base a área de Organização e Representação do Conhecimento. Apesar da satisfatória quantidade de literatura publicada em artigos e trabalhos de eventos sobre indexação e sub-temas afins, poucos autores conseguiram sistematizar teórica e metodologicamente a indexação. Todavia, quando se fala em indexação precisamos nos reportar à Foskett (1973, 1996) e seu livro “A abordagem temática da informação” [2] e a Lancaster (1993) e seu livro “Indexação e Resumos: teoria e prática” [3]. Saindo do âmbito da linha inglesa de indexação (*indexing*), mais recente é a publicação do livro “Manual de Indización: teoría y práctica” de Isidoro Gil Leiva, pesquisador espanhol que, ao contrário dos pesquisadores da linha espanhola de análise documental de conteúdo, segue a linha inglesa de *indexing*.

Para compreensão da abrangência estrutural, teórica e metodológica da indexação utilizamos como parâmetro a organização textual e o conteúdo dos livros de Lancaster (1993) e Gil Leiva (2008).

A organização textual do livro de Lancaster (1993) compreende, conforme nossa análise e síntese, os seguintes aspectos da indexação:

- prática de indexação: exaustividade e especificidade;
- princípios da indexação: etapas da indexação, tematicidade, vocabulários controlados (especificidade de vocabulário, tradução);
- métodos adotados em serviços de indexação e resumos;
- indexação automática;
- avaliação da indexação: qualidade e consistência de indexação.

Em Gil Leiva, os aspectos abordados sobre indexação, em nossa análise e síntese, são:

a) indexação:

- Qualidades da indexação (exaustividade, especificidade, correção e consistência);

- Normas sobre indexação;
 - Relação entre indexação e recuperação;
 - Indexação na internet;
- b) ferramentas para a indexação:
- listas de cabeçalhos de assuntos e tesouros;
 - interoperabilidade entre vocabulários controlados
- c) prática de indexação
- processo de indexação
 - indexação de documentos audiovisuais, sonoros, gráficos e textuais
 - política de indexação
- d) indexação automática
- e) avaliação da indexação: intrínseca e extrínseca

Em quadro comparativo, é possível observar aspectos em comum mesmo que a ordenação seja diferente:

Lancaster (1993)	Gil Leiva (2008)
- prática de indexação: exaustividade e especificidade;	a) indexação: - Qualidades da indexação (exaustividade, especificidade, correção e consistência); - Normas sobre indexação; - Relação entre indexação e recuperação; - Indexação na internet;
- princípios da indexação: etapas da indexação, tematicidade, vocabulários controlados (especificidade de vocabulário, tradução);	b) ferramentas para a indexação: - listas de cabeçalhos de assuntos e tesouros; - interoperabilidade entre vocabulários controlados
- métodos adotados em serviços de indexação e resumos;	c) prática de indexação - processo de indexação; - indexação de documentos audiovisuais, sonoros, gráficos e textuais; - política de indexação
- indexação automática;	d) indexação automática
- avaliação da indexação: qualidade e consistência de indexação.	e) avaliação da indexação: intrínseca e extrínseca

Quadro 1: Abrangência estrutural, teórica e metodológica da indexação segundo Lancaster e Gil Leiva

Tendo em vista Lancaster (1993) e Gil Leiva (2008) é possível sistematizar a abrangência estrutural, teórica e metodológica da indexação com os seguintes aspectos:

- prática de indexação: processo de indexação, tematicidade, normas sobre indexação, manuais de indexação, indexação automática;
- qualidades da indexação: exaustividade, especificidade, correção e consistência
- ferramentas para a indexação: listas de cabeçalhos de assunto e tesouros
- avaliação da indexação

Entretanto, o cerne da indexação, propriamente dita, está no processo de indexação para representação documentária que inclui a tematicidade (Hutchins, 1977,

Wilson, 1985, Beghtol, 1986, Albrechtsen, 1993, Todd, 1992, Naves, 1996, Fujita, 2003a, Tálamo, 1987, Moraes, 2012). A literatura sobre o processo de indexação tem qualidade comprovadas por estudiosos como Metcalfe (1959), Fairthorne (1969) Foskett (1973), Austin (1974), Farradane (1977), Aitichinson, Gilchrist (1979), Lancaster (1993) e outros que criaram os marcos teóricos fundamentais da indexação, motivo pelo qual foi possível a elaboração das normas como a ISO-5693 (International Organization For Standartization, 1985) e a ABNT-12.676 (Associação Brasileira De Normas Técnicas, 1992) sobre o processo de indexação cujo conceito foi generalizado para a determinação de assuntos em conteúdos documentários a fim de criar uma amplitude maior de aplicabilidade.

Ao discutir o progresso científico de uma área é importante observar a abrangência teórico-prática mesmo que o conjunto não seja considerado ciência, mas dotado de aspectos científicos importantes. Fujita (2008) analisa a perspectiva de Dahlberg (2006) acerca dos aspectos essenciais da ciência (teorias, modelos, hipóteses) e foi possível considerar como principais fundamentações teóricas da área de Organização e Representação do Conhecimento o trabalho de Ranganathan referente à análise em facetas para a categorização e às regras de combinação de conceitos, parcialmente antecipadas pela invenção dos auxiliares de Otlet e La Fontaine, que propiciaram a revisão e expansão do Dewey Decimal Classification. A partir disso é possível considerar na área de Indexação, contribuições à sua fundamentação teórica que reforçam e dão consistência à área de Organização e Representação do conhecimento. Nesse sentido é importante que as linhas investigativas em Ciência da Informação realmente façam a sistematização e análise dessas contribuições que podem ser avaliadas pela repercussão teórica e metodológica que obtiveram, como por exemplo, o Sistema de Indexação PRECIS.

Em 1974, Derek Austin, Diretor da British Library e pesquisador integrante do Classification Research Group (CRG), publica o manual do Sistema de Indexação PRECIS [4] (Austin, 1974) que, embora desatualizado, em termos de aplicabilidade prática, oferece base teórica e metodológica aprofundada, interdisciplinar e inovadora ao considerar o respaldo teórico da linguística gerativo transformacional e de Casos, bem como trabalhar na perspectiva da abordagem cognitiva e sociocognitiva. A abordagem cognitiva do PRECIS estava na análise conceitual do processo de indexação realizado pelo indexador humano para atribuição de operadores de função que geravam automaticamente as entradas de assunto dos índices da British National Bibliography. A análise conceitual aplicada ao documento pelo indexador propunha uma construção de

conhecimentos a partir do contexto do documento que se associava à lógica construtivista. A abordagem sociocognitiva estava presente na filosofia de preservação do contexto do conteúdo do texto e do trabalho cognitivo do indexador tendo em vista que a análise conceitual indicava o uso de operadores de estrutura superficial e de estrutura profunda que equivaliam ao conteúdo explícito e implícito, respectivamente. O PRECIS foi além da proposta de aplicação de um sistema de indexação, realizou inovação e é um marco teórico e metodológico de relevância na área de Indexação para o aprimoramento do processo de indexação.

Outros marcos teóricos dotados de componente de inovação sobre o processo de Indexação surgiram:

Décadas	Marcos teóricos
70	Na década de 70 abordando aspectos lingüísticos e cognitivos tais como Austin (1974) Jones (1976), Borko (1977), Cooper (1978) e Fugmann (1979) que forneceram lastros teóricos para investigações posteriores.
80	Ainda na década de 80, Fugmann (1984) desenvolveu uma teoria de indexação a partir de 5 axiomas relacionados com a estratégia de busca e recuperação em bases de dados e inicia uma longa discussão sobre política de indexação que antecipa resultados da busca e recuperação da informação para “ganhar tempo” de busca do usuário; Soergel (1985) utiliza concepção da indexação orientada pelo documento que não considera o contexto e as necessidades de usuários; No Brasil, o grupo Temma (SMIT, 1989) é responsável, na década de 80 em diante, por trabalhos em Indexação, na linha de Análise Documentária que aliam aportes teóricos da Lingüística, Lógica e da Psicologia Cognitiva.
90	Na década de 90, Fidel (1994) aborda a indexação a partir de concepções orientadas a usuários e suas demandas na recuperação da informação, diferente das concepções mais exclusivamente orientadas ao documento. Na perspectiva da concepção orientada pelo usuário, os estudos de avaliação da indexação usam as características ou qualidades de exaustividade, especificidade, correção e consistência que podem ser medidas pela recuperação e, ao mesmo tempo, modificar e aprimorar a recuperação; Albrechtsen (1993) propõe a existência de três concepções de análise de assunto no processo de indexação: a simplista, a orientada para o conteúdo e a orientada para a demanda; Farrow (1991), com um modelo do processo cognitivo da indexação de documentos, além de Bertrand e Cellier (1995), enfatizam a perspectiva cognitiva da Indexação. Ainda na década de 90, Lancaster (1993) e Fugmann (1993) aliam teoria e prática e Frohman (1990) propõe regras para indexação sem a perspectiva do que ele denomina “mentalismo”.
2000	A partir de 2000 surgem com Mai (2001) a perspectiva da semiótica para análise da natureza do processo de indexação de assunto e a perspectiva de domínio na indexação orientada pelo domínio (MAI, 2004). No Brasil, as pesquisas de perspectiva cognitiva na Indexação são desenvolvidas por Fujita (2003b), Naves (2000) e Neves (2004). Fujita e Rubi (2006) avança na proposta de um Modelo de Leitura documentária para indexação de artigos científicos com base em resultados de coleta de dados introspectiva do Protocolo Verbal para observação do processamento mental da tarefa de indexação por indexadores proficientes e aprendizes.

Quadro 2: Evolução da teoria da indexação
Fonte: Almeida, Fujita, Reis, 2013

O conteúdo deste quadro demonstra uma síntese da evolução da teoria da indexação sem avaliação crítica. Porém, é importante destacar, a partir desta síntese, que a Indexação tem uma evolução teórica realizada a partir de abordagens com enfoque interdisciplinar como é o caso da linguística, da psicologia cognitiva e da semiótica porque contribuem na investigação de aspectos da natureza do processo de indexação, bem como de concepções cujos enfoques se destinam às finalidades e elementos do processo de indexação, o usuário, o documento ou o domínio.

A evolução da teoria da indexação demonstra bem que as abordagens estão intrinsecamente interligadas à natureza do processo de indexação, de fato, podem parecer pouco claras por estarem ligadas a processos mentais pouco esclarecidos ainda pelos estudos cognitivos, linguísticos e semióticos, mas consideramos que as abordagens ajudam a iluminar determinados pontos e aspectos advindos da prática que teoricamente ainda não conseguimos esclarecer. Também, pode-se inferir que os estudos interdisciplinares realizados por pesquisadores da área de Indexação ou de Ciência da Informação dependem, em grande medida, de boa compreensão das teorias dessas áreas cuja aplicabilidade não foi testada. De toda maneira, pode-se creditar às abordagens, o avanço teórico alcançado pela Indexação.

É importante observar que a natureza do processo de indexação, embora empírico e sujeito à uma aplicabilidade intuitiva demanda investigações científicas que se proponham a aprimorar o conhecimento profissional de quem o executa bem como os instrumentos, métodos e técnicas utilizados para tal. Significa de outro modo que os objetos de pesquisa da área de Indexação são resultados dos usos e práticas que, como tal, devem ser investigados cientificamente e incorporados ao universo de pesquisa e, em sequência, retornarem aos usos e práticas com aprimoramentos e adequações que sejam válidos para várias outras necessidades, domínios de assunto e ambientes variados, cumprindo, assim, o ciclo de generalização científica fundamental ao progresso social.

[1] BVS – iniciativa do Ministério da Saúde do Brasil e BIREME/OPS/OMS em parceria com instituições nacionais (<http://brasil.bvs.br/>)

[2] Originalmente publicada em 1969 com o título “The approach of subject information” e no Brasil, foi traduzida por A.A. Briquet de Lemos como “A abordagem temática da informação” (Foskett, 1973) e atualmente está na 5ª edição (Foskett, 1996)

[3] Originalmente publicado em 1991 com o título “Indexing and abstracting in theory and practice”, e 2. edição publicada em 2004

[4] Acrônimo que significa *PR*eserved *C*ontext *I*ndexing *S*ystem

Referencias

Aitichison, J., Gilchrist, 1979. *A. Manual para construção de tesouros*. Trad de Helena Medeiros Pereira Braga. Rio de Janeiro: BNG/BRASILART, 142p.

Albretchtsen, H. 1993. Subject analysis and indexing: from automated indexing to domain analysis. *The Indexer*, London, 18: 4, 219-24, Oct.

Almeida, C. C. de, Fujita, M. S. L., Reis, D. M. dos. 2013. Peircean semiotics and subject indexing: contributions of speculative grammar and pure logic. *Knowledge Organization*. (artigo aceito para publicação)

Associação Brasileira De Normas Técnicas. 1992. *NBR 12676: Métodos para análise de documentos - determinação de seus assuntos e seleção de termos de indexação*. Rio de Janeiro, 4p.

Austin, D. 1974. *PRECIS: a manual of concept analysis and subject indexing*. London: Council of the British National Bibliography, 551 p.

Beghtol, C. 1986. Bibliographic classification theory and text linguistics: aboutness analysis, intertextuality and the cognitive act of classifying documents. *Journal of Documentation*, London, 42: 2, 84-113.

Bertrand, A., Cellier, J.M. 1995. Psychological approach to indexing: effects of the operator's expertise upon indexing behavior. *Journal of Information Science*, 21:6, 459-72.

Borko, H. 1977. Toward a theory of indexing. *Information Processing & Management*, 13:335-365.

Coates, E. J. 1960. *Subject catalogues: headings and structures*. London: Library Association.

Cooper, W.S. 1978. Indexing documents by Gedanken experimentation. *Journal of the American Society for Information Science*, may:107-119.

Cutter, C. A. 1904. *Rules for a dictionary catalog*. 4.ed. Government Printing Office: Washington.

Dahlberg, I. 1993. Knowledge organization: its scope and possibilities. In: *Knowledge Organization*, Frankfurt, 20:4, 211-222.

_____. 2006. Knowledge organization: a new science? *Knowledge Organization*, Frankfurt, 33:1, 11-19.

Fairthorne, R. A. 1969. Content analysis, specification, and control. *Annual Review of Information Science and Technology*, Medford, NJ, 4: 73-109.

Farradane, J. A. 1977. A comparison of some computer produced permuted alphabetical subject indexes. *International Classification*, Munich, 4:2, 94-101.

Farrow, J. 1991. A cognitive process model of document indexing. *J. Doc.*, London. 47: 2, 149-166.

Fidel, R. 1994. User-centered indexing. *Journal of the American Society for Information Science*, 45:8, 572-576.

Foskett, A.C. 1973. *A abordagem temática da informação*. Tradução de Antônio Agenor Briquet de Lemos. São Paulo: Polígono; Brasília: Ed.UnB.

Frohmann, B. 1990. Rules of indexing: a critique of mentalism in information retrieval theory. *Journal of documentation*, 46:2, 81-101.

Fugmann, R. 1984. The five-axiom theory of indexing of indexing and information supply. *Journal of the American Society for Information Science*, 36:116-129,

_____. 1979. Toward a theory of information supply and indexing. *International Classification*, 6:1, 3-15.

Fujita, M. S. L. 2003a. A identificação de conceitos no processo de análise de assunto para indexação. *Revista Digital de Biblioteconomia e Ciência da Informação*, 1:1,60-90.

_____. 2003b. *A leitura documentária do indexador: aspectos cognitivos e lingüísticos influentes na formação do leitor profissional*. 321f. Tese (Livre-Docência em Análise Documentária e Linguagens Documentárias Alfabéticas) – Faculdade de Filosofia e Ciências da UNESP, Marília.

_____. 2008. Organização e representação do conhecimento no Brasil: análise de aspectos conceituais e da produção científica do ENANCIB no período de 2005 a 2007.. *Tendências da Pesquisa Brasileira em Ciência da Informação*, 1:1 – 32.

_____, Rubi, M. P., Boccato, V. R. C. 2009. O contexto sociocognitivo do catalogador em bibliotecas universitárias: perspectivas para uma política de tratamento da informação documentária. *Datagramazero*, Rio de Janeiro, abril.

_____.; Rubi, M. P. 2006. Modelo de lectura profesional para la indización. *Scire (Zaragoza)*, 12: 47-69.

GIL LEIVA, I. 2008. *Manual de indización*. Teoría y práctica. Gijón: Trea.

Guimarães, J. A. C. 2008. A dimensão teórica do tratamento temático da informação e suas interlocuções com o universo científico da International Society for Knowledge Organization (ISKO). *Revista Ciência da Informação*, Brasília, 1: 77-99, jan./abr.

Hjørland, B. 2008. What is knowledge organization? *Knowledge Organization*, 35:2/3, 86-101.

Hutchins, W. K. 1977. On the problem of aboutness in document analysis. *Journal of Informatics*, 1: 17-35.

International Organization For Standartization. 1985. *Documentation – methods for examining documents, determining their subjects, and selecting indexing terms*. 1.ed. Suíça: ISO, 5p.

Jones, K. P. 1983. How do we index? A report of some Aslib Information Group activity. *J. Doc.*, London, 39: 1-23.

Kaiser, J. O. 1911. *Systematic indexing*. London: Pitman.

- Lancaster, F. W. 1993. *Indexação e resumos: teoria e prática*. Tradução de Antonio Agenor Briquet de Lemos. Brasília: Briquet de Lemos/Livros, 347p.
- Mai, J-E. 2001. Semiotics and indexing: an analysis of the subject indexing process. *Journal of Documentation*, 57:5, 591-622.
- Mai, J-E. 2004. The role of domains, documents, and decisions in indexing. In: *Knowledge Organization and the Global Information Society. Proceedings of the Eight International ISKO Conference*. Advances in Knowledge Organization, 9: 207-213
- Metcalfe, J. *Subject classifying and indexing of libraries and literature*. New York: Scarecrow, 1959.
- Moraes, J. B. E. 2012. Perspectivas metodológicas para la identificación de la tematicidad en los textos narrativos de ficción. *Scire (Zaragoza)*, 18: 57-66,
- Naves, M. M. L. 2000. *Fatores interferentes no processo de análise de assunto: estudo de caso de indexadores*. 257 f. Tese (Doutorado em Ciência da Informação) – Escola de Ciência da Informação, Universidade Federal de Minas Gerais, Belo Horizonte.
- _____. 1996. Análise de assunto: concepções. *Revista de Biblioteconomia de Brasília*, Brasília, 20:2, p. 215-226, jul./dez.
- Neves, D. A. de B. 2004. *Aspectos metacognitivos na leitura do indexador*. 2004. 131 f. Tese (Doutorado em ciência da informação) Escola de ciência da informação, Universidade Federal de Minas Gerais, Belo Horizonte.
- Olson, H. A.; Boll, J. L. 2001. *Subject analysis in online catalog*. 2.ed. Englewood: Libraries Unlimited.
- Šauperl, A. 2002. *Subject determination during the cataloging process*. Lanham: Scarecrow Press.
- Smit, J.W. 1989. *Análise documentária : a análise da síntese*. 2. ed. Brasília: IBICT.
- Soergel, D. 1985. *Organizing information: principles of data base and retrieval systems*. Orlando: Academic Press.
- Tálamo, M. F. G. M. 1987. *Elaboração de resumos*. Escola de Comunicação e Artes, 14 f. Datilografado.
- Todd, R. T. 1992. Academic indexing: what's it all about? *The Indexer*, London, 18: 2, 101-104.
- Wilson, P. 1985. Subject and the sense of position. In: CHAN, C. et al. *Theory of subject analysis: a manual*. Littleton, Colorado: Libraries Unlimited, p. 306-23.

UMLS como suporte ao processamento semântico de artigos em ciências biomédicas

Carlos H. Marcondes
Universidade Federal Fluminense
marcon@vm.uff.br

1. Problema

A criação do “Philosophical Transactions” em 1665 marca o surgimento do periódico científico. Desde então, este vem cumprindo importantes papéis nos marcos da ciência. BJÖRK et al. (2009) estimam que, no ano de 2006 foram publicados 1.346.000 artigos científicos. Na área Biomédica o PubMed (www.ncbi.nlm.nih.gov/pubmed/) contém hoje cerca de 22 milhões e 500 mil referências, que crescem aceleradamente.

Attwood et al. (2009) reclamam da premência de ferramentas computacionais para encontrar conhecimento “perdido” na literatura científica publicada eletronicamente. Novas áreas de pesquisa endereçam esta questão, como “Literature-based discovery” (SWANSON et al., 2006), (KOSTOFF et al., 2008), “text mining” (TANABE et al., 1999), (BATH, 2002), (SPASIC et al., 2005).

160

2. Modelo semântico de artigos científicos e seus requisitos

O conhecimento científico consiste em afirmações no texto das conclusões do artigo, expressando *relações entre fenômenos* (MARCONDES, 2011). Para raciocinar sobre fenômenos cientistas usam conceitos, expressos através de termos, organizados em terminologias como UMLS – Unified Medical Language System.

Kuhn (2007) teorizou sobre a mudança de paradigmas científicos. Ressalta que é necessário um novo sistema conceitual para descrever um paradigma emergente e que isto requer um intervalo de tempo para sua representação nas terminologias científicas.

A partir destas considerações o modelo endereça duas funcionalidades: 1- representar o conhecimento contido nas conclusões de artigos em formato semanticamente “rico”, processável por máquinas; 2- permitir a identificação de indícios de descobertas científicas (MALHEIROS, 2010), ao comparar o conteúdo das

conclusões de artigos, com o conteúdo de termos no UMLS. Estas funcionalidades foram implementadas através do protótipo de interface de submissão de artigos (COSTA, 2010), criando um “registro semântico”.

O trabalho objetiva discutir a experiência do uso do UMLS para atender aos requisitos descritos.

3. Metodologia

São usados os seguintes parâmetros de análise: a- expressividade ontológica: categorias, suporte à relações; b- apoio ao processamento lingüístico, c- formalismo lógico e computacional: definições expressas em Lógica, representadas em formatos legíveis por programas.

4. Resultados e Conclusões

O UMLS é uma base terminológica no domínio das ciências biomédicas, englobando mais de 100 fontes²⁰. É composta de três bases de conhecimento integradas: o Metathesaurus, contendo 2.886.423 termos²¹; a “Semantic Network”, que estrutura termos biomédicos em 154 categorias, denominadas “*semantic types*”, relacionadas entre si por 54 “*semantic relations*”; e o Specialist Lexicon, com informações sintáticas, morfológica e ortográficas sobre termos em inglês encontrados no UMLS. Rotinas em linguagem de programação Java usam o Specialist Lexicon para tratar textos biomédicos.

Em relação aos parâmetros de análise, constata-se o seguinte:

a- expressividade ontológica: o UMLS, ao contrário de vocabulários que fazem parte dele e são utilizados para indexar a literatura científica como o MeSH, incorpora *relações semânticas* na “Semantic Network”, nas quais são especificados tipos

20

http://www.nlm.nih.gov/research/umls/knowledge_sources/metathesaurus/release/source_vocabularies.html

21

http://www.nlm.nih.gov/research/umls/knowledge_sources/metathesaurus/release/statistics.html

semânticos que podem fazer parte de cada uma delas. Num modelo de representação do conhecimento como o proposto, esta característica é essencial, uma vez que o objetivo do processamento é que as conclusões do artigo sejam formatadas em relações, onde cada um dos relata é mapeado para termos do “MetaTheaurus” e a relação é mapeada para uma das “semantic relations” do “Semantic Network”; estas trazem somente definições textuais, não lógicas nem formais, o que impede o processamento por programas.

b- apoio ao processamento lingüístico: o Specialist Lexicon e as ferramentas de software associadas apóiam a identificação da estrutura sintática das conclusões dos artigos e a identificação no texto de termos biomédicos. As conclusões incluem texto com finalidade retórica, que tem que ser tratado pela atribuição de pesos aos termos biomédicos candidatos a relata, de acordo com sua ocorrência em partes específicas do artigo, como título, resumo, palavras-chave, objetivo, conclusões (COSTA, 2010).

c- suporte à identificação de descobertas: para que “registros semânticos” de um artigo possam ser comparados por programas com o UMLS é necessário que ambos estejam em *formatos legíveis por programas*. Marcondes e Malheiros (2009), Malheiros (2010) e Marcondes (2011) propõe representar “registros semânticos” usando RDF. Existem experiências para a representação do UMLS usando tecnologias da Web Semântica; Segundo Kashyap e Bordiga (2003), o UMLS é se ressentido de formalismo, o que impediria a adoção destas tecnologias. Por exemplo, as relações do Semantic Network são sempre binárias (entre dois *relata*), mas não trazem especificada a *cardinalidade* de cada um dos relata, isto é, em que número cada um dos relata participa da relação; segundo o exemplo dado (KASHYAP e BORGIDA, 2003, p. 6), uma relação como <bactéria><causa><infecção>, sem especificação da cardinalidade, poderia ter interpretação ambígua, como “All bacteria cause {each/only/some} infection(s)”.

Quanto ao requisito b-, a falta de formalismo do UMLS torna-se vantajosa na tarefa de validação pelos autores da formatação das conclusões em relações feita pelo sistema, uma vez que pode lhes ser exibida a definição textual dos termos e relações.

No UMLS cada conceito possui um identificador único, o CUI – Concept Unique Identifier -, que permite a criação de URIs (RFC 2396, 1998), essenciais para identificar recursos segundo os padrões da Web Semântica.

Existem também formalismos como SKOS, voltados especificamente para representação de tesouros, esquemas de classificação e taxonomias, com o uso de tecnologias da Web Semântica; seu uso, que poderia representar uma transição para linguagens formais como OWL. Dado aos aspectos institucionais envolvidos na gestão da UMLS, a adoção destes formalismos e padrões parece ainda distante.

Referências

- ATTWOOD, T. K.; KELL, D. B.; MCDERMOTT, P.; MARSH, J.; PETTIFER, S. R.; THORNE, D. Calling international rescue: knowledge lost in literature and data landslide! *Biochemical Journal*, [S.l.], dec. 2009.
- BATH, P. Data Mining in Health and Medical Information, *Anual Review of Information Science and Technology*, v. 38, p. 331–369, 2002.
- BJÖRK, B. C., ROOS, A.; LAURI, M. Scientific journal publishing: yearly volume and open access availability. *Information Research*, v. 14, n. 1 paper 391, 2009. Disponível em: <<http://InformationR.net/ir/14-1/paper391.html>>. Acesso em: 22 fev. 2013.
- COSTA, Leonardo Cruz. Da. Um proposta de processo de submissão de artigos científicos à publicações eletrônicas semânticas em Ciências Biomédicas, Tese (doutorado), Programa de Pós-graduação em Ciência da Informação UFF-IBICT. Niterói, (2010).
- KASHYAP, Vipul; BORGIDA, Alex. Representing the UMLS® semantic network using OWL. In: *The Semantic Web-ISWC 2003, 2003, Proceedings...* Springer, 2003. p. 1-16, Disponível em: <<http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.94.3296&rep=rep1&type=pdf>>. Acesso em: 28 fev. 2013.
- KOSTOFF, R. N.; BRIGGS, M. B.; SOLKA, J. L.; RUSHENBERG, R. L. (2008). Literature-related discovery (LRD): Methodology. *Technological Forecasting and Social Change*, v. 75, n. 2, p.186–202. Disponível em: <[doi:10.1016/j.techfore.2007.11.010](https://doi.org/10.1016/j.techfore.2007.11.010)>. Acesso em: 20 jul. 2010.
- KUHN, Thomas S. *A estrutura das revoluções científicas*. São Paulo: Perspectiva; 2007.
- MARCONDES, Carlos Henrique. Um modelo semântico de publicações eletrônicas. *Liinc em revista*, v. 7, n. 1, 2011. Disponível em <<http://revista.ibict.br/liinc/index.php/liinc/article/viewFile/404/262>>. Acesso em: 30 maio 2011.
- MARCONDES, Carlos Henrique; MALHEIROS, Luciana Reis. Identifying traces scientific discoveries by comparing the content of articles in biomedical sciences with web ontologies. In: *ISSI - International Conference on Informetrics and Scientometrics, 2009, Rio de Janeiro. 12, Proceedings...* São Paulo: BIREME/PAHO/WHO, UFRJ, 2009. v. 1. p. 173-177.
- MALHEIROS, Luciana Reis. *A identificação de traços de descobertas científicas pela comparação do conteúdo de artigos em Ciências Biomédicas com uma ontologia pública*. Tese (Doutorado em Ciência da Informação)-Programa de Pós-Graduação em Ciência da Informação convênio UFF/IBICT, Niterói, 2010.

RFC 2396 - Uniform Resource Identifiers (URI): Generic Syntax, Berners-Lee T., Fielding R., Masinter L., IETF, August 1998. Disponível em: <<http://www.isi.edu/in-notes/rfc2396.txt>>. Acesso em: 15 dez. 2011.

SKOS. Simple Knowledge Organization System, <http://www.w3.org/2004/02/skos/>.

SPASIC, I.; ANANIADOU, S.; MCNAUGHT, J.; KUMAR, A. Text mining and ontologies in biomedicine: making sense of raw text. *Briefings in bioinformatics*, v. 6, n. 3, p. 239–51, 2005. Disponível em: <<http://www.ncbi.nlm.nih.gov/pubmed/16212772>>. Acesso em 27 jul. 2010.

SWANSON, D.R.; SMALHEISER, N. R.; TORVICK V. I. Ranking indirect connections in literature based discovery. The role of Medical Subject Headings, *Journal of the American Society for Information Science and Technology*, v. 57, n.11, p. 1427–1439, 2006.

TANABE, L.; SCHERF, U.; SMITH, L. H.; LEE, J. K.; HUNTER, L. ; WEINSTEIN, J. N. MedMiner: an Internet text-mining tool for biomedical information, with application to gene expression profiling. *BioTechniques*, v. 27, n. 6, p. 1210–4, 1216–7, 2009. Disponível em: <<http://www.ncbi.nlm.nih.gov/pubmed/10631500>>. Acesso em 4 abr. 2009.

A norma ISO 25964 e a semântica latente das folksonomias: a interoperabilidade semântica em questão

Juliana de Assis
Universidade Federal de Minas Gerais
jayaweb@gmail.com

Maria Aparecida Moura
Universidade Federal de Minas Gerais
cidamoura@gmail.com

1 Introdução

Na *web* são cada vez mais requeridas estratégias e metodologias que integrem tanto a formalização e estruturação semântica do conhecimento quanto o uso dinâmico da linguagem dos sujeitos que se agregam em inúmeras redes sociais mediante o compartilhamento de narrativas e conteúdos.

O foco na estruturação e formalização das metalinguagens voltadas para a representação e a organização do conhecimento com o objetivo de contribuir para a consolidação da Web 3.0 é observado na norma ISO 25964.

A norma ISO 25964, intitulada “*Informação e documentação - Tesouros e interoperabilidade com outros vocabulários*” representa um avanço para o campo, visto que é elaborada em diálogo com suas características e desafios recentes. A norma tem como principal objetivo promover a interoperabilidade entre as metalinguagens.

A partir da análise desta norma problematiza-se a busca pela interoperabilidade semântica entre diferentes estruturas conceituais, construídas com propósitos diferentes, para públicos e contextos diferentes.

Argumenta-se que a interoperabilidade semântica pressupõe acordos de sentido que podem ser explorados a partir da semântica latente nas folksonomias, devido a isso, as mesmas podem contribuir para o desenvolvimento de metodologias e instrumentos de representação e recuperação da informação nos mais diversos contextos.

2 A ISO 25964 e a interoperabilidade semântica: desafios e possibilidades

A norma ISO 25964 é dividida em duas partes, a primeira publicada em 2011 e a segunda publicada em 2013. A primeira parte da norma (ISO 25964-1) abrange a estruturação de tesouros monolíngues e multilíngues, a aplicação da análise facetada aos tesouros, a modelagem de dados a partir do *XML Schema*, diretrizes para manutenção e o desenvolvimento dos tesouros e requisitos necessários a *softwares* para atuarem nesse campo (CLARKE; ZENG, 2012).

A segunda parte (ISO 25964-2), aborda aspectos adicionais para a interoperabilidade, focando diretrizes para mapeamentos conceituais entre tesouros e outras metalinguagens, como as taxonomias, as classificações bibliográficas e as ontologias, que são consideradas estruturas informacionais da Web 3.0.

A ISO 25964 assume a recomendação da W3C (*World Wide Web Consortium*) ao adotar um conjunto de especificações comuns para a elaboração desses vocabulários denominado SKOS (*Simple Knowledge Organization System*), derivado do RDF (*Resource Description Framework*) e do RDFS (*RDF Schema*), o SKOS possibilita a estruturação e formalização das metalinguagens com foco na interoperabilidade entre as mesmas.

Por interoperabilidade entende-se a capacidade de um sistema se comunicar e trabalhar no intercâmbio de dados com outro sistema de modo efetivo (REITZ, 2004). Esse tipo de cooperação entre sistemas distintos envolve dois níveis básicos de interoperabilidade, o primeiro nível (sintático) refere-se aos formatos, protocolos, sistemas de segurança e outros que garantam a troca de mensagens; o segundo nível (semântico) refere-se aos acordos semânticos e a interpretação da informação (MOREIRA; LARA, 2012, *on-line*).

O foco deste trabalho recai sobre a interoperabilidade semântica definida por Patel *et.al* (2005) como a capacidade de diferentes sistemas de informação comunicarem informações de acordo com o significado pretendido da informação codificada. Acredita-se que os acordos semânticos observados nos espaços em que a folksonomia ocorre podem constituir elementos que auxiliem no desenvolvimento de metodologias de construção das metalinguagens em diálogo com a norma ISO 25964.

Folksonomias constituem espaços sociais semânticos em que ocorre a justaposição de representações subjetivas, objetivas e práticas dos objetos informacionais (QIN, 2008). Isso faz com que sejam frágeis em relação ao grau de formalização da linguagem, mas confere às mesmas um elevado potencial de semanticidade. Esse potencial, denominado “semântica latente”, tem sido explorado como uma alternativa a

concepção de ontologias com vistas ao desenvolvimento da Web 3.0 (BASSO; SILVA, 2009), (ALVES, 2012).

Observa-se a partir de 2005, com a criação da *Tag Ontology*, proposta por *Thomas Gruber* um crescimento considerável de pesquisas voltadas para a formalização e integração das *tags* através de ontologias. Tal movimento pode ser incorporado à criação e gestão de outras metalinguagens visando a formação e o compartilhamento de acordos semânticos entre as mesmas.

Ontologias pressupõem um modelo conceitual estruturado em um conjunto de definições formais e processáveis por máquina e o compartilhamento de conhecimento consensual entre determinado grupo de indivíduos (BORST, 1997). Consideramos que a obtenção desse consenso é um dos desafios não apenas para a concepção de ontologias, mas também para alcançar o objetivo de promover a interoperabilidade semântica entre distintas metalinguagens conforme estabelece a norma ISO 25964.

3 Comunidades virtuais de prática e regularidades semânticas

A fase inicial deste estudo analisou as semânticas latentes oriundas de três comunidades virtuais de prática, que se agregam em torno dos temas Saúde e Jogos Digitais nos seguintes serviços que utilizam folksonomia: *Delicious*, *Diigo* e *Stumble Upon*.

Pela complexidade desses ambientes, que envolvem redes sociais e redes semânticas, utilizou-se uma perspectiva metodológica composta pela Netnografia, a Análise de Redes Sociais e a Semiótica.

Os temas Saúde e Jogos Digitais foram escolhidos por apresentarem características gerais e específicas, respectivamente, atendendo ao critério de diversidade das *tags*. Já as ferramentas escolhidas apresentaram funcionalidades e interfaces favoráveis à organização e compartilhamento dos conteúdos informacionais.

A partir da identificação e formalização dos laços relacionais entre os atores que constituem as comunidades foi possível identificar e monitorar os acordos semânticos estabelecidos entre os mesmos através das *tags*.

Constatou-se que, de um modo mais amplo, as *tags* são entidades que indicam as trocas geradas pelos laços cognitivos entre os atores sociais (MIKA, 2007). Ou seja, remetem a ligações que surgem a partir do compartilhamento de narrativas, linguagens e signos comuns promovendo a troca de conhecimentos. A Figura 1 representa um extrato

dos dados coletados no âmbito da temática “Jogos Digitais”, em que é possível observar as regularidades semânticas nas interseções entre as comunidades virtuais analisadas.

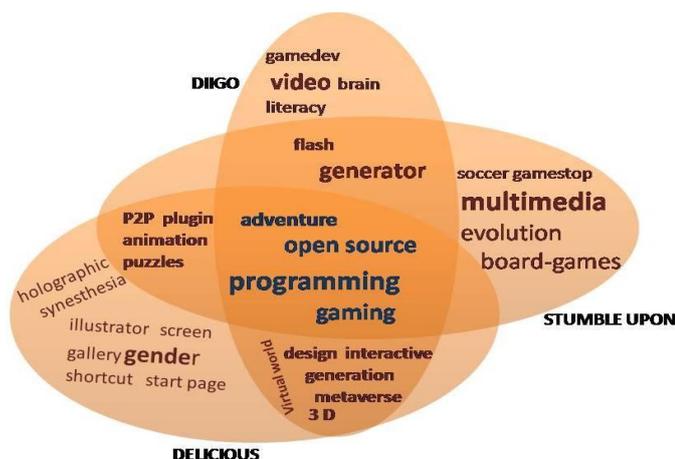


Figura 1: Comunidades virtuais e regularidades semânticas.

Nas interseções entre estes ambientes observamos um exemplo do que se almeja através da interoperabilidade semântica, visto que as *tags* destacadas resultam de significados pactuados pelas distintas comunidades virtuais. Dessa forma, tais contextos apresentam potencialidades para a elaboração de metalinguagens que compartilhem um modelo comum baseado em acordos semânticos, conforme a Figura 2.

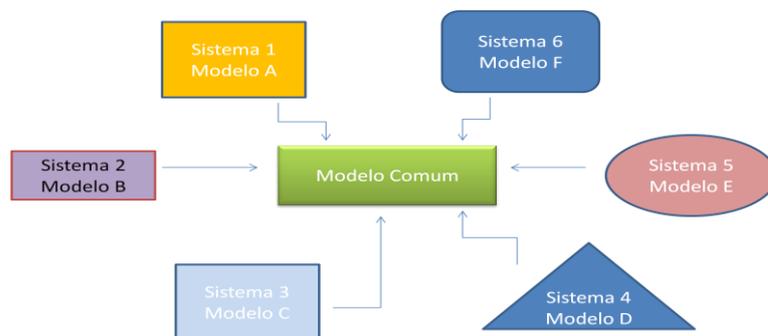


Figura 2: Interoperabilidade semântica. Fonte: Lima (2010)

Um dos próximos desenvolvimentos deste estudo prevê análise e reuso de ontologias baseadas em *tags* (*Tag Based Ontologies*) que utilizam elementos SKOS para explicitar as relações entre as *tags* a fim de promover a estruturação e a integração entre as semânticas latentes observadas nas comunidades analisadas.

De acordo com esta proposta, destaca-se a ontologia de alto nível UTO (*Upper Tag Ontology*) que foi elaborada por um grupo de pesquisadores liderados por *Ying Ding* e *Elin Jacob* da *Indiana University* em 2008 com o objetivo de contornar a potencial complexidade e redundância inerente aos vocabulários gerados pelos usuários de

sistemas folksonômicos, já que a UTO busca promover a integração entre dados de marcação social oriundos de fontes distintas (DING, *et al.* 2010).

Considerações finais

A utilização da linguagem contextualizada que emerge das folksonomias implica em desafios que envolvem a formalização e a interoperabilidade entre os diferentes sistemas de representação da informação e do conhecimento que constituem a *web* e a investigação de novas metodologias e modelos conceituais no âmbito da organização da informação.

Acredita-se que as perspectivas metodológicas da Análise de Redes Sociais e da Semiótica contribuam para estudos futuros nesse âmbito, visto que, enquanto a primeira permitiu a investigação de estruturas relacionais heterogêneas como redes sociais e redes semânticas, a segunda foca os fenômenos de linguagem e compartilhamento de significados.

A observação e a análise dessas dinâmicas de compartilhamento contribuem para a ampliação das metodologias de organização intelectual da informação visto que os ambientes em que a folksonomia ocorre geram extensões compartilháveis da memória e da cultura informacional de seus utilizadores e propiciam a elaboração de camadas semânticas pactuadas, contestadas e atualizadas através das práticas colaborativas.

A interoperabilidade semântica deve ser preconizada nas metodologias de construção das metalinguagens e não apenas na formalização/operacionalização das mesmas.

Referências

Alves, H. A. 2012. Ontologias Folksonomizadas: uma abordagem para fusão de ontologias e folksonomias. 123p.

Basso, Carlos A. M.; Silva, Sérgio R. P. da. 2008. Uma proposta para a evolução de ontologias a partir de folksonomias. In: Workshop de teses e dissertações (WTD). Anais do Simpósio Brasileiro de Sistemas Multimídia e Web (WebMedia), 14., 2008. p.197-200.

Borst, Willem Nico. 1997. *Construction of engineering ontologies for knowledge sharing and reuse*. Dutch Graduate School for Information and Knowledge Systems, Enschede.

Clarke, S.G.D.; Zeng, M.L. 2012. From ISO 2788 to ISO 25964: the evolution of thesaurus standards towards interoperability and data modeling. *Information Standards Quarterly*, v.24, n.1.

Ding, *et al.* 2010. Upper tag ontology for integrating social tagging data. *Journal of the American Society for Information Science and Technology*. v. 61, n. 3. p. 505-521.

Lima, João. 2010. Interoperabilidade semântica no governo eletrônico: informação legislativa e jurídica.

Mika, Peter. 2007. *Social Networks and the Semantic Web*. New York: Springer.

Moreira, W.; Lara, M. 2012. Ontologias, categorias e interoperabilidade semântica. *Data Gram Zero*, v. 13, n. 4.

Patel *et al.* 2005. Semantic Interoperability in Digital Library Systems, Report of DELOS2 Network of Excellence in Digital Libraries.

Qin, Jian. 2008. Folksonomies and taxonomies: where the two can meet. In: Nworkshop, Washington.

Reitz, Joan M. 2004. *Dictionary for Library and Information Science*. Westport: Libraries Unlimited.

Folksonomias na organização do conhecimento: Tags atribuídas a artigos sobre medicina

Luis Fernando Herbert Massoni
Universidade Federal do Rio Grande do Sul
fernandomassoni@hotmail.com

Regina Helena van der Laan
Universidade Federal do Rio Grande do Sul.
rhvdl@ufrgs.br

Rita do Carmo Ferreira Laipelt
Universidade Federal do Rio Grande do Sul
ritacarmo@yahoo.com.br

Sônia Elisa Caregnato
Universidade Federal do Rio Grande do Sul
sonia.caregnato@ufrgs.br

1 Introdução

Um dos principais problemas nas unidades de informação ocorre quando o usuário não consegue recuperar o conteúdo que deseja. Muitas vezes, essa falha de comunicação entre sistema e usuário se deve a divergência presente entre a combinação de termos utilizados como unidades de indexação e a verbalização empregada pelo usuário ao tentar recuperar determinado documento.

Nessa perspectiva, acredita-se que a folksonomia é uma possível fonte de identificação de termos mais próximos do acervo linguístico dos usuários. Segundo Brandt e Medeiros, “Folksonomia é o resultado do processo de etiquetagem, também chamado de classificação social, de recursos da *Web*. Isso significa dizer que as próprias pessoas, no caso, usuários da informação, classificam os documentos.” (2010: 112). Ou seja, é a prática da atribuição de palavras a documentos, chamadas de *tags* (etiquetas), pelos usuários de alguns ambientes *online* que, na sua concepção, são como palavras-chave para as informações contidas neles.

Tendo em vista as características da folksonomia, acredita-se em seu potencial para o desenvolvimento de estudos terminológicos, uma vez que a Terminologia é uma ciência que estuda a composição e o comportamento dos termos, tendo como objeto de estudo o termo técnico-científico. É ele que marca a identidade da área, embora a

fraseologia especializada e a definição terminológica também tenham passado a integrar seus horizontes de pesquisa. (Krieger; Finatto, 2004). Para Cabré (1999), a Teoria Comunicativa da Terminologia (TCT) busca explicar os termos como unidades singulares, admitindo variação de gênero conceitual e denominativo, considerando a dimensão textual e discursiva dos termos.

Para os estudos terminológicos, o número de ocorrências de cada termo no discurso do especialista não é parâmetro para a designação de termos. Embora essa pesquisa utilize os pressupostos teóricos da Terminologia como base, a frequência das *tags* foi considerada relevante, tendo em vista que a teoria de elaboração de tesouros considera esse aspecto importante para a determinação dos descritores constituintes de um tesouro.

2 O Estudo

A pesquisa parte dos pressupostos teóricos da Terminologia e da Organização do Conhecimento, tendo como objetivo geral analisar a pertinência da utilização da folksonomia como fonte de identificação de candidatos a termos, na elaboração de linguagens documentárias.

O objetivo geral da pesquisa desdobra-se em quatro objetivos específicos, que são: averiguar se as *tags* atribuídas a artigos científicos, por usuários do CiteUlike, podem ser considerados termos a partir do ponto de vista dos fundamentos teóricos da Terminologia; identificar, a partir de um *corpus* elaborado para a pesquisa, outras palavras-chave além daquelas atribuídas pelos autores dos artigos e pelos usuários do CiteUlike; avaliar o potencial da contribuição da folksonomia nos processos de indexação em bibliotecas; e propor alternativas que viabilizem a incorporação das folksonomias como ferramenta de apoio nos processos de indexação de assunto.

Dentre os *sites* que utilizam a folksonomia, foi escolhido o CiteUlike, ambiente *online* e gratuito que é utilizado para compartilhar publicações acadêmicas, sendo desenvolvido e mantido pela editora Springer. A escolha desse ambiente se deu por ser o único que visa especificamente o compartilhamento de conteúdo científico, como artigos de periódicos, constituindo-se, assim, no mais adequado para a proposta e metodologia da pesquisa.

O *corpus* textual é constituído por artigos de periódicos originais e de revisão, publicados em língua inglesa, na íntegra, da área de Medicina, os quais receberam *tags*

de, no mínimo, dois usuários diferentes do ambiente CiteULike. O termo utilizado para pesquisa no ambiente foi “*medicine*”. Recuperou-se 127 artigos com essas características, no entanto, obteve-se acesso a apenas 96 artigos na íntegra, pois os outros 31 não se encontravam disponíveis para *download* no CiteULike ou em outros *sites* da *Web*. Dos 96 textos, foram coletadas todas as *tags* atribuídas pelos usuários, sendo com elas construído um *corpus* de pesquisa em arquivo *Excel*.

A pesquisa possui caráter inovador, contribuindo para a formação do conhecimento e entendimento das formas de recuperação da informação utilizadas pelos usuários de sistemas de informação. Utiliza métodos qualitativos para a limpeza dos artigos e *tags*, e quantitativos para a validação das mesmas em situação discursiva e vocabulário controlado.

3 Validação Das *Tags*

Nessa seção são apresentadas as maneiras pelas quais realizaram-se as validações no discurso do especialista e em vocabulário controlado das *tags* atribuídas pelos usuários aos artigos na área de Medicina que compõe o *corpus*. Antes disso, cabe explicar o significado de cada uma dessas validações:

- a) validação no discurso do especialista: consiste em averiguar se as *tags* encontram-se em situação discursiva, ou seja, se os autores dos artigos utilizaram essas *tags* como termos ao redigirem seus textos. É, em linguagem estatística, a frequência com que essas *tags* ocorrem nos textos. Para a realização dessa validação, utilizou-se o e-Termos;
- b) validação em vocabulário controlado: consiste em analisar se essas *tags* figuram como descritores ou remissivas (variantes) apontados por um vocabulário controlado. Para a realização dessa validação, utilizou-se o DeCS – Descritores em Ciência em Saúde.

3.1 E-Termos

O e-Termos é um ambiente *online* gratuito resultante da tese de doutorado de Leandro Henrique Mendonça de Oliveira, destinado ao estudo terminológico. Para que ele pudesse ser utilizado na validação das *tags* em situação discursiva, realizou-se o *upload* do *corpus* textual em seu sistema. Foram considerados os 96 artigos na íntegra

(*corpus* textual) e suas respectivas *tags* (*corpus* de pesquisa). Aos artigos coletados, foram atribuídas pelos usuários um total de 1611 *tags*. Eliminadas as repetições, restaram 657 *tags* distintas, das quais 575 foram descartadas, por não possuírem nenhum sentido claro quando empregadas ao assunto Medicina, restando, portanto, 82 *tags* que constituem o *corpus* de pesquisa.

Foram consideradas todas as formas pelas quais se apresentam os termos, desde que possuam o mesmo sentido, por exemplo: o termo *medicine* ocorre 1225 vezes ao longo do *corpus*, o termo *MEDICINE* (em caixa alta) ocorre 20 vezes e o termo *Medicine* (com a primeira letra maiúscula) 328, porém todos foram contados como constituintes de apenas um termo, pois possuem o mesmo sentido, apenas foram escritos de formas diferentes pelos autores. As *tags* e termos no plural e no singular foram separados e se adicionaram às siglas os seus respectivos significados entre parênteses. As siglas possuem duas frequências de ocorrência: a primeira diz respeito à sigla e foi pesquisada com a ferramenta Contador de Frequência. Já a segunda diz respeito ao significado da sigla, tendo sido buscada através da ferramenta Consulta Termos no E-Termos.

As *tags* foram pesquisadas no vocabulário controlado DeCS – Descritores em Ciências da Saúde, que fornece uma linguagem de indexação que pode ser empregada em artigos de revistas científicas, livros, anais de congressos, relatórios técnicos, entre outros, da área da saúde (Biblioteca Virtual em Saúde, 2012).

4 Resultados

Das 82 *tags* atribuídas pelos usuários, 79 encontram-se em situação discursiva, ou seja, apenas 3 *tags* (aproximadamente 4%) foram apropriadas pelos usuários, mas não foram citadas pelos autores nos artigos que compõem o *corpus* textual. Esse dado indica que, dentre as *tags* validadas, cerca de 96% são compostas por termos presentes no discurso do especialista.

Os dados da pesquisa apontam que aproximadamente 79% das 82 *tags* referidas pelos usuários estão presentes de alguma forma no DeCS, sendo como descritores ou variantes. Apenas 17 (aproximadamente 21%) *tags* foram citadas pelos usuários, mas não se encontram dentre os descritores ou variantes apontados pelo DeCS.

Ao comparar o percentual total de *tags* validadas no discurso do especialista com o respectivo percentual no Descritores em Ciências da Saúde, é visível a discrepância

entre ambos, uma vez que 96% das *tags* ocorrem em situação discursiva, enquanto que 79% delas são validadas pelo DeCS. Tendo em vista que, para um termo passar a ser apontado por um vocabulário controlado, ele tem que antes ser utilizado no discurso dos especialistas da área, parece mais correto afirmar que o vocabulário controlado está desatualizado do que defender a ideia de que determinada palavra, por não constar em sua lista de descritores e variantes, não configura termo de uma área.

5 Considerações Finais

A pesquisa revela que quase todas as *tags* atribuídas pelos usuários estão presentes no discurso do especialista e/ou são utilizadas pelo DeCS como descritores ou variantes, o que indica que as *tags* apropriadas por usuários de sistemas de informação que utilizam a folksonomia são representativas das informações contidas nos respectivos documentos, podendo ser utilizadas como fontes de extração de prováveis termos, tornando possível sua utilização na elaboração de linguagens documentárias. Quanto à consulta às *tags* para a elaboração de tesouros, poderia se dar tanto na forma de descritores, como também na construção de um eficaz sistema de remissivas. Essa intervenção possui o benefício de aproximar a linguagem de indexação utilizada pelos bibliotecários da linguagem utilizada pelo usuário, facilitando, assim, a recuperação da informação por parte destes.

A utilização da folksonomia revela um horizonte promissor para a área de Organização do Conhecimento, pois possibilita o reconhecimento da linguagem dos usuários, de modo a enriquecer o processo de construção de tesouros. Conseqüentemente, é mais um recurso que pode auxiliar na solução de um problema que acompanha os estudos biblioteconômicos há muito tempo: a dificuldade de recuperação dos documentos por parte dos usuários.

Referências

Biblioteca Virtual em Saúde. DeCS: Descritores em Ciências da Saúde. Endereço: <<http://decs.bvs.br/P/decsweb2012.htm>>. Acesso em: 15 jul. 2012.

Brandt, Mariana; Medeiros, Marisa Brascher Basílio. maio/ago. 2010. Folksonomia: esquema de representação do conhecimento?, *TransInformação*, Campinas, v. 22, n. 2, p. 111-121. Disponível em: <<http://revistas.puc-campinas.edu.br/transinfo/search.php?op=search&query=Folksonomia%3A+esquema+de+representa%E7%E3o+do+conhecimento%3F&limit=title>>. Acesso em: 16 dez. 2011.

Cabré, Maria Teresa. 1999. La Terminología: representación y comunicación. Elementos para una teoría de base comunicativa y otros artículos. Barcelona: IULA; Universitat Pompeu Fabra.

Empresa Brasileira de Pesquisa Agropecuária. 2009. E-Termos. Disponível em: <<http://www.etermos.cnptia.embrapa.br/>>.

Krieger, Maria da Graça; Finatto, Maria José Bocorny. 2004. Introdução à Terminologia: teoria & prática. São Paulo: Contexto, 223 p.

Springer. CiteULike. Disponível em: <<http://www.citeulike.org/>>.

Ontologias a partir de parâmetros da norma ANSI/NISO Z39.19

Ana Cristina de Albuquerque
Universidade Estadual de Londrina
albuanati@yahoo.com.br

Brígida Maria Nogueira Cervantes
Universidade Estadual de Londrina
brigidacervantes@gmail.com

Acerca das ontologias

Como conceito, ontologia é um termo que se mostra com uma aplicação variada dependendo de seu uso nas diferentes áreas do conhecimento. Desde as formas de representação de informações, passando por seu sentido filosófico e sendo também utilizada na recuperação de informações, este termo está na Filosofia, Ciência da Computação, Inteligência Artificial, Web Semântica e Ciência da Informação. Este caminho foi trilhado graças ao avanço rápido das tecnologias da informação na sociedade pós-industrial, que permitiu sua inserção no contexto tecnológico e a fez presente em inúmeras áreas do conhecimento, como por exemplo, na Ciência da Informação. De acordo com Cunha e Cavalcanti (2008, p. 268), a ontologia apresenta a possibilidade de visualizar o “domínio da hierarquia” e conseqüentemente os relacionamentos que se formam entre os conceitos. Os autores acrescentam ainda que a ontologia específica “[...] uma conceitualização, podendo incluir a descrição de objetos, conceitos e outras entidades de um conceito ou parte dele, bem como as relações entre eles. Fornece significado para descrever explicitamente uma conceitualização atrás de um conhecimento representado em uma base de conhecimento.” (Cunha; Cavalcanti, 2008, p. 268). Neste âmbito, a ontologia, na Ciência da Informação, assume características distintas da que assumia no campo filosófico, afinal o cerne da questão deixa de ser o universo e o ser, e passa a ser a informação, em como representá-la e organizá-la no ambiente tecnológico.

De acordo com Currás (2010), os estudos sobre ontologias começaram a ser citados em 1954, mas sua frequência em eventos científicos é maior a partir da segunda metade da década de 1990, quando definições e discussões já poderiam ser encontradas. A autora oferece um panorama destas definições que vão de 1984, apresentada por Sava, que diz ser ontologia “um catálogo de qualquer coisa que serve para construir um

mundo possível”, até a definição de García Marco (2003 apud Currás, 2010, p. 41), que, na visão da autora, oferece uma definição mais ampla e de entendimento adequado, pois a caracteriza como uma “[...] descrição sistemática das entidades e modalidades e das regras que permitem descrever um domínio específico de acordo com as entidades e processos [...]” (Currás, 2010, p. 40), desta forma, as ontologias possibilitaram “descrever ‘todas’ as coisas e processos.” (Currás, 2010, p. 40).

Uma ontologia pode ser entendida como um artefato computacional para representação do conhecimento de domínios específicos ou genéricos (Mendonça et al, 2010) e suas contribuições podem ser no sentido de colaboração, compartilhamento, padronização e confiabilidade para representar o conhecimento de um domínio (Mendonça et al, 2010). É neste contexto tecnológico, que a ontologia é explicada como “[...] uma especificação explícita dos objetos, conceitos e outras entidades que assumimos existirem em uma área de interesse [...]” (Oliveira et al, 2004, p. 2), ou seja, uma forma de representação e descrição do objeto, baseada em um vocabulário controlado que será utilizado por um sistema lógico, apresentada hierarquicamente e com termos de busca definidos, permitindo assim a redução de ambiguidades.

Em artigo sobre o assunto, Almeida e Bax (2003) explicam que há características e componentes básicos na maioria das ontologias, o que não significa que estas sejam sempre formadas por uma mesma estrutura. Os autores destacam ainda que as ontologias são compostas basicamente por: *classes* – apresentação dos conceitos em hierarquia, *relações* – apresenta a relação entre esses conceitos, *axiomas* – utilizados para apresentar regras constantemente verdadeiras e *instâncias* – representam os elementos específicos da área trabalhada.

A diferença entre ontologia e tesauro, de acordo com Currás (2010, p. 42), se dá na estrutura. Os tesouros, com seu caráter de vocabulário controlado, se apresentam através de uma estrutura que inclui termos que se relacionam de forma semântica. Desta forma, são capazes de proporcionar o controle terminológico destes termos, assim como sua estruturação é manifestada como hierárquica ou associativa. Como ferramenta de organização e recuperação de informações, o tesauro proporciona a normalização da linguagem e o relacionamento entre os termos, fazendo com que haja um maior aprimoramento em relação às buscas. Seguindo Currás (2010, p. 42): “Nos tesouros se parte de uma ordenação dos termos que o compõe, em hierarquias e relações semânticas e sintáticas. Nas ontologias, concebe-se a ordenação de uma maneira diferente, levando em conta certas peculiaridades e propriedades dos termos.”

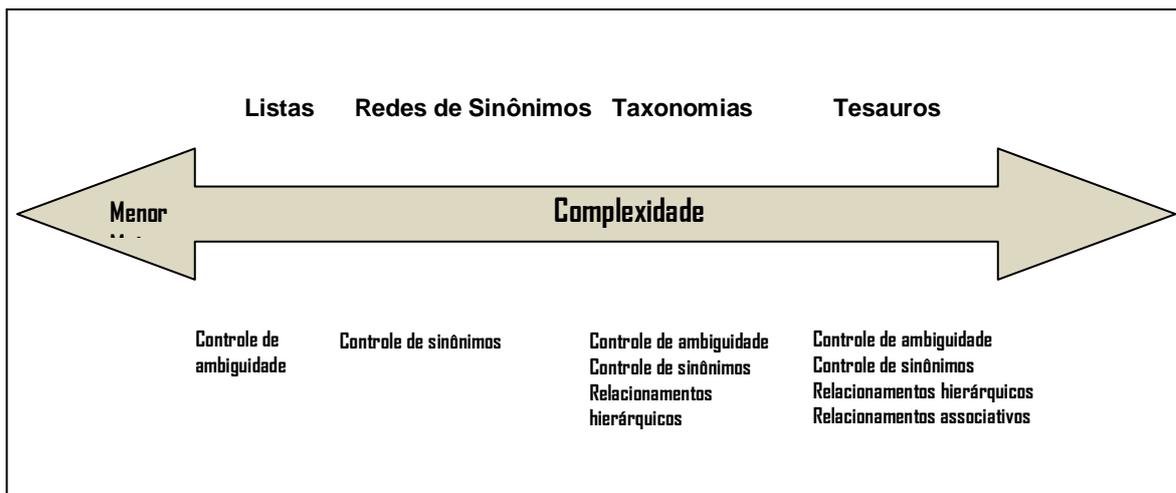
É possível construir uma base de conhecimentos sobre a estrutura básica de uma ontologia, diz-se então que a ontologia desempenha neste caso o papel de esquema. Ou seja, a base de conhecimentos utiliza os termos fornecidos pela ontologia para a descrição da realidade. A diferença entre base de conhecimento e ontologia se dá pelo fato da realidade descrita em uma base de conhecimento poder ser alterada sem que a ontologia mude, ou seja, se o domínio da ontologia não sofrer alteração, esta será a mesma ontologia. Contudo, da mesma forma que uma ontologia desempenha a um dado momento o papel de esquema, também é possível encontrarmos esquemas que possam ser ontologias.

A norma ANSI/NISO Z39.19 como um parâmetro para construção de ontologias

Percebe-se que muitas ontologias têm seu desenvolvimento sem o aporte de um método sistematizado, que contemple suas especificidades. Se comparadas com os tesouros, para estes tem-se a ANSI/NISO Z39.19, como um exemplo que, em suas proposições indica metodologias de construção de vocabulários controlados, abrangendo “[...] estruturas de complexidade crescente que vão das menos complexas como listas e redes de sinônimos, até estruturas de maior complexidade como taxonomias e tesouros. A diferença entre essas estruturas é a complexidade envolvida no processo de construção [...]” (Cervantes; Fujita, 2012, p. 131).

A ANSI/NISO Z39.19, em versão preliminar de 2005, nos remete a uma estrutura que apresenta de forma crescente, parâmetros para a construção de vocabulários controlados evidenciando características para a sua construção desde os menores graus de complexidade para Listas, Redes de sinônimos até os maiores graus de complexidade para a construção de Taxonomias e Tesouros (Cervantes; Fujita, 2012). As autoras, em tradução própria, apresentam uma figura com os graus de complexidade e com a estrutura que se mostra crescente entre os vocabulários controlados:

Figura 1 – Relação das estruturas mais simples às estruturas mais complexas



Fonte: ANSI/NISO Z39.19 (2005, p. 17, tradução das autoras apud Cervantes; Fujita, 2012, p. 132)

Os graus de complexidade gradativamente aumentam conforme a direção da seta, onde pode-se visualizar claramente que o distanciamento que vai sendo formado a partir dos primeiros elementos propostos, a listas e redes de sinônimos até as taxonomias e tesouros que se apresentam com tipos de relacionamentos mais complexos. Cervantes e Fujita (2012, p. 132) analisam a figura da seguinte maneira:

180

[...] pelas evidências de uma estrutura crescentemente complexa de vocabulários controlados, ditada pelas exigências dos tipos de relações que cada um tem que acomodar. Também mostra que os vocabulários mais complexos (taxonomias, *thesauri*), incluem as estruturas mais simples (listas, redes de sinônimos). Por exemplo, um tesouro inclui dispositivos explícitos por controlar sinônimos, organizar hierarquias, e criar relações associativas, enquanto uma lista é um conjunto de termos simples que não contém relação de qualquer tipo, não distingue “**termo** de **palavra** [...] também não apresenta relações estruturais entre os elementos”. (Gomes, 1990, p. 15 grifo da autora).

Os graus de complexidade, menor ou maior, demonstrados de forma elucidativa pela ANSI/NISO Z39.19, se apresentam como controle de ambiguidades, controle de sinônimos, relacionamentos hierárquicos e relacionamentos associativos (Cervantes; Fujita, 2012, p. 132). Para Shiessl e Shintaku (2012), o fator determinante para que a complexidade aconteça é marcado justamente pelo relacionamento que existe nos diferentes tipos de vocabulários controlados. Isto posto, indaga-se, quais seriam os graus de complexidade para elaboração de ontologias, uma vez que estas, tanto quanto os tesouros possibilitam a definição de uma rede de relações, os termos e as combinações destes termos com o fim de atuar em uma determinada área do

conhecimento. Mercè Lorente Casafont (2005) explica que as ontologias para domínios específicos oferecem uma diversidade de classes de conceitos, relações conceituais de diversas origens e principalmente a categorização, relativa ao campo específico “En esto superan sin lugar a duda a los tesauros documentales, en los que generalmente predominan las relaciones jerárquicas y asociativas, y en los que la función primordial es relacionar conceptos con categorías determinadas previamente.” (Lorente Casafont, 2005, p. 03). A ontologia especifica uma conceitualização, ou seja, possui um padrão comum de estruturação para que as informações possam ser representadas e acessadas. No mesmo sentido Almeida e Bax (2003), argumentam que, com as ontologias, é possível definir formas de representação fundamentadas na lógica, o que indica que novos conhecimentos podem ser criados a partir dos descritos e já existentes, fato que representaria uma “evolução em relação a técnicas tradicionais” (Almeida; Bax, 2003, p. 17). As relações conceituais e as aproximações com o domínio do conhecimento são possíveis por meio das ontologias, que têm um caráter de extrema relevância na construção do conhecimento.

Na visão de Shinessl e Shintaku (2012, p. 57) “Talvez a característica que mais se destaca e diferencia as ontologias dos demais tipos de vocabulário controlado seja a automação de seus relacionamentos, axiomas e regras de inferência”.

Diante do exposto, considera-se que o aporte teórico-metodológico da norma ANSI/NISO Z39.19 (2005) para a construção de vocabulários controlados poderia contribuir para a fundamentação da construção de ontologias de um domínio, uma vez que os elementos necessários para compor e promover a estruturação de ontologias de um domínio, são “os termos que denotam conceitos e que estão inter-relacionados em dado domínio.” (Shinessl e Shintaku, 2012, p. 57). A representação do conhecimento, através das ontologias, é passível de entendimento que estas se dispõem como vocabulários controlados em ambientes tecnológicos.

Considerações

As ontologias se estruturam e podem ser compreendidas como um domínio que têm em seu cerne a comunicação entre a pessoa que busca a informação e os sistemas que oferecem estas informações sendo que podem ser utilizadas e compartilhadas em diferentes situações e aplicações dentro da organização do conhecimento. É necessário definir suas metodologias e proposições e conceber seu entendimento como parte de

uma estrutura tão complexa quanto a dos tesouros, com elementos relevantes que, se bem formulados, se aplicam e se relacionam entre si e com um campo determinado. As ontologias determinam uma combinação em relação ao vocabulário que é compartilhado, que desta forma se torna mais consistente na medida em que se define coerente com o compartilhamento sugerido. Assim, para futuros estudos, propõem-se o aprofundamento de pesquisa quanto aos graus de complexidade referentes aos tesouros e congregar as ontologias como um novo escopo e ferramenta, analogamente, complexa para a organização do conhecimento no âmbito das tecnologias de informação.

Referências

Almeida, Mauricio de.; Bax, Marcello P. Uma visão geral sobre ontologias: pesquisa sobre definições, tipos, aplicações, métodos de avaliação e de construção. **Ciência da Informação**, Brasília, v.32, n.3, p.7-20, set./dez. 2003.

American National Standards Institute; National Information Standards Organization. Z39.19-2005: guidelines for the construction, format, and management of monolingual controlled vocabularies. Bethesda, Ma:NISO Press, 2005. 184p. Disponível em: <<http://www.niso.org/standards/resources/Z39-19-2005.pdf>>. Acesso em 10 de jan. 2013.

Cervantes, Brígida Maria Nogueira; Fujita, Mariângela Spotti Lopes. Uma perspectiva diacrônica de parâmetros e modelos de construção de tesouros. In: Cervantes, Brígida Maria Nogueira (Org.). **Horizontes da organização da informação e do conhecimento**. Londrina: Eduel, 2012, p. 119-138.

Cunha, Murilo Bastos da; Cavalcanti, Cordélia Robalinho de Oliveira. **Dicionário de biblioteconomia e arquivologia**. Brasília, DF: Briquet de Lemos, 2008.

Currrás, Emilia. **Ontologias, taxonomia e thesaurus em teoria de sistemas e sistemática**. Brasília: Thesaurus, 2010.

Lorente Casafonte, Mercè. Ontologia sobre economia y recuperación de información. **Hipertext.Net**, Barcelona, n.3, 2003. <Disponível em http://www.upf.edu/hipertextnet/numero-3/ontologia_ri.html>. Acesso em: 10 de maio de 2013.

Mendonça, Fabricio Martins; Cardoso, Ana Maria Pereira; Drumond, Eliane. Ontologia de aplicação no domínio de mortalidade: uma ferramenta de apoio para preenchimento da declaração de óbitos. **Ciência da Informação**, Brasília, v.39, n.3, 2010. Disponível em <<http://revista.ibict.br/ciinf/index.php/ciinf/article/view/1870>>. Acesso em 22 de out. de 2012.

Schiessl, Marcelo; Shintaku, Milton. Sistemas de organização do conhecimento. In: Alvares, Lilian (Org.). **Organização da informação e do conhecimento: conceitos, subsídios interdisciplinares e aplicações**. São Paulo: B4 Editora, 2012.

Oliveira, Ismênia Ribeiro de; Sousa Neto, Rafael; Girardi, Rosario. Uma ontologia para a especificação de sistemas de padrões. **Quarta conferência Latino-Americana em linguagens de Padrões para programação**, 2004. Disponível em <http://sugarloafplop2004.ufc.br/index_pt.html>. Acesso em: 22 de out. de 2012.

Reflexões sobre processamento e representação automática de conhecimento.

Viviane Neves dos Santos
Universidade de São Paulo
vivianeneves81@gmail.com

Nair Yumiko Kobashi
Universidade de São Paulo
nairkobashi@gmail.com

Introdução

A busca de procedimentos automáticos e semiautomáticos ou algorítmicos de indexação (Hjørland, 2008) são motivados pela quantidade e velocidade de produção de informação no mundo contemporâneo. Apresentamos, neste texto, uma breve exposição de modelos de processamento automático e representação de textos (indexação automática) propostos por diferentes grupos de pesquisa, que merecem olhar atento da área de Organização e Representação do Conhecimento (ORC).

As operações de ORC são, por definição, processos forjadores de sentido. O ato receptivo para produzir representações requer compreensão do texto de partida. Nesse processo, se estabelecem filtros, cuja legitimidade se relaciona ao contexto situacional. Só assim, se fixa a relação pragmática frente ao texto e ao contexto. Dito de outro modo, cada ato de recepção é orientado por acordos pragmáticos construídos segundo os horizontes de expectativas do produtor e do receptor. É possível objetivar e automatizar esse processo? Em caso positivo, como objetivá-lo? A ORC tem procurado respostas a esses problemas desde meados do século passado.

183

Modelos De Indexação Automática E Grupos De Pesquisa

Os atuais modelos de indexação automática recebem variados nomes: Classificação automática, Categorização automática e Clustering, entre outros (Gil Leiva, 1999). Todavia, as denominações se referem ao ato de atribuir automaticamente, a um documento, um ou mais assuntos, com o uso ou não de Linguagens Documentárias. Lamarca Lapuente (2007) afirma que os modelos propostos se fundamentam na equação “Linguística + Estatística+Informática+Base de

conhecimento”. Integram essa equação a Terminologia, a Linguística Computacional, a Inteligência Artificial e a Mineração de Textos.

Méndez Rodríguez e Moreiro González (1999) afirmam que os sistemas de indexação automática adotam os pressupostos do Processamento da Linguagem Natural (PLN) e dos Sistemas Inteligentes. Estes últimos têm três componentes fundamentais: uma base de conhecimento, uma ferramenta de inferência e a interface com o usuário (Rodríguez Pedrojo; Ronda León, 2006). A evolução desses sistemas é nítida: são construídos, atualmente, algoritmos complexos que combinam referências de base matemática e modelos linguísticos, longe, portanto, dos modelos ancestrais que se baseavam em estatísticas de reconhecimento de formas e sua frequência.

No Brasil, há vários grupos que estudam o processamento automático de textos e a indexação automática. São constituídos por pesquisadores da Ciência da Computação, da Ciência da Informação e da Linguística, em equipes multidisciplinares. A Fig. 1 mostra os grupos cadastrados no Diretório de Grupos Pesquisa do CNPq (2013), que se dedicam ao tema.

Figura 1: Grupos de Pesquisa (Brasil) e Abordagens de Indexação Automática

GRUPO	INSTITUIÇÃO	ÁREA	ABORDAGEM
Classificação de Informação e Processamento de Linguagem (CIPL)	Instituto Federal de Alagoas - (IFAL)	Ciência da Computação	Sistemas Inteligentes
Estudos Linguístico-Computacionais da Linguagem (CELIC)	UNESP	Linguística	PLN
Ferramentas em Linguística Computacional	PUC-Rio	Linguística	PLN
Gerência e Análise Inteligente de Dados	UFOP	Ciência da Computação	Sistemas Inteligentes
Grupo de Banco de Dados e Recuperação de Informação	UFAM	Ciência da Computação	PLN+Sistemas Inteligentes
Inteligência Computacional	USP / São Carlos	Ciência da Computação	PLN+Sistemas Inteligentes
Laboratório de Engenharia de Software (LES)	PUC-Rio	Ciência da Computação	PLN
Laboratório de Linguística Computacional (LaLiC)	UFSCar	Ciência da Computação	PLN+Sistemas Inteligentes
Modelagem Conceitual Para Organização Hipertextual De Documentos (MHTX)	UFMG	Ciência da Informação	PLN
Núcleo Interinstitucional de Linguística Computacional (NILC)	USP/São Carlos	Ciência da Computação	PLN+ Sistemas Inteligentes

Representação do Conhecimento, Ontologias e Linguagem	UFMG	Ciência da Informação	PLN
Representação do Conhecimento, Ontologias e Linguagem	UFMG	Ciência da Informação	PLN

Fonte: CNPq, 2013.

Os sistemas não linguísticos, aqueles que adotam procedimentos de base quantitativa (estatísticos, probabilísticos e bibliométricos) não têm presença isolada nas pesquisas dos grupos acima mencionados. No entanto, pode-se afirmar que os procedimentos quantitativos estão presentes nos sistemas linguísticos, principalmente na etapa de mapeamento de ocorrência e frequência de termos.

O Processamento de Linguagem Natural (PLN) é um sistema linguístico, adotado por 6 Grupos de pesquisa (UNESP, PUC-Rio e UFMG). Recorre ao processamento computacional de textos, nos níveis morfo-sintático e semântico. Para a desambiguação são utilizados, de um lado, vocabulários controlados (de termos da língua especializada) e, de outro, a identificação de sintagmas nominais, que contextualizam as unidades do texto.

Os chamados Sistemas inteligentes, utilizados pelo Grupo da Universidade Federal de Ouro Preto (UFOP) têm por base o Aprendizado de Máquina, que supõe a construção de bases de conhecimentos. Os algoritmos do sistema realizam inferências automáticas, sendo eles construídos com base em classificação prévia de documentos por humanos.

No levantamento apresentado acima observa-se que 4 grupos de pesquisa (Universidade Federal do Amazonas-UFAM; Universidade de São Paulo-USP-São Carlos e Universidade Federal de São Carlos-UFSCar) adotam a combinação PLN+Sistemas inteligentes. Estes vêm sendo utilizados em experimentos de indexação e sumarização automáticas. Distinguem-se dos chamados Sistemas inteligentes pelo fato de as bases de conhecimentos serem geradas com base em processamento automático (morfo-sintático, semântico e pragmático) de textos.

Outro aspecto que chama a atenção é a filiação institucional dos grupos de pesquisa. Dos 12 Grupos identificados no Diretório de Grupos de pesquisa (CNPq, 2013), 7 pertencem a instituições de Ciência da Computação, 3 à Ciência da Informação e 2 a Departamentos de Linguística. Observa-se, portanto, que as pesquisas sobre sumarização e indexação automática adquirem contornos interdisciplinares. Uma limitação deste estudo refere-se à identificação da formação dos membros desses

grupos. Esse aspecto será estudado em outra etapa da pesquisa, para verificar a formação dos especialistas que integram esses grupos.

Considerações Finais

O processamento automático de textos é realizado, atualmente, com base em diferentes modelos que combinam técnicas de Processamento de Linguagem Natural (PLN) e Inteligência Artificial. São realizados experimentos de sumarização automática, representação do conhecimento por meio de ontologias, modelagem e organização conceitual de textos, análise e recuperação automática de informação. Não foram identificados grupos que se baseiam apenas em modelos de base estatística (ocorrência e frequência de palavras), como ocorria até os anos 1980. Outro aspecto importante refere-se à composição dos grupos de pesquisa: são formados por pesquisadores de diferentes áreas do conhecimento, com maior presença da Ciência da Computação, da Linguística Computacional e da Ciência da Informação.

A área da ORC acumula conhecimentos teóricos e práticos expressivos sobre indexação e representação automática de documentos. A principal preocupação do campo se relaciona aos complexos processos inferenciais, de natureza semântica e lógica, necessários ao processamento de textos e sua representação. A área reconhece, porém, que os modelos e experimentos atuais carecem da compreensão do que Wittgenstein (1975) chama de diversidade dos jogos linguísticos cotidianos. Assim, não há, ainda, sistemas que respondam satisfatoriamente à questão: como dotar os sistemas de “inteligência” suficiente para interpretar, condensar e representar textos para os fins da recuperação?

Outro problema que merece atenção: as experiências dos grupos são feitas majoritariamente em textos em inglês, pela facilidade de constituição de *corpora* e de acesso a softwares de processamento de textos para esse idioma. Justificam-se, no quadro atual de demanda de informações, as pesquisas sobre *corpora* em língua portuguesa, por várias razões: o português é falado em 8 países, sendo ela a quinta língua mais falada no mundo (Nóbrega, 2013). Adensar os estudos em *corpora* de língua portuguesa, no contexto da ORC é, certamente, uma das formas de internacionalizar a pesquisa, se compreendemos a internacionalização como processo de mão-dupla, que implica a ida de pesquisadores brasileiros ao exterior, mas também a vinda de estudantes do exterior, para inserção em programas de pós-graduação, e realização de pós-doutoramento de pesquisadores do exterior junto aos grupos brasileiros. Hoje, o

Brasil, pela presença e influência econômica, cultural e política, no cenário internacional, vem despertando o interesse gradativo de estudantes e pesquisadores de diferentes países da Europa. Esta não parece, ser, portanto, uma proposta utópica.

REFERÊNCIAS

Borges, Graciane Silva Bruzanga; Maculan, Benildes Coura Moreira dos Santos; Lima, Gercina Angela Borem de Oliveira. 2008. Indexação automática e semântica: estudo da análise do conteúdo de teses e dissertações. *Informação & Sociedade*, 18:181-193. Disponível em: < <http://www.ies.ufpb.br/ojs2/index.php/ies/article/download/1759/2129> >. Acesso em: 15 fev. 2012.

Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq). 2013. *Directório dos Grupos de Pesquisa no Brasil* [base de dados]. [Brasília]. Disponível em: < <http://dgp.cnpq.br/buscaoperacional/> >. Acesso em: 10 fev. 2013.

Gil Leiva, Isidoro. 2007. A indexação na Internet. *Brazilian Journal of Information Science* 2:47-68. Disponível em: < <http://www.bjis.unesp.br/pt/include/getdoc.php?id=72&article=21&mode=pdf> >. Acesso em: 20 out. 2012.

Gil Leiva, Isidoro. 1999. *La automatización de la indización de documentos*. Gijón: Trea.

Hjorland, Birger. 2008. Automatic Indexing. In: _____. *Lifeboat for Knowledge Organization*. [s.l.]:[s.n.]. Disponível em: < http://www.db.dk/bh/lifeboat_ko/CONCEPTS/automatic_indexing.htm >. Acesso em: 5 dez. 2008.

Lamarca Lapuente, María Jesús. 2007. *Hipertexto: El nuevo concepto de documento en la cultura de la imagen*. Tesis doctoral. Un. Complutense de Madrid. Disponível em: < http://www.hipertexto.info/documentos/indiz_automat.htm >. Acesso em: 27 out. 2007.

Méndez Rodríguez, Eva M.; Moreiro González, José A. 1999. Lenguaje natural e indización automatizada. *Ciencias de la Información*, 30:1-23. Disponível em: < <http://www.bib.uc3m.es/~mendez/publicaciones/articulos/indizacion99.pdf> >. Acesso em: 15 maio 2008.

Moreiro González, José Antonio. 2002 Aplicaciones al análisis automático del contenido provenientes de la teoría matemática de la información. *Anales de Documentación*, 5:273-286. Disponível em: < <http://revistas.um.es/analesdoc/article/viewFile/2101/2091> >. Acesso em: 15 maio 2009.

Rodríguez Pedrojo, K.; Ronda Leon, R. 2006. Organización y recuperación de la información: un enfoque desde la perspectiva de la automatización. *ACIMED*: 14, n. 1. Disponível em: < http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S1024-94352006000100004&lng=es&nrm=iso&tlng=es >. Acesso em: 15 maio 2008.

SANTOS, Viviane Neves dos. 2009. *Indexação automática documentos textuais: iniciativas dos grupos de pesquisa de universidades públicas brasileiras, 2009*. 72 p. (Trabalho de Conclusão de Curso - Escola de Comunicações e Artes - Universidade de São Paulo)

Sebastiani, F. Machine Learning in Automated Text Categorization (2002). *ACM Computing Surveys*, 34, n. 1:1-47. Disponível em: <
<http://citeseer.ist.psu.edu/article/sebastiani99machine.html> >. Acesso em: 15 maio 2008.

Souza, Renato Rocha. 2006. Uma proposta de metodologia para indexação automática utilizando sintagmas nominais. *Encontros Bibli*, n. esp: 42-59. Disponível em: <
<http://www.periodicos.ufsc.br/index.php/eb/article/view/329/385> >. Acesso em: 15 abr. 2009.

Wittgenstein, Ludwig. 1975. *Investigações Filosóficas*. São Paulo: Abril Cultural (Coleção "Os Pensadores").

Nóbrega, M. H. 2013. Internacionalização da língua portuguesa. *Jornal da USP*, 998: 2.

Análisis de Dominio Como Abordaje Teórico en la Construcción de Ontologías. Estudio de Caso, Desarrollo de Ontologías Para la IDE-Uruguay

Diana Comesaña
Universidad de la República
diana.comesana@prodic.edu.uy

Alicia Díaz
Universidad de la República
alicia.diaz@prodic.edu.uy

Introducción

El altísimo costo de la Información Geográfica (IG), su dispersión y su valor estratégico nos lleva a crear las mejores condiciones de recuperación de la información.

80% de la información generada por las instituciones es IG, en su mayoría digital, diseminada en la web. Necesitamos contar con herramientas que permitan una eficiente recuperación.

189

Recuperación de la información, el leit motiv de la Organización y Representación del Conocimiento.

¿Por qué detenernos en analizar el concepto de recuperación de la información en un trabajo sobre ontología?

Si entendemos la ORC como el “*área de Tratamiento temático de la información que persigue la identificación, el procesamiento y la disponibilidad del contenido informacional de los documentos*” (Guimaraes (1998) en Barité, 2001: 38), queda implícita que la función última de la ORC es permitir recuperar la información en sus dos dimensiones: el acceso y la apropiación del contenido.

Esto nos permite afirmar que al recuperar información estamos hablando de una información que se materializa de alguna forma a través de documentos y que son esos documentos de los que nos apropiamos, es esa realidad objetiva de los fondos documentales de la que nos habla Barité (2001: 51) con la que lidiamos en la ORC.

La relevancia de la información que se recupera está vinculada a las necesidades específicas del que busca inserto éste en un “habitus” entendido como “*un ámbito en el que cabría suponer que los agentes actúan de acuerdo con unas intenciones conscientes y calculadas, de acuerdo con unos métodos y unos programas conscientemente elaborados*” (Bourdieu, 2003: 72).

La información que se recupera es subjetiva, depende de la interpretación del usuario, visto éste no como sujeto aislado sino como perteneciente a una comunidad.

De ahí nuestro papel como mediador entre información-conocimiento-documento y usuarios-comunidad, entendiendo que el conocimiento es social, se materializa y responde a procesos sociales y culturales y el usuario es un sujeto con necesidades sociales de información. En ese contexto es imposible obviar el lenguaje como vehículo del conocimiento, sin él no es posible la comunicación de ideas.

Ontologías. Definición, características.

Para optimizar la recuperación de la información deben desarrollarse modelos conceptuales de organización y representación del conocimiento.

La construcción del modelo conceptual, se basa en que éstos deben ser capaces de representar un contexto y se construyen a través de relaciones y atributos, consistentes con la realidad.

Con el desarrollo de ontologías, obtenemos una representación de los conceptos de un área del conocimiento, basándonos en categorías ontológicas que proporcionan una estructura axiomática para su desarrollo según el modelado conceptual.

Para Gruber de 1993 (SANZ & JIMENEZ – RUIZ, 2007):

“Una ontología es una especificación explícita de una conceptualización”.

Según Borst (1997):

- Es conocimiento compartido, fruto del consenso dentro de un grupo.
- Debe ser especificada usando un lenguaje formal, procesable por ordenadores.

Es un modelo estructurado por nociones de dominio, clases, subclases, propiedades, axiomas e individuos que poseen relaciones entre sí.

Es convertir la información en conocimiento mediante estructuras que referencian los datos, (metadatos) según un esquema común normalizado respecto de algún dominio del conocimiento.

Los esquemas consensuados simplifican el intercambio de datos digitales, para que puedan ser reutilizados, mediante estructuras de conocimiento formalizadas (ontologías).

Desarrollar una Ontología es desarrollar un lenguaje formal para representar cierto conocimiento, relaciones entre sus conceptos y significado de sus términos, de forma que éste lenguaje pueda ser “leído por máquinas”.

María Jesús Lamarca (2011) señala cuatro características:

Multiplicidad: pueden combinarse dos o más ontologías

Distintos niveles de abstracción estableciendo una topología de ontologías.

Multiplicidad de la representación: coexisten múltiples representaciones del mismo concepto.

Mapeo de ontologías que permite establecer relaciones entre ontologías para generalizaciones y especializaciones.

El Análisis de Dominio como herramienta teórica para la construcción de ontologías

El análisis de dominio como paradigma de la CI parte de un enfoque sociocognitivo. Se centra en conceptos tales como comunidad, discursos, contextos. Trabaja desde una visión centrada en lo social que privilegia el contexto como modo de limitar y definir necesidades, búsquedas y recuperación de la información por parte de la comunidad discursiva del dominio.

Hjørland y Albrechten la conceptualizan como *“el estudio del dominio de conocimiento como pensamiento o comunidad discursiva que forma parte de una división del trabajo en la sociedad”* (Hjørland, B & Albrechtsen, H, 1995: 400)

Siguiendo a los autores se pueden marcar los siguientes rasgos distintivos : la prioridad está dada por la comprensión de las necesidades del usuario desde una perspectiva social centrándolo en el contexto de su disciplina, la visión está influenciada por la sociología del conocimiento siendo conceptos claves la comunidad científica,

tipos de documentos, disciplinas, estructuras de información, paradigmas, organización del conocimiento.

Importa identificar cuáles son los productores de información de ese dominio, qué canales de comunicación usan, qué tipo de documentos producen, cuál es el lenguaje utilizado, si existe un “etos” común a la disciplina, desde qué visión epistemológica trabajan, si hay transdisciplinaridad, qué instituciones están implicadas, qué conflictos teóricos soportan. Desde ese lugar el análisis de dominio nos acerca una mirada crítica que nos obliga a no ser ingenuos con respecto al proceso de producción del conocimiento y a cómo se construyen los contenidos.

Una mirada al enfoque sobre Estudios Terminológicos, lenguaje para un propósito especial, base de datos semántica y estudios de discursos

Es necesario para la estructuración de la ontología, conocer los términos del lenguaje que caracteriza el área específica de datos espaciales. Creemos que la herramienta de Estudios Terminológicos nos permitirá un primer acercamiento al abordaje teórico y también práctico en la elaboración de la ontología.

Hjorland (2002: 443) citando a Hirschman & Sager (1982) nos introduce en el concepto de sublenguajes “*un lenguaje particular usado en un cuerpo de textos que trata sobre un área temática concreta (a menudo reportes o artículos de una especialidad técnica o subcampos de la ciencia), en cual los autores de los documentos comparten un vocabulario común y un uso común de las palabras*”.

La existencia de un sublenguaje compartido por la comunidad discursiva, acorta la brecha a la hora de definir y agrupar qué términos elegir.

Poder identificar el sublenguaje que caracteriza a la comunidad de usuarios que trabajan con datos espaciales es el primer paso para la construcción de ontologías.

Hjorland (2002: 445) entiende el lenguaje con un propósito específico conectado a 4 tópicos:

- Comunidades discursivas entendidas como el conjunto de signos.
- Las diferentes comunidades desarrollan tipos de documentos específicos cuyos elementos pueden ser potencialmente puntos de acceso en recuperación electrónica
- Las comunidades discursivas están influenciadas por aspectos epistemológicos que a su vez influyen en la construcción de símbolos, significados y conocimiento

➤ Generación de bases de datos semánticas que minimicen la pérdida de información. Para completar el análisis terminológico es oportuno combinarlo con estudios bibliométricos y estudios históricos.

Estudio de caso: una ontología para la IDE-Uruguay

El manejo de la IG se centra en la capacidad de almacenamiento y recuperación de información del computador. El costo de su producción, su variedad y dispersión, han dado origen a las Infraestructuras de Datos Espaciales (IDE).

Una IDE se compone de tecnologías, políticas y acuerdos institucionales destinados a facilitar el acceso a la IG, traducidos en servicios web y donde se destaca el Catálogo por Metadatos.

Guarino (1988) define ontología como *“un producto de ingeniería consistente en un vocabulario específico usado describir una realidad más un conjunto de asunciones relacionadas con el significado del vocabulario”*

El desarrollo de ontologías para la IDE-Uruguay, transformará la información en información semántica que permitirá alcanzar nuevo conocimiento.

María Jesús Lamarca (2011) indica distintos tipos de ontologías:

Según el ámbito del conocimiento al que se apliquen, pueden ser **Generales** (describen conceptos generales), **De dominio** (describen un dominio concreto del conocimiento) o **Específicas** (describen los conceptos para una aplicación concreta).

Según los destinatarios, serán: **Lingüísticas** (vinculan aspectos lingüísticos, gramática, semántica y sintaxis destinados a su utilización por seres humanos), **No lingüísticas**, destinadas a ser utilizadas por máquinas o **Mixtas**.

Según el nivel de abstracción, serán **Descriptivas** (descripciones, taxonomías, relaciones) y **Lógicas** (permiten inferencias lógicas).

Para la IDE-Uruguay, proponemos desarrollar una ontología específica, no lingüística o mixta y lógica.

Surgen algunos problemas derivados de la diversidad de la comunidad participante, de la falta profesionales de la información trabajando en el área y de no haber adoptado normas para la representación del conocimiento.

No hay mayores dificultades al determinar los conceptos que la integran pues los objetos geográficos están descritos en el Catálogo de Objetos (norma ISO-19.126), pero la influencia del portugués en la frontera este, aporta diferencias en la denominación de

algunos objetos geográficos. Más complejo es establecer la taxonomía que determinan las relaciones entre los objetos pues algunos términos de tienen definiciones ambiguas.

6. Reflexión final

Pensarnos como mediadores entre el conocimiento socialmente producido y la comunidad de usuarios, es a la vez de un desafío, una responsabilidad ética. No podemos seguir concibiendo la Organización y Representación del Conocimiento como una actividad técnica, de puertas adentro, objetiva; no hay actividad humana que sea neutral.

Referencia Bibliográfica

Barité, M (2001). Organización del conocimiento: un nuevo marco teórico-conceptual en Bibliotecología y Documentación. – En: Carrara, K (org) Educacao, universidad e pesquisa. – Marilia : Unesp, p. 35-60.

Borst, P. et al. (1997). Engineering ontologies. *International Journal of Human-Computer Studies*, 46 (2-3)

Bourdieu, P. (2003). El oficio de científico: ciencia de la ciencia y reflexividad. – Barcelona : Anagrama, 212 p

Hjorland,B; Albrechtsen,H (1995). Toward a new horizon in Information Science: domain analysis. – En: Journal of American Society for Information v. 46, n. 6: 400-425.

Hjorland, B (2002). Domain analysis in information science: eleven approaches –traditional as well as innovative. – En: Journal of Documentation v. 58, n. 4: 422-462.

Guarino, Nicola. Formal Ontology and Information Systems. Proceedings of FOIS '98. Recuperado de: <http://citeseer.ist.psu.edu/guarino98formal.html>

ISO TC 211 Geographic Information/geomatics. N 834. New work item proposal: Geographic information profile. FACC Data Dictionary. Oslo: NTS, 2006, 336 p.

Lamarca Lapuente, María Jesús (2011). *Hipertexto, el nuevo concepto de documento en la cultura de la imagen*. (Tesis doctoral, Universidad Complutense de Madrid).
Recuperado de: <http://www.hipertexto.info>

SAINZ, Ismael & JIMENEZ - RUIZ, Ernesto (2007). Ontologías en Informática. Recuperado de: <http://krono.act.uji.es/publications/techrep/Book-Chapter-Protege-report2007.pdf>

Uso de BISAC como sistema de organización del conocimiento en el catálogo.

Daniel Martínez-Ávila
Universidad Carlos III de Madrid
dmartine@bib.uc3m.es

Margaret E.I. Kipp
University of Wisconsin-Milwaukee
kipp@uwm.edu

Hope A. Olson
University of Wisconsin-Milwaukee
holson@uwm.edu

BISAC, Book Industry Standards and Communications, es el sistema de organización del conocimiento desarrollado y mantenido por el principal grupo de la industria del libro en los Estados Unidos, el Book Industry Study Group (BISG). La Lista de Encabezamientos BISAC, también conocida como Lista de Códigos de Materia BISAC, está definida como “un estándar usado por muchas compañías en la cadena de suministro del libro para categorizar libros en base a su contenido temático. El Encabezamiento de Materia aplicado a un libro puede determinar dónde está colocado en la estantería de una tienda física o el/los género/s por los que puede ser buscado en una base de datos interna” (Book Industry Study Group, 2012a).

En este sentido, algunas de las “compañías” y “tiendas físicas” que también han comenzado a utilizar BISAC han sido algunas bibliotecas públicas de los Estados Unidos como sustituto de la Clasificación Decimal de Dewey (CDD), en gran medida bajo los auspicios de OCLC y con prospecciones de expansión globales (Martínez-Ávila et al., 2012a) (Martínez-Ávila et al., 2012b).

Una de las ventajas desprendidas de la adopción de BISAC por bibliotecas es que, al contrario que el sistema de clasificación gestionado por OCLC, BISAC está disponible en línea sin coste alguno para su consulta (Book Industry Study Group, 2012b). Algunas de las características de BISAC incluyen el ser una lista alfabética de categorías, con 51 grandes secciones y otras tantas subsecciones de hasta tres niveles de profundidad, y que los descriptores están representados por un código alfanumérico que, aunque en la primera versión del sistema tenían un significado jerárquico, ya no lo tienen y por lo tanto deben ser tratados como simples códigos subrogados (Book Industry Study Group, 2013a). En la práctica, las principales diferencias entre la CDD y

BISAC vendrán dadas por las diferentes filosofías que guían el desarrollo de los sistemas, basada en el desarrollo de disciplinas científicas en el caso de la CDD y siguiendo una organización por temáticas de interés según el mercado en el caso de BISAC.

La denominación de esta filosofía dependiente del mercado que guía el desarrollo de BISAC ha sido propuesta como garantía del mercado ("market warrant") (Martínez-Ávila, 2012). Algunos ejemplos que ilustran las diferencias respecto a organización y recuperación con BISAC incluyen "mascotas" o la "Guerra Civil" estadounidense. En el primer caso, la CDD tiene más categorías con mascotas pero éstas se encuentran esparcidas entre "Ganadería" y otras áreas, no existiendo una sección de mascotas que las agrupe; en el caso de BISAC sí que existe una categoría sobre mascotas que, sin embargo, se concentra principalmente en perros y gatos debido a la garantía del mercado.

En el segundo caso, la CDD tiene más de sesenta categorías para la Guerra Civil, incluyendo Historia Diplomática, Unidades Navales, Historia Militar y Relatos Personales; BISAC, en cambio, tiene una sola categoría estando luego los elementos organizados alfabéticamente por autor. Las consecuencias de las diferentes filosofías en el desarrollo de los sistemas, como en el caso de estos ejemplos, afectarán no sólo al modo en el que los documentos son agrupados físicamente en las estanterías sino también la recuperación y organización de los elementos en el catálogo.

El uso de BISAC para el catálogo es algo contemplado por el BISG e implementado por numerosas organizaciones, tanto comerciales como bibliotecas. Según el BISG, "BISAC fue desarrollado para estandarizar la transferencia electrónica de información sobre la materia. Los Encabezamientos pueden ser usados para transmitir información entre socios comerciales, como términos de búsqueda en grandes bases de datos comerciales, como puntos de acceso para búsquedas en bases de datos y como guías de estanterías. Muchos grandes negocios requieren que los editores usen BISAC cuando envían datos" (Book Industry Study Group, 2013b).

Algunas de las organizaciones listadas en la página web del BISG que usan BISAC "en una variedad de maneras" incluyen Amazon, Baker & Taylor, Barnes & Noble, Bookscan, Booksense, Bowker, Indigo, Ingram y la mayoría de los principales editores. Por otra parte, el uso de BISAC para la categorización de libros en Google no ha estado exento de críticas. En agosto del 2009, Geoff Nunberg (UC Berkeley) criticó la calidad de los metadatos de Google, señalando numerosos errores de clasificación y

fechas (Nunberg, 2009a) (Nunberg, 2009b). Nunberg también cuestionó la decisión de Google de utilizar BISAC, un esquema de la industria del libro y no de las bibliotecas. Algunos de los errores señalados por Nunberg incluían "*Moby Dick*", categorizado como Informática, "*The Century Dictionary*", categorizado como Familia y relaciones, y "*Catalogue of Copyright Entries*", categorizado como Drama. Aunque muchos de estos problemas han sido arreglados con el tiempo, BISAC ha seguido siendo utilizado por BISAC en Google Books (alrededor del mundo), Google ebookstore, y ahora Google Play.

Por otra parte, otras tiendas de libros electrónicos como Amazon.com, SONY ebookstore o Itunes también están usando categorías de BISAC en sus catálogos. Tanto Google como Amazon y Sony son miembros del grupo de estudio de la industria del libro (BISG), al igual que otras entidades del ámbito bibliotecario como la American Library Association (ALA), la Library of Congress y OCLC (Book Industry Study Group, 2013c). Entre otras cosas, OCLC ha participado activamente en los últimos lanzamientos de BISAC, mapeado Dewey con BISAC, incluido BISAC en WebDewey 2.0, introducido BISAC en MARC (en los campos 600-657 ‡2, 072 ‡2, 073 ‡z), y mapeado MARC con ONIX (Martínez-Ávila et al., 2012b). La Biblioteca del Congreso también ha mantenido desde 2005 un mapeado de ONIX con MARC 21 (Library of Congress, 2005), estando ONIX definido como "la mejor manera de comunicar códigos BISAC entre socios" (Book Industry Study Group, 2013d). Actualmente, ONIX for books es usado por organizaciones como Google, Amazon.com, Baker & Taylor, Barnes & Noble, Bowker, y la Library of Congress. Respecto a las bibliotecas que adoptado BISAC, la mayoría de ellas muestran las categorías de BISAC en el catálogo en el campo de localización (Call number).

Para el presente trabajo se ha analizado el uso que hacen las principales bibliotecas que han adoptado BISAC en sus catálogos con el objetivo de corregir y mejorar las características que se pudieran estar desaprovechando. Algunas de las bibliotecas analizadas individualmente y de forma detallada en el estudio incluyen Maricopa County Library District en Arizona, Rangeview Library District en Colorado, Frankfort Public Library en Illinois, Phoenix Public Library, Topeka and Shawnee County Public Library en Kansas, Gail Borden Public Library District en Illinois, Albany Public Library System en New York, Darien Library en Connecticut, Henry County Public Library en Kentucky, Langley-Adams Library en Groveland (Massachusetts) y la Julia L. Butterfield Memorial Library en New York.

Algunas de las conclusiones extraídas del estudio son las siguientes: en las bibliotecas que utilizan BISAC en el catálogo, BISAC no está siendo usado en todas las sucursales del sistema bibliotecario, y cuando es usado, no se está utilizando una adopción estricta de BISAC (tal como se proporciona en los registros de ONIX/MARC) ni se está modificando de manera homogénea por todas las sucursales. Esto significa que las categorías de BISAC (las modificaciones) son mostradas en casi todos los casos en el campo de localización (Call number) del registro, y no en los campos habilitados para materias (y por lo tanto la información de BISAC en el registro ONIX/MARC importado no es válida).

Como consecuencia, las categorías de BISAC no están siendo recuperadas en las búsquedas por materia en los catálogos de bibliotecas cuando no están siendo incluidas en los campos MARC 6xx adecuados (los campos de materias), y cuando están siendo incluidos, normalmente se tratan de las categorías "canónicas" de BISAC y no de las modificaciones que realmente se están usando en las bibliotecas. Respecto a la función de agrupación de BISAC, tampoco está siendo aprovechada en los catálogos en línea, ya que los registros asignados a las categorías de BISAC (en el campo de Localización / Call number) no están siendo vinculados electrónicamente a otros registros con las mismas categorías de BISAC, a excepción de cuando comparten las mismas "materias" en el campo de materia (Subject field) normalmente tratándose de la LCSH, etc. Las ventajas técnicas de usar un estándar en el catálogo para la recuperación de información están siendo perdidas por las necesarias modificaciones de BISAC en las bibliotecas, y BISAC está siendo modificado en las bibliotecas porque se trata de un sistema diseñado para librerías, las cuales son entidades con unas muy diferentes características. Por último, los literales de BISAC (o sus modificaciones) no están añadiendo ninguna familiaridad o amigabilidad a la recuperación y organización de la información en los catálogos, por la simple razón de que no pueden ser recuperados cuando están siendo usados como sustitutos de Dewey.

Mientras que las ventajas de usar BISAC para la organización física de libros en bibliotecas es también cuestionable, si tampoco se están adaptando las ventajas en el catálogo del uso de sistemas provenientes de entornos comerciales del libro, cabe replantearse la adecuación de utilizar este tipos de sistemas en sustitución de otros bibliotecarios hasta que no se justifiquen de alguna manera sus ventajas.

Referencias:

Book Industry Study Group. 2012a. Standards & Best Practices - Classification Schemes - BISAC Subject Headings 2012 Edition - Book Industry Study Group, Inc. Disponible <http://www.bisg.org/what-we-do-20-73-bisac-subject-headings-2010-edition.php>.

Book Industry Study Group. 2012b. BISAC Subject Headings List, Major Subjects. Disponible <http://www.bisg.org/what-we-do-0-136-bisac-subject-headings-list-major-subjects.php>.

Book Industry Study Group. 2013a. What terms are used to describe elements in the BISAC Subject Heading list?. Disponible <http://www.bisg.org/activities-programs/activity.php?n=d&id=47&cid=20#What%20terms%20are%20used%20to%20describe%20elements%20in%20the%20BISAC%20Subject%20Heading%20List%20%28BSHL%29?>.

Book Industry Study Group. 2013b. How are BISAC Subject Headings used?. Disponible <http://www.bisg.org/activities-programs/activity.php?n=d&id=47&cid=20#What%20are%20the%20BISAC%20Subject%20Headings%20used%20for?>

Book Industry Study Group. 2013c. BISG Members | Book Industry Study Group. Disponible <http://www.bisg.org/directory/>

Book Industry Study Group, 2013d. How do I tell my trading partner which code applies to my books?. Disponible <http://www.bisg.org/activities-programs/activity.php?n=d&id=47&cid=20#How%20do%20I%20tell%20my%20trading%20partner%20which%20code%20applies%20to%20my%20books?>

Library of Congress. 2005. *ONIX to MARC 21 Mapping*. Disponible <http://www.loc.gov/marc/onix2marc.html>.

Martínez-Ávila, Daniel. 2012. *DDC-BISAC switching as a new case of reader-interest classification*. Ph.D. dissertation, Universidad Carlos III de Madrid.

Martínez-Ávila, Daniel, Olson, Hope, and Kipp, Margaret. 2012a. New Roles and global agents in information organization in Spanish libraries. *Knowledge organization* 39: 125-36.

Martínez-Ávila, Daniel, Kipp, Margaret, and Olson, Hope. 2012b. DDC or BISAC: the changing balance between corporations and public institutions. *Knowledge organization* 39: 309-19.

Nunberg, Geoff. 2009a. Google Books: The Metadata Mess. Presentado en *Google Book Settlement Conference UC Berkeley, August 28*. Disponible <http://people.ischool.berkeley.edu/~nunberg/GBook/GoogBookMetadataSh.pdf>.

Nunberg, Geoff. 2009b. Google Books: a Metadata Train Wreck. *Language Log* August 29. Disponible <http://languagelog.ldc.upenn.edu/nll/?p=1701>.

A consistência de um banco de dados como fonte de pesquisa e recuperação da informação - Projetos sobre a Região Platina/MERCOSUL

Gilberto Fladimar Rodrigues Viana
Universidade Estadual Paulista
vrfGPC@gmail.com

Telma Campanha de Carvalho Madio
Universidade Estadual Paulista
telmaccarvalho@marilia.unesp.br

1 Introdução

Este trabalho tem como tema a verificação do cadastramento da produção acadêmica relativa ao desenvolvimento de projetos na área de Artes e Letras na Universidade Federal de Santa Maria – UFSM, que envolvam assuntos relativos à Região Platina e/ou Mercado Comum do Sul – MERCOSUL. O objetivo é levantar no banco de dados – SIE (Sistema de Registro de Projetos na UFSM, implantado a partir de 1992 na referida instituição) o que se produz em termos de projetos de ensino, pesquisa e extensão no âmbito do CAL da UFSM, no período de 1992 a 2012, enfocando os assuntos já mencionados e que se constituem como parte do patrimônio documental da instituição. Visa-se também averiguar a consistência das informações arquivísticas dispostas no sistema.

Os procedimentos para a elaboração do presente trabalho constituem-se da coleta e análise de dados de relatórios referentes aos projetos desenvolvidos e dos grupos e linhas de pesquisa no CAL. No que tange à questão métrica das informações trabalhadas neste artigo, a tipologia é Informétrica e tem como objeto de estudo o banco de dados da UFSM/SIE, cuja variável é a medição da recuperação da informação, que quanto ao método, é por meio de modelos booleanos de recuperação da informação.

Pelo fato das etapas desenvolvidas para realização deste trabalho terem envolvido tanto aspectos quantitativos como qualitativos, define-se sua abordagem como quali-quantitativa, conforme Vieira (2004:15).

2 Revisão Bibliográfica

A partir da análise dos dados apresentados e das considerações a respeito dos números, passa-se a analisar esses elementos que se destacam dentro do panorama a que se propõe, ou seja, verificar se são desenvolvidos projetos que envolvam assuntos pertinentes à Região Platina/MERCOSUL, criado em 1991.

Suscita-se a investigação de como acontecem as redes de conexão da intelectualidade latino-americana, se através de projetos via convênios ou através de colaborações informais, denominadas colégios invisíveis, conforme Solla Price (1976), que destaca que muito da produção científica mundial acontece no âmbito desses.

Desenvolve-se este artigo também sob o suporte teórico arquivístico, no que se refere ao suporte digital, enfatizando a premência de dotar este banco de dados de requisitos e normatizações que propiciem a consistência das informações arquivísticas nele armazenadas.

Quanto à questão das tecnologias analógicas, cita-se Lemos (2008) quando trata da inflexibilidade das tecnologias analógicas como tecnologia ultrapassada e, em contrapartida, enfatiza o aspecto positivo da flexibilidade das tecnologias digitais e suas facilidades, afirmando:

Com o digital, a forma de distribuição e de armazenamento são independentes multimodais, onde a escolha em obter uma informação sob a forma textual, imagética ou sonora é independente do modo pelo qual é transmitida. Nesse sentido, as redes eletrônicas constituem uma nova forma de publicação (a eletrônica), onde os computadores podem produzir cópias tão perfeitas quanto a original (Lemos, 2008:69).

Mesmo dentro do universo digital, é importante destacar as funções arquivísticas: produção, avaliação, aquisição, conservação, classificação, descrição e difusão, pois todas essas funções são consideradas muito importantes. Enfatiza-se a função de classificação onde o princípio elementar é o “da proveniência” e o da “ordem original”, conforme ICA 16 (2005). Destaca-se, também, duas características de documento de arquivo consideradas muito importantes: a sua autenticidade – “significa que o documento de arquivo é o que pretende ser”, e a sua fidedignidade – “servir de prova digna de crédito”.

Quanto à autenticidade, pode-se dizer que o documento deverá manter suas características originais quanto ao contexto, à estrutura e ao conteúdo, sendo isso definido como coerência, com o fim para o qual o documento foi criado.

Essas considerações sobre autenticidade também estão expressas sob a estruturação de normas, através do Conarq – Conselho Nacional de Arquivos, da e-ARQ Brasil e da Norma ISO n. 15.489-1, ao tratar dos requisitos de gestão de documentos de arquivo.

Quanto à fidedignidade, essa é a qualidade atribuída ao documento referente à sua capacidade de “servir de prova digna de crédito”.

Constata-se também a falta do recurso de trilha de auditoria, como recomenda a e-ARQ Brasil, para preservar o histórico das possíveis alterações das informações no sistema. Observa-se que o sistema foi concebido sob a estrutura de metadados. Através da operação do sistema constata-se também que a sequência do registro de projetos é totalmente eletrônica e armazenada somente em suporte digital.

Pode-se dizer que essa questão da consistência das informações é de extrema importância e, como subsídio para expressar tal relevância, cita-se:

A exigência de que o registro seja produzido no curso regular das transações assegura que o documento em questão não é um fato casual ou isolado, mas foi criado, mantido e preservado porque a pessoa ou organização necessita dele para funcionar adequadamente (Duranti 1994:53).

202

O registro de projetos na instituição, ou seja, o que já se fez e se faz em termos de projetos de ensino, extensão e pesquisa, constitui-se como patrimônio documental, conforme o que consta na Constituição 1988, em seu artigo 216, que diz:

Constituem patrimônio cultural brasileiro os bens de natureza material e imaterial [...] IV – as obras, objetos, documentos, edificações e demais espaços destinados às manifestações artístico-culturais [...] (Constituição da República Federativa do Brasil, 1988).

3 Análise Dos Resultados

Constatou-se, conforme busca no Banco de Dados do SIE, a existência de duas linhas de pesquisa, porém nenhuma do CAL e de três grupos de pesquisa, sendo um do CAL que abrange o assunto Região Platina/MERCOSUL. Cabe enfatizar que a busca dessas informações no SIE, desenvolveu-se através do acesso ao banco de cadastros de

linhas e grupos de pesquisa – metadados – o qual dispõe de um recurso de recuperação da informação através de captura do banco na íntegra e por filtro.

Quanto aos projetos, a busca foi realizada através do título dos projetos, associando-os ao assunto pesquisado com uma ferramenta do sistema onde a busca é realizada por palavra inicial ou do meio, ou seja, não existe um vocabulário controlado, é uma busca aberta a todos os títulos de projetos que fazem parte do sistema.

Mesmo assim, verifica-se que existem projetos registrados que envolvem o assunto, conforme tabela 1.

Tabela 1 – Projeto registrados no CAL, relacionados ao assunto Região Platina/MERCOSUL

INFORMAÇÕES		TIPO DE PROJETO			SITUAÇÃO DO PROJETO		VINCULO ACADÊMICO	
ÁREAS	TOTAL	PESQUISA	ENSINO	EXTENSÃO	CONCLUÍDO	ANDAMENTO	GRADUAÇÃO	PÓS-GRAD
ARTES	8	5		3	8		4	4
LETRAS	20	16		4	12	8	11	9
MÚSICA	2			2	2		2	
TEATRO	1		1		1		1	
TOTAL	31	21	1	9	23	8	18	13

Fonte: SIE/UFSM

Esse quantitativo representa o universo coletado em relação aos assuntos propostos, Região Platina/MERCOSUL, que perfazem um total de 31 projetos, conforme tabela 1. Em relação ao total de projetos por áreas, verifica-se que a de Letras apresenta os números mais expressivos, com 64,52% dos projetos registrados. Considerando os tipos de projetos, percebe-se que os de pesquisa são os que se apresentam em maior número, com 67,74%.

Já, em relação à situação dos projetos, constata-se a situação “concluído” como sendo a mais expressiva, com 74,19%. E em relação ao vínculo dos projetos com a pós-graduação, verifica-se um percentual de 41,94.

4 Conclusão

No decorrer deste trabalho, fez-se uma retrospectiva sucinta do contexto histórico da Região Platina/MERCOSUL, onde se tangencia esse tema em relação ao

cadastro e desenvolvimento de projetos, linhas e grupos de pesquisa na UFSM e, mais especificamente, no CAL.

Chegou-se a constatações numéricas que demonstram o pouco envolvimento acadêmico com o tema, considerando o universo institucional de projetos desenvolvidos da UFSM.

Concluindo, faz-se necessário que as informações pertinentes ao tema, projetos referentes à região Platina/MERCOSUL, assim como os demais projetos constantes do sistema de registros dos projetos e armazenados no Banco de Dados da UFSM sejam consistentes, através de requisitos que garantirão a sua fidedignidade e autenticidade, assim como sua acessibilidade em termos de pesquisa e recuperação das informações à academia e à sociedade.

REFERÊNCIAS

Brasil. 1988. *Constituição da República Federativa do Brasil de 1988*. Disponível em: <http://www.planalto.gov.br/ccivil_03/Constituicao/Constitui%C3%A7ao.htm>. Acesso em: 08 out. 2008.

Conarq. Conselho Nacional de Arquivos, Câmara Técnica de Documentos Eletrônicos. 2006. *Modelo de requisitos para sistemas informatizados de gestão arquivística de documentos: e-ARQ Brasil*. Disponível em: <<http://www.conarq.arquivonacional.gov.br/Media/publicacoes/earqbrasilv1.pdf>>. Acesso em: 08 out. 2008.

Duranti, Luciana. 1994. Registros documentais contemporâneos como provas de ação. *Estudos Históricos*, Rio de Janeiro, 7(13): 49-64. Disponível em: <<http://www.cpdoc.fgv.br/revista/arq/134.pdf>>. Acesso em 10 out. 2008

International Council on Archives. Committee on current records in an electronic environment. 2005. *Electronic records: workbook for archivists*. Paris (France): ICA. (ICA Studies 16). Disponível em: <http://www.ica.org/sites/default/files/ICAEstudo16_PT_4.pdf>. Acesso em: 08 out. 2008.

Lemos, André. 2002. *Cibercultura, tecnologia e vida social na cultura contemporânea*. Porto Alegre: Sulina.

Solla Price, Derek J. 1976. *O desenvolvimento da ciência: análise histórica, filosófica, sociológica e econômica*. Rio de Janeiro: Livros Técnicos e Científicos. 77 p.

Ufsm/SIE – Sistema de Informações para o Ensino. *Banco de dados*.

Vaz, Alcides Costa. 2003. Mercosul: educação, justiça, cultura e desenvolvimento social em perspectiva. In: Álvarez, Gabriel; Furtado, Celso (orgs.). *Indústrias culturais no Mercosul*. Brasília, DF: IBRI. Disponível em: <<http://www.ibri-rbpi.org/livros/>>.

Vieira, Marcelo Milano F.; Zouain, Deborah M. (orgs.). 2004. *Pesquisa qualitativa em administração*. 2. ed. Rio de Janeiro: FGV.

Elaboração de vocabulário controlado em formato SKOS usando Tematres: Implicações metodológicas e web-semântica

Cibele Araújo Camargo Marques dos Santos
Universidade de São Paulo
cibeleac@usp.br

Michely Jabala Mamede Vogel
Universidade de São Paulo
michelyvogel@usp.br

Vânia Mara Lima
Universidade de São Paulo
vamal@usp.br

João Carlos Holland de Barcellos
Universidade de São Paulo
jocax@usp.br

Introdução

As normas de tesouros desenvolvidas na década de 1970 não atendem mais as necessidades da organização e recuperação da informação em um mundo de recursos interconectados, não deixando claras as diferenças entre conceitos que envolvem uma informação e os termos expressos na busca (Clarke e Zeng, 2012).

Do ponto de vista da OCR, a escolha de um termo preferido em um tesouro é feita através de critérios pré-estabelecidos, porém sabemos que nem sempre o termo escolhido será o termo utilizado na busca.

Embora atualmente, os modelos de metadados permitam interoperabilidade entre vocabulários, as relações semânticas entre seus termos, continuam sendo identificadas e operacionalizadas através da análise cuidadosa de profissionais da informação para determinar os campos semânticos a que os termos e conceitos pertencem, o nível de relação existente entre eles, preferências de termos, identificação de sinônimos e resolução de termos ambíguos.

Objetivos

Este trabalho apresenta o desenvolvimento experimental de um micro-tesauro sobre Cervejas e analisa implicações metodológicas na construção, utilização e conversão em vocabulário estruturado para a web-semântica.

Metodologia

Foi utilizado o software Multites para o tesauro e o software Tematres, executado na Internet e compatível com demais softwares de tesouros, na conversão para o vocabulário estruturado, permitindo a importação de termos, para campos adaptados às normas atuais e gerando listas em formatos diversos.

Vocabulário estruturado e web-semântica

Na web-semântica é preciso mais que um termo associado a um conceito, o software precisa de modelos de dados explicitados com a distinção entre termo/conceito (Clarke e Zeng, 2012). O tesauro é ferramenta de suporte para o acesso à informação, em contraposição à abordagem da pesquisa por texto completo, limitada pela expressão de um assunto por diferentes palavras e combinações de palavras.

Sendo a web-semântica constituída por documentos hipertextuais existe a necessidade de estabelecer a compatibilidade semântica entre o conteúdo destes documentos, pois podem existir diferentes estruturas conceituais para a sua representação. Esta situação pode aproximar assuntos que estão correlacionados entre si, ou documentos através de diferentes termos e contextos e a criação de um mecanismo que permita a inter-relação entre estas estruturas facilitando a recuperação da informação pelos usuários (Silva e Lima, 2012: 433).

Os tesouros e vocabulários na busca de compatibilidade para representação semântica, ou seja, cada termo deve possuir um termo correspondente com o mesmo significado em outro vocabulário permitindo a conversão sem alteração da significação, tiveram que encontrar soluções para questões linguísticas como a sinonímia e a polissemia. A primeira pode ser resolvida com as relações de equivalência dos termos presentes nas normas, mas na polissemia um termo com vários conceitos precisou nas normas do artifício do uso de qualificadores para definição do campo semântico a que se refere o que dificulta ao usuário a identificação do termo correto para a busca, sendo que para processamento por máquina esta solução também é insatisfatória.

O tesauro usado em contextos controlados por humanos ou com mediação de processo de pesquisa tem boa resposta, pois o usuário consegue interpretar os resultados realizando a inferência entre o termo e conceito (Clarke e Zeng, 2012). Esta inferência deve-se principalmente ao reconhecimento pelo usuário do campo semântico onde o termo se encontra, isto é, quais são as relações de significados entre os conceitos que permitem precisar a significação de cada termo. O estabelecimento de categorias no interior de uma linguagem documentária como o tesauro facilita esta inferência, pois a organização dos descritores em categorias delimita o campo semântico compartilhando as relações de significado (Fujita, 2001: 49).

A evolução das normas de tesauro levou, a partir de 2005, à quarta revisão da norma ANSI/NISO Z39.19 para vocabulários controlados cobrindo vários tipos de vocabulários cujas abordagens e estruturas lidam com as mesmas dificuldades que os tesouros como as listas de termos controlados, os anéis de sinônimos e as taxonomias, mas esta norma ainda não traz o modelo de dados que torne clara e específica a relação termo/conceito (Clark e Zeng, 2012).

A norma britânica BS8723 já voltada para vocabulários estruturados foi publicada em cinco partes entre 2005 e 2008 e tem ênfase no tesauro. Apresenta um guia para mapeamento entre vocabulários, traz um modelo de dados explícito com esquema em XML para intercâmbio de um tesauro inteiro ou suas partes e separa as classes de conceito e de termos.

Em 2005, o W3C (World Wide Web Consortium) apresentou a recomendação SKOS (Simple Knowledge Organization Systems) desenvolvida para publicação de vocabulários estruturados e traz no modelo de dados explicitamente a distinção entre conceito e as etiquetas usadas para representar conceitos. Na SKOS os relacionamentos semânticos entre conceitos correspondem de forma bem próxima aos relacionamentos hierárquicos e associativos recomendados nas normas de tesauro (Clarke e Zeng, 2012).

Existem diversas iniciativas para a representação do tesauro usando RDF que visam a aplicação na web-semântica, mas a SKOS é a principal proposta pois organiza os conceitos em diagramas (Pastor-Sanchez et al, 2009). A noção de termos descritores e não descritores foi substituída pela associação de conceitos de etiquetas preferidas e alternativas, apresentando também as relações de correspondência entre conceitos pertencentes a diferentes esquemas. A SKOS é usada para expressar a estrutura básica e o conteúdo dos conceitos nos esquemas como tesouros, sistemas de classificação, listas de cabeçalho de assuntos, taxonomias, terminologias, e glossários (Milles et al, 2005).

Foi revisada em 2009 e nesta versão a SKOS Core Data Model foi complementada com um apêndice contendo várias etiquetas que estendem as relações entre termos, denominado SKOS-XL (Clarke e Zeng, 2012).

A norma ISO também foi revisada por um grupo de trabalho e criou-se a Norma ISO 25964 parte 1 e a parte 2 que está em análise para publicação. Esta norma é rigorosa na distinção de termos e conceitos e embora não seja gratuita, o modelo de dados e o esquema XML estão no site da NISO.

Elaboração do micro-tesauro

O assunto Cervejas é bastante discutido atualmente, foram levantados diversos livros e quase duzentos sítios na internet, que discutem tipos existentes, aromas, cores, consistência, degustação, processos de fabricação, industrialização e comercialização das cervejas. Nessas fontes encontra-se de forma mais organizada alguns glossários e fichas de degustação.

A macro-estrutura foi pensada em relação aos ingredientes das cervejas, à degustação, ao aroma, cor e corpo, à origem, produção, tipos de cervejas, chegando a 399 descritores, 73 não-descritores e 126 notas de escopo.

No intuito de explorar o formato SKOS, foram identificadas equivalências de conceitos com a LC, que está “skosificada” e apresenta alguns termos da área de cerveja. Não foi possível fazer o mesmo com o Vocabulário da USP, DeCS (Descritores em Ciências da Saúde), e o Mesh (Medical Subject Headings), pois os dois primeiros não estão em formato SKOS, e o Mesh embora esteja, não deixa disponível os links.

O Tematres foi o software escolhido por conter recursos para geração de listas em formatos como SKOS, XML/BS8723 e XML/DC, com vistas a promover a interoperabilidade. Apesar de gratuito, sua instalação é complexa e exige servidor de internet, portanto optou-se pela versão de demonstração do software Multites para a inserção de descritores, explorando então no Tematres seus recursos de importação.

Para o processo de importação e conversão de formatos foi fundamental o aporte técnico da área de informática.

Resultados

O micro-tesauro criado no Multites e convertido para o Tematres permitiu a geração de lista em formato SKOS como previsto. No entanto, percebeu-se que o Tematres não oferece recurso para definição de facetas, assim, utilizou-se o recurso de inserir as mesmas entre colchetes, de modo a diferenciar esses termos dos demais. Também não permite a indicação de categorias ou macro-estruturas e optou-se por inseri-las em caixa alta.

Quanto ao processo de migração dos dados do Multites para o Tematres ocorreu que: a acentuação das palavras se perdeu, apesar de ambos os softwares aceitarem diacríticos, foi necessário refazer manualmente a acentuação; o arquivo de exportação do Multites não traz elementos suficientes para o reconhecimento no Tematres e desse modo, relações associativas, de equivalência e notas de escopo tiveram de ser manualmente indicadas no arquivo, antes da importação. Além disso, são raros os manuais disponíveis sobre o software e o trabalho com o servidor, requisito imprescindível para o sucesso da instalação do Tematres. É importante salientar que foram necessários vários testes para determinar a consistência do vocabulário na conversão.

Considerações finais

209

A conversão se mostrou relevante, na medida em que é necessário aproveitar todo o investimento já feito para o desenvolvimento de tesauros padronizados utilizando a norma antiga. Acredita-se que o sucesso obtido na importação dos termos foi significativo, porém demandou grande esforço, suporte técnico de informática, muitos testes, mudança na lógica do arquivo gerado a partir da outra norma, e muitos acertos manuais de termos e até de acentuação.

Deste experimento pode-se reforçar que as relações entre os termos e seus conceitos precisam ser desenvolvidas pelos especialistas em vocabulários com aportes da linguística, da lógica, da terminologia e da informática, pois estes conhecimentos são importantes para os vocabulários estruturados na web-semântica. Assim, os estudos nessas áreas continuam sendo fundamentais para estes formatos, mesmo quando interoperáveis.

Referências

Clarke, S. G.; Zeng, M. L. 2012 From ISO 2788 to ISO 25964: the evolution of thesaurus standards towards interoperability and data modeling. *Information Standards Quarterly*, 24: 20-26,

Fujita, M. S. Modelos de categorização para a construção de tesauros: metodologia de ensino. 2011. In. Boccato, V.R.C; Gracioso, L de S. (org) *Estudos de linguagem em Ciência da Informação*. Campinas: Alínea, p.35-67.

Miles, A.; Matthews, B.; Wilson, M.; Brickley, D. 2005. SKOS Core: Simple knowledge organisation for the Web .In: DCMI International Conference on Dublin Core and Metadata Applications, Proceedings. Madrid.

Pastor-Sanchez, J. A.; Martinez-Mendez, F. J.; Rodríguez-Munóz, J. V. 2009. Advantages of thesaurus representation using the Simple Knowledge Organization System (SKOS) compared with proposed alternatives. *Information research*, 14. n. 4,

Silva, G.B.S; Lima, G.A.B.de O. 2012. Using topic maps in establishing compatibility of semantically structured hypertext contents. *Knowledge Organization*, 39: 432-445.

O trabalho de descrição de acervos arquivísticos no Brasil

Eliezer Pires da Silva
Universidade Federal do Estado do Rio de Janeiro
eliezerpires@gmail.com

Evelyn Goyannes Dill Orrico
Universidade Federal do Estado do Rio de Janeiro
evelynorrico@unirio.br

Introdução

A descrição arquivística é o processo em que o arquivista cria representações de um determinado acervo arquivístico, apresentando seu contexto e conteúdo. É uma atividade intelectual que demanda competências de interpretação de texto, conhecimento histórico e habilidade para redigir descrições dos acervos. O objetivo é o controle dos documentos arquivísticos, tendo em vista a promoção do acesso.

Evidentemente essa é uma função do arquivista associada à Organização do Conhecimento no contexto dos arquivos. O conceito de descrição arquivística apresenta relativo isolamento em relação ao desenvolvimento da noção de representação no âmbito Organização do Conhecimento, ainda que, tanto neste quanto naquele campo as funções técnicas se configuram em torno do objetivo de dar acesso aos acervos, mas com trajetórias separadas.

Esta comunicação objetiva problematizar as referências norteadoras do trabalho de descrição de acervos arquivísticos no Brasil. Os procedimentos metodológicos foram a revisão da literatura sobre o assunto e a utilização da publicação “Descrição arquivística: referências bibliográficas”, produzida pela Câmara Técnica de Normalização da Descrição Arquivística, no âmbito do Conselho Nacional de Arquivos-CONARQ, servindo de base empírica.

Panorama Internacional

As iniciativas de padronização internacional da descrição arquivística revelam uma tendência em que o conceito de descrição arquivística se confunde com os instrumentos normativos propostos pelo Conselho Internacional de Arquivos. Atualmente existem normas internacionais para descrever os acervos arquivísticos

(ISAD/G: General International Standard Archival Description), os produtores de arquivos (ISAAR/CPF: International Standard Archival Authority Record for Corporate Bodies, Persons and Families), as funções dos produtores de arquivos (ISDF: International Standard for Describing Functions) e as instituições custodiadoras de acervos arquivísticos (ISDIAH: International Standard for Describing Institutions with Archival Holdings).

O processo de construção das normas de descrição arquivística surge no final da década de 1980 como uma iniciativa que envolveu comunidade arquivística internacional, a partir do Conselho Internacional de Arquivos - CIA. Inicialmente houve a criação de um grupo de trabalho, posteriormente transformado em Comissão Ad Hoc de Normas de Descrição do CIA. Em 1993, essa Comissão aprovou a Norma geral internacional de descrição arquivística, ISAD(G). Essa norma estruturou sete áreas de informação para representar e explicar o contexto e o conteúdo de documentos de arquivo a fim de promover seu acesso aos interessados.

Desse modo, foram sistematizadas as seguintes áreas de informação: **identificação** da unidade de descrição informando referências sobre o controle físico, nome do que está sendo representado, datação, dimensões do suporte e o nível dessa unidade de descrição em relação ao fundo de arquivo que forma ou faz parte; **contextualização** identificando a origem arquivística e a história da custódia dessa unidade de descrição; **conteúdo e estrutura** em que se destina à caracterização do assunto contemplado pela unidade de descrição; **condições de acesso** informando sobre possíveis restrições legais, critérios de reprodução, instrumentos de pesquisa e idioma; **fontes relacionadas** indicando a existência de outras unidades de descrição associadas a que está sendo descrita; **notas** apresentando informações que não forem enquadradas nos campos previstos pela norma; e, **controle da descrição** identificando quem a elaborou.

Panorama Nacional

As práticas e o saber arquivístico têm uma trajetória no país que remonta, por exemplo, a um tempo em que o Brasil é melhor denominado de América Portuguesa. O que chamamos hoje de Arquivo Nacional é uma criação de 1838. Apesar desse histórico, há um movimento recente de institucionalização específica da Arquivologia dentro do aparelho de ciência no Brasil, tratando-se da perspectiva clara sobre a

configuração do campo disciplinar para os estudos arquivísticos (ORRICO;SILVA, 2011).

Na arquivologia há (para além das interfaces e zonas em que se mobilizam diferentes disciplinas na compreensão do fenómeno arquivístico) um núcleo conceitual historicamente referenciado pela trajetória das práticas em torno do processamento técnico dos arquivos, o que configura uma estrutura intelectual, ou epistemologia, dessa disciplina. Assim, a Arquivologia é mais do que apenas uma mobilização específica de diferentes saberes em torno das questões dos arquivos, há um território de problemas socialmente reconhecidos sobre os quais pessoas e instituições encontram-se envolvidas, sobretudo atores de ciência.

É nesse quadro que a Câmara Técnica de Normalização da Descrição Arquivística (CTNDA), criada pela portaria n. 56, de 30/9/2001, do Conselho Nacional de Arquivos, em conformidade com a ISAD(G) e a ISAAR(CPF), propõe a NOBRADE: Norma Brasileira de Descrição Arquivística. Tal normalização das práticas arquivísticas no Brasil é a valorização dos padrões a partir das recentes iniciativas do Conselho internacional de Arquivos. Segundo Fonseca (2004), o saber arquivístico apresenta uma tradição “manualística” (compêndios de doutrinas, princípios e definições), no entanto, assiste-se ao surgimento de um quadro, em movimento, no qual há uma espécie de aposta na perspectiva das normas.

Ações institucionais, técnicas e políticas de largo alcance também contribuem para estabelecer uma cultura técnica comum, como por exemplo, aquelas que visam identificar e controlar massas documentais ainda não recolhidas, estabelecer canais de articulação e intercâmbio de informação com a criação e implementação de sistemas de arquivo, interferir no sentido da racionalização da produção e conservação de documentos de fase corrente e intermediária. (FONSECA, 1999, p.2)

Para Fonseca (1999), a necessidade das normas seria um consenso, além de urgente, nas instituições arquivísticas brasileiras, porque estas serviriam para equacionar os principais problemas no processamento técnicos dos acervos arquivísticos. O estado da questão no Brasil aponta para a ausência de pesquisas sobre os acervos arquivísticos, especificamente, quanto à adoção das normas de descrição e característica dos instrumentos de pesquisa.

Referências da Descrição Arquivística no Brasil

A seguir são apresentados os gráficos que revelam quais são as referências norteadoras do trabalho de descrição dos acervos arquivísticos, realizado pelos arquivistas no Brasil. Esta pesquisa estabeleceu indicadores que revelam os países de origem e as datas de publicação, bem como uma categorização das obras de referência desse assunto entre os seguintes enfoques: normalização da descrição arquivística; fundamentos arquivísticos; temáticas afins à descrição arquivística.

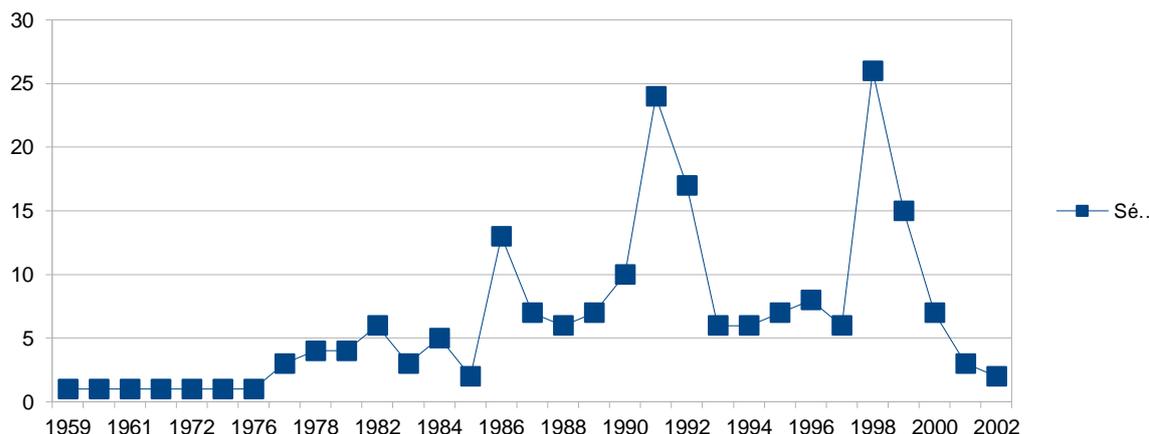


Gráfico 1: Distribuição das referências sobre descrição arquivística por ano

Fonte: Elaborado pelos autores.

O gráfico acima ratifica a importância dos anos 1990 para os debates sobre a descrição arquivística num período que coincide com o surgimento das normas propostas pelo Conselho Internacional de Arquivos. O percentual das publicações com data entre 1990-1999 é de 61%, com destaque para o ano de 1998 com 13%.

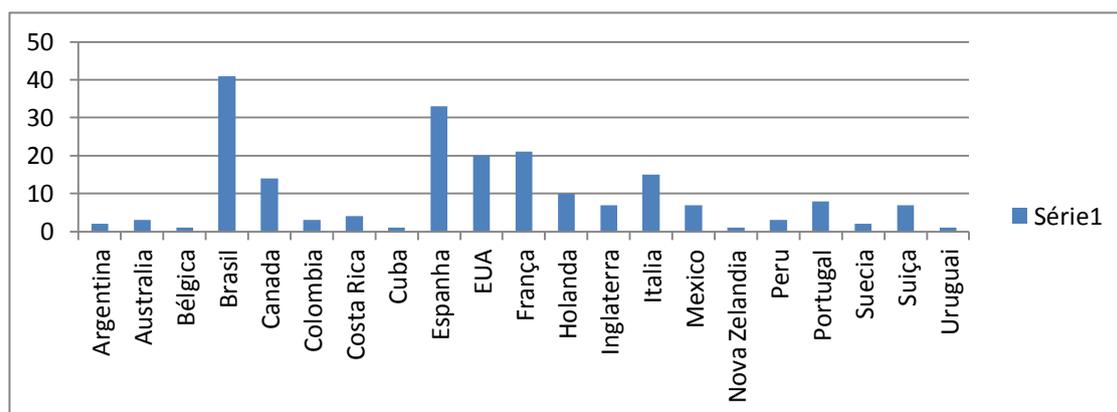


Gráfico 2: Referência bibliográfica sobre descrição arquivística por país

Fonte: Elaborado pelos autores.

A perspectiva da identificação do país de origem das obras de referência revelou a relevância da produção nacional de conhecimento arquivístico, alcançando o

percentual de 20%, seguida da influência espanhola com 16%. Chama a atenção o fato de que a Espanha possui números superiores aos da França e dos Estados Unidos, países que tradicionalmente têm exercido maior influência sobre a comunidade arquivística brasileira.

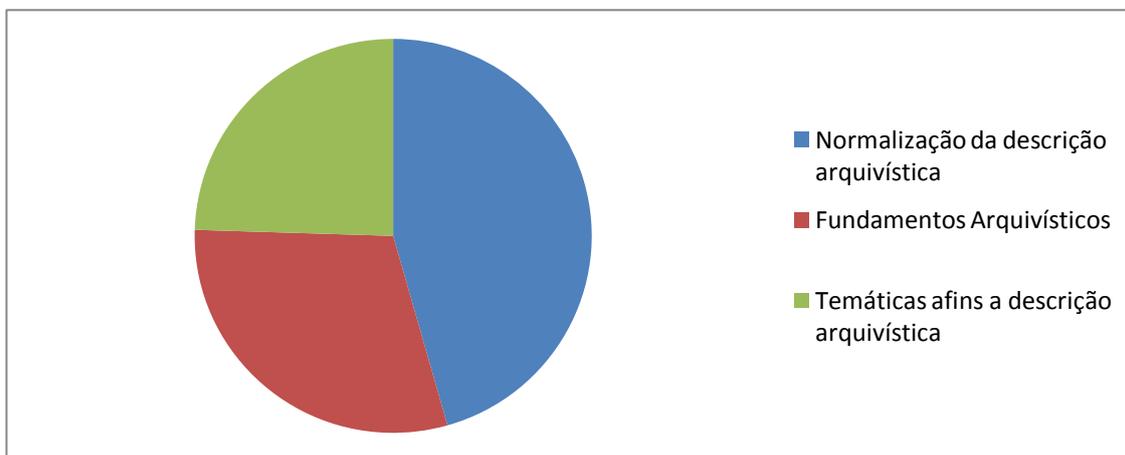


Gráfico 3: Categorização dos enfoques das referências sobre a descrição arquivística

Fonte: Elaborado pelos autores.

O gráfico 3 resulta do exercício de análise das obras de referência do trabalho de descrição de acervos arquivísticos no Brasil. Os percentuais são: normalização da descrição arquivística (45%); fundamentos arquivísticos (30%); temáticas afins a descrição arquivística (25%). A concentração do enfoque da normalização nas referências da descrição arquivística confirma a perspectiva de que o debate conceitual sobre a descrição arquivística vem sendo absorvido pelo processo de sua normalização nos planos internacional e nacional.

Considerações Finais

O processo de construção das normas de descrição arquivística surge no final da década de 1980, como uma iniciativa do Conselho Internacional de Arquivos, que representa a comunidade arquivística internacional. Em 1993 foi divulgada a Norma Geral Internacional de Descrição Arquivística – ISAD(G). Esse instrumento normativo estrutura sete áreas de informação para representar e explicar o contexto dos documentos de arquivo a fim de promover o acesso aos mesmos.

O surgimento das normas de descrição arquivística representa um movimento de normalização das práticas arquivísticas, sugerindo uma inovação em relação aos

manuais (uma espécie de tradição “manualística”), mas que dá continuidade ao caráter prescritivo do saber arquivístico. Hoje se questiona em que medida as normas permitiriam a interoperabilidade de sistemas e a criação de descrições consistentes.

As obras de referência da descrição arquivística no Brasil revelam que há uma significativa produção nacional sobre o assunto, além da concentração dessas publicações na década de 1990. É importante ressaltar que 45% dessas referências são textos que apresentam o enfoque da normalização do trabalho arquivístico de descrição.

Referências

ARQUIVO NACIONAL (Brasil). Conselho Nacional de Arquivos. **Descrição arquivística:** referências bibliográficas. Rio de Janeiro: Arquivo Nacional, 2004.

FONSECA, Maria Odila. **Arquivologia e Ciência da Informação:** (Re) definição de marcos interdisciplinares. Tese (Doutorado em Ciência da Informação)- IBICT, Escola de Comunicação. Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2004.

FONSECA, Vitor Manoel Marques da. **A normalização da descrição arquivística:** avanços internacionais e a situação do Brasil. Rio de Janeiro: Arquivo Nacional do Brasil, 1999.

ORRICO, Evelyn Goyannes Dill; SILVA, Eliezer Pires da. Network of specialists in the archival field and the impact on knowledge organization: the case of Brazil. In: **X Congreso ISKO-España**, 2011, Ferrol. ISKO Congreso-España, 2011.

A dimensão social, cultural e política da ORC

A presença brasileira no cenário de produção científica da ISKO

José Augusto Chaves Guimarães
Universidade Estadual Paulista
guima@marilia.unesp.br

Antecedentes

Para que se possa abordar teoricamente a organização do conhecimento enquanto área de estudos, inicialmente, há de se diferenciar duas concepções de conhecimento.

a) enquanto processo individual, constitui-se, na concepção da Ingetraut DAHLBERG (1995, p.11) em uma *certeza subjetiva ou objetivamente conclusiva da existência de um fato ou do estado de um caso, não sendo transferível e somente podendo ser adquirido por meio da reflexão;*

b) enquanto algo sobre o qual existe um certo consenso social, trabalhando-se aqui com o conhecimento registrado e divulgado (BARITÉ, 2001).

Nesse panorama, a organização do conhecimento enquanto disciplina surge, como recorda GARCIA MARCO (1997, p.8) na década de 90, “na encruzilhada das denominadas ciências cognitivas, no campo de encontro entre as Ciências do Conhecimento Humano (Neurociência, Psicologia e Epistemologia), Ciências da Informação e da Comunicação (incluindo a Semiótica e a Linguística), matemática (incluindo a Lógica e as Linguagens formais) e a Ciência da Computação”

Referindo-se à concepção de organização do conhecimento adotada pela ISKO, afirma STRAIOTO (2001, p.79) que:

“para a ISKO, toda a forma de organização do conhecimento deve estar baseada em unidades de conhecimento ou conceitos e nas diferentes possibilidades de combinar palavras/termos. Conhecendo-se os princípios teóricos que vêm sendo desenvolvidos, desde Ranganathan, passando pelo Classification Research Group, poder-se-á construir sistemas de classificação ou tesouros facetados de uma maneira mais fácil”.

Procurando dar uma dimensão mais aplicada a tais domínios, esclarecem ESTEBAN NAVARRO e GARCIA MARCO (1999) que a Organização do Conhecimento, enquanto disciplina, refere-se aos processos de “planejamento, construção, gestão, uso e avaliação de sistemas de descrição, catalogação, ordenação, classificação, armazenamento, comunicação e recuperação dos documentos”

notadamente a partir do conteúdo desses, para que a informação neles contida e representada possa gerar novo conhecimento.

STRAIOTO (2001, p.78), ressalta que a Organização do Conhecimento, centra sua ênfase investigativas na compreensão das formas pelas quais o conhecimento pode ser organizado e representado de modo a que possa estar amplamente disponível e acessível.

Dada sua natureza eminentemente interdisciplinar, ressalta GARCIA MARCO (1997, p.220) que os espaços acadêmicos (ensino e pesquisa) devem voltar suas atenções para a fundamentação do caráter científico da disciplina, notadamente no que tange aos conteúdos que lhe são inerentes e aos métodos que lhe são aplicáveis, para a aplicação de tais conteúdos e métodos em realidades e contextos diversificados, de modo a integrar um *know why* com um *know how*, e para o estabelecimento de quadros de referência de interface.

Nesse âmbito, a experiência internacional da área encontra seu maior representante na ISKO – International Society for Knowledge Organization, criada em 1989, na Alemanha (Frankfurt) tendo à frente pesquisadores como Ingetraut Dahlberg e Dagobert Soergel, a partir de uma dissidência ocorrida no seio da Society for Classification, criada em 1977.

Cumprido ressaltar, como influência teórica marcante, nos trabalhos da ISKO, os trabalhos realizados por Ranganathan e, em continuidade, pelo Classification Research Group (CRG) os quais *podem ser utilizados na organização de todos os tipos de sistemas gerais e específicos de organização e representação do conhecimento* (STRAIOTO, 2001, p.77).

Tendo por objetivos a promoção da pesquisa, o desenvolvimento de aplicação de métodos, o intercâmbio científico de informações e a construção e aplicação de instrumentos de organização e representação do conhecimento, a ISKO possui hoje uma estrutura organizada em capítulos (Alemanha/Áustria/Suíça, Brasil, Canadá/Estados Unidos, China, Espanha, França, Índia, Irã, Itália, Maghreb, Polônia e Reino Unido) e conta com a realização de congressos internacionais (nos anos pares) e congressos nacionais ou regionais (nos anos ímpares), bem como dispõe da revista Knowledge Organization e da série Advances in Knowledge Organization como veículos de divulgação de sua produção científica.

Especificamente no caso brasileiro, a pesquisa em Organização e Representação do Conhecimento – ORC - tem evidenciado um crescente espaço no meio acadêmico,

em especial no âmbito da Ciência da Informação, em que a formação e massa crítica vem sendo, ao longo das últimas quatro décadas, um processo gradual e ascendente.

Tal aspecto pode ser observado pelo seguinte elementos que, de forma mais direta, vêm contribuindo para a consolidação científica da ORC no Brasil, a saber:

A criação e consolidação de programas de pós-graduação *stricto sensu* em Ciência da informação. Iniciado nos anos 70, em nível de mestrado, e nos anos 90, em nível de doutorado, os referidos programas, desde sua gênese, revelaram uma forte preocupação com as temáticas relativas à ORC, como demonstram programas pioneiros como do IBICT, da UFMG, da USP e da UnB, com linhas de pesquisa nessa área. Posteriormente, outros programas vieram enriquecer esse universo, como os programas da UNESP, da UNIRIO e da UFF. Tal aspecto pode ser observado em ementas de linhas específicas ligadas à OC nos citados programas, a saber (vide destaque em negrito no que tange aos aspectos relacionados à OC)²²:

IBICT (Comunicação, Organização e Gestão da Informação e do Conhecimento)

Ementa: Estudos históricos e epistemológicos da Ciência da Informação e metodologias das Ciências Sociais e Aplicadas. Comunicação e divulgação em Ciência e Tecnologia; análises e aplicações bibliométricas, informétricas, webmétricas e cientométricas. **Sistemas de organização e representação do conhecimento, ontologias, web semântica e contribuições da lingüística.** Processos de busca, acesso, recuperação e uso da informação. Dimensões conceituais e semióticas das estruturas e dos fluxos da informação e do conhecimento em diferentes contextos. Informação e gestão, monitoramento tecnológico, gestão estratégica da informação e do conhecimento nas organizações e nas políticas públicas. Cultura organizacional.

220

UFMG (Organização e Uso da Informação)

Ementa: A linha “Organização e Uso da Informação” preocupa-se com estudos de duas das funções básicas de sistemas de informação: a recuperação da informação e **a organização** e o uso **de informação**. Foi estruturada com base no pressuposto de que o estudo e a reflexão sobre qualquer das duas funções são potencializados a partir da interação/inter-relação existente entre as duas, procurando explorar as teorias correspondentes, de forma a consolidar núcleos teóricos relevantes para as áreas envolvidas. Entre os grandes temas da linha destacam-se: **Representação da informação (classificação, descrição e modelagem) em contextos digitais, análise de assunto**, Bibliometria, estudos de usos e usuários de sistemas de informação.

UNESP (Produção e Organização da Informação)

²² Dados disponíveis em <http://www.ancib.org.br/pages/pos-graduacoes-em-ci.php>. Acesso em 20.04.2013.

Considerando a informação registrada e institucionalizada como insumo básico para a construção do conhecimento no contexto da Ciência da Informação, destaca-se o desenvolvimento de referenciais teóricos e aplicados, de natureza interdisciplinar, acerca da produção e da **organização da informação**. A produção da informação é abordada sob os eixos da produção científica (avaliação do comportamento da ciência) e da produção documental (Diplomática contemporânea) e a **organização da informação é abordada a partir dos processos de análise, síntese, condensação, representação, e recuperação do conteúdo informacional, bem como das competências e comportamentos informacionais do usuário inerentes a tais processos. A dimensão teórica que fundamenta a produção e a organização da informação encontra subsídios na organização do conhecimento (notadamente em suas relações interdisciplinares com a Lógica, a Linguística, a Terminologia, a Semiótica e a Análise de Domínio) e na teoria da ciência**, enquanto a dimensão aplicada se efetiva a partir dos estudos métricos (Informetria, Cienciometria, Bibliometria e Webometria), de tipologia documental, dos instrumentos e produtos de organização da informação, e das questões de formação e atuação profissional na área.

UNIRIO: Memória e linguagem

Ementa: Estudos sobre as **relações entre linguagem, representações sociais**, identidade e memória. A produção, circulação e apropriação de sentidos em práticas de informação e discurso. Discursos fundadores e a construção da identidade e da diferença. Os embates sócio-históricos nas construções discursivas. Os discursos na manutenção e nas transformações sociais, como *locus* de lutas dos sujeitos/instituições e projetos sócio-culturais. As representações sociais nos meios midiáticos, no contexto das novas configurações sociais e das inovações tecnológicas.

221

USP: Organização da Informação e do Conhecimento

Ementa: Estudos teóricos e metodológicos relativos à organização do conhecimento e da informação e de sua circulação para fins de acesso, recuperação e uso. Compreende a análise dos objetivos, processos e instrumentos que caracterizam as distintas possibilidades de **organização da informação**, considerando - se ainda a sua inserção histórica e sócio-cultural e as condições de interação face à diversidade da produção e dos públicos da informação. Compreende, também, abordagens históricas e epistemológicas da **organização do conhecimento e da informação**. As pesquisas que a integram distribuem-se nos seguintes eixos complementares: a) **teorias e métodos de construção e organização da informação documentária para distintos receptores. Observam-se os aspectos textuais/discursivos dos objetos informacionais e os diferentes modelos de leitura, análise, condensação e representação, incluídos os modelos computacionais;** b) **a construção de linguagens documentárias e outras ferramentas de organização da informação para o acesso, recuperação e uso, observando-se características linguísticas, semióticas, terminológicas e comunicacionais, dos conteúdos documentários e dos grupos receptores, bem como de insumos tecnológicos;** c) **estudos históricos e epistemológicos relativos à organização social do**

conhecimento e sua relação com as propostas de organização da informação; d) análise e proposição de políticas de organização da informação no escopo da sua distribuição e recepção.

UnB: Organização da Informação

Ementa: A Linha de Pesquisa Organização da Informação busca propor conhecimentos nos níveis epistemológico, científico e prático relativos à origem, coleta, **organização**, estocagem, recuperação, interpretação, transmissão, transformação e uso **da informação**. Neste contexto, relaciona-se com a natureza da informação, **a terminologia e modelos de tratamento e recuperação de informações**; as necessidades dos usuários de informação e suas implicações; a identificação dos recursos necessários a partir dos tipos e formatos; a identificação, o tratamento e a recuperação de informações adequadas para o usuário; a formulação de políticas, estratégias, planejamentos, normas e processos relacionados a diferentes espaços de informação.

UFF: Fluxos e mediações sócio-técnicas da informação

Ementa: Investiga os processos informacionais e comunicacionais, considerando as relações entre as tecnologias da informação e da comunicação e os diferentes campos do conhecimento científico e técnico, seus padrões, demandas e uso de informação. Estudam-se a geração, **a organização**, a representação e a gestão da informação, com especial enfoque nas mediações sócio-técnicas **da informação** e da comunicação nestes processos. Inclui estudos de Informativa.

Outro importante espaço acadêmico nesse contexto reside na ANCIB – Associação nacional de Pesquisa em Ciência da informação, criada em 1989, e que há mais de uma década, vem desenvolvendo atividades científicas a partir da atuação de um conjunto de grupos que congregam pesquisadores e produção científica em temáticas específicas, como é o caso do GT-2, denominado *Organização e Representação do Conhecimento*, cuja ementa assim se enuncia:

Teorias, metodologias e práticas relacionadas à **organização** e preservação de documentos e da informação, enquanto conhecimento registrado e socializado, em ambiências informacionais tais como: arquivos, museus, bibliotecas e congêneres. Compreende, também, os estudos relacionados aos **processos, produtos e instrumentos de representação do conhecimento** (aqui incluindo o uso das tecnologias da informação) e as relações inter e transdisciplinares neles verificadas, além de aspectos relacionados às políticas de organização e preservação da memória institucional.

Vale recordar que o GT-2 da ANCIB constitui, ao longo de sua história, um dos grupos que congrega maior número de pesquisadores e para o qual se submete – para fins de apresentação no ENANCIB - um enorme espectro de trabalhos, o que reflete o “fôlego investigativo” dessa área, no Brasil. Essa realidade, se reflete, por exemplo, no universo das agências públicas de fomento à pesquisa. A título de exemplo, verifica-se

que, hoje, 20 dos 45 bolsistas de Produtividade em Pesquisa do CNPq na área na Ciência da Informação têm atividades acadêmicas relacionadas à área de OC o que corresponde a aproximadamente 45% do total de bolsas em curso na área de Ciência da Informação, atestando a centralidade essa.

Ainda nesse contexto, destaca-se a criação da ISKO-Brasil, cujas discussões iniciaram-se no âmbito do GT-2 da ANCIB, por ocasião do ENANCIB de 2005, em Florianópolis. Naquela ocasião, um conjunto vinte pesquisadores da área de OC, a exemplo do que ocorria em distintos países e sentindo a necessidade de um espaço científico mais verticalizado para agregar a pesquisa na área, manifestou sua intenção de criar um capítulo brasileiro da ISKO. Para tanto, o grupo contou, no ENANCIB seguinte, realizado em Marília em 2006, com o apoio incondicional da Profa. Dra. Maria José Lopez Huertas, então presidente da ISKO internacional, que orientou o grupo na confecção do estatuto. Dessa forma, no ENANCIB de 2007, realizado em Salvador, foi oficialmente criada a Sociedade Brasileira da Organização do Conhecimento, filiada à ISKO internacional, como capítulo brasileiro, o primeiro da América Latina.

Passado mais de um ano de esforços para dar conta da burocracia de se constituir uma sociedade científica no Brasil, no ENANCIB de 2009, em João Pessoa teve-se efetivamente, a primeira atividade científica da ISKO-Brasil, quando o Prof. Dr. Joseph T. Tennis, da University of Washington (EUA) proferiu a palestra *Engaged knowledge organization*, atentando para uma dimensão eminentemente cultural da área de OC.

Paralelamente a esse desenvolvimento da ISKO-Brasil, observou-se uma crescente presença de pesquisadores brasileiros em espaços regionais da ISKO (Espanha, América do Norte e França).

Dessa forma, e com um *know how* acumulado em distintas experiências internacionais, a ISKO-Brasil Pôde promover, até o momento, dois Congressos Brasileiros de Organização e Representação do Conhecimento, respectivamente em Brasília, em 2011 e no Rio de Janeiro, em 2013, cujos temas versaram sobre os *Desafios e perspectivas para a organização do conhecimento na atualidade* (em suas dimensões epistemológica, aplicada e cultural) e sobre a *Complexidade e Organização do Conhecimento: desafios de nosso século* (igualmente nas três dimensões anteriores).

Hoje, com pouco mais de cinco anos de existência, a ISKO-Brasil congrega 43 pesquisadores, sendo o capítulo de maior dimensão dentre os onze da ISKO.

Metodologia

Considerando essa questão, vale analisar como vem se dando a inserção internacional da pesquisa brasileira em ORC no contexto da ISKO. Para tanto, as seguintes questões se colocam: a) Como vem se desenvolvendo, ao longo do tempo, a presença brasileira no universo científico da ISKO, a partir de sua produção científica? b) Que universidades e pesquisadores brasileiros vêm ocupando centralidade nesse universo? c) Que parcerias estrangeiras se manifestam? e d) Que grau de articulação ocorre, nessa produção científica, entre os congressos da ISKO e a revista KO?

Para tanto, analisaram-se as coleções completas da revista Knowledge Organization (1993-2012) e da série Advances in Knowledge Organization (1990-2012), selecionando-se aqueles trabalhos que tivessem a participação de autor(es) e/ou instituição (ões) brasileiro(s). Dessa forma, de um conjunto de 381 artigos publicados na revista Knowledge Organization, foram selecionados 14 (3,6%) com a participação de brasileiros, assim como de um total de 716 *papers* publicados na série Advances in Knowledge Organization (relativas aos congressos internacionais da ISKO) foram selecionados 39 (5,4%) com a participação e autores brasileiros).

Análise e discussão dos dados

Um primeiro aspecto a destacar, decorrente da conjuntura anteriormente apresentada, é a crescente participação brasileira na literatura científica da ISKO. Tal aspecto pode ser observado no gráfico a seguir, em que se observa o efetivo crescimento dessa participação, seja em artigos na revista Knowledge Organization, seja em *papers* publicados em anais dos congressos internacionais da ISKO.

A esse respeito, merecem menção as considerações feitas por Smiraglia (2011, p. 179-180) acerca do Congresso Internacional da ISKO realizado em Roma em 2010, ao afirmar que, relativamente aos *papers* publicados a liderança coube à participação norte-americana (29%), “but the second largest contribution came from Brazil (13.8%), whose chapter was just formed in 2009”. Essas considerações se reiteram acerca do congresso subsequente, realizado em Mysore, em 2011 (Smiraglia, 2013, p.4): “Also impressive was the Brazilian presence, which accounted for nearly a third of the papers, more than doubling their presence in 2010.”

Observa-se, outrossim, que essa crescente presença de *papers* brasileiros os congressos internacionais da ISKO também se reflete na submissão – e publicação - de

artigos de autores brasileiros na revista Knowledge Organization, aspecto que denota não ser um fenômeno isolado mas, sim, uma tendência.



Gráfico 1: Participação brasileira na ISKO
Fonte: Elaborado pelo autor.

É interessante observar que essa produção brasileira nos veículos de comunicação científica da ISKO reflete os espaços tradicionais de produção de conhecimento na área, nomeadamente as universidades que possuem linhas de pesquisa de pós-graduação sobre essa temática, como é o caso da UFF, da UFMG, da UNESP, da UNIRIO e da USP, como se pode observar a seguir:

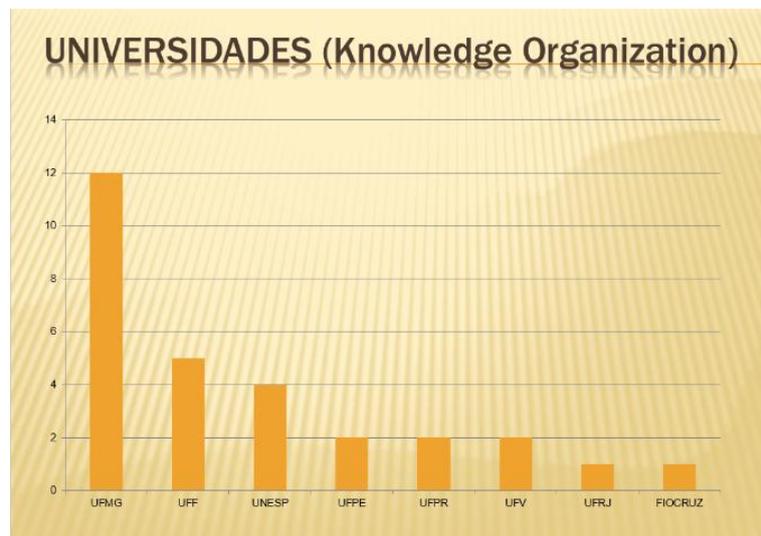


Gráfico 2: Universidades com artigos publicados na revista Knowledge Organization.
Fonte: Elaborado pelo autor.



Gráfico 3: Universidades com *papers* publicados em *proceedings* dos congressos da ISKO.
Fonte: Elaborado pelo autor.

Nesse universo, destaca-se um grupo de 20 pesquisadores, considerados mais produtivos no universo ISKO (com 2 ou mais publicações):

Autor	Instituição
Almeida, M. B.	UFMG
Alvarenga, L.	UFMG
Barros, T.H.B.	UNESP
Coelho, F.C.	FGV
Corrêa, C.A.	USP
Dodebei, V.L.D.L.M.	UNIRIO
Gontijo, M.	UNIRIO
Gracio, M.C.C.	UNESP
Gracioso, L.S.	UFSCar
Guimarães, J.A.C.	UNESP
Kobashi, N.Y.	USP
Lima, G.A.B.O.	UFMG
Milani, S.O.	UNESP
Moraes, J.B.E.	UNESP
Murguia, E.I.	UFF/UNESP
Oliveira, E.F.T.	UNESP
Orrico, E.G.D.	UNIRIO
Pinho, F.A.	UFPB
Souza, R.R.	FGV/UFMG
Tognoli, N.B.	UNESP

Quadro 1- Pesquisadores mais produtivos
Fonte: Elaborado pelo autor.

Considerações finais

A vista do exposto, pode-se observar que o processo de consolidação científica brasileiro na área de organização e representação do conhecimento revela uma coerência entre o crescimento acadêmico e visibilidade da produção científica

internacional, notadamente em espaços científicos mais especializados, como é o caso da ISKO. Esse crescimento, por sua vez, parece seguir uma “aceleração constante”, notadamente nos últimos cinco anos, o que, inclusive, coincide com a existência da ISKO-Brasil.

As instituições e pesquisadores mais produtivos confirmam os espaços acadêmicos de construção científica na área, em geral ligados a programas de pós-graduação com tradição de pesquisa na área (linhas ou temas de pesquisa).

No *corpus* analisado foi ainda possível perceber que a as parcerias científicas de autores brasileiros com pesquisadores internacionais é ainda bastante tímida e merece ser intensificada. De toda forma, as parcerias existentes (com países como Espanha, Estados Unidos, Índia e País de Gales) confirmam a histórica tradição brasileira da área de Ciência da Informação no tocante a diversificadas influências e diálogos científicos.

Referências

- BARITÉ, M. Organización del conocimiento: um nuevo marco teórico-conceptual em Bibliotecologia y documentación. In: CARRARA, K. (org.) *Educação, universidade e pesquisa*. Marília ; Unesp-Publicações : FAPESP, 2001, p.35-60.
- DAHLBERG, Ingetraut. Current trends in knowledge organization. In: GARCIA MARCO, F.J. (org.). Organización del conocimiento en sistemas de información y documentación. Zaragoza : Universidad de Zaragoza, 1995. v.1 p:7-26
- ESTEBAN NAVARRO, Miguel Ángel , GARCIA MARCO, Francisco Javier. Las primeras jornadas sobre organización del conocimiento.: organización del conocimiento e información científica. Scire, v.1, n.1, p.149-157, Ene./Jun. 1995.
- GARCIA MARCO, Francisco Javier. Avances en organización del conocimiento en España: los II encuentros sobre organización del conocimiento en los sistemas de información y documentación. In: GARCIA MARCO, F.J. (ed.) Organización del conocimiento en sistemas de información y documentación. Zaragoza : Librería General, 1997. v.1 p:7-12 (Actas del II ENCUENTRO DE ISKO-ESPAÑA, Getafe, 16-17 nov. 1995).
- SMIRAGLIA, R.P. ISKO'10 diverse Bookshelf: na editorial. Knowledge organization, v.38, n.3, p.179-186, 2011.
- SMIRAGLIA, R.P. ISKO'12 bookshelf: involving intension: em editorial. Knowledge organization, v.40, n.1, p.3-10, 2013.
- STRAIOTO, Ana Cláudia. Classificação de *shampoos*: elementos para um sistema facetado. Marília : FFC-UNESP, 2000. P: Relatório Geral de Qualificação (Mestrado em Ciência da Informação).

Facets, views and frames: enhancing connectedness in information organization.

Kathryn La Barre
University of Illinois at Urbana Champaign
klabarre@illinois.edu

Library and information science (LIS) has long been concerned with providing optimal access to cultural heritage resources, yet serious obstacles to this goal remain. This presentation will discuss ongoing work to address the difficulties in creating effective access for two types of cultural heritage materials – films and folktales. This discussion will center upon the role that facets can play in enhancing access to these kinds of complex materials through evaluating the lessons learned from two projects: *Films and Facets* (La Barre and Cordeiro, 2012) and *Folktales and Facets* (La Barre and Tilley, 2012). Both projects used facet analysis as the centerpiece for user-focused and task-focused approaches which sought to better understand the needs of librarians, scholars and storytellers working with folktales (*Folktales and Facets*) and casual film enthusiasts (*Films and Facets*).

Chief among the lessons that will be discussed is the abiding positive valence of frames of reference, viewpoints and context as guideposts during searching for both groups of users in these projects. Here, frames and views or viewpoints represent those ideas or assumptions that determine how a resource might be approached, perceived or understood. Frames, a familiar concept to LIS, are structures that divide knowledge into ‘stereotyped situations’ and can be connected together to form a complete idea (Minsky, 1977). Today we most often think of frames in the context of structure-based knowledge representations. The concept of framing is integral to the way in which the user groups in *Folktales and Facets* interacted with cultural heritage materials, and the source of greatest disconnect in search systems which did not provide these users with the means to search according to context or viewpoint. Provision of a variety of views of films, including descriptions that included subject access and pointed to related items, were highly desired among the casual film enthusiast group in *Films and Facets*

Reaching toward the provision of optimal access to information resources, several researchers including Maria López-Huertas and Hope Olson have placed feminist theory at the center of their research as positive step towards the creation of access and discovery structures which are deeply situated in a context of social and cultural differences. These two projects continue in this vein and build upon Olson’s

(2007) observations of the four ‘traces of connectedness’ in existing systems of information access and discovery. These traces consist of associative relationships in thesauri and subject headings, the existence of facets and synthetic devices in subject access systems, the entity relationship model of FRBR and collaborative tagging practices. These four areas are identified as especially fruitful approaches that may well enhance access to information resources. Pointing towards the need for increasing the connectedness of subject access is part of the long trajectory of Olson’s work which builds upon Maron’s (1977) three types of aboutness: S-about (subjective), O-about (objective) and R-about (retrieval), though Robert Fairthorne (1969) is credited with the first use of that term.

Themes of aboutness and the need for connectedness are deeply embedded in the results from the folktale project in the form of the *viewpoint facet* which relates to the worldview, ethnic origins, ethical concerns or possible methodological approaches to a given item. *Viewpoint* encompasses whether or not a teller can recount a tale from a different culture without express permission. Awareness of viewpoint can extend to the circumstances of the time of year during which a tale can be told, and to whom. Themes of connectedness are also a critical part of the results of the films project which found a need for an *associations facet* which could support searching for awards, derivations. *Mise en scenè* with correlate search extensions such as setting, time and theme are also highly desirable features of online search systems for digital films. Olson’s observations about the prospects for using switching languages, metathesauri and other approaches that help better situate and map such syntagmatic relationships are a key part of the future directions for both research projects to be discussed.

References

- Fairthorne, R. A. (1969). Content analysis, specification and control. *Annual Review of Information Science and Technology*, 4, 73-109.
- La Barre, K; Cordeiro, R. (2012). That obscure object of desire: Facets for film access and discovery. In Diane Rasmussen Neal (Ed.) *Indexing and retrieval of non-text information*. Berlin: DeGruyter Saur. pp. 234-262.
- La Barre, K.; Tilley, C. (2012). The elusive tale: Leveraging the study of information seeking and knowledge organization to improve access to and discovery of folktales. *Journal of the American Society for Information Science and Technology*, 63(4):687-701.
- López-Huertas, M., de Torres, I., & Barité, M. (2004). Terminological representation of specialized areas in conceptual structures: The case of gender studies. In I. C. McIlwaine (Ed.),

Knowledge Organization and the Global Information Society Proceedings of the Eighth International ISKO Conference, 13–16 July 2004, London, UK (pp. 35–39). Würzburg, Germany: Ergon.

Maron, M. E. (1977). "On indexing, retrieval and the meaning of about." *Journal of the American Society for Information Science*, 28, 38-43.

Minsky, M. (1974). A Framework for Representing Knowledge. MIT-AI Laboratory Memo 306. Reprinted in *The Psychology of Computer Vision*, P. Winston (Ed.), McGraw-Hill, 1975. <http://web.media.mit.edu/~minsky/papers/Frames/frames.html>

Olson, Hope A. (2007). How we construct subjects: A Feminist analysis. *Library Trends* 56 (3), 509-541.

Imagens e filmes: o potencial informativo dos documentos visuais para o acesso em ambientes de informação

Rosa Inês de Novais Cordeiro
Universidade Federal Fluminense
igneznovais@uol.com.br

Nossa intenção nesta comunicação é retomar algumas questões que dizem respeito diretamente a Arquivística e a Ciência da Informação, no enfoque da Organização da Informação e do Conhecimento, mas precisamente da imagem em movimento, marcadamente o filme cinematográfico, o qual temos estudado sobre três perspectivas e em momentos diferentes: o filme e suas fotografias de cenas; o filme e sua família de documentos; o filme e sua recepção. Assim, nossa proposta se situará na terceira perspectiva (o filme e sua recepção), onde está a análise do indexador e a busca de informação pelo o demandante (usuário). Procuraremos, assim, compreender as possíveis trajetórias de pensamento na leitura do filme e os elementos do documento visual que são acionados pelo indexador e usuário como pontos de acesso ao conteúdo dos documentos imagéticos.

231

Tentaremos apontar alguns aspectos que poderão ser considerados na análise de imagens e filmes, visando o estabelecimento de princípios para a sua política de indexação e recuperação em ambientes de grandes estoques de informação ou acervos de documentos (repositórios, arquivos, bibliotecas).

Inicialmente é fundamental acentuarmos que esse universo de documentos visuais constitui-se em ambientes informacionais de naturezas diversas (ambiente web ou não; documentos digitais ou não, entre outros). Soma-se a isto o cenário atual, no qual fica cada dia mais instável a identificação pelos serviços de informação dos possíveis perfis de interesse dos demandantes (usuários), bem como o seu espaço cognitivo no processo da busca de itens de informação, no qual observamos as necessidades expandidas de informação do usuário em uma era de “linguagens líquidas”, expressão que é uma metáfora da “liquidez” usada por Santaella a partir de Novak (1991) e Bauman (2000).

Diante disso, acreditamos, cada vez mais, na importância do desenvolvimento de diretrizes relacionadas com a possibilidade do estudo sobre o potencial informativo dos documentos a partir de cartografias que deem conta de mecanismos concernentes a tal

perspectiva, como categorias visando à expansão da leitura do analista-indexador acerca do objeto informacional imagético e incentivando-se, por exemplo, o desenvolvimento e a aplicação de procedimentos que objetivam o estudo do comportamento de busca e uso da informação na dimensão de Kuhlthau (2003), ou seja, durante o processo de trabalho do profissional. Isto é, considerando o comportamento do demandante na busca e uso da informação nos espaços e práticas de sua atuação e a atividade que está sendo executada em determinado momento, por exemplo, na realização de um trabalho artístico.

Portanto, essa informação demandada por esse amplo universo de profissionais faz parte de um repertório singular e resultante do seu conhecimento, seja ele de natureza tácita, explícita ou científica e, no qual, diferentes fontes de informação serão consultadas. Esses estoques de informação/documentos necessários nos diferentes momentos de trabalho do profissional necessitarão de organização e, por conseguinte, entre outros aspectos, da análise e síntese do seu conteúdo para inclusão nos serviços de informação para o acesso e o uso dos documentos (imagens, audiovisuais, textos), ou seja, “espaços de fluxos”.

Compreende-se a complexidade da diversidade das fontes de informação quanto a sua produção, formatos, suportes e linguagens (formatos analógico ou digital, conteúdos textuais, imagens ou audiovisuais, e outros), entretanto, nos cabe estudar como essas fontes em um ambiente de informação e visando à polirrepresentação das mesmas, poderão ser analisadas, indexadas e recuperadas de modo a permitir que o demandante tenha acesso ao conteúdo facetado, pois, como sabemos, a recuperação da informação, seja ela imagem, audiovisual ou texto, faz parte de um processo comunicacional interativo, estando, portanto, inserida e sujeita a um contexto situacional, mutável e transitório.

As necessidades dos usuários em relação aos documentos são variantes, fazendo com que um documento possa ser múltiplo-indexado para permitir a busca por diferentes facetas. No caso de documentos imagéticos, o usuário provavelmente fará a escolha da imagem desejada, a partir do oferecimento, pelo sistema, de um painel de imagens contendo várias opções de seleção.

Desse modo, nos parece possível, ponderarmos sobre princípios convergentes para a ampliação do espectro do tratamento da informação, tais como a “polirrepresentação²³” e a “busca interdisciplinar²⁴”.

²³ Hjørland (1997, p.47) usa o conceito de “*polyrepresentation*” a partir de Ingwersen.

É fundamental que também seja considerada a interatividade na recuperação da informação entre Espaço Cognitivo do Usuário e Espaço de Informação do sistema, resultando em uma dada situação (variável), ou seja, contexto situacional de necessidade de informação do usuário.

Referências

HJORLAND, Birger. *Information seeking and subject representation*. Westport: Greenwood Press, 1997.

INGWERSEN, Peter. Cognitive perspectives of information retrieval, interaction: elements of a cognitive IR theory. *Journal of Documentation*, London, v. 52, n.1, p.3-50, Mar. 1996.

KNAUTZ, Kathrin. Emotion felt and depicted: consequences for multimedia retrieval. In: NEAL, Diane Rasmussen. *Indexing and retrieval of non-text information*. Berlin: DeGruyter, 2012. cap. 14, p.343-375.

KUHLTHAU, Carol C. *Seeking meaning: a process approach to library and information services*. 2.ed. Westport, Libraries Unlimited, 2003.

LA BARRE, Kathryn; CORDEIRO, Rosa Inês de Novais. That obscure object of desire: facets for film access and discovery. In: NEAL, Diane Rasmussen. *Indexing and retrieval of non-text information*. Berlin: DeGruyter, 2012. cap. 10, p.234-262.

SANTAELLA, Lúcia. *Linguagens líquidas na era da mobilidade*. São Paulo: Ed. Paulus, 2007.

²⁴ Hjørland (1997, p.48) emprega esta expressão a partir de Klein (1990) e Bartolo; Smith (1993)

Comunicação da mesa: Panorama da pesquisa em OC em sua dimensão social, cultural e política (formação profissional, ética, cultura e identidade, contextos, sustentabilidade)

María J. López-Huertas
Universidad de Granada
mjlopez@ugr.es

La OC en la actualidad se enfrenta a nuevos retos debido a los cambios sufridos por la sociedad, la cultura, la manera de producir conocimiento, representada por la inter y transdisciplinariedad, y la gestión de las empresas, entre otros (López-Huertas, 2006a, 2006b, 2007, 2009, 2010). Asimismo, la necesidad de ofrecer información en sistemas transversales a los usuarios potenciales; esto es, sistemas que sean capaces de atender una determinada petición independientemente del lugar del mundo en el que residan dichos/as usuarios/as, plantea de igual manera la necesidad de la integración cultural y de la fundamentación del diseño y la construcción de Sistemas de Organización del Conocimiento en supuestos que den respuesta a estas realidades.

Ante esta situación, nos encontramos con varios temas que demandan reflexión y acción:

1. Las funciones de la OC exceden en gran medida el ámbito tradicional de la recuperación de la información en sentido estricto y hay que prepararse para dar respuesta desde aquí a esos otros contextos que se van a mencionar a continuación. La OC es una actividad neurálgica en temas como la toma de decisiones en organizaciones y empresas para mejorar sus actividades y su poder competitivo (Rodríguez y López-Huertas, 2013). Este trabajo demuestra cómo fundamentar la toma de decisiones en el conocimiento organizado de una organización o empresa y cómo este procedimiento redundará en beneficios para la organización, su personal y el contexto social al que se dirige.

2. La OC es, asimismo, una magnífica fuente de información y un método efectivo para evaluar contextos determinados que requieren la identificación de nuevos indicadores de calidad. Este es el caso que se desarrolla en un artículo de Prévil, López-Huertas et al. (2012), donde se demuestra cómo las categorías de la estructura de un tesoro de acciones de enfermería en cirugía vascular sirven no sólo para organizar el conocimiento sino también para la detección de nuevos indicadores, generados desde dentro de la especialidad.

3. El problema de la transculturalidad ha sido recientemente abordado por diversos autores que han reivindicado la inclusión de las perspectivas culturales como un elemento imprescindible en los Sistemas de Organización del Conocimiento (Beghtol, 1998, 2002, López-Huertas, 2008a). El concepto de inclusión cultural se refiere no sólo a la integración de áreas culturales propiamente dichas sino también a las distintas “otras culturas” que coexisten con la estándar o mayoritaria López-Huertas, 2008b). Sin duda, el hecho de que la transversalidad sea ya un imperativo en el diseño de sistemas de información hace que este tema cobre una gran dimensión. La investigación en este tema habría que dirigirla primero hacia un análisis de casos para ver cuáles son los principales problemas que impiden esta integración para luego intentar darles una respuesta adecuada y lo más general posible para que se pueda aplicar en diversos contextos. Creo que las posibles respuestas pasarían necesariamente por una revisión de las bases teóricas y metodológicas actualmente establecidas.

Muy relacionada con la transculturalidad está la ética profesional porque, desde mi punto de vista, este tema tiene dos aristas: una la ya comentada más arriba y la otra está relacionada con la responsabilidad que tenemos de intentar que los sistemas que creamos sean lo más inclusivos posible socialmente

4. Por todo lo dicho hasta ahora y, es especial por la transversalidad que reclaman los sistemas, es necesario un replanteamiento de los fundamentos que rigen la construcción de sistemas de organización del conocimiento. Es necesario encontrar alternativas que pasen por una revisión creativa de los fundamentos que rigen estos sistemas. Esto significa un replanteamiento de:

- a) La concepción y el ámbito de los conceptos
- b) El concepto de categoría, su representatividad y funciones
- c) La estructura conceptual que los organiza, que es el motor y último responsable de la eficiencia del sistema.

Todos los cambios mencionados han ido generando una corriente de pensamiento general que algunos autores han llamado post-epistemológica (García Gutiérrez, 2011) o después de la epistemología (Harris, 2009). Otros no le asignan un nombre concreto pero en sus teorías subyacen ideas parecidas (Nicolescu, 2011; Gibbons, 1994 and Nowotni, 2001). Aquí sólo se va a comentar una de sus críticas por ser la que más directamente afecta a los elementos constitutivos OC: referents, conceptos, categorías y

la organización estructural de los mismos. Se trata de la *crítica al límite en la conceptualización de la naturaleza y la realidad*.

Conceptos. Actualmente, establecen límites (exclusiones), son concebidos como entidades cerradas. Esto genera la concepción de conceptos binarios (lo que pertenece o no a un concepto), heredado del positivismo. El cambio iría hacia el tratamiento del concepto como una entidad abierta y porosa de manera que sea capaz de incluir diferentes contextos

Categorías. Seguirían el mismo camino que el señalado para los conceptos

Estructura conceptual. Igualmente, en este caso, no debe entenderse siguiendo un planteamiento dicotómico, sino que, siguiendo la dinámica de conceptos y categorías indicada, debe ser más flexible basándose en otras lógicas. Aquí Nicoliescu en su visión de la realidad abierta y estructurada en niveles no subordinados sino en coexistencia y regidos por otra lógica, la del tercero incluido, nos da alguna pauta a seguir. Una estructura capaz de:

- a) alojar distintos contextos, ya presentes en los conceptos que aloja, y, por tanto, abierta
- b) reinterpretar la jerarquización como criterio esencial estructurante. La jerarquización crea estructuras cerradas que favorecen poco la comunicación y menos la transversalidad. En un estudio sobre la clasificación de instrumentos musicales en distintas culturas, se ha demostrado que el orden de cita de las categorías y conceptos es lo que más aleja a unos esquemas de otros (López-Huertas, 1999)

Esto no significa olvidarse de la jerarquización sino incluirla como un elemento más en el sistema en lugar de ser su parte esencial

Referencias

Beghtol, Clare (1998). Knowledge domains: Multidisciplinarity and Bibliographic classification Systems. *Knowledge Organization*, 25 (1/2), 1-12

Beghtol, Clare (2002). Universal concepts, cultural warrant and cultural hospitality. En: M.J. López-Huertas (ed.). *Challenges for Knowledge Representation and Organization for the 21st Century. Integration of knowledge across boundaries*. Proceedings of the 7th International ISKO Conference. Würzburg, Ergon Verlag, pp. 45-49

García Gutiérrez, Antonio (2011). *Epistemología de la Documentación*. Barcelona, Stonberg.

Gibbons, M. et al. (1994). *The new production of knowledge. The dynamics of science and research in contemporary societies*. London, Sage

Harris, Roy (2009). *After Epistemology*. Sandy, Brightpen, Gamlingay

López-Huertas, María J. 1999. Potencialidad evolutiva del tesoro: hacia una base de conocimiento experto. In: María López-Huertas and Juan C. Fernández Molina, eds. *La representación y organización del conocimiento en sus distintas perspectivas: su influencia en la recuperación de la información*. Actas del IV Congreso ISKO-España. Granada 1999. Granada, Universidad-ISKO, pp. 129-135

López-Huertas Pérez, María J. (2006a). Análisis del dominio interdisciplinar para la representación y organización del conocimiento. En: Congreso Memória, Informação e Organização do Conhecimento: Cruzando as fronteiras da identidade. Río de Janeiro, UNIRIO-IBICT, pp. 209-236

López-Huertas Pérez, María J. (2006b). Aproximación a un modelo para la construcción de mapas temáticos en dominios interdisciplinares. En: I International Conference on Multidisciplinary Information Sciences and Technologies, InSciT2006. Badajoz, Instituto Abierto del Conocimiento, pp. 548-552

López-Huertas Pérez, María J. (2007). Gestión del conocimiento multidimensional en los sistemas de organización del conocimiento. En: 8º Congreso ISKO España. La interdisciplinariedad y la transdisciplinariedad en la organización del conocimiento científico. León, Universidad

López-Huertas, María J. (2008a) Cultural Impact on Knowledge Representation and Organization in a Subject Domain. Xth Internacional ISKO Conference. Montreal, 2008. Würzburg, Ergon, 2008

López-Huertas, María J. (2008b). Some current research questions in the field of Knowledge Organization. Knowledge Organization,

López-Huertas Pérez, María J. (2009) La terminología como método para representar y organizar dominios multidimensionales. En: 1er Simposio sobre Organización del Conocimiento: Bibliotecología y Terminología. México, UNAM

López-Huertas Pérez, María J. (2010). Nuevo conocimiento, innovación y sociedad: Retos para la gestión de la información. V Reunión Internacional de Gestión de Investigación y Desarrollo. Florianópolis, AVEGID-AIGID

Nicolescu, Basarab (2011). Transdisciplinariedad. Manifiesto 1. Available at: <http://es.scribd.com/doc/38437874/Bassarab-Nicolescu-La-Transdisciplinariedad-Manifiesto1>

Nowotni, Helga et al.. (2001). *Rethinking science. Knowledge and the public in an Age of uncertainty*. Cambridge, Polity Press

Previl, Ketty, López-Huertas, María J. et al. (2012). Indicadores para la evaluación de la satisfacción de los pacientes de cirugía vascular en tiempos de crisis. En: XXIV Congreso Nacional de Enfermería Vascular. Barcelona, 2012

Problemas da organização do conhecimento na contemporaneidade

Marilda Lopes Ginez de Lara
Universidade de São Paulo
larama@usp.br

1 Introdução

“Perguntamo-nos como encaixar em algo que pareça real, tão real como um mapa, este feixe de comunicações distantes e incertezas cotidianas, atrações e desenraizamentos, que se nomeia como globalização. (...) onde encontrar a teoria que organize as novas diversidades?” (Canclini, 2007).

Enfrentamos hoje, na pesquisa em Organização e Representação do Conhecimento, dificuldades de tratar documentos em face à manifestação das diversidades sociais, culturais e políticas. Que ferramentas teremos de desenvolver para dar conta da produção do conhecimento e de sua recepção quando verificamos que as referências da documentação ainda estão assentadas nas hierarquias, no privilégio dos substantivos (as essências) e, muitas vezes, em parâmetros universais? Como não incorrer nas falácias das tecnologias que, muitas vezes, se restringem a codificar os vocabulários sem colocar em questão os problemas originais de estruturação dos vocabulários? Como deveremos encaminhar a formação de novos profissionais sabendo das restrições que o uso de referências consolidadas colocam para a organização e a recepção das informações?

Não pretendemos dar respostas a esses problemas, mas enumerar alguns deles para a reflexão considerando o tema do evento: a *Complexidade e organização do conhecimento: desafios de nosso século*.

O problema hoje colocado – e em grande parte tornado mais explícito pela globalização e pelas tecnologias de comunicação e informação – diz respeito à multiplicidade de referências sócio-culturais que dão respaldo às escolhas dos pontos de partida para a organização do conhecimento. Procuraremos, neste trabalho, caracterizar brevemente o momento contemporâneo em que a ciência moderna é questionada e discutir, em seguida, as dificuldades enfrentadas para contemplar as novas demandas a partir da utilização de instrumentos de organização e metodologias clássicas.

Buscaremos problematizar princípios de organização do conhecimento tradicionalmente utilizados de modo a colocar em evidência seus limites para propor, em seguida, uma reflexão sobre caminhos que possam ser mais adequados para dar conta da produção e da recepção nos dias de hoje.

2 O momento contemporâneo

A ‘sociedade da informação’, ou ‘sociedade do conhecimento’, sem entrar nos problemas dessas denominações, enunciam um momento de transições que tem colocado inúmeras perguntas sobre “as relações entre a ciência e a virtude”, sobre o valor do conhecimento ordinário ou vulgar frente ao conhecimento científico, sobre o “contributo positivo ou negativo da ciência para nossa felicidade” etc. (SANTOS, 1988). “Estamos no fim de um ciclo de hegemonia de uma certa ordem científica” quando começa a ruir distinção entre ciências naturais e sociais (1988) e, com ela, seus parâmetros. Vivemos a crise do modelo de racionalidade – que fundamentou a Documentação como a maior parte do pensamento sobre Organização e Representação do Conhecimento – que coloca em xeque nossa confiança epistemológica assentada na redução da complexidade.

Para a racionalidade, “conhecer significa dividir e classificar para depois determinar relações sistemáticas entre o que se separou” (SANTOS, 1988), significa procurar pela causalidade formal. Suas bases estão assentadas no pressuposto metateórico da ordem e da estabilidade do mundo, do privilégio do modelo das ciências naturais sobre as ciências sociais. Tais bases começam a ruir começando pelos questionamentos colocados pela física desde Eistein que, propondo a relatividade da simultaneidade, revolucionou as concepções sobre espaço e tempo (dois acontecimentos só são simultâneos no mesmo sistema de referência), passando por Heisenberg e Bohr, que demonstraram que a observação não é isenta de interferências do observador (as leis da física são probabilísticas, a totalidade do real não se reduz à soma das partes observadas, a dissolução da clara distinção sujeito/objeto), e por Gödel, com o teorema da incompletude e da impossibilidade da prova de consistência em proposições que não se podem demonstrar nem refutar. A ‘crise paradigmática’ também atingiu a química, a biologia (com Ilya Prigogine, discutindo a ‘ordem através de flutuações e a lógica de auto-organização em situações de não-equilíbrio que, em síntese, resultam no reconhecimento da importância da história, da imprevisibilidade, da interpenetração, da

espontaneidade, da auto-organização, da irreversibilidade, da desordem, da criatividade, e e do acidente). Tais teorias se alinham a tantas outras, propostas por cientistas nas ciências da natureza e também nas ciências sociais, contribuindo não só para o questionamento sobre o conhecimento das coisas, como para mostrar a necessidade de considerar as condições sociais, os contextos culturais e os modelos da investigação científica que culminam no questionamento da ciência como visão privilegiada do mundo. Na visão de Boaventura Santos, o paradigma emergente, fruto da crise, reconhece que: “Todo conhecimento científico-natural é científico-social”, que “Todo o conhecimento é local e total”, que “Todo conhecimento é autoconhecimento”, e que “Todo conhecimento científico visa constituir-se num novo senso comum” (SANTOS, 1988).

Numa outra perspectiva, e mais à frente, Harvey (2002), refletindo sobre o momento contemporâneo, mostra que as tecnologias de comunicação e informação, as estruturas em rede, a inexistência de parâmetros únicos e centrais corroboram a emergência de uma nova sensibilidade, onde tem lugar o efêmero, o transitório, o descontínuo e o aparente caos. Conceitos-chave para a contemporaneidade – interculturalidade, heterogeneidade, diversidade, diferença, instabilidade – apontam para deslocamentos em relação à situação vivida desde o iluminismo e exigem novas práticas que considerem a negociação, os intercâmbios, os trabalhos colaborativos.

Do ponto de vista da cultura, Canclini mostra que “Os aspectos cognitivos e socioculturais estão distribuídos e são apropriados de modos muito diversos”, gerando “diferenças, desigualdades e desconexões” (CANCLINI, 2007, p.225). Diferenças, porque as comunidades têm valores diferentes, desigualdades, porque as sociedades promovem uma distribuição desigual dos bens econômicos, e desconexões, porque nem todos têm acesso às tecnologias da informação e comunicação. As estruturas tradicionais, tais como as unidades familiares, a escola e o bairro, são atravessadas por novas e diferentes relações: as formas de socialização já não são as mesmas do momento anterior à internet de meados do século XX; o pertencimento das pessoas transcende a seus grupos sociais: o ‘juntar-se’ não se realiza exclusivamente pela tradição, mas por interesses temáticos que são, por sua vez, múltiplos, voláteis e temporários.

Da mesma forma, as estruturas e categorias tradicionais não mais dão conta da complexidade contemporânea, o que leva a uma parafernália de conceitos que tentam dar nome e enquadramento àquilo que escapa à categorização. Na tentativa de explicar

as ciências, as artes, as técnicas, recorre-se a conceitos como interdisciplinaridade, transdisciplinaridade e outros, mas à medida em que são, ainda, disciplinares, acabam por reiterar as formas tradicionais de estabelecer as relações entre os fenômenos. Brian Holmes sugere um novo ‘tropismo’ e um novo tipo de reflexividade além dos limites que tradicionalmente se atribuem às atividades com a intenção expressa de enfrentar o desenvolvimento de uma sociedade complexa.

O termo ‘tropismo’ expressa o desejo e a necessidade de girar-se em direção a outra coisa, para um campo ou disciplina exteriores, enquanto a noção de reflexividade indica agora um regresso crítico ao ponto de partida, um intento de transformar a disciplina inicial, acabar com seu isolamento, abrir novas possibilidades de expressão, análise, cooperação e compromisso. Este movimento para frente e para trás, ou melhor, esta espiral transformadora é o princípio operativo do que chamarei *pesquisas extradisciplinares* (HOLMES, 2008, p. 203).²⁵

O conjunto dessas reflexões permite mostrar que o momento que vivemos é complexo, não só porque a ciência e a vida social mudou, mas porque não se apagaram, ainda, as referências de uma ciência moderna. Seus reflexos se fazem sentir também na produção teórica e prática da Organização e Representação do Conhecimento que, de um lado, se repora ao que já foi construído em bases iluministas, positivistas; de outro, a problemas difíceis de resolver na falta de parâmetros.

As tecnologias de comunicação e informação parecem abrir novas possibilidades para a organização e acesso às informações mas, na maior parte das vezes, têm restrições de uma lógica formal. Buscamos, ainda, por modelos claros e seguros, mas eles não mais existem. Que restrições nossos modelos de organização de conhecimento têm que dificultam a operação de dar conta de uma multiplicidade de formas de compreender e organizar o mundo? Que problemas eles apresentam em face às diferentes condições de apropriação do conhecimento?

3 Dificuldades/Problemas com os quais se depara hoje o ‘campo’ da Organização e Representação do Conhecimento

a) Começemos por assinalar a dificuldade de delimitar os campos do conhecimento, uma vez que constatamos que seus limites estão ‘borrados’: reconhecemos que eles não são naturais, embora tenham sido desenhados na suposição de sua autonomia.

²⁵ Tradução livre.

Admitimos que eles se entrelaçam, que têm inúmeras interconexões, porém carecemos de ferramentas para enfrentar tais configurações;

b) Os parâmetros dos sistemas de classificação bibliográfica, como o Dewey Decimal Classification e a Classificação Decimal Universal, além de prescreverem formas universais para o tratamento dos documentos, também supõem um usuário universal. As escolhas dos dois sistemas para a organização são etnocêntricas, fato evidente, entre outros, quando reservam poucos números para contemplar a literatura e as línguas não ocidentais. Propostas de sistemas abrangentes, enciclopédicos, são inexequíveis frente às inter-relações entre os campos, a diversidade de pontos de vista e de públicos;

c) Os objetos a organizar são complexos e, à exceção dos objetos concretos – móveis, computadores, rochas etc. – não são convenientemente tratados em sistemas baseados em características autoexcludentes, como é o caso dos objetos teóricos e dos fenômenos comportamentais das Humanidades e das Ciências Sociais. Os próprios princípios de interdisciplinaridade, transdisciplinaridade, como também de interculturalidade já enunciam uma configuração complexa;

d) Para operacionalizar a rede entre conceitos, ainda nos mantemos presos à cadeia. À cadeia corresponde o processo dedutivo, que tem na implicação (se, então) seu motor de funcionamento, derivando daí operações de encaixe em classes e formação de hierarquias. Nesse processo, parte-se de uma definição do tipo aristotélico – gênero próximo, diferença específica – subsumando as espécies ao termo superordenado. A hierarquia, assim, responde às essências e às classes e dá origem a itens que se excluem mutuamente. Porém, as essências “não podem ser atualizadas para adequar-se a uma nova ou diferente situação” (OLSON, 2012, p.109-110), sendo imunes ao contexto. As essências, aplicadas a grupos de pessoas, podem torná-los estereotipados (2012). A abordagem essencialista naturaliza o convencional. Além disso, “Toda lógica de árvore é uma lógica do decalque e da reprodução” (DELEUZE; GUATTARI, 2007, p.21);

e) A nova norma de construção de tesouros e vocabulários (ISO 25964-1/2, 2011/2012) ainda recorre à observação dos princípios de divisão a partir de características e essências. Essa lógica, embora adequada para contemplar objetos concretos, mostra-se problemática para representar, por exemplo, ‘objetos fronteiriços’, denominação utilizada para falar daqueles objetos que podem ser aproximados, mas fogem à caracterização pelas essências por (BOWKER & STAR, 1999). Que outras lógicas poderiam ser adequadas para falar de objetos de uma realidade complexa?

f) A principal motivação da norma ISO 25964-1/2 (2011/2012) para a construção de vocabulários é a interoperabilidade entre vocabulários na web semântica. Buscar atender a esse objetivo é, de fato, muito importante, à medida em que se oportuniza a oferta de vocabulários para a consulta e o reuso. No entanto, resta resolver os problemas de incompatibilidades entre vocabulários, não só pelos seus diferentes níveis de granularidade, mas principalmente pelas distintas opções de partida. Além disso, a simples codificação dos vocabulários não garante, por si só, bons resultados para sua reutilização.

4 Reflexões e possíveis encaminhamentos

Ao lado dos problemas, há, também, iniciativas e reflexões importantes.

a) A parcelização do conhecimento, pressuposto moderno, é substituída pela organização “ao redor de temas que em dado momento são adotados por comunidades interpretativas concretas como projetos de vida locais (...) A fragmentação pós-moderna não é disciplinar e sim temática” (SANTOS, 1988). O objeto se amplia mesmo recorrendo à árvore “pelo alastramento das raízes em busca de novas e mais variadas interfaces” (1988);

b) Os arranjos passam a ser mais flexíveis, observando a linguagem das comunidades interpretativas (expressas nos discursos escritos ou orais). O recurso às metodologias da Terminologia passam a considerar os termos nos contextos de uso, mesclando procedimentos semasiológicos e onomasiológicos de modo a contemplar, também, a variação de definições e conseqüentemente, variações conceituais e de designação. Essa flexibilidade pode ser apresentada como leque de opções;

c) Numa rede, os nós, interligados por diferentes tipos de relações, se aproximam de um rizoma, abrindo a possibilidade de conexões entre quaisquer pontos, sem posições hierárquicas, sem início ou fim, podendo ser rompidas em qualquer lugar para retomar seu percurso segundo outras linhas (DELEUZE & GUATTARI, 1995). Mais do que a hierarquia, o rizoma se presta melhor às representações do real, que não se organiza necessariamente de maneira lógica. “Não se pode fazer coincidir uma ordem pluridimensional (o real) ... com uma ordem unidimensional (a linguagem)” (BARTHES, 1971, p.22). A rede e o rizoma também são mais permeáveis à atualização, permitindo a representação de contextos e circunstâncias de enunciação. No limite, a rede pode recorrer à hierarquia sem lhe dar privilégio. A rede de associação de ideias,

mais do que o ‘encadeamento’ é a base da noção do Thesaurus de Roget que inspirou o tesouro documentário;

d) O mapeamento de vocabulários previsto pela nova norma de tesouros pode ser aproveitado para identificar as diferentes opções de arranjo, o que recoloca objetivos semelhantes às propostas anteriores de compatibilização de vocabulários e leva à possibilidade de escolha entre opções de organização.

5 Considerações finais

“O ideal que a função nomeadora/classificadora se esforça por alcançar é uma espécie de arquivo espaçoso que contém todas as pastas que contêm todos os itens do mundo – mas confina cada pasta e cada item num lugar próprio, separado (...) É a inviabilidade de tal arquivo que torna a ambivalência inevitável. E é a perseverança com que a construção desse arquivo é perseguida que produz um suprimento sempre renovado de ambivalência” (BAUMAN, 1999, P. 10-11).

Dividir para classificar pelas essências exclui a possibilidade de outras derivações. A dicotomia, a classificação binária, reduz artificialmente a complexidade do real, reduzindo também a possibilidade do sentido.

Assim, se quisermos enfrentar a complexidade na Organização e Representação do Conhecimento, temos de buscar lógicas mais adequadas. Temos de aprender a conviver com a ambivalência que, apesar de ser uma desordem da linguagem, é também condição normal da própria linguagem. Temos de conviver com a contradição (GARCÍA GUTIÉRREZ, 2011) e compreender que as perturbações no equilíbrio reconhecidas pela ciência contemporânea não podem ser enfrentadas pela racionalidade. “Nenhuma classificação binária usada na construção da ordem pode se sobrepor inteiramente à experiência contínua e essencialmente não discreta da realidade” (BAUMAN, 1999, p.70). Marcas do momento contemporâneo, a ambivalência e a instabilidade não podem ser apagadas ou reduzidas sob o olhar dos nossos instrumentos de trabalho.

Referências

- BARTHES, Roland. *Elementos de semiologia*. São Paulo: Cultrix, 1971.
- BAUMAN, Zygmunt. *Modernidade e ambivalência*. Rio de Janeiro: Jorge Zahar Ed., 1999.
- BOWKER, Geoffrey C. & STAR, Susan Leigh. *Sorting things out: classification and its consequences*. Cambridge: MIT Press, 1999 (Introduction, cap.1-2, 9-10). Disponível em: <http://epl.scu.edu:16080/~gbowker/classification/> . Acesso em: 12 abril 2012.
- CANCLINI, Néstor García. *Diferentes, desiguais e desconectados: mapas da interculturalidade*. 2.ed. Rio de Janeiro: Ed. UFRJ, 2007.
- DELEUZE, Gilles & GUATTARI, Félix. Introdução: rizoma. In: _____. *Mil platôs: capitalismo e esquizofrenia*, v.1. São Paulo: Ed. 34, 1995. p. 11-37.
- GARCÍA GUTIÉRREZ, Antonio. *Pensar en la transcultura*. Madrid: Plaza y Valdés, 2011.
- HARVEY, David. *Condição pós-moderna: uma pesquisa sobre as origens da mudança cultural*. São Paulo: Loyola, 2002.
- HOLMES, Brian. Investigaciones extradisciplinares: hacia una nueva crítica de las instituciones. In: PRODUCCIÓN cultural y prácticas instituyentes: líneas de ruptura en la crítica institucional. Madrid: Traficantes de Sueños, 2008, p.203-215.
- INTERNATIONAL STANDARD ORGANIZATION. *ISO 25964: thesauri and interoperability with other vocabularies*. Part 1: Thesauri for information retrieval. Geneve: International Standard Organization, 2011.
- INTERNATIONAL STANDARD ORGANIZATION. *ISO 25964: thesauri and interoperability with other vocabularies*. Part 2: Interoperability with other vocabularies. Geneve: International Standard Organization, 2012.
- OLSON, Hope A. Distilling essence, enforcing Shibboleth. In: GUIMARÃES, J. A. C.; DODEBEI, V., orgs. *Desafios e perspectivas científicas para a representação do conhecimento na atualidade*. Marília: ISKO-BRASIL; FUNDEPE, 2012.
- SANTOS, Boaventura de Sousa. Um discurso sobre as ciências na transição para uma ciência pós-moderna. *Estudos avançados*, São Paulo, v.2, n.3, maio/ago. 1988. Disponível em: http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0103-40141988000200007 . Acesso em: 03 abr. 2013.

Metáfora e Ortofemismo na Representação de Assunto

Fabio Assis Pinho
Universidade Federal de Pernambuco
fabiopinho@ufpe.br

Suellen Oliveira Milani
Universidade Estadual Paulista
suellenmilani@marilia.unesp.br

Introdução

O indexador, no desenvolvimento de suas atividades profissionais, organiza a informação, *rotulando-a*, o que, no mais das vezes, ocorre permeado por certa subjetividade, que decorre do contexto cultural e ideológico do próprio indexador e do sistema de informação em que se insere. Ao representar assuntos e atribuir nomes ou rótulos ao conteúdo tópico dos documentos, os indexadores constroem substitutos documentais (por exemplo: índices, resumos e notações de classificação) e, para tanto, valem-se daquilo que Olson (2002:4) denomina como *poder de nomear*, voltado a “[...] controlar a representação de assunto e, por tanto, seu acesso”. Dessa forma, tornam-se necessários procedimentos que possam contribuir para minimizar a subjetividade na indexação - mormente no que se refere às possibilidades de danos ao usuário dela advindas - e encontrar garantias para que o rótulo a ser atribuído àquele documento seja técnica e culturalmente adequado.

Um dos aspectos que evidenciaríamos a subjetividade na indexação, como aponta a literatura, são as metáforas na indexação as quais, como destacam Berman (1993), Olson (2002), Guimarães (2006) e López-Huertas (2008), entre outros, podem levar a vieses na representação, envolvendo problemas relativos ao preconceito, proselitismo, questões de gênero, categorizações dicotômicas e ausência de precisão terminológica, sendo que este último é objeto da presente discussão.

Indo além das discussões relativas à precisão como medida de recuperação da informação (Chaumier, 1988, Lancaster, 2003), observa-se que a falta de suficiente precisão terminológica pode disfarçar o preconceito na representação e, conseqüentemente, influenciar negativamente a recuperação da informação. Uma vez que uma linguagem de indexação deveria ser elaborada, dentre outros aspectos, a partir da terminologia específica do domínio a que se destina, tem-se, na contramão, que um termo errônea ou equivocadamente incluído em uma linguagem de indexação pode

levar a uma representação preconceituosa, evidente ou dissimulada. Isso também pode ocorrer no momento da análise de assunto, quando o indexador identifica conceitos no documento a partir dos quais irá consultar a linguagem de indexação.

Pelo fato de os termos selecionados para representar os conceitos extraídos dos documentos possuírem características ortofêmicas, isto é, buscam referenciar expressões diretas ou neutras, pouca atenção tem sido dada, no contexto da indexação, às figuras de linguagens utilizadas por comunidades discursivas específicas. No entanto, e com respaldo nos estudos da Sociolinguística relativamente às inerentes variações da língua, que dela fazem algo não rígido, torna-se necessário levar em conta as relações existentes entre a língua e a condição social do falante, analisando os fatores socioculturais (Preti, 2000), o que leva a que, no presente estudo, se busquem alternativas para o enquadramento das metáforas no âmbito da representação de assunto.

A metáfora e o ortofemismo

A literatura científica de domínios do conhecimento específicos encontra-se permeada por metáforas, a exemplo do que ocorre na linguagem cotidiana, como é o caso das temáticas relativas às mulheres ou à homossexualidade masculina que, por vezes, têm sido representadas de maneira genérica, prejudicando o reconhecimento, em instrumentos e atividades de representação, dessa subcultura e, conseqüentemente, uma representação incorreta.

Metáfora (do grego *metá* que significa trans e *phérein* que quer dizer levar) consiste em uma mudança de sentido, uma transferência ou transposição (Pereira, 1976). As metáforas podem adotar várias formas, dependendo do efeito que se deseja alcançar, do conteúdo que se quer veicular, do tempo disponível, do interlocutor ou do grupo de ouvintes.

Sob o rótulo de metáfora, Aristóteles coloca todo o tipo de figura de transferência de significado, isto é, as sinédoques particularizantes e generalizantes, a metonímia e a metáfora propriamente dita. Contudo, a denominação metafórica também opera nos processos de eufemismo e de disfemismo, uma vez que, sua característica é “[...] a motivação que designa seres ou objetos pouco conhecidos e os associa a outros mais conhecidos que se assemelham pela forma” (Guiraud, 1975: 61).

Eufemismo (do grego *euphemismos* que significa bem dizer) é uma figura de retórica que visa à substituição de um termo desagradável por outro que o amenize. Usa-se o eufemismo para disfarçar os sentimentos desagradáveis, os pensamentos cruéis

ou as palavras consideradas como tabu. De alguma forma, o eufemismo retrata uma forma puritana de expressão que quer fugir da representação da realidade que pode vir a ser cruel ou agradável (Ceia, 2005). Dessa forma, eufemismo é o “[...] modo de dizer ou sugerir de forma dissimulada ou com decoro ideias cuja reta e franca expressão seriam duras ou soariam mal” (Casares, 1959: 372).

Por outro lado, o disfemismo consiste em uma expressão com conotações ofensivas, quer sobre o *denotatum* (objeto ao qual se refere à expressão linguística) ou sobre o público, ou ambos. Segundo Mascarenhas (2005), “*Dysphemia* era para os gregos a palavra de mau agouro, ato de pronunciar palavras de mau agouro, palavras más, de desgraça. Considera-se disfemismo o uso de palavras ou expressões de caráter rude, repugnante, desagradável, agressivo ou horrível”.

Chamizo Domínguez (2004: 45) explica que uma determinada palavra será entendida como um eufemismo a partir do momento em que o contexto permita isso, não dependendo da palavra em si, mas do contexto e do uso que se faz dela através das intenções dos emissores. Estende-se essa explicação para os disfemismos, uma vez que a fronteira entre tais conceitos é tênue. Do ponto de vista sincrônico, tal fator implica dizer que um eufemismo não pode ser utilizado em qualquer outro contexto e continuar surtindo os mesmos efeitos cognitivos, estilísticos e sociais (Chamizo Domínguez, 2004).

Considerando que qualquer comunidade discursiva pode fazer uso da metáfora enquanto um recurso linguístico, Orrico (2001) discute que o uso da figura de linguagem pode aumentar a precisão das respostas de uma demanda específica por informação em bibliotecas. Para tanto, ela propõe um metafiltro (Orrico, 2001: 125) cujo eixo norteador é o semântico, pois entende que o ser humano representa a si mesmo e o mundo usando metáforas. Com isso, a autora baseia-se em duas esferas fundamentais que é a do discurso e a da identidade, ou seja, é no discurso e através das figuras de linguagem que o grupo pode construir sua identidade. A metáfora pode, então, estabelecer relações de significação.

Os discursos são permeados por eufemismos e disfemismos, os quais foram apropriados pelos indivíduos de um grupo para descaracterizar palavras que, então, assumiriam novos significados. Essa apropriação de termos significa o entendimento de que o grupo é marginalizado e esse fator é combatido com uma nova interpretação do termo. Ignorar isso, seria ignorar a própria formação de identidade desse grupo. Por

isso, eufemismos e disfemismos devem ser considerados pelos bibliotecários no momento da indexação.

Já o ortofemismo é o processo pelo qual as palavras atuam em seu sentido mais estrito, o que é relativamente comum na maioria das esferas sociais. Allan e Burrige (2006:2) criaram o termo ortofemismo (grego *ortho* - 'adequado, direto, normal', conforme ortodoxo) “[...] com o objetivo de referenciar expressões diretas ou neutras que não são delicadas, evasivas ou demasiadamente educadas (eufemística) nem ásperas, contundentes ou ofensivas”.

Ortofemismo pode ser entendido simplesmente como expressões diretas. Esse conceito traz à baila os conceitos de adequado, apropriado e normal aos atos linguísticos. Tanto o ortofemismo quanto o eufemismo referem-se a palavras ou frases usadas como alternativas a expressões não preferidas ou desfavoráveis. Essas palavras ou frases são entendidas como sendo mais apropriadas em contraste ao disfemismo, que pode ser considerado como tabu.

Assim, o ortofemismo é mais direto, formal e literal do que os seus correspondentes eufemísticos, os quais são mais coloquiais, figurativos e indiretos. Um simples exemplo sobre as metáforas aqui discutidas seria relacionado ao assunto tabu da morte, amparado pela expressão eufemística *passar dessa para melhor* e disfêmica *bater as botas*. O equivalente ortofêmico à essas metáforas seria *morrer*.

Conclusão

As figuras de linguagem, no âmbito de determinadas comunidades discursivas ou domínios do conhecimento, são importantes e podem desempenhar papel decisivo - seja positiva ou negativamente - na representação de assunto. Por isso, o que propomos, em última instância, é que as representações oriundas de figuras de linguagem sobre determinada área do conhecimento, como forma de o ser humano identificar o mundo, obedeçam a algumas regras estabelecidas pela comunidade que com ela lida. Essas representações, portanto, 'sujeitam-se' a regularidades que podem servir para organizar os conceitos da área (Orrico, 2001).

O uso das metáforas é impulsionado por várias razões, como por exemplo, o desejo de não ofender, a necessidade de demonstrar uma identidade de grupo ou, ainda, para dissimular a realidade e, dessa maneira, evitar consequências danosas. O ortofemismo, por sua vez, consiste em uma palavra ou expressão usada como alternativa a expressões não preferidas.

Uma vez que “[...] o que em dada época é entendido como eufemístico, pode mais tarde ser considerado disfemístico” (Mascarenhas, 2005) e vice-versa, a presença de termos eufemísticos e ortofêmicos na prática da representação de assunto apresentaria indícios de como determinadas comunidades discursivas estão organizadas, inclusive encorajando a sua preservação cultural. É importante ressaltar que se os eufemismos e os disfemismos podem conter preconceitos, os ortofemismos podem conter indícios excessivos do politicamente correto.

No âmbito da representação de assunto, o uso de eufemismos e ortofemismos podem auxiliar a recuperação da informação de necessidades mais específicas, conferindo mais garantia cultural ao sistema nos moldes propostos por Beghtol (2002).

Referências

Allan, Keith e Burridge, Kate. 2006. *Forbidden words: taboo and the censoring of language*. New York: Cambridge University Press.

Beghtol, Clare. 2002. A proposed ethical warrant for global knowledge representation and organization systems. *Journal of Documentation* 58n5: 507–32.

Berman, Sanford. 1993. *Prejudice and antipathies: a tract on the LC subject heads concerning people*. Jefferson: McFarland & Company Inc. Publishers.

Casares, Julio. 1959. *Diccionario ideológico de la lengua española: desde la idea a la palabra; desde la palabra a la idea*. 2. ed. Barcelona: Ed. Gustavo Gili.

Ceia, Carlos. 2005. Eufemismo. In: CEIA, C. (Org.). *E-dicionário de termos literários*. Disponível em: <<http://www2.fcsh.unl.pt/edtl/verbetes/E/eufemismo.htm>>. Acesso em: 02 maio 2008.

Chamizo Domínguez, Pedro J. 2004. La función social y cognitiva del eufemismo y del disfemismo. *Panace@* 5n15: 45-51.

Chaumier, Jacques. 1988. Indexação: conceitos, etapas e instrumentos. *Revista Brasileira Biblioteconomia e Documentação* 21n1/2: 63-79.

Guiraud, Pierre. 1975. *A semântica*. São Paulo: Difel.

Lancaster, Frederick Wilfrid. 2004. *Indexação e resumos: teoria e prática*. 2. ed. rev. atual. Brasília: Briquet de Lemos.

López-Huertas, Maria José. 2008. Some current research questions in the field of knowledge organization. *Knowledge Organization* 35n2/3: 113-136.

Mascarenhas, Isabel B. Disfemismo. In: CEIA, C. (Org.). *E-dicionário de termos literários*. 2005. Disponível em: <<http://www2.fcsh.unl.pt/edtl/verbetes/E/eufemismo.htm>>. Acesso em: 02 maio 2009.

Orrico, Evelyn Goyannes Dill. 2001. *Binômio Linguística-Ciência da Informação: abordagem teórica para elaboração de metafiltro de recuperação da informação*. Doctoral dissertation, Instituto Brasileiro de Informação em Ciência e Tecnologia / Universidade Federal do Rio de Janeiro, Rio de Janeiro, Brazil.

Olson, Hope A. 2002. *The power to name: locating the limits of subject representation in libraries*. Dordrecht: Kluwer Academic Publisher.

Pereira, Isidro. 1976. *Dicionário grego-português e vice-versa*. Porto: Livraria apostolado da Imprensa.

Preti, Dino. 2000. *Sociolinguística: os níveis da fala: um estudo sociolinguístico do diálogo na literatura*. São Paulo: EDUSP.

Organização e Representação do Conhecimento na Web: Desafios para a construção colaborativa de uma ontologia do samba

Marcos Miranda
Universidade Federal do Estado do Rio de Janeiro
mlmiranda@unirio.br

Jair Miranda
Universidade Federal do Estado do Rio de Janeiro
jairmm@criar.art.br

1 Introdução

Este trabalho é fruto do diálogo entre os projetos de pesquisa “Genealogia do Samba” e “Preservação da Memória do Samba Carioca: Escolas de Samba, sua história sua gente” vinculados à Universidade Federal do Estado do Rio de Janeiro – UNIRIO.

Para os estudos da representação, conceituação, classificação e o registro sobre a história e memória do samba, lançamos mão de um fato muito marcante na história do tráfico atlântico de africanos para se tornarem escravos: o da árvore do esquecimento.

Conta-se que os escravos antes de embarcar nos tumbeiros na cidade de Ajudá, em Benin (antigo Daomé), para a viagem de destino ao Brasil, eram obrigados pelos traficantes a caminharem várias vezes em círculos, em torno de uma árvore que tinha o poder de apagar todo o seu passado, a sua memória e toda a sua resistência cultural (Araújo, 2009: 127-128) Esse ritual, que certamente não foi de todo eficiente, tinha o objetivo de minimizar as perdas financeiras dos escravagistas pelo alto grau de mortes de escravos no Brasil, causadas pelo banzo, uma doença decorrente de uma tristeza profunda, acometida aos escravos pela saudade da terra de origem, uma depressão aguda originada das condições adversas do novo *habitat*.

Essa história, vista a partir deste trabalho, que busca rememorar o samba para representá-lo, nos remete imediatamente à significação que tem as árvores na cosmovisão africana, ou seja: à da própria noção de tempo e ancestralidade, muito presente no Inquice Tempo (da Nação Angola) ou no Orixá Iroco (da Nação Gege/Nagô), aquela sacraliza a árvore como uma entidade de memória, como uma guardiã de lembranças, de testemunho de um tempo, ao mesmo tempo passado e futuro, que, diferentemente da noção linear do mito ocidental “Cronos”, indicam um ciclo

permanente que une a vanguarda à velha guarda, um olhar sempre contemporâneo sobre a tradição e a memória.

Portanto, essas referências iniciais, ao contrário da ideologia subjacente naquela árvore do esquecimento, provavelmente a milenar Baobá, serve como preâmbulo para fundamentar histórica e filosoficamente este texto, ao sugerir simbolicamente uma árvore da lembrança para essa tão significativa manifestação da cultura afro-brasileira, na sua transposição para a árvore baniana de Ranganathan.

As dificuldades encontradas em diversas unidades de informação no Rio de Janeiro para a representação do conhecimento do samba nos despertou o interesse em discutirmos as possibilidades de configurações do etnoconhecimento para além das estruturas arborescentes.

A partir dessas constatações, empreendemos este estudo para investigar o etnoconhecimento e sua representação em sistemas de organização do conhecimento no universo do Samba.

Neste sentido, buscamos analisar a representação do conhecimento afrodescendente, especificamente do samba, nos sistemas de organização do conhecimento; possibilitar novas formas de representar a partir do conhecimento dos saberes milenares das culturas afrodescendentes; estimular a criação de um fórum permanente congregando os atores do processo de construção de sistemas de organização do conhecimento (afrodescendentes, profissionais da informação e classificacionistas); contribuir para a organização do conhecimento no Portal do Carnaval e superar preconceitos e discriminações nos sistemas de organização do conhecimento.

2 Os desafios na configuração do Samba para fins de organização do conhecimento na Web

A fundamentação teórica deste estudo reflete conceitos oriundos da Teoria da Classificação Facetada, da Teoria do Conceito, dos aspectos éticos da Organização do Conhecimento, da tese sobre Garantia e Hospitalidade Culturais, da Epistemografia Interativa e da Teoria do Rizoma conforme estudos de Miranda (2011: 153).

Considerando o universo do conhecimento e sua dinâmica, a Teoria da Classificação Facetada apresenta os elementos de uma estrutura que abrange a unidade classificatória, as categorias, as facetas, as cadeias e os renques. Os princípios da Teoria do Conceito, por outro lado, nos permitem identificar qualquer objeto no universo

empírico (referente), atribuindo-lhe um conjunto de características que visam construir enunciados verdadeiros acerca de tal objeto. Este será posteriormente nomeado, chegando assim à definição de conceitos. Este modelo de formação possibilita também a identificação de semelhanças e diferenças em relação a outros objetos, o que permite estabelecer as relações conceituais existentes.

Em se tratando de uma pesquisa sobre o samba, as primeiras questões que nos desafia respostas, são de ordem ontológica, ou seja: afinal, o que é samba? Um gênero musical, uma dança ou uma festa? Um desfile de carnaval, um espetáculo musical ou uma manifestação cultural? Ele seria o mesmo no Recôncavo Baiano, em Salvador, Recife, Manaus, no Rio de Janeiro ou em Tóquio e Berlin? Quais e quantas são as espécies ou sotaques do samba? samba-reggae é samba? bossa nova é samba ? afro-samba é samba ? Jongo é samba ? Côco é samba ? Cabula é samba? São questões complexas que antecedem qualquer tentativa de conceituar, categorizar, classificar ou designar uma música, uma performance ou seus registros documentais como samba.

No âmbito desta pesquisa, nossa hipótese, no entanto, é de que o samba, muito mais que um gênero musical e uma matriz cultural, se constitui, também, como nos diz Ligiéro (2011: 25), como uma força motriz, presente nas muitas manifestações culturais afro-brasileiras que integram o canto, a dança e a música na sua performance, uma vez que cantar, dançar e batucar nas culturas de origem africana são indissociáveis. Como legado dessa força e da ancestralidade africana, o samba é também visto aqui como uma árvore frondosa de muitos ramos, mas de raiz forte que se movimenta subterrânea e subliminarmente em rede como um rizoma, fazendo nascer, sob uma ótica "deleuziana", outras árvores da mesma família em vários pontos da terra como uma árvore baniana. Deleuze&Guattari (1980). O samba também é entendido aqui, contemporaneamente, como uma potência cultural, afetada pelo sentimento positivo da alegria, como nos indica Spinoza (2003); é, ao mesmo tempo, um capital cultural, como nos fala Bourdier (1982), que agrega valor à própria marca Brasil; como também, é uma forma de vida extremamente sedutora, uma ocupação virtuosa e um trabalho imaterial prazeroso Virno (2009), que arrebatava multidões e constitui um império, segundo os conceitos "negrianos" de multidão e império. Hardt & Negri (2003, 2005); além de se constituir como o maior e mais longo exemplo de movimento de resistência biopolítica no Brasil. Foucault (1988), já que para lutar contra uma morte anunciada do corpo, da alma e da memória, toda uma legião de escravos desvalidos optou por celebrar a vida cantando e dançando como estratégia de sobrevivência, resistência cultural e exercício de fé, uma

quase religião, já que nascido nos rituais sacros dos terreiros de Batuques e Candomblés, como observado por Caciatore (1977), Sodré (1998), D'Ávila (1982), Lopes (2005), Spirito Santo (2011) e Ligiéro (2011).

O fato, porém, é que, independente das nossas hipóteses, todas as teses e conceitos já produzidos sobre o samba, suas várias manifestações e sua vasta produção, ainda é cercada de contradições, desconhecimentos e alguns “propositais esquecimentos”.

Para contrapor, então, essas teses com nossas hipóteses e responder minimamente a questão: o que é samba, é que nos propusemos nesta pesquisa a criação de uma “Ontologia” para o Samba”, como um recurso metodológico extraído do campo de estudo das ciências da informação e da computação, que à semelhança das taxonomias, vocabulários controlados e tesouros, visa organizar, representar, compreender e compartilhar o conhecimento de um dado domínio, como nos aponta Silva (2009) ao citar Jurisica, Mylopoulos e Yu (1999):

as ontologias podem ser usadas como conhecimento comum de um domínio, viabilizando a comunicação entre uma comunidade de interesse... na perspectiva da ciência da computação e da ciência da informação, uma ontologia pode ser útil na organização e representação de conhecimento, tendo a tecnologia como apoio na viabilização de uma infra-estrutura para gerência de conhecimento (Silva, 2008:)

255

Considerando que a gerência desse conhecimento comum sobre o domínio “Samba”, especialmente na web, deve ser baseada nessa infra-estrutura tecnológica das redes sociais, o grande desafio desta pesquisa, no nosso entender, deve-se à prevista dificuldade para se estabelecer um consenso nas comunidades de interesse sobre o samba, em relação ao que viria a ser a sua definição, sua história, suas vertentes, seus sotaques, seus conceitos e as designações de elementos de seu repertório em determinadas categorias, uma vez que essas comunidades, especialmente aquelas formadas por musicólogos, antropólogos, etnomusicólogos, cronistas, jornalistas e outros estudiosos do samba, nem sempre ratificam as conceituações e categorizações usadas no cotidiano por aqueles que praticam, produzem e reproduzem o samba nas suas variadas formas. Dessa forma, entendemos que a organização e representação do conhecimento do domínio “Samba”, através do simbolismo de uma árvore (de memória e lembrança), pode facilitar a compreensão e participação dessas comunidades.

3 Metodologia

Para alcançarmos os objetivos aqui propostos procuramos observar, descrever, analisar e sintetizar a forma como a comunidade do samba com diferentes culturas identificam e reconhecem os fenômenos que ocorrem em suas realidades nos seus entrelaçamentos culturais e historicamente construídos e de que maneira, a partir desses conhecimentos, e verificar suas representações na LCSH, na CDD e na CDU.

Dessa forma, realizamos a análise dos sistemas de organização do conhecimento (SOC), relativo: às classes e categorias de assuntos que sob algum aspecto representem o conhecimento do samba; identificamos e extraímos segmentos dos SOC que representassem o conhecimento do samba e um levantamento de pressupostos e fundamentos teóricos inerentes ao samba, para tornar patente a multiplicidade étnica e a pluralidade cultural.

Em seguida, efetivamos buscas de informação no ambiente Web utilizando o termo samba, com o seguinte resultado: 136 milhões de itens no Google, 240 mil no Google Books, 46,5 mil no Google Vídeos e 29,5 mil itens no Youtube.

Posteriormente, trabalhamos uma amostra aleatória de 44 termos, cujos conceitos foram definidos e organizados: Afoxé. Agueré. Alujá. Barravento. Batucada. Batuque de Umbigada. Bossa-nova. Cabula. Calango. Carimbó. Congo de Ouro. Ijexá. Jongo. Lundu. Maracatu. Maxixe. Nimbu. Opanijé. Pagode. Sambalada. Samba-canção. Samba-choro. Samba-chula. Samba-enredo. Samba-funk. Samba de breque. Samba de bumbo. Samba de carnaval. Samba de côco. Samba de embolada . Samba de gafieira. Samba de morro. Samba de partido-alto. Samba de Roda. Samba de Terreiro. Samba-Jazz. Samba-Lenço. Sambalanço. Samba Raiado. Samba-Reggae. Samba-Rock. Sambolero. Tambor de Crioula. Tambor de Mina.

A sistematização dos conceitos se deu considerando quatro grandes categorias Canto, Dança, Festa e Ritmo. Porém, para dar conta das relações conceituais que não as arborescentes configuramos os conceitos de maneira rizomática.

Em decorrência, propusemos também a utilização do editor de ontologias “Protégé” e o repositório digital ICA-AtoM, tendo em vista que o editor de ontologias Protégé é uma ferramenta recomendada pelo consórcio W3C, para atender ao objetivo de construção coletiva de um conhecimento comum compartilhado num determinado domínio, além de ser uma plataforma livre de código aberto, mundialmente conhecido e; que o ICA-AtoM (acrônimo de Access to Memory) é um sistema de repositório

digital para submissão, organização, descrição, acesso e recuperação de registros e documentos, credenciado pelo Conselho Internacional de Arquivos (ICA), baseado na Internet, *open source*, multiusuário, multirepositório e compatível com os novos padrões internacionais de armazenagem, intercâmbio e descrição de acervos.

4 Considerações finais

Uma análise preliminar das buscas realizadas na Web sobre o samba nos leva a constatar que esse enorme acervo de textos, livros, imagens, sons e vídeos e, esse vasto conhecimento acumulado sobre o samba, não está devidamente organizado de forma que os conteúdos do e sobre o que é samba possam ser utilizados: precisamos de uma web semântica. O uso de uma ontologia facilitaria o fornecimento de respostas mais apuradas.

Por outro lado, a observação do grau em que esses conhecimentos estão sub-representados nos SOC analisados nos permitiu entrever o caráter de urgência, mas também a viabilidade, de construir sistemas de organização do conhecimento inferências, tais como as ontologias, que assegurem a continuidade das tradições do samba.

Observamos, entretanto que para isso torna-se indispensável que os profissionais da informação trabalhem em conjunto com as comunidades e as culturas representadas, procurando desmistificar mitos raciais e étnicos e também conhecer mais de perto as experiências riquíssimas e milenares que essas comunidades preservam oralmente.

Concluimos que a organização do etnoconhecimento deva ser realizada por meio de sistemas de organização do conhecimento colaborativos construídos por metodologia participativa, promovendo uma nova ética para a representação do conhecimento de grupos em desvantagem.

Referências Bibliográficas

ALVARENGA, O. 1946. *A Influência negra na música brasileira*. Boletim Latino-Americano de música, VI.

ALVES, Henrique. 1976. *Sua Ex.^a o samba*. São Paulo. Símbolo.

ARAÚJO, Ana Lucia. 2009. *Caminhos atlânticos memória, patrimônio e representações da escravidão na Rota dos Escravos*. *Varia hist.* [online], vol.25, n.41, pp. 129-148.

BASTIDE, R. *As Religiões Africanas no Brasil*. Vols. I e II. S.Paulo: Pioneira-USP, 1971.

BOURDIEU, P.; PASSERON, J. C. 1982. *A economia das trocas simbólicas*. São Paulo: Perspectiva.

CACCIATORE, Olga G. 1977. *Dicionário de Cultos Afro-Brasileiros*. Rio de Janeiro: Ed. Forense Universitária; S.E.E.C.

D'ÁVILA, N. R. 1982. *Le Samba et Percussions: batucada Brésilienne*. Paris: Courmontagne.

DIAS, Paulo. 2011. Diásporas musicais africanas no Brasil. Disponível em: http://www.cachuera.org.br/cachuerav02/index.php?option=com_content&view=article&id=297:diaporasmusicaisafricanasnobrasil&catid=80:escritos&Itemid=89 Acessado em 07/ 2011

FOUCAULT, Michel. 1988. *História da Sexualidade: a vontade de saber*. Rio de Janeiro: Graal.

JURISICA, I.; MYLOPOULOS, J.; YU, E. 1999. Using Ontologies for Knowledge Management: an Information Systems Perspective. Annual Conference of the American Society for Information Science, Washington, D.C.,

LAZZARATO, Maurizio, NEGRI, Antonio. 2001. Trabalho imaterial. Rio de Janeiro: DP&A,

LIGIÉRO, Zeca. 2011. *Corpo a corpo: estudo das performances brasileira*. Rio de Janeiro: Garamond.

LOPES, Nei. 2005. A presença africana na música popular brasileira. *Revista Espaço Acadêmico* nº 50, Universidade Federal de Uberlândia, julho de

MIRANDA, M.L.C. de; PARANHOS, J.P.B.; OLIVEIRA, J.X. DE; PAES, M.S. A organização e representação do conhecimento em religiões yorubanas na Library of Congress Subject Heading. In: GUIMARÃES, J.A.C; DODEBEL, V. (Orgs.). *Desafios e perspectivas científicas para a organização e representação do conhecimento na atualidade*. [recurso eletrônico]. Marília: ISKO-Brasil : FUNDEPE, 2012. 285p. p. 153-159. e-Book

258

MORAES, Eneida Moraes. 1987. *História do carnaval carioca*. Rio de Janeiro, Record.

MUKUMA, K. W. A. *Contribuição Bantu na Música Popular Brasileira*. São Paulo: Global.

MUNIZ JR., J. 1976. *Do Batuque a Escola de Samba: subsídios para a história do samba*. São Paulo: Símbolo.

RODRIGUES, Nina. 1977. *Os Africanos no Brasil*. 5. ed, São Paulo: Companhia Ed. Nacional. (Brasiliana, v. 9)

SANDRONI, Carlos. 2001. *Feitiço decente: transformações do samba no Rio de Janeiro (1917-1933)*. Rio de Janeiro: J. Zahar : UFRJ.

SANTOS, Regina Maria Meirelles. [S.d.]. *Matrizes Africanas no Samba*.

SILVA, Daniela Lucas da. 2008. Uma proposta metodológica para construção de ontologias: uma perspectiva interdisciplinar entre as Ciências da Informação e da Computação. 286 f. Dissertação (Mestrado em Ciência da Informação) - Escola da Ciência da Informação, Universidade Federal de Minas Gerais, Belo Horizonte.

SODRÉ, M. 1998. *Samba: o dono do corpo*. Rio de Janeiro: Mauad.

SPINOZA, B. 2003. *Ética: demonstrada à maneira dos geômetras*. São Paulo: Afiliada.

SPIRITO SANTO. 2011. *Do Samba ao Funk do Jorjão: ritmos, mitos, e ledos enganos no enredo de um samba chamado Brasil*. Petrópolis: KRB.

TINHORÃO, J. R. 1972. *Música Popular de índios, negros e mestiços*. Petrópolis: Vozes.

Democracia digital e as suas implicações para a organização participativa do conhecimento

Míriam Gontijo de Moraes
Universidade Federal do Estado do Rio de Janeiro
miriam.gontijo.moraes@gmail.com

Introdução

A abordagem participativa na construção de políticas públicas é uma tendência em sociedades que se caracterizam pela complexidade e fragmentação decorrentes, conforme aponta Wersig (1993), do uso diferenciado das Novas Tecnologias de Informação e Comunicação (TICs). Este uso possibilitou a produção de um conhecimento cada vez mais complexo, sofisticado e, principalmente, fragmentado.

Como pontua Wersig (1993), o conhecimento se torna fragmentado em razão da impossibilidade humana de acumulação de todo o conhecimento disponível, dependendo cada vez mais do uso das TICs. No entanto, situamos para efeito desta reflexão, o conceito de computabilidade como a capacidade de realização e solução de problemas a partir do uso de computadores. A noção de computabilidade está relacionada a dois importantes aspectos. O primeiro é consequência do trabalho de Kurt Gödel (o Teorema da Incompletude) e consiste na constatação de que nem todo sistema formal é completo a ponto de resolver problemas engendrados por ele. Entendendo a computação como um sistema formal, constata-se, a partir do Teorema de Gödel, que esta não possa responder todas as questões colocadas por ela, como por exemplo, a exclusão digital. O segundo aspecto, abordado por Alan Turing (1936), aponta a impossibilidade de caracterizar a classe de problemas computáveis, aspecto este que reforça a impossibilidade das TICs resolverem os problemas engendrados por elas.

Nesta perspectiva, enfrentamos o paradoxo de se trabalhar com os referenciais de Democracia e a interface excludente das TICs. A exclusão digital, citada anteriormente, é um problema ensejado pelo próprio desenvolvimento tecnológico implicando em uma sofisticação que cada vez mais exclui a grande maioria como apontado por Gontijo (2005). Este paradoxo impacta cada vez mais a gestão de políticas públicas de informação pois esta realidade torna impossível pensar estratégias de governo simplesmente dentro de um único sistema formal, reforçando o

aspecto levantado por Turing (1936).

No âmbito do Projeto Ágora de Democracia Digital, em andamento no Programa de Pós-Graduação em Informática da Unirio, financiado pelo CNPq (Edital CT-INFO), foi delineado um modelo de apoio à participação democrática baseado em três aspectos principais: a) colaboração entre os participantes do processo (Araújo, 1997); b) memória do processo de discussão e deliberação (Conklin, 1996); e, c) transparência necessária tanto às informações envolvidas no processo quanto às ações e decisões tomadas ao longo de sua execução (Vattimo, 1992).

Conforme o projeto, esta discussão vem utilizando o termo "Democracia Digital" (Hague, 1999), (Wilhelm, 2000:192) com o sentido de, segundo avaliação apresentada por Araújo (2011), discutir a dimensão de relações possíveis entre Cidadão e Sistemas de Governo através de TICs. Esta relação envolve tanto os aspectos de administração pública e prestação de serviços ao cidadão, comumente conhecido como e-governo ou governo eletrônico, quanto os aspectos de participação social nas decisões políticas e governamentais, comumente conhecidos como e-democracia ou democracia digital (Hindman, 2009:160).

Ainda no escopo do Projeto Ágora, pretende-se investigar como a geração de conhecimento de maneira participativa, voluntária e altamente conectada abre a discussão sobre as condições relacionadas ao patrimônio no ambiente virtual (Dodebei, 2006), (Gondar e Dodebei, 2005 :160), (Dodebei e Gouveia, 2008). Nesta perspectiva, a memória, a partir de registros digitais, pode ser patrimonializada para a ampliação da cidadania. Quais seriam os critérios e os métodos para se atingir este patamar que, apesar de parecer secundário, é um tema fundamental para a democracia digital possível?

No âmbito do projeto, os contornos do conceito de Patrimônio Digital apontam para as seguintes definições :

- Nascido digital ou posteriormente digitalizado, o patrimônio deve ser apreendido como um objeto e também como um valor de informação sobre o objeto, seja de natureza material ou imaterial;
- O ambiente digital possibilita a renovação participativa da documentação e registro que vão compor o patrimônio digital;
- O ambiente digital proporciona, por meio de comunidades de prática, a negociação constante, através da qual se mantém um caráter dinâmico do que se pretende patrimonializar, e não mera celebração da sobrevivência do passado.

Objetivo

O caminho apontado pela ideia de uma Democracia Digital, conforme o Projeto Ágora (2009), nos incita a investigar como a memória, a partir de registros digitais, pode ser patrimonializada para a ampliação da cidadania. Por se tratar de uma proposta que parte de um problema no bojo do desenvolvimento tecnológico, cujo enfoque transdisciplinar é o mais adequado, pretende-se produzir conhecimento em âmbito transdisciplinar a partir da identificação de memórias de interação e de registros digitais em comunidades de práticas.

Comunidades de práticas e a patrimonialização do conhecimento no ambiente digital

A proposta de trabalhar com o conceito de comunidades de prática é oportuna. Uma primeira noção sobre Comunidade de Prática designa um grupo de pessoas que se unem em torno de um mesmo tópico ou interesse. Essas pessoas trabalham juntas para achar meios de melhorar o que fazem, ou seja, na resolução de um problema na comunidade ou no aprendizado diário, através da interação regular. O conceito de comunidades de prática é utilizado no estudo do capital social nas empresas.

Segundo Terra (2005), o conceito de *Comunidades de Prática - CdPs* é essencial na era do conhecimento. O conhecimento humano não faz sentido sem o contexto de comunidades., e evolui a partir do reconhecimento e validação de nossos pares nas comunidades das quais participamos.

Avançando na constituição do conceito como uma categoria de análise na perspectiva tanto dos estudos sociais como da cognição, Jean Lave e Etienne Wenger (1991) apontam para vários aspectos que nos interessam. O primeiro diz respeito ao fato de que esta categoria pode ser usada para descrever como atividades locais são organizadas e se interligam no sistema social mais amplo, ou como é que ao constituirmo-nos como membros de comunidades de prática acabamos por negociar e experienciar o significado de pertença a organizações mais amplas . No entanto, estes mesmos autores admitem que “o termo comunidade não implica necessariamente co-presença, um grupo bem identificável, ou fronteiras socialmente visíveis” (LAVE;

WENGER, 1991,p. 98), apresentando-se então como uma categoria que pode ser usada para a análise de fenômenos em rede.

Um outro aspecto que vai nos interessar, é quanto a identificação de três dimensões que caracterizam as comunidades de prática, a partir da análise de vários estudos etnográficos feita por Wenger (1998). As três dimensões inter-relacionam-se e ao pensar cada uma delas é necessário ter presente a interação com as outras. São elas :empenhamento mútuo (mutual engagement); empreendimento conjunto (joint enterprise); repertório partilhado (shared repertoire).

A constituição de Comunidades de Prática para fins do projeto deve se orientar por estas três dimensões com destaque para a dimensão do repertório compartilhado, em que os participantes desenvolvem um capital simbólico que além de garantir uma coerência da comunidade, vai de encontro à operacionalização da proposta de patrimonialização digital.

O procedimento metodológico aqui previsto se inscreve no tipo pesquisa-ação: quando os pesquisadores e os participantes envolvem-se no trabalho de pesquisa de modo participativo ou cooperativo, interagindo em função de um resultado esperado, e as CdPs são o instrumento.

Neste sentido, o trabalho com a figura das comunidades de prática se dará da seguinte maneira:

- a) Identificação de registros passíveis de ser considerados “memória” (conhecimento e práticas) no contexto universitário federal público (Unirio);
- b) Identificação de comunidades de práticas a partir de um repertório partilhado: memória de Interações, habitus e memória coletiva.
- c) Realização de workshops e seminários de sensibilização: objetivando fomentar a participação e discussão ampla das questões do projeto, no âmbito transdisciplinar, e dos seus resultados.

Considerações Finais

A complexidade dos novos problemas que se apresentam para a Ciência não encontra suas soluções em fragmentos do conhecimento. A produção de conhecimento transdisciplinar pretendida por esta proposta poderá impactar as áreas de conhecimento envolvidas, em um primeiro momento, de várias formas, mas em síntese representará um avanço para o ensino e pesquisa em nível de graduação e pós-graduação. Para as

áreas de conhecimento no campo das ciências humanas e sociais vislumbram-se como inputs:

1. a consolidação de conceitos;
2. a ampliação da cidadania no ambiente digital;
3. a preservação de conteúdos da produção intelectual e da produção cultural para gerações futuras;
4. maior integração entre memória social e ciência da informação.

Finalizando, destaca-se que a participação, a pesquisa-ação e a construção do saber são estabelecidas como os fundamentos teóricos de uma abordagem pedagógica. O conceito de participação a que nos referenciamos e que irá conduzir a implementação das comunidades de prática envolve a criação de oportunidades concretas para as pessoas tomarem iniciativas e terem acesso à produção do conhecimento.

Referências

- Araújo, R. M.; Borges, M.R.S.; Dias, M. S. 1997. A Framework for the classification of computer supported collaborative design approaches. In: *III CYTED-RITOS International Workshop on Groupware*.
- Araujo, R. M. 2011. *Relatório Projeto Ágora*.
- Conklin, E.J. 1996. *Designing organizational memory: preserving intellectual assets in a knowledge economy*, white paper, Group Decision Support Systems, Washington, D.C.
- Dodebei, Vera. 2006. Patrimônio, informação e memória digital. Rio de Janeiro. *Revista Morpheus*, v. 6, p. 1-10.
- Dodebei, Vera; Gouveia, Inês. 2008. Contribuições das teorias da memória para o estudo do patrimônio na web. In: FUGITA, Mariângela S. L.; MARTELETO, Regina M.; LARA, Marilda G. de. (Org.). *A dimensão epistemológica da Ciência da Informação e suas interfaces técnicas, políticas e institucionais nos processos de produção, acesso e disseminação da informação*. 1 ed. São Paulo: Cultura Acadêmica, v. 1, p. 87-102.
- Gondar, J.; Dodebei, Vera (Orgs.) 2005. *O que é memória social?* 1. ed. Rio de Janeiro: Contra Capa Livraria Ltda., v. 1. 160 p.
- Gontijo, Míriam. 2005. Sujeito, tecnologia e recepção: contribuição aos estudos de uso de novas tecnologias de informação e comunicação. *Perspectivas em Ciência da Informação*. Belo Horizonte, v.10 n.1, p. 70-85, jan./jun.
- Hague, B.N. 1999. *Digital democracy: discourse and decision making in the Information age*. Routledge, 1. Ed.
- Hindman, Matthew. 2009. *The myth of digital democracy*. Princeton University Press. 181 p.
- Lave, J., & Wenger, E. (1991). *Situated learning: legitimate peripheral participation*. Cambridge,

USA: Cambridge University Press.

Terra, J. Cláudio C. 2005. *Comunidades de práticas: conceitos, resultados e métodos de gestão*. Biblioteca Terra Fórum Consultores.

Turing, Alan. M .1936-7), *On Computable Numbers with an Application to the Entscheidungs problem*, Proceedings of the London Mathematical Society, Series 2, Vol. 42, pp. 230-265.

Vattimo, G. 1992. *The transparent society*. The Johns Hopkins University Press.

Wenger, E. (1998). *Communities of Practice: learning, meaning and Identity*. Cambridge,USA: Cambridge University Press

Wersig, G., Neveling, U. 1993. *Information Science: the study of postmodern knowledge usage*. In: *Information Processing & Management*. Oxford, v. 29, n. 2, p. 229-239.

Wilhelm, A. 2000. *Democracy in the digital age: challenges to political life in the cyberspace*. Routledge, 1. Ed.192 p.

Documentação Jurídica: reflexões sobre a função social do documento legislativo

Simone Torres
Assembleia Legislativa do Estado de Minas Gerais
simone.torres@almg.gov.br

Maurício Barcellos de Almeida
Universidade Federal de Minas Gerais
E-mail: mba@eci.ufmg.br

1 Introdução

O documento compreende não apenas objetos criados pelo homem especificamente para este fim (livros, cartas, jornais, revistas, sítios na Internet, registros em bases de dados, etc.), mas também objetos que foram preservados em virtude de sua informatividade, como artefatos em museus e peças de colecionadores. Conforme o conceito proposto por Suzzane Briet e ratificado por outros autores, por documento entendemos “qualquer elemento concreto ou simbólico, conservado, ou registrado para fins de representar, reconstituir ou provar um fenômeno físico ou intelectual” (Briet, 1951:10, tradução nossa).

Pretende-se neste trabalho tecer reflexões sobre a função social do documento jurídico, mais especificamente o documento legislativo. Para atingir estes objetivos, o trabalho apresenta na seção 2 e 3 uma breve revisão de literatura sobre a Filosofia da Linguagem e Documentação Jurídica, em que busca fornecer subsídios teóricos para a compreensão do tema; na seção 4, o documento legislativo é analisado em relação à teoria dos atos dos documentos e finalmente na seção 5, são apresentadas as considerações finais.

2 A Filosofia da Linguagem

Foi a linguagem que possibilitou à sociedade, o seu estabelecimento propriamente dito, transformando as relações transitórias do encontro nas relações duradouras da convivência. Segundo Vendryès (1921) a sociedade é, em grande parte, uma expressão da sua linguagem, e esta possui como função suprema a comunicação e resulta dos contatos sociais. A linguagem varia de acordo com os grupos sociais, porque os exprimem em sua mais profunda realidade. Segundo Wittgenstein (2002), é

incoerente pensar em uma linguagem privada, pois a linguagem é uma prática pública, onde as regras e convenções são compartilhadas pelos falantes.

Na virada do século XIX para o século XX, ocorreu no âmbito da Filosofia, a chamada “virada linguística”, com grandes mudanças em decorrência do surgimento da Filosofia Analítica (Souza Filho, 1990: 8). A Filosofia da Linguagem é uma especialidade da Filosofia Analítica que tem como objeto de estudo a forma como as pessoas se comunicam. Investiga como as pessoas dizem o que querem dizer e o conteúdo da fala que pode ser constituída de perguntas, ordens, promessas, desculpas. Searle (1981: 10) a define como “tentativa de fornecer uma descrição filosoficamente esclarecedora para certos traços gerais da linguagem, tais como a referência, a verdade, a significação e a necessidade”.

2.1 A teoria dos atos da fala e a teoria dos atos dos documentos

No bojo da Filosofia da Linguagem, Austin (1990: 21) criou a teoria dos atos da fala que propõe que expressões como “Eu sei que...” e “Eu prometo...” são usadas não para fazer descrições de um ato mental, mas para fazer algo, para realizar atos. O autor estuda a linguagem a partir de seu uso, como forma de ação, ou seja, os efeitos e as consequências produzidas pelo uso de determinadas expressões linguísticas em determinadas situações. Essa proposta de Austin tem como consequência o surgimento de um novo paradigma teórico, onde a linguagem é considerada como uma forma de atuação sobre o real e não mais sua representação (Souza Filho, 1990: 9).

Desenvolvida por Smith (2005) como uma extensão da teoria dos atos da fala, a teoria dos atos dos documentos se concentra na forma como as pessoas usam documentos, não só para registrar a informação, mas também para gerar uma variedade de fenômenos sociais. Tem como objetivo fornecer uma melhor compreensão do papel desempenhado pelos documentos na coordenação de ações humanas, possibilitando novos tipos de relações sociais. Smith (2005) ressalta que sua abordagem não limita a visão do documento apenas em sua função de fornecer elementos de prova ou informações, mas também na sua capacidade de criar uma variedade de tipos de poderes sociais e institucionais, chamados por Searle (1995) de “poderes deônticos”.

Em grandes sociedades, que lidam com muitas relações sociais complexas e com as interações que podem evoluir ao longo do tempo, os poderes mnemônicos dos indivíduos foram estendidos através dos documentos. Almeida, Cendon e Kerr (2011),

afirmam que enquanto os atos da fala existem somente quando de sua execução, os atos dos documentos se mantêm ao longo do tempo, e geram efeitos que nem sempre são diretamente observáveis.

3 O documento jurídico

Cada uma das fontes do Direito produz, de forma peculiar, informação. E para que esta informação possa ser comunicada, é necessária sua materialização em um documento. Para Nascimento e Guimarães (2004: 33), por documento jurídico entende-se o “conjunto de espécies documentais geradas pelo e/ou para o Direito”, que “diz respeito às relações jurídicas existentes entre os indivíduos ou destes para com o Estado e vice-versa”. A documentação jurídica é comumente dividida em três categorias doutrina, jurisprudência e legislação.

A doutrina jurídica, segundo Guimarães (1988: 66), “consiste na teorização do conhecimento jurídico, feita por especialistas da área e expressa em publicações monográficas ou seriadas”. Já a jurisprudência, segundo Atienza (1979: 51), é “o conjunto uniforme e constante das decisões judiciais sobre casos semelhantes”.

A documentação legislativa, por sua vez, representa o conjunto de documentos gerados durante o processo legislativo, compreendendo as proposições legislativas e as normas jurídicas propriamente ditas. Segundo Atienza (1979: 22), são produzidas por autoridade competente e contém “preceitos, regulamentos ou instruções, cuja observância se circunscreve à determinada jurisdição”. As características principais das normas jurídicas, segundo Bittar e Almeida (2009), são a prescrição da conduta considerada normal numa determinada circunstância e a prescrição de uma sanção, caso esta conduta não se concretize. Podemos citar como exemplos de documentos legislativos a constituição, as leis, decretos, resoluções, portarias, deliberações, etc.

Os documentos legislativos geram efeitos sociais ao estabelecerem a conduta, direitos, deveres e sanções, ao criar órgãos públicos e estabelecer suas respectivas competências. Geram efeitos ainda, ao estabelecer as normas para os atos negociais, que possibilitam a criação de relações jurídicas entre pessoas (físicas e jurídicas). Portanto, o documento legislativo não é apenas um registro, é a norma jurídica em si, que perdura durante todo o seu período de vigência, com total independência de seus autores, produzindo os efeitos sociais a que se destina em contextos temporais e ou geográficos distintos.

Os documentos legislativos possuem regras rígidas de elaboração, redação, alteração e consolidação como aquelas previstas no art. 59 da Constituição Federal de 1988 e nos dispositivos da Lei Complementar nº 95, de 26 de fevereiro de 1998. São obrigatoriamente publicados em diários oficiais, possuem diferentes tipos como Constituição, Lei, Decreto, Resolução, Portaria, etc. e podem, dependendo do tipo, receber contribuição de diversos atores sociais, como parlamentares, entidades públicas e civis e cidadãos em geral.

4 Análise da documentação legislativa à luz da teoria dos atos dos documentos

Ao analisar a documentação legislativa podemos comprovar a aplicabilidade da proposta de Smith (2005) na teoria dos atos dos documentos. É possível observar que a função precípua dos documentos legislativos não é o registro de informações, mas a criação de uma variedade de tipos de poderes sociais e institucionais, para atender aos objetivos do Direito, ao estabelecer a conduta em sociedade.

A eficácia do Direito, ou seja, sua aptidão para produzir efeitos jurídicos, tem nas normas jurídicas seu principal instrumento. As normas jurídicas geradas pelo processo legislativo são manifestações da linguagem produzidas exclusivamente através de documentos, que dependendo de sua finalidade, conteúdo e contexto de produção, geram efeitos jurídicos, administrativos e sociais. Uma vez promulgada uma lei que estabelece regras de conduta, torna-se obrigatória sua observância por parte de todos os cidadãos afetos à sua jurisdição, determinando não apenas a conduta permitida, mas também a sanção em caso de descumprimento. Desta forma, pode-se observar que os documentos legislativos se constituem em um instrumento concreto de atuação sobre a realidade.

5 Considerações finais

O Direito é constituído de complexos processos que incluem a produção de normas de conduta para a sociedade civil e para a Administração Pública, a solução de litígios, o estabelecimento de negócios jurídicos entre pessoas, o registro de fatos jurídicos, além dos estudos teóricos sobre o assunto. Esses processos geram documentos não apenas para fins probatórios, mas documentos dotados de “poderes deônticos”, que geram efeitos no âmbito da sociedade a que se destinam. Desse modo, o documento legislativo é um exemplo de aplicação da teoria dos atos dos documentos.

Conclui-se que além de preservar a memória socialmente compartilhada, favorecer a construção da cultura e atuar como uma ferramenta de controle e prova, o documento é um instrumento para o estabelecimento de regras de conduta social, necessário à geração de uma variedade fenômenos sociais. Analisar a documentação gerada em diferentes áreas do conhecimento em relação à teoria dos atos dos documentos está além dos objetivos do presente trabalho, mas espera-se que essa tarefa seja contemplada em pesquisas futuras.

Referências

Almeida, Maurício B.; Cendon, Beatriz V.; Kerr, Marta P. 2011. *Princípios metodológicos para a caracterização da dimensão pragmática de documentos no desenvolvimento de ontologias biomédicas*.

Atienza, C.A. 1979. *Documentação jurídica: introdução à análise e indexação de atos legais*. Rio de Janeiro: Achiamé.

Austin, John Langshaw. 1990. *Quando dizer é fazer: palavras e ação*. Porto Alegre: Artes Médicas.

Bittar, Eduardo C. B.; Almeida, Guilherme Assis de. 2009. *Curso de filosofia do direito*. 7.ed. São Paulo: Atlas.

Briet, Suzanne. 1951. *Qu'est-ce que la documentation?* Paris: Édit - Éditions Documentaires Industrielles et Techniques.

Guimarães, José Augusto Chaves. 1988. *A recuperação temática da informação em direito do trabalho no Brasil: propostas para uma linguagem de indexação na área*. Dissertação (Mestrado em Ciências da Comunicação)-Escola de Comunicações e Artes, Universidade de São Paulo, São Paulo.

Nascimento; Lúcia Maria Barbosa; Guimarães, José Augusto Chaves. 2004. *Documento jurídico digital: a ótica da diplomática*. In: Passos, Edilenice (Org). *Informação jurídica: teoria e prática*. Brasília: Thesaurus.

Searle, John R. 1981. *Os atos de fala: um ensaio de filosofia da linguagem*. Coimbra: Almedina.

Searle, John R. 1995. *The Construction of Social Reality*. New York: Free Press.

Smith, Barry. 2005. *Document acts*. Disponível em: <http://ontology.buffalo.edu/document_ontology/document_acts.doc>. Acesso em: 10 mar. 2012.

Souza Filho, Danilo Marcondes de. 1990. A filosofia da linguagem de J.L. Austin. In: Austin, John Langshaw. *Quando dizer é fazer: palavras e ação*. Porto Alegre: Artes Médicas.

Vendryès, Joseph. 1921. *Le langage: introduction linguistique à l'histoire*. Paris: Le Renaissance du Livre.

Wittgenstein, Ludwig. 2002. *Tratado lógico-filosófico; Investigações filosóficas*. 3.ed. Lisboa: Fundação Calouste Gulbenkian.

Breve reflexão sobre a ética na representação do conhecimento

Dulce Maria Baptista
Universidade de Brasília
dmbp@unb.br

1. Introdução: Organização do conhecimento e ética profissional

A breve reflexão a seguir baseia-se na consideração da ética como uma questão indissociável dos processos de representação e organização do conhecimento. Embora haja uma percepção mais clara dessa questão enquanto associada à recuperação e à disseminação da informação (a qual abrange desde o acesso aberto a conteúdos de cunho científico, à censura e aos direitos do autor, entre outros aspectos) ela se torna também presente em processos técnicos, tais como classificação e indexação, que, típicos de biblioteca, passam por adaptações e modificações, em virtude dos avanços tecnológicos da atualidade. Tais processos passam a fazer parte de um conjunto mais abrangente de práticas que hoje se referem à organização do conhecimento.

Ao se abordar o conceito de representação, no contexto da organização do conhecimento, e levando-se em conta a multiplicidade de estudos que procuram compreender esse conceito – amplo e polissêmico, em sua natureza – torna-se pertinente não perder de vista algumas indagações básicas, principalmente na perspectiva de vinculá-lo a preocupações de ordem ética. Tais preocupações derivam da percepção de que, da mesma forma que a ciência não é neutra, conforme argumenta Japiassu (1981), a atividade profissional e especializada – tal como vem sendo impactada pelos avanços da tecnologia – tampouco é neutra, no sentido da ausência de interferências ideológicas ou sócio-políticas.

A propósito da atividade profissional do bibliotecário, e ao identificar problemas éticos na representação do conhecimento, Milani e Guimarães se referem ao caráter *excessivamente prescritivo* do Código de Ética, citando a única disposição que poderia, em seu entender, ser aplicada à organização e representação do conhecimento:

Seção IV – Das proibições Art. 12 – Não se permite ao profissional de Biblioteconomia, no desempenho de suas funções: ... h) deturpar, intencionalmente, a interpretação do conteúdo explícito ou implícito em documentos, obras doutrinárias, leis, acórdãos e outros instrumentos de apoio técnico de exercício da profissão, com intuito de iludir a boa fé de outrem... (Milani; Guimarães, 2011: 3)

2. Alguns questionamentos

O que é, afinal, a ética? A resposta, evidentemente, não é simples, dado que se trata de conceito filosófico, porém é certo que, em linhas muito gerais, relaciona-se com a concepção e a prática do bem, como um valor moral a ser cultivado, tanto em nível individual como coletivo. Chauí considera que

[...] a ética exprime a maneira como uma cultura e uma sociedade definem para si mesmas o que julgam ser o mal e o vício, a violência e o crime e, como contrapartida, o que consideram ser o bem e a virtude, a brandura e o mérito. (Chauí *apud* Pinho, 2006: 51)

No que se refere à representação da informação, não há como fugir da questão ética, a qual, no entanto, nem sempre tem merecido a devida atenção, justamente em decorrência de uma certa inversão de valores, que tende a priorizar os meios em relação aos fins. Em outras palavras, se a ética não é, em si, inerente aos meios (processos técnicos), o uso de tais meios exige tomada de decisão, e toda decisão é necessariamente ética. Assim, torna-se oportuno questionar: por que organizar a informação e o conhecimento? Para esta pergunta específica, existem várias respostas que funcionam como justificativas *a priori*: explosão e ubiquidade de informação; diversidade do conhecimento; produção exponencial de literatura científica; diversidade de mídias e públicos, etc. É justamente com base nessa realidade, e se valendo de aportes multidisciplinares oriundos das tecnologias da informação e da comunicação (TICs), da biblioteconomia, da filosofia, da linguística e da semântica, que a ciência da informação se ocupa da organização da informação (como objeto) e do conhecimento (como conceito), e em tal nível de complexidade, que motiva um artigo da autoria de Dahlberg (2006), com o sugestivo título: *Knowledge organization: a new science?*

Nessa perspectiva, outras indagações se tornam legítimas: O que representar? Para quem representar? A partir de que critérios se representa algo? Tomando-se como exemplo a literatura referente à história do livro e das bibliotecas, como reflexo da própria história da civilização em seu imperativo de registrar, preservar e transmitir conhecimento, e considerando o suporte *livro* como o registro mais permanente do pensamento, parece desde sempre haver consenso quanto à necessidade de organizar as publicações armazenadas na biblioteca, sob pena de se inviabilizar o acesso ao conhecimento. A esse propósito, Naudé apresenta uma interessante metáfora:

[...] *sem essa ordem e disposição, qualquer amontoado de livros que possa existir, até mesmo de 50 mil volumes, não mereceria o nome de*

biblioteca, da mesma forma que um ajuntamento de 30 mil homens não mereceria o nome de exército, se eles não estivessem ordenados em diversas divisões sob a direção de seus chefes e capitães. (Naudé, apud Mckitterick, 2006: 95)

Tal ordenação inclui, por sua vez, a necessidade de representação, tanto no sentido macro da representação, em bloco, dos assuntos tratados nos livros – os quais, com base em classificações filosóficas (Araújo, 2006), deram origem aos mais engenhosos sistemas de classificação bibliográfica, servindo também à localização dos livros nas estantes – como no sentido da representação dos conteúdos temáticos de cada unidade documentária, por meio da extração de conceitos e sua tradução em palavras chave, e da elaboração de resumos, como formas de representação condensada da informação. Nesse aspecto, é de se considerar que a importância atribuída à representação traz em si, desde os primórdios da atividade bibliotecária, a preocupação com o leitor do livro que se destina a ser lido, e sucessivamente, com o receptor da mensagem que precisa ser decodificada, com o usuário da informação a ser recuperada, utilizada, disseminada, e transformada em conhecimento.

Por outro lado, conforme diz Manguel,

As categorias que um leitor traz para uma leitura e as categorias nas quais essa leitura é colocada – as categorias cultas sociais e políticas e as categorias físicas em que uma biblioteca se divide – modificam-se constantemente umas às outras, de maneira que parecem, ao longo dos anos, mais ou menos arbitrárias ou mais ou menos imaginativas. Cada biblioteca é uma biblioteca de preferências e cada categoria escolhida implica uma exclusão. (Manguel, 2006: 226)

3. Considerações finais

Tanto na pesquisa como na prática, observa-se, em grande medida, uma tendência a se privilegiar os processos técnicos, em detrimento de questionamentos de ordem ética, os quais podem contribuir a um aperfeiçoamento, inclusive, nos próprios sistemas de organização.

Não caberia no espaço deste texto uma revisão de literatura. Entretanto, uma busca aleatória por autores que tratam do assunto, revela que, embora numa escala menor que o foco nos Sistemas de Organização do Conhecimento (SOCs), em modelos e processos técnicos em geral, a preocupação com a ética no tratamento da informação

encontra-se presente em diversos estudos. Kobashi (2007), por exemplo, identifica contribuições de González de Gómez, Beghtol, Lopez-Huertas, entre outros. Em sua pesquisa de mestrado, Pinho (2006) empreende uma análise pormenorizada da questão, que abrange desde a ética na indexação, por exemplo, e a identificação de desvios (ou *bias*) na construção de tesouros, aos princípios norteadores da ética profissional do bibliotecário. Utiliza-se de estudos da Garcia Gutiérrez, Hudon e Beghtol como ponto de partida na busca por uma convergência de ideias sobre o assunto. Em extenso ensaio, Floridi (1999) explora, entre outros aspectos, a ética do computador e a ética da informação.

Encerrando esta reflexão, e entendendo que a organização do conhecimento envolve agentes, objetos e processos, é oportuno lembrar que, como elemento propulsor de toda a dinâmica informacional, existe o profissional da informação. Cabe a ele buscar respostas a uma série de desafios, tais como: Compatibilizar diversidade (de instrumentos) com consistência (na representação); Considerar o contexto de produção de informação *versus* o contexto de uso da informação; Estabelecer relações entre passado e presente na compreensão de conceitos; Resolver a dualidade entre o que se considera como politicamente correto e o real significado das coisas. Certamente, a ética deverá estar presente no bojo de todas essas discussões.

4. Referências

Araújo, Carlos Alberto Ávila. 2006. Fundamentos teóricos da classificação. *Encontros Bibli*, Florianópolis, n. 22, 2º sem. 2006. Disponível em: <http://www.periodicos.ufsc.br/index.php/eb/article/viewFile/296/296/368> Acesso em: janeiro 2013.

Dahlberg, Ingetraut. 2006. Knowledge organization: a new science? *Knowledge Organization*, 33, 1: 11-19.

Floridi, Luciano. 1999. Information ethics: on the philosophical foundation of computer ethics. *Ethics and Information Technology*, 1, 1: 37-56.

Japiassu, Hilton. 1981. *O mito da neutralidade científica*. Rio de Janeiro: Imago.

Kobashi, Nair Fumiko. 2007. Fundamentos semânticos e pragmáticos da construção de instrumentos de representação da informação. *DataGramaZero*, 8, 6: 1-10. Disponível em: http://www.dgz.org.br/dez07/Art_01.htm Acesso em: janeiro 2013.

Manguel, Alberto. 1997. *Uma história da leitura*. São Paulo: Companhia das Letras.

Mckitterick, David. 2006. A biblioteca como interação: a leitura e a linguagem da bibliografia. In Baratin, Marc; Jacob, Christian orgs. *O poder das bibliotecas: a memória dos livros no ocidente*. Rio de Janeiro: Editora UFRJ, p. 94-107.

Milani, Suellen Oliveira; Guimarães, José Augusto Chaves. 2011. Problemas éticos em representação do conhecimento: uma abordagem teórica. *DataGramaZero*, 12, 1: 3.

Pinho, Fabio Assis. 2006. *Aspectos éticos em representação do conhecimento: em busca do diálogo entre Antonio Garcia Gutiérrez, Michele Hudon e Claire Beghtol*. Dissertação (Mestrado em Ciência da Informação), Marília, UNESP.

O caráter técnico, social e mediador da organização do conhecimento

Aida Varela Varela
Universidade Federal da Bahia
varela@ufba.br

Marilene Lobo Abreu Barbosa
Universidade Federal da Bahia
marilene@ufba.br

1 Introdução

A organização da informação, do ponto de vista da biblioteconomia e da ciência da informação, transcende às questões operacionais e logísticas de processar/tratar a informação, na medida em que os princípios da disciplina estão atrelados ao ideal de tornar acessível a todos o conhecimento produzido. Com efeito, analisando-se a trajetória histórica da área, evidenciam-se que os vários processos de organização e disseminação da informação estão centrados nos eixos da criação e padronização das ferramentas de representação da informação, recursos mediadores do acesso ao conhecimento, marcando o caráter social, cognitivo e mediador, em essência, da disciplina. Neste estudo, objetiva-se, por meio de um estudo teórico-investigativo, defender este argumento.

277

2 Dos Primórdios aos Tesouros

Ao longo da evolução das formas de registro do conhecimento produzido pela humanidade, a missão da biblioteconomia tem sido permitir a todas as gerações o acesso ao conhecimento produzido e acumulado. Deste modo, grandes bibliotecas foram criadas desde a Antiguidade até a Idade Contemporânea, como a biblioteca grega de Alexandria, no Egito, A biblioteca de Nínive, na Babilônia, A Biblioteca Nacional da França, a Biblioteca do Congresso, em Washington, A Biblioteca Nacional, no Rio de Janeiro e outras. Todas motivadas por um único princípio: a acessibilidade da informação.

Cedo, no entanto, os primeiros bibliotecários descobriram que não basta reunir o acervo, é preciso organizá-lo, para permitir o acesso à informação nele registrada. Nesta perspectiva, Calímaco, poeta e bibliotecário grego, produziu o Pinakes, catálogo

que referencia grande parte do acervo da biblioteca de Alexandria, organizado seguindo a lógica da classificação aristotélica; já na era da imprensa escrita, Conrad Gesner, bibliotecário suíço, organizou e publicou a bibliografia *Bibliotheca Universalis*; e Gabriel Naudé, bibliófilo francês, escreveu, em 1627, o *Advis pour dresser un bibliothéque*, fixando os primeiros princípios da biblioteconomia, como a ideia de ordem bibliográfica e, em 1643, publicou o *Bibliotheca Cordesiana Catalogus*, com um esquema de doze classes e, destacando o significado dos instrumentos de representação, na busca e recuperação dos assuntos (MEY, 1995:17).

Em 1595, o bibliógrafo inglês, Andrew Maunsell, introduziu o conceito de uniformização na descrição dos itens documentais, ao publicar o *Catalogue of English Printed Books*, promulgando as primeiras regras de catalogação (CAMPELLO, 2006: 57). A invenção da imprensa móvel, por Guttenberg, e a contínua melhoria dos processos de registro do conhecimento ampliaram muito a produção de livros, enquanto que as revoluções científicas e industriais impulsionaram o desenvolvimento do conhecimento e fizeram surgir os periódicos científicos especializados. Livros, periódicos, jornais etc. passaram a ser fonte de conhecimento e a biblioteconomia foi levada a desenvolver métodos e técnicas mais eficientes para coletar, organizar e disseminar o acervo produzido. O conceito era de que, para permitir o acesso ao conhecimento é preciso catalogar, classificar e indexar as publicações.

Em 1870, Harris apresentou um sistema de classificação baseando-se na ordem indireta da classificação de Bacon - razão, imaginação e memória – motivando o surgimento de estudos e outros esquemas de classificação, como a *Classificação Decimal Universal*, de Dewey; a *Expensive Classification*, de Cutter, construída a partir da inversão da classificação de Bacon, (PIEIDADE, 1983: 74); a *Classificação Decimal Universal*, de Otlet e La Fontaine, cuja finalidade era organizar, por assunto, o *Répertoire Bibliographique Universel*, bibliografia que arrolava o acervo do *Mundaneum*.

Destacam-se os estudos do matemático e bibliotecário indiano Ranganathan, que revolucionaram a noção de classificação, ao publicar a *Filosofia da Classificação Bibliográfica*, com considerações sobre filosofia e epistemologia e a relação destas com a classificação e o *Sistema de Classificação Facetada*, cuja maior contribuição é a possibilidade de inter-relação entre os conceitos (PIEIDADE, 1983:198) Bliss realizou expressivos estudos sobre organização do conhecimento e divulga esta expressão em seus livros *The Organization of Knowledge and the System of Sciences*, de 1929, e

Organization of Knowledge in Libraries and the Subject Approach to Books, de 1933. Criou também a Bibliographic Classification (PIEDADE,1983:74).

Os estudos sobre classificação produziram consistentes esquemas de classificação de assunto, baseados nas classificações filosóficas e científicas do conhecimento. Na catalogação, merece distinção a continuidade dos estudos sobre padronização da representação descritiva dos itens documentais e a associação desta à representação temática na perspectiva da regulação e do controle da representação da informação e consequente organização do conhecimento produzido. Neste sentido, o bibliotecário italiano Antonio Pannizzi, atuando no British Museum, em 1835, apresentou 91 regras de catalogação, e, em 1841, publicou as Regras para a Compilação de um Catálogo, nas quais já introduzia a noção de vocabulário controlado (PINHO, 2006:10).

No esforço para a padronização e simplificação da descrição bibliográfica devem ser citadas as Anglo-American Cataloguing Rules- AACR, publicadas em duas edições, elaboradas pela American Library Association(ALA) e pela Library Association(Reino Unido) e os estudos de Seymour Lubetzky, que publicou Cataloguing Rules and Principles, motivando a discussão sobre os fundamentos teóricos da disciplina., a Conferência Internacional sobre Princípios de Catalogação, da IFLA e UNESCO, em 1961 e, em 1969, a Reunião Internacional de Especialistas em Catalogação, que resultou no acordo para a adoção das International Standard Bibliographic Descriptions . O Programa de Controle Bibliográfico, com a adoção dos ISBN e ISSN, marcou o ápice da catalogação (CAMPELLO, 2006:58-60).

Na indexação, surgiram listas de cabeçalhos de assunto, como a Lista de Sears e a da Biblioteca do Congresso americano, e os vocabulários controlados, como o tesouro geral da UNESCO, e especializados, como o da OCDE. Também os Sistemas de Recuperação da Informação ampliaram o acesso ao documento, com o aperfeiçoamento de tesouros especializados.

A criação da International Society for Knowledge Organization (ISKO), em 1989, sob a influência do Classification Research Group(CRG) e dos estudos sobre conceito, de Ingetraut Dahlberg, contribuiu para o aprofundamento e intensificação dos estudos sobre organização do conhecimento (PINHO, 2006:13).

3 Os Meios, processos e ferramentas emergentes

Atualmente, nas redes informacionais, os conteúdos se entrelaçam, maximizando e dificultando, as possibilidades de busca e recuperação, como a mediação na Internet por motores de busca, que revelam preciosidades ou conteúdos respostas sem significado. Estudiosos de renome do campo da indexação e recuperação da informação, como Lancaster e Tefko Saracevic, ressaltam a necessidade de que princípios de relevância e precisão sejam preservados para garantir qualidade nas pesquisas bibliográficas.

A Web Semântica surgiu para a superação das barreiras mencionadas por meio das várias soluções tecnológicas implementadas em sua arquitetura, como metadados, linguagem de marcação XML e a arquitetura RDF-Resource Description Framework, que garantem a interoperabilidade dos dados em nível semântico, estrutural e sintático; além das ontologias, que proporcionam a definição semântica; e os agentes inteligentes ou motores de busca, que, espera-se, proporcionarão uma busca e recuperação mais efetiva a partir do estabelecimento de regras e do acesso a coleções de recursos devidamente estruturados, representados e definidos semanticamente (PICKLER, 2007: 17).

4 Mediação

Os processos e ferramentas de organização e representação do conhecimento, no âmbito da biblioteconomia e ciência da informação, têm funcionalidades nitidamente mediadoras, tendo em vista que objetivam a localização da informação na perspectiva do acesso pleno ao conhecimento.

Sobre esta relação intrínseca da cognição com os processos de tratamento e difusão da informação, já na década de 70 do século XX, Shera (1973) chamava a atenção sobre as questões de mediação que envolviam o usuário, os registros gráficos e a cognição. Conforme Castells(2003) os modos como as informações são construídas e apresentadas são diversificados de acordo com os esquemas de quem as disponibiliza e não de quem os busca, o que pressupõe a necessidade de mediação e de conhecimento de teorias cognitivas.

Para Latour (2001:350) a mediação ocorre no espaço social de interações e produção do conhecimento, incluindo entidades humanas e artificiais num mesmo

continuum; já Vygotsky (2012) assevera que as funções mentais superiores do ser humano devem ser consideradas produtos de uma atividade mediada e que, para acessar a informação a mediação é desempenhada pelos instrumentos materiais e psicológicos, construídos artificialmente, observando que ambos têm caráter social por natureza. Estes pensamentos remetem ao papel das ferramentas naturais e artificiais representativas do conhecimento que atuam como pontes para acesso às fontes de informação.

Ingwersen (2002), na concepção do modelo cognitivo do processo de recuperação da informação, identifica os processos de cognição que ocorrem durante o processamento da informação. Por seu turno, nos anos 90, Hjørland (1995) acrescenta que a indexação é um processo intelectual dependente da cognição e do domínio do contexto físico, psicológico e sociocognitivo, supondo que o indexador necessita realizar a identificação e a seleção de conceitos na concepção orientada para o conteúdo e para a demanda do usuário.

5 Conclusão

De todo o exposto, neste espaço muito reduzido e simplificado, confirma-se, a premissa de que a organização do conhecimento sempre associou o caráter de utilidade desta atividade à aspiração do sujeito social de ter acesso ao conhecimento. Neste sentido a biblioteconomia, a documentação, a arquivologia e a ciência da informação, acompanhando a evolução da ciência e da sociedade, procuraram produzir meios, processos e ferramentas mediadoras, que otimizassem esta relação do homem com a produção intelectual universal, isto desde os tabletes de argila da Biblioteca de Nínive, na Babilônia, passando pelo Mundaneum, na Bélgica, até as modernas redes sociais e a computação em nuvem.

REFERÊNCIAS

CASTELLS, Manuel. 2003. Internet e sociedade em rede. In: MORAES, Denis de (org.). *Por uma outra comunicação: mídia, mundialização cultural e poder*. Rio de Janeiro: Record, p. 255-288.

CAMPELLO, Bernadete. 2006. *Introdução ao controle bibliográfico*. 2.ed. Brasília: Briquet de Lemos/ Livros.

HJØRLAND, Birger. 1995. Toward a new horizon in information science: domain analyses. *Journal of the American Society for Information Science*, 46 (6): 412.

INGWERSEN, P. 2002. *Information retrieval interaction*. London: Taylor Graham, 246 p. ISBN: 0-947568-54-9. PDF Version. Retrieved December 13, 2005. Disponível em: <www.db.dk/pi/iri>.

LATOURE, Bruno. 2001. *A esperança de Pandora: ensaios sobre a realidade dos estudos científicos*. Bauru: EDUSC.

MEY, Eliane Serrão Alves. 1995. *Introdução à catalogação*. Brasília: Briquet de Lemos/Livros.

PICKLER, Maria Elisa Valentim. 2007. Web Semântica: ontologias como ferramentas de representação do conhecimento. *Perspectivas em Ciência da Informação* 12 (1): 65- 83, jan./abr.

PIEIDADE, Maria Antonieta Requião. 1983. *Introdução à teoria da classificação*. 2. ed. rev. aum. Rio de Janeiro: Interciência.

PINHO, Fábio Assis. 2006. *Aspectos éticos em representação do conhecimento: em busca do diálogo entre Antonio García Gutiérrez, Michèle Hudon e Clare Beghtol*. 123 f. Dissertação (Mestrado em Ciência da Informação) – Faculdade de Filosofia e Ciências, Universidade Estadual Paulista, Marília.

SHERA, Jesse. 1973. Toward a theory of librarianship and information science. *Ciência da Informação*, 2 (2): 87-97.

VYGOTSKY, Lévi. 2012. *Pensamiento y lenguaje*. Barcelona: Paidós. Nova edição sob a responsabilidade de Alex Kozulin.

Modelagem, Organização e Representação do Conhecimento: proposta de elaboração da Biblioteca Virtual Temática em Vigilância Sanitária dos serviços de estética do município do Rio de Janeiro.

Jorge C. de Almeida Biolchini
Instituto Brasileiro de Informação em Ciência e Tecnologia/
Universidade Federal do Rio de Janeiro
jorge.biolchini@gmail.com

Aline Vieira do Nascimento
Universidade Federal do Ceará
vieiraaline@yahoo.com.br

1 Introdução

As ontologias têm se evidenciado cada vez mais no uso das técnicas de organização da informação, como um importante recurso na tentativa de melhorar os sistemas de busca de informação. No desenvolvimento dos estudos sobre ontologia, segundo Ushold e Jasper (1999) a ontologia “assume uma variedade de formas, mas isto irá necessariamente incluir um vocabulário de termos e alguma especificação de seu significado”. Ainda expressando o pensamento de Ushold e Jasper (1999) esses vocabulários expressam “definições e uma indicação de como os conceitos estão inter-relacionados, cuja coletividade impõe uma estrutura no domínio e restringe as interpretações possíveis dos temas”.

283

Vejam os abaixo duas definições para ontologias:

Autor	Definição
Sowa (1999)	É o estudo das categorias de coisas que existem ou podem existir, em alguns domínios.
Merriam-Webster; Gove (2002)	Teoria que se refere a entidades específicas, abstratas, que são aceitas em um sistema com uma linguagem.

As ontologias têm sido utilizadas, na área da gestão do conhecimento, como uma técnica para auxiliar a representação da informação. De acordo com Medeiros (2010

apud GOMÉZ-PERÉZ, 2004) “Ontologias se referem a uma conceituação consentida sobre uma determinada área de conhecimento”, com o intuito de trazer a possibilidade de manipulação dos significados ou semântica num sistema de representação de informação num ambiente, seja ele na web ou em outro lugar.

A modelagem se constitui de uma atividade que envolve o mapeamento do setor de forma que as ontologias tenham uma representação única. Na medida em que evolui, uma ontologia pode ter várias formas como um catálogo até uma taxonomia. Quanto ao tipo de componentes de uma ontologia, estes variam de acordo com o uso de frames e lógica de primeira ou descritiva e podem ser de acordo com (GOMEZ-PEREZ et al, 2004): Classes, relações, funções, axiomas e instâncias.

Desde os primórdios a sociedade preocupa-se em tentar exercer controle sobre os mais diversos problemas existentes na vida coletiva e que é considerado riscos à saúde pública. Essas práticas contra o que é “nocivo” visam elaborar medidas de controle sanitário ambiental, alimentício, do exercício da farmácia e da medicina, de produtos, tecnologias e serviços, todos envolvidos no círculo saúde, doença, cuidado e qualidade de vida. Essa preocupação foi se desenvolvendo ao longo do tempo pela sociedade e aos poucos através das leis, sendo posta em prática pelo poder público, dando à população o direito à saúde de qualidade.

De acordo com COSTA (1999, p.7) “os serviços de saúde e certas categorias de produtos sob o controle sanitário são indispensáveis à consecução do direito à saúde”.

Neste sentido, a lei 9.782 de janeiro de 1999 definiu o Sistema Nacional de Vigilância Sanitária e criou Agência Nacional de Vigilância Sanitária, com o objetivo de regularizar, normatizar, controlar e fiscalizar no âmbito da vigilância sanitária produtos, substâncias e serviços de interesse para a saúde; acompanhar e coordenar as ações estaduais, distritais e municipais de vigilância sanitária, atuando nos setores de medicamentos, alimentos, bebidas, cosméticos, saneantes, equipamentos médico-hospitalares, imunobiológicos, transplantes, fumígeno e quaisquer produtos que envolvam a possibilidade de riscos à saúde.

Dentre estes ramos, mapeamos os serviços de embelezamento, estética e congêneres, controlados e encontrados no sítio da VISA com foco no âmbito municipal, neste estudo a cidade do Rio de Janeiro. O setor possui os serviços de cabeleireiro, barbearia, depilação de todos os tipos, manicure, pedicure, podologia, estética facial, corporal, massagem, banho de ofurô, drenagem linfática, massagem estética e tatuador.

O objetivo deste trabalho é propor, através da construção de ontologias, a elaboração de uma Biblioteca Virtual Temática em Vigilância Sanitária de Serviços de Estética. Inicialmente será identificado o setor dos serviços de estética, posteriormente serão elencadas as ontologias a serem utilizadas no sítio através da elaboração do mapa conceitual. A intenção é de proporcionar a armazenagem e manipulação do conhecimento no sistema de informação de tal maneira que estas possam ser usadas em qualquer tempo.

Abaixo, uma tabela com os atores envolvidos no setor e suas necessidades informacionais (TABELA 1).

Tabela 1 – Atores envolvidos nos serviços de estética

Ator	Necessidades informacionais
Consumidores	<ul style="list-style-type: none"> - Direito do consumidor; - Canais de reclamação, sugestão; - Prevenção de Acidentes; - Links com informações sobre cuidados pessoais, tratamento de pele, alimentação saudável; - Publicidade;
Profissionais	<ul style="list-style-type: none"> - Diretrizes para fiscalização de clínicas; - Normas e Leis da VISA; - Quais os serviços e produtos deste setor; - Quais os profissionais que podem atuar; - Vídeos educativos relacionados ao assunto; - Ter cursos virtuais; - Normas e Leis da VISA; - Formas de manuseio de material de trabalho.
Vigilantes sanitários	<ul style="list-style-type: none"> - Funcionamento dos serviços da VISA nestas três esferas; - Atribuições da VISA em cada esfera. -
Empreendedores	<ul style="list-style-type: none"> - Diretrizes para fiscalização de clínicas; - Ter cursos virtuais; - Aspectos sanitários na administração de empresas do setor de estética; - Normas e Leis da VISA; - Alvarás e licenças;
Pesquisadores	<ul style="list-style-type: none"> - Periódicos, livros, artigos; - Estatísticas do setor

Fonte: Elaboração própria.

3 Etapas do processo

3.1 Objetivo

Obter informações sobre serviços de estética e atividades similares abrangidas pela VISA em ambiente digital, para elaborar uma ontologia que represente os conceitos visando a contribuir para a facilitação da recuperação da informação de forma mais precisa.

3.2 Justificativa

Proporcionar aos envolvidos nos serviços de estética, incluindo os consumidores e pesquisadores, através do uso de ontologia, uma estrutura de informações que subsidie, de forma esclarecedora, como funcionam tais serviços, informações sobre controle e fiscalização, bem como as normas e regras que regulam este tipo de serviço em âmbito municipal, estadual e federal.

286

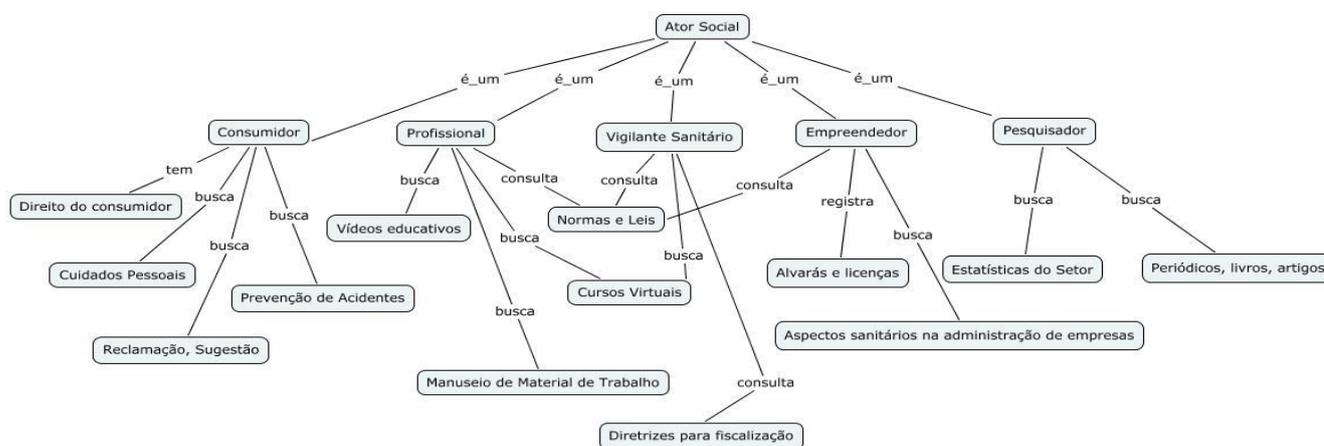
3.3 Metodologia

- a) Mapeamento dos serviços de estética que a VISA abrange e possui, nos sítios da ANVISA e em links de escolas de estética;
- b) Levantamento dos atores sociais e das necessidades de informação para elaboração do mapa conceitual;
- c) Construção do mapa conceitual com a estrutura da BVE, a partir dos conceitos e das relações conceituais mapeados;
- d) Entrevistas com os envolvidos;
- e) Comparação entre as informações disponíveis no sítio da VISA com a ontologia elaborada (neste momento será testada a precisão representacional da ontologia).

4 Considerações Finais

Diante das informações e problemas encontrados ao longo da elaboração do projeto da BVE, foram abordados aspectos teórico-práticos relevantes para a modelagem e organização do conhecimento. Dentre eles estão: a) o levantamento das informações referentes à situação dos serviços prestados à sociedade pela Vigilância Sanitária; b) a modelagem da organização do conhecimento através dos conceitos organizados hierarquicamente, na forma do mapa conceitual; e c) a disposição na qual as informações foram organizadas se direcionando para cada ator social envolvido no processo.

Detectou-se que o uso de ontologia na construção de um ambiente digital, no caso a Biblioteca de Vigilância Sanitária em Serviços de Estética, pode garantir ao usuário um aumento da precisão da informação desejada, apoiando uma estrutura web que abrange um vasto campo de informações a cerca dos serviços e abrangência do setor, relacionado ao sítio da VISA.



Referências

Agência nacional de vigilância sanitária. 2012. Assuntos de interesse: cosméticos. Disponível em: <<http://portal.anvisa.gov.br/wps/content/Anvisa+Portal/Anvisa/Inicio/Cosmeticos>> Acesso em: 30 nov. 2012.

Agência nacional de vigilância sanitária. 2009. Referencia técnica para o Funcionamento dos Serviços de Estética e Embelezamento sem Responsabilidade Médica. Brasília: ANVISA, 2009. Disponível em: <<http://portal.anvisa.gov.br/wps/wcm/connect/527126804745890192e5d63fbc4c6735/Servicos+de+Estetica+e+Congeneres.pdf?MOD=AJPERES>>

Asunción Gómez-Pérez, A.; Fernández-López, A.; Corcho, O. 1967. Top-level ontologies. In: **Ontological Engineering**: with examples from the areas of Knowledge Management, e-Commerce and the Semantic Web. London: Springer-Verlag.

Brasil. Lei nº 8080 de 19 de setembro de 1990. Dispõe sobre as condições para a promoção, proteção e recuperação da saúde, a organização e o funcionamento dos serviços correspondentes e dá outras providências.

Gómez-Pérez; Fernádes; Corcho. 2006. *Ontology Engineering: With Examples From The Areas of Knowledge Management, e-Commerce and the Semantic Web*. New York: Springer.

Medeiros, Luciano Frontino de. Framework para engenharia e processamento de ontologias utilizando computação quântica. 2010, 204 f. Tese (Doutorado em Engenharia e Gestão do Conhecimento) – Centro Tecnológico, Universidade Federal de Santa Catarina, Florianópolis, 2010.

Prefeitura Municipal Do Estado Do Rio De Janeiro. 2012. Vigilância e Fiscalização Sanitária do Estado do Rio de Janeiro. Disponível em: <http://www2.rio.rj.gov.br/vigilanciasanitaria/info_saude_saloes.cfm> . Acesso em: 30 nov. 2012.

Sowa, John F. 2000. *Knowledge representation : logical, philosophical, and computational foundations*. Pacific Grove : Brooks Cole Pub., 2000.

Costa, Ediná Alves. 1999. *Vigilância sanitária e proteção da saúde*. São Paulo, Hucitec/Sobravime, 1999.

Tessler, Marga Inge Barth. 2004. A vigilância sanitária e os princípios da precaução e da prevenção. In: Encontro Internacional dos Profissionais em Vigilância Sanitária, 4., 2004, Foz do Iguaçu. Palestra... Foz do Iguaçu: ABPVS, 2004, 16 p. Disponível em: <http://www2.trf4.jus.br/trf4/upload/arquivos/curriculo_juizes/vigilancia_sanitaria_principios__precaucao__prevencao_foz_iguacu.pdf> Acesso em: 03 dez de 2012.

Oliveira, Iraci Sobral de. 2009. Notas sobre Engenharia Ontológica. Disponível em: <<http://e-reality-home.blogspot.com.br/2009/04/notas-sobre-engenharia-ontologica.html>> Acesso em: 03 dez de 2012.

Sistema de Organização do Conhecimento sobre Intensificação Agropecuária: agregando valor às redes de conhecimento da Embrapa

Ivo Pierozzi Junior
Embrapa Informática Agropecuária
ivo.pierozzi@embrapa.br

Leandro Henrique Mendonça de Oliveira
Embrapa Informática Agropecuária
leandro.oliveira@embrapa.br

Maria de Cléofas Faggion Alencar
Embrapa Meio Ambiente
cleofas.alencar@embrapa.br

Gladis Maria de Barcellos Almeida
Universidade Federal de São Carlos
gladis@ufscar.br

1 Introdução

A Embrapa gera conhecimentos para solução de questões agropecuárias complexas, desenvolvendo ações de PD&I em projetos multi/inter/transdisciplinares e que envolvem redes de conhecimento e competências (Torres *et al.*, 2010, 2011). Soluções eficazes, efetivas e sustentáveis passam pela construção de conceitualizações técnico-científicas coletivamente negociadas e consensuais. Uma das dificuldades mais generalizadas desse tipo de arranjo de trabalho em rede se refere à criação, uniformização e compartilhamento conceituais, originadas por entraves e conflitos nos processos de intercâmbio de informações.

Os sistemas de organização de conhecimento (SOCs) são ferramentas muito úteis para superação dessa dificuldade, inerente aos processos de gestão do conhecimento em contextos de trabalhos colaborativos (Brascher e Café, 2008; Souza, Tudhope e Almeida, 2010). Por meio do tratamento terminológico, conhecimento é codificado em vocabulários especializados, que podem evoluir em estrutura e função para modelos de organização e representação mais complexos, viabilizando inclusive a transposição da linguagem natural para as linguagens artificiais e permitindo que aplicações computacionais possam ser derivadas (Campos, Campos e Medeiros, 2011; Carlan e Medeiros, 2011).

O presente trabalho relata a experiência desenvolvimento e aplicação, no âmbito de um projeto de pesquisa, de um modelo de organização e representação do conhecimento sobre os processos de intensificação agropecuária, contextualizados por observações empíricas em regiões produtoras de *commodities* agrícolas no Brasil. Tal iniciativa partiu da preocupação de oferecer soluções adequadas e oportunas para sistematização, organização e disseminação do conhecimento sobre esse assunto, visando facilitar o seu compartilhamento pela comunidade científica e atores do setor agropecuário.

2 Intensificação agropecuária: do mundo real ao SOC

Os processos de intensificação agropecuária (IA) (Boserup, 1965), em regiões brasileiras com expressivas produções de grãos e cana de açúcar, foram objeto do Projeto Intagro, desenvolvido no âmbito dos programas de pesquisa, desenvolvimento e inovação (PD&I) da Embrapa. O projeto foi executado no período de 2008-2011 por uma equipe multidisciplinar envolvendo abordagens ambientais, agronômicas e socioeconômicas e exigindo ferramentas de organização da informação e do conhecimento, como suporte ao processo de produção e integração de resultados, culminando em uma proposta de construção de um sistema de organização do conhecimento (SOC).

Parte do itinerário metodológico de construção do SOC está descrito em Pierozzi Jr., Oliveira e Souza (2010): construção, compilação e gestão do *córpus* textual; extração semiautomática de candidatos a termos; construção do mapa conceitual e de recursos para sua visualização. A partir dessa etapa, o SOC evoluiu e atualmente consiste de um vocabulário de termos/conceitos categorizados sobre IA, complementado por um mapeamento no tesauro AGROVOC (<http://aims.fao.org/standards/agrovoc/about>), recurso terminológico específico para agricultura e áreas do conhecimento afins. A Figura 1 apresenta como as etapas metodológicas desse trabalho são relacionadas aos resultados sucessivos que resultaram no SOC aqui relatado.

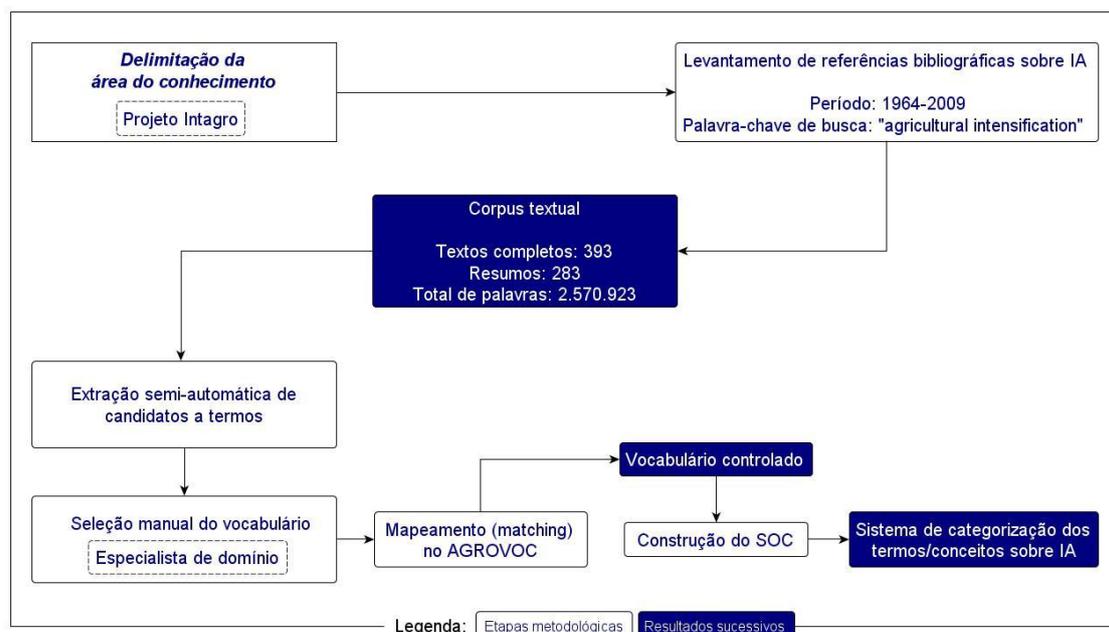


Figura 1: Etapas metodológicas e resultados sucessivos do processo de elaboração do SOC sobre intensificação agropecuária para a Embrapa.

3 Mapeamento terminológico

A base de dados do AGROVOC, disponível em <<http://aims.fao.org/standards/agrovoc/functionalities/download>>, em fevereiro de 2010 foi utilizada como referencial para o mapeamento do vocabulário extraído do cópús textual e selecionado por especialista de domínio. Para essa atividade, foi desenvolvida uma ferramenta computacional que comparou automaticamente ambos os vocabulários e os reorganizou em uma planilha, indicando se determinada palavra já estava registrada no tesouro e, em caso positivo, recuperando suas relações associativas já estabelecidas, além de sua tradução em espanhol, francês e português. Uma consulta sucessiva ao AGROVOC, desta vez manual, serviu para alinhar a escolha dos termos/conceitos representativos de cada categoria do SOC sobre IA àqueles já presentes no tesouro.

Vale a pena mencionar que o modelo conceitual do AGROVOC é o de um vocabulário em inglês, com tradução para outras 21 línguas, uma das quais o português. No entanto, a terminologia em língua portuguesa nesse tesouro esteve, até há pouco tempo, sob custódia de profissionais lusitanos. Dessa forma, refletem a realidade linguística de Portugal em relação à agricultura, pesca e silvicultura praticadas naquele país e que é muito diferente do contexto tropical e subtropical brasileiro. Uma grande quantidade de termos necessita então ser adaptada para o português falado e escrito no Brasil.

4 Categorização conceitual

O SOC sobre IA foi organizado partindo-se do conceito principal *Agricultural Intensification* (intensificação agropecuária) no primeiro nível de categorização. Abaixo dele, organizam-se sete subcategorias ou subníveis conceituais. Quatro deles referem-se a outros domínios do conhecimento que emprestam conceitos (e, portanto, terminologias) ao entendimento do processo de IA, quais sejam: *Environment* (Meio Ambiente); *Agronomy* (Agronomia); *Territoriality* (Territorialidade) e *Socioeconomy* (Socioeconomia). O que se pretendeu com o SOC foi indicar que o conhecimento global sobre IA depende essencialmente do entendimento da significação e da inter-relação de conceitos desses quatro domínios e de suas interfaces. O modelo ainda inclui três outras categorias que agregam valor ao entendimento da IA: *Methodology* (Metodologia): métodos utilizados nos estudos dos processos; *Geographic Localization* (Localização Geográfica): onde esses processos são observados no mundo real e *Institution* (Instituição): organizações que estão envolvidas com o assunto sob os pontos de vista de PD&I, financeiro ou mercadológico, governamental ou não governamental, etc.

A Figura 2 apresenta o SOC completo em seu estágio atual de desenvolvimento, em inglês e português, no formato de *folder tree*, mostrando o arranjo hierárquico das categorias e algumas de suas subdivisões abertas até seus respectivos últimos níveis de granularidade. Sob essa formatação, o modelo estará disponível no ambiente e-Termos (<http://www.etermos.cnptia.embrapa.br/index.php>) após ser publicado, mas já pode ser visualizado em formato de grafo, acessado em <http://www.cnptia.embrapa.br/~leandro/intagro/>, onde cada termo/conceito é apresentado com suas relações associativas, além de sua tradução, quando disponível, para as línguas espanhola, francesa e portuguesa, conforme mapeamento no tesauro AGROVOC.

5 Conclusão e continuidade do trabalho

O estágio atual do trabalho contempla cerca de 600 termos/conceitos reunidos e organizados no modelo, apresentados em inglês com as respectivas traduções para o português brasileiro. Obviamente, esse SOC é um “sistema aberto”, de maneira que a partir de sua utilização, novos conceitos e termos podem e devem ser continuamente

agregados na sua arquitetura, respeitando-se ou reajustando-se as relações hierárquicas ou associativas ali desenhadas preliminarmente. Entretanto, os elementos fundamentais necessários à representação do conhecimento sobre IA já foram aí reunidos e organizados.

Em seu atual estágio de construção, o SOC sobre IA é um sistema de categorização de termos/conceitos e, seguindo os modelos de tipologia dos SOCs (Souza, Tudhope e Almeida, 2010; Zeng, 2008), pode evoluir para formatos cada vez mais aprimorados em estrutura e função. No futuro, mais evoluído, poderá ser traduzido para linguagens formais de representação do conhecimento, aproximando-o das aplicações computacionais.

Do ponto de vista conceitual, esse exercício de organização e representação do conhecimento permitiu reproduzir o caráter multifacetado e multidisciplinar dos processos de IA, representando sua natureza conforme observados no mundo real (Boserup, 1965) e indicando que seu entendimento deve considerar a interação e integração de diferentes perspectivas de variáveis ambientais, agronômicas, territoriais e socioeconômicas, considerando ainda a necessidade de analisar tais variáveis sob metodologias adequadas e em contextos institucionais específicos.

Como proposta de continuidade do trabalho, além de incrementar quantitativamente a terminologia, pretende-se proceder a definição dos termos e validar mais amplamente tal atividade com o apoio de especialistas, para enfim disponibilizá-los para consulta em um *website* específico com viés colaborativo. Adicionalmente, este trabalho integra outras pesquisas de mesma natureza, cujo objetivo é alcançar num futuro próximo um SOC amplo e representativo da agricultura brasileira.

- ☐ agricultural intensification (intensificação agropecuária)
 - ☐ methodology (metodologia)
 - ☐ field observations (observações de campo)
 - ☐ remote sensing (sensoriamento remoto)
 - ☐ satellites and sensors (satélites e sensores)
 - ☐ models and modeling (modelos e modelagem)
 - ☐ geographical information systems (sistemas de informação geográfica)
 - ☐ databases (bases de dados)
 - ☐ secondary data (dados secundários)
 - ☐ statistical analysis (análises estatísticas)
 - ☐ environment (meio ambiente)
 - ☐ physical environment (ambiente físico)
 - ☐ natural resources (recursos naturais)
 - ☐ biotic environment (ambiente biótico)
 - ☐ biodiversity (biodiversidade)
 - ☐ species composition (composição de espécies)
 - ☐ species richness (riqueza de espécies)
 - ☐ dominant species (espécie dominante)
 - ☐ exotic species (espécies exóticas)
 - ☐ wild species (espécies selvagens)
 - ☐ species diversity (diversidade de espécies)
 - ☐ native species (espécies nativas)
 - ☐ rare species (espécies raras)
 - ☐ biological diversity (diversidade biológica)
 - ☐ species diversity (diversidade de espécies)
 - ☐ environmental conditions (condições ambientais)
 - ☐ environmental variables (variáveis ambientais)
 - ☐ ecosystem (ecossistema)
 - ☐ agronomy (agronomia)
 - ☐ agricultural production (produção agrícola)
 - ☐ management practices (práticas de manejo)
 - ☐ farming systems (sistemas de cultivo)
 - ☐ pests and diseases (pragas e doenças)
 - ☐ pest management (manejo de pragas)
 - ☐ crops (culturas)
 - ☐ perennial crop (cultura perene)
 - ☐ cyclic crops (culturas cíclicas)
 - ☐ sugar cane (cana-de-açúcar)
 - ☐ crop cycle (ciclo da cultura)

(continua...)

(continuação...)

- ☐ annual crops (culturas anuais)
 - ☐ cereal crops (culturas de cereais)
 - ☐ grain crops (culturas de grãos)
- ☐ livestock (pecuária)
- ☐ agricultural products (produtos agropecuários)
- ☐ biomass (biomassa)
- ☐ extension services (serviços de extensão)
- ☐ territoriality (territorialidade)
 - ☐ total land (terra total)
 - ☐ land tenure (posse da terra)
 - ☐ private property (propriedade privada)
 - ☐ property rights (direitos de propriedade)
 - ☐ access to land (acesso à terra)
 - ☐ land scarcity (escassez de terra)
 - ☐ common property (propriedade comum)
 - ☐ estate farmers (proprietários rurais)
 - ☐ tenure security (segurança de posse)
 - ☐ land management (manejo das terras)
 - ☐ land use (uso das terras)
 - ☐ land use change (mudança de uso das terras)
 - ☐ land use patterns (padrões de uso das terras)
 - ☐ land use types (tipos de uso das terras)
 - ☐ national park (parque nacional)
 - ☐ land cover (cobertura das terras)
 - ☐ agricultural frontier (fronteira agrícola)
- ☐ socioeconomy (socioeconomia)
 - ☐ population dynamics (dinâmica populacional)
 - ☐ population density (densidade demográfica/populacional)
 - ☐ population growth (crescimento populacional)
 - ☐ population pressure (pressão populacional)
 - ☐ population size
 - ☐ goods and services (bens e serviços)
 - ☐ human population (população humana)
 - ☐ agricultural sector (setor agrícola)
 - ☐ development (desenvolvimento)
 - ☐ economic benefits (benefícios econômicos)
 - ☐ work/job (trabalho/emprego)
 - ☐ household income (renda familiar)
 - ☐ political economy (economia política)
 - ☐ private sector (setor privado)
 - ☐ public sector (setor público)
 - ☐ transaction costs (custos de transação)
- ☐ institution (instituição)
- ☐ geographic localization (localização geográfica)

Figura 2: Sistema de categorização dos termos/conceitos relacionados à organização e representação do conhecimento sobre intensificação agropecuária no Brasil (visualização em formato *folder tree*).

6 Referências

- Boserup, E. 1965. *The conditions of agricultural growth: the economics of agrarian change under population pressure*. London: G. Allen and Unwin; Chicago: Aldine.
- Brascher, M. e Café, L. 2008. Organização da informação ou do conhecimento? In *Anais do 9º Encontro Nacional de Pesquisa em Ciência da Informação*. São Paulo: ANCIB.
- Campos, M. L. D. A., Campos, L. M. e Medeiros, J. D. S. 2011. A representação de domínios de conhecimento e uma teoria de representação: a ontologia de fundamentação. *Informação & Informação*, 16, 2: 140–164.
- Carlan, E. e Medeiros, M. B. B. B. 2011. Sistemas de organização do conhecimento na visão da Ciência da Informação. *Revista Ibero-Americana de Ciência da Informação*, 4(2): 53-73.
- Pierozzi Jr., I., Oliveira, L. H. M. e Souza, K. X. S. 2010. Construindo ontologias de domínio: o (re)conhecimento da intensificação agropecuária no Brasil. In *Anais do 3o. Seminário de Pesquisa em Ontologia no Brasil*. Florianópolis, SC.: UFSC. pp. 100-109.
- Souza, R. R., Tudhope, D. e Almeida, M. B. 2010. The KOS spectra: a tentative typology of knowledge organization systems. *Advances in Knowledge Organization*, 12: 122-128.
- Torres, T. Z., Pierozzi Jr., Bernardes, R. M. e Vacari, I. 2010. Collaborative environments in RD&I institutions of the brazilian agricultural sector. *Journal of Technology Management & Innovation*, 5(3): 69-70.
- Torres, T. Z., Pierozzi Jr., I., Pereira, N. R. e Castro, A. 2011. Knowledge management and communication in brazilian agricultural research: an integrated procedural approach. *International Journal of Information Management*, 31(2): 121-127.
- Zeng, M. L. 2008. Knowledge organization systems (KOS). *Knowledge Organization*, 35(2-3): 160-182.

A representação e a organização do conhecimento em uma instituição pública de pesquisa: a implementação do Repositório do Conhecimento do Instituto de Pesquisa Econômica Aplicada (RCIpea)

Fábio Ferreira Batista
Instituto de Pesquisa Econômica Aplicada (Ipea)
Universidade Católica de Brasília (UCB)
fabio.batista@ipea.gov.br

Veruska da Silva Costa
Instituto de Pesquisa Econômica Aplicada - Ipea
veruska.costa@ipea.gov.br

Introdução

Neste artigo, apresentamos o processo de implementação de uma prática de Gestão do Conhecimento (GC) conhecida como repositório digital no Instituto de Pesquisa Econômica Aplicada (Ipea).

Entendemos GC neste trabalho como “... um conjunto de técnicas e ferramentas para identificar e utilizar os ativos de informação e conhecimento” (BATISTA, 2012) e “... uma abordagem sistemática e organizada para melhorar a capacidade da organização de mobilizar conhecimento para aumentar o desempenho (KPMG, 2003).

O processo de implementação do repositório no Ipea iniciou-se com a escolha do modelo de GC. Em seguida, utilizou-se de um roteiro – baseado nesse modelo – para elaborar o Plano de GC. O foco desse plano foi precisamente a implementação do repositório digital para eliminar uma importante lacuna de conhecimento, assim como para melhorar o desempenho institucional. Passamos, a seguir, a narrar como ocorreu a elaboração e implementação do RCIpea.

1. Modelo Teórico e Conceitual de Gestão do Conhecimento (GC)

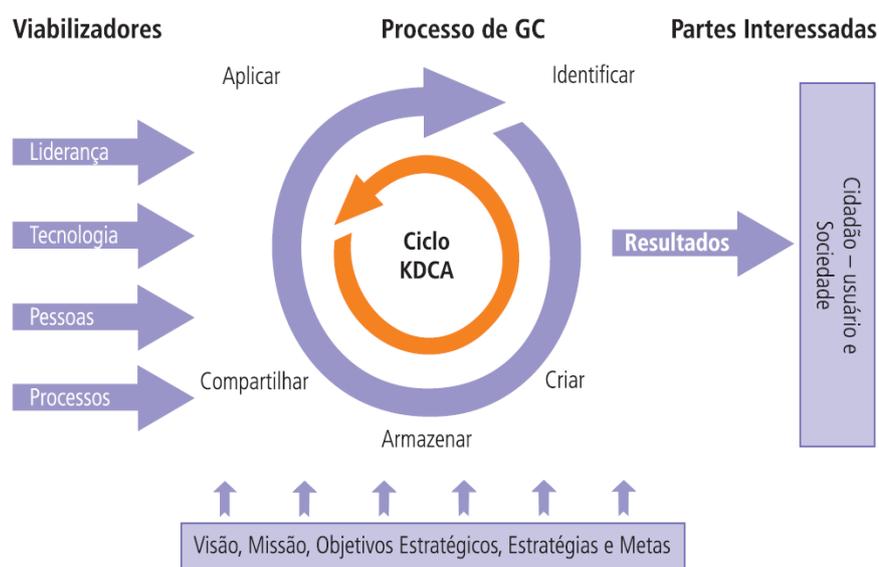
1.1. Modelo de Gestão do Conhecimento para a Administração Pública Brasileira

A **Figura 1** mostra os seis componentes do Modelo de Gestão do Conhecimento para a Administração Pública Brasileira proposto e adotado pelo Ipea, a saber: i) direcionadores estratégicos: visão, missão, objetivos estratégicos, estratégias e metas; ii) viabilizadores: liderança, tecnologia, pessoas e processos; iii) Ciclo de GC: identificar,

criar, armazenar, compartilhar e aplicar; iv) Ciclo KDCA; v) resultados de GC; e vi) partes interessadas: cidadão-usuário e sociedade (BATISTA, 2012).

As principais características deste modelo são: i) genérico (serve para todas organizações públicas); ii) holístico (permite um entendimento integral de GC); iii) com foco em resultados (visa alcançar os objetivos estratégicos e melhorar o desempenho); e iv) específico para a administração pública brasileira.

Figura 1. Modelo de Gestão do Conhecimento para a Administração Pública Brasileira



Fonte: BATISTA, 2012.

1.2. Roteiro para a Elaboração do Plano de Gestão do Conhecimento (GC)

O Quadro 1 mostra o Roteiro de Elaboração do Plano de GC utilizado pelo Ipea. São cinco as etapas: i) Onde estamos agora? ii) Onde queremos estar? iii) Como chegaremos lá? iv) Quais são as métricas ou indicadores para avaliar o grau de sucesso? e v) Estamos alcançando nossos objetivos e visão de GC?

Quadro 1. Objetivos dos Passos do Roteiro de Elaboração do Plano de GC

Etapa 1: Onde estamos agora?	
Passos	Objetivos
1	Autoavaliação do Grau de Maturidade em GC. A autoavaliação é baseada no Instrumento para a Avaliação do Grau de Maturidade em GC.
2	Identificação dos pontos fortes (PF) e oportunidades de melhoria (OM). Objetivo: identificar o que está bom (PF) e o que precisa ser melhorado (OM) para que a efetiva institucionalização da GC na organização.
3	Identificar as lacunas de conhecimento. Uma lacuna do conhecimento é a

	diferença entre o que a organização sabe e o que ela deveria saber para alcançar os objetivos estratégicos da organização. É onde ocorrerá a intervenção da GC.
Etapa 2: Onde queremos estar?	
4	Definir a visão de GC, isto é, a situação futura após a implementação da GC em que a lacuna de conhecimento já terá sido eliminada.
Etapa 3: Como chegaremos lá?	
5	Definição dos objetivos e estratégia de GC
6	Elaboração do Plano de Gestão do Conhecimento PGC
7	Elaboração do Plano de Melhoria do Grau de Maturidade em GC – PMM. O PMM aponta para as ações a serem executadas para eliminar as OMs identificadas na autoavaliação e, conseqüentemente, para elevar o grau de maturidade em GC na organização.
Etapa 4: Quais são as métricas para avaliar o grau de sucesso?	
8	Elaborar os Indicadores dos Resultados da Estratégia para verificar o sucesso na eliminação das lacunas do conhecimento.
9	Para avaliar se está havendo ou não elevação no grau de maturidade em GC, a organização comparará a pontuação obtida no diagnóstico atual com o diagnóstico anterior.
Etapa 5: Estamos alcançando nossos objetivos e visão de GC	
10	Monitorar, avaliar e elaborar relatórios.

Fonte: BATISTA, 2012

2. Implementação da Gestão do Conhecimento (GC) no Ipea

2.1 Roteiro para a Elaboração do Plano de GC

O Quadro 2 apresenta uma breve descrição de como o Ipea seguiu o roteiro apresentado na seção anterior para elaborar seu Plano de GC.

Quadro 2. Elaboração do Plano de GC

Passo	Comentários
1. Autoavaliação do grau de maturidade em GC	O Ipea obteve um total de 75,8 pontos (de um total de 210) na autoavaliação do grau de maturidade em GC. Esse resultado coloca a instituição no primeiro nível (“reação”). Os demais níveis são: i) iniciação (125 pontos); ii) Introdução (expansão – 146 pontos), refinamento (188 pontos) e maturidade (210 pontos).
2. Pontos Fortes e Oportunidades	No Critério 5.0 Processo de Conhecimento, identificou-se a OM “inexistência de um processo sistematizado de armazenamento, renovação, acesso e utilização do conhecimento produzido na

de Melhoria	instituição”. Essa OM ajudou a identificar a lacuna do conhecimento onde a intervenção de GC ocorreu.
3. Lacuna Estratégica de Conhecimento	Lacuna identificada: O conhecimento produzido pelo Ipea não está representado, organizado e armazenado de forma estruturada e não segue as recomendações internacionais para a disseminação, acesso e utilização da informação, visando aumentar a visibilidade da pesquisa e a interoperabilidade em rede.
4. Visão de GC	“Conhecimento produzido pelo Ipea representado, organizado e armazenado de forma estruturada, seguindo as recomendações internacionais para a disseminação, acesso e utilização da informação permitindo, assim, sua utilização adequada pelo público interno e externo.
5. Objetivos e Estratégia de GC	<u>Objetivo principal</u> : aumentar a visibilidade e o acesso à pesquisa e a interoperabilidade de toda a produção do Ipea em rede. <u>Estratégia</u> : implantar processo de representação, organização, armazenamento, renovação, disseminação, acesso e utilização do conhecimento produzido nos projetos e processos de apoio e finalísticos do Ipea.
6. Plano de GC	O Plano de GC teve como foco a construção do protótipo do Repositório do Conhecimento do Ipea (RCIpea) e será apresentado na próxima seção.
7. PMM	Ainda não elaborado
8. Indicadores	O Ipea elaborou indicadores referentes aos seguintes processos: i) armazenamento/disseminação (4); ii) acesso (6); e iii) utilização (4).

Está prevista a realização anual da autoavaliação do grau de maturidade para se obter a pontuação atualizada (Passo 9). Além disso, o Ipea deverá realizar mensalmente o monitoramento, avaliação e elaboração de relatórios (Passo 10).

O Ipea seguiu o Roteiro para a Elaboração do Plano de GC para eliminar a lacuna do conhecimento referente aos processos de representação, organização, disseminação, acesso e utilização do conhecimento produzido. Na próxima seção, detalharemos como o Ipea executou o projeto piloto, isto é, como a entidade construiu o protótipo do RCIpea.

2.2 Repositório do Conhecimento do Ipea (RCIpea)

Como vimos, essa prática foi implementada para eliminar a lacuna estratégica do conhecimento identificada no passo 3 do Roteiro para a Elaboração do Plano de GC e está ancorada no modelo de GC adotado pela instituição.

O protótipo do Repositório do Conhecimento do Ipea foi construído por um grupo de trabalho (GT) multidisciplinar. Além de legitimar o projeto, o GT envolveu áreas cujo apoio foi fundamental para o êxito da iniciativa.

A elaboração e implementação do RCIpea teve sete marcos importantes. O primeiro foi a definição das diretrizes para o desenvolvimento do acervo. Essas diretrizes serviram de base para o planejamento das coleções e ofereceram parâmetros de consistência e qualidade ao acervo. O segundo marco foi a instalação e customização do software Dspace e aplicação da identidade visual. Em seguida, realizou-se o estudo do domínio do repositório e definiu-se sua estrutura hierárquica (arquitetura informacional). O quarto marco foi a definição do padrão de metadados e o formulário de entrada de dados. O conjunto de metadados do RCIpea foi definido segundo o Padrão Dublin Core. O quinto marco foi a definição do fluxo geral do repositório e o fluxo de submissão dos objetos digitais. A definição da Política de Acesso Aberto e Direitos Autorais foi o sexto marco do projeto. Essa política servirá para garantir a não violação dos Direitos Autorais e para nortear a implementação do acesso aberto e gestão de tais direitos. Finalmente, o sétimo e último marco foi a definição de regras de preenchimento dos metadados.

3. Considerações Finais

Com a implementação do RCIpea, o conhecimento produzido pelo Ipea pode hoje ser representado, organizado e armazenado de forma estruturada. 2013 será o ano da consolidação do repositório digital quando parte significativa da produção técnica e científica da instituição será inserida no RCIpea, permitindo, assim, seu acesso e utilização pelo público interno e externo. É importante destacar, finalmente, o modelo inovador de implementação de GC adotada pela instituição onde a prática adotada (repositório digital) servirá para eliminar uma lacuna estratégica do conhecimento e para o cumprimento da missão institucional.

REFERÊNCIAS

BATISTA, Fábio Ferreira. Modelo de gestão do conhecimento para a administração pública brasileira: como implementar a gestão do conhecimento para produzir resultados em benefício do cidadão. Brasília, DF: Instituto de Pesquisa Econômica Aplicada (Ipea), 2012.

KPMG. Insights from KPMG's European Knowledge Management Survey 2002/2003.

Disponível em: <http://ep2010.salzburgresearch.at/knowledge_base/kpmg_2003.pdf> Acessado em: 10 dez. 2011.

(*) Ipea: Instituto de Pesquisa Econômica Aplicada

A integração museu/biblioteca na organização e representação do conhecimento de documentos artísticos

Giovana Deliberali Maimone
Universidade de São Paulo
bci.gdm@gmail.com

Maria De Fátima Gonçalves Moreira Tálamo
Universidade de São Paulo
mfgmtala@usp.br

1. INTRODUÇÃO

A integração entre museu e biblioteca dentro das instituições museológicas é de extrema importância, devido às possibilidades de resultados mais completos para a curadoria do museu na organização de exposições, para a biblioteca como fonte de referência e informação e também para o público que o visita, pela percepção de complementaridade das obras.

Associam-se, portanto, a obra material e suas diferentes formas de representação, possibilitando a observação da relação entre acervo museológico e bibliográfico – sob um ponto de vista dinâmico e mais eficiente em nível comunicacional. Segundo Smit (2010: 34)

[...] um museu não é composto unicamente do acervo museológico, pois ele ainda necessita de documentos de gestão do acervo, de apoio para a pesquisa e para seu funcionamento institucional.

Dentre os diferentes tipos de documentos (aqui denominados genericamente de documentos artísticos) presentes nos museus, há interesse por dois conjuntos em particular: os do acervo museológico, que contém os objetos (peças, pinturas, esculturas); e os do acervo bibliográfico, que são compostos de materiais que servem para apoiar pesquisas desenvolvidas no museu. Esse último justifica-se por sua função de complemento das necessidades informacionais de artistas, pesquisadores, curadores, do público genérico interessado, entre outros, funcionando como fonte de informações (Smit, 2010).

É fato que a gestão do acervo museológico não opera com o tratamento temático das obras e sim com o que é mostrado por elas (exposição, curadoria, pesquisa). Porém,

sua integração com o acervo bibliográfico é importante, visto que as mesmas são fontes para diversos tipos de documentos, como livros, catálogos, periódicos e demais. A documentação deste tipo de acervo, ou seja, das obras, é referente ao levantamento informacional sobre a origem dos objetos, sua descrição (materiais, métodos, inscrições), pesquisas de contextualização, localização física e ações de preservação, entre outras (Smit, 2010).

Em vista disso, o presente estudo trabalha com as informações do documento artístico, no intuito de proporcionar melhor comunicação das obras de e sobre artes tendo como base a investigação terminológica dessa área de conhecimentos. Embora a relação sensível existente entre obra e público em todo museu de arte seja fundamental para o alcance de seus objetivos, esta pesquisa enfoca o trabalho com a informação das obras que admitem o fluxo e a geração de conhecimentos por meio de sua codificação/decodificação.

A delimitação entre o trabalho com o acervo do museu (objetos, pinturas, esculturas etc.) e o material impresso sobre as obras (livros, catálogos, periódicos etc.) está centrada na função que cada material deve desenvolver. O acervo museológico deve reproduzir informações descritivas sobre as obras, como título, autor, técnica, época, estilo, entre outras, que permitam conhecê-las. O tratamento temático sobre as mesmas restringiria a beleza polissêmica inerente a elas. Porém, a exploração de informações contextualizadas sobre o autor, o título da obra e o movimento artístico ao qual pertence, por exemplo, seria possível e viável por meio do tratamento terminológico.

Já para materiais bibliográficos, a representação temática (classificação, indexação) seria pertinente para identificar de que tratam as obras impressas, servindo-se, então, de linguagens documentárias como os tesouros para o relacionamento dos/entre assuntos.

Um tipo de mediação diferenciada é adequado, devido à condição de efetivar pontes entre documentos e usuários (diversos). Sob esse ponto de vista, o vocabulário controlado (tesauro) funcionaria como um *software* que disponibilizaria uma rede de representações através de sua função comunicativa; função esta melhor realizável através do apoio terminológico.

Identificar e estabelecer os espaços favoráveis à geração do conhecimento na relação museu/biblioteca, para que ambos trabalhem de modo integrado, encaminham-

se ao encontro de um dos objetivos cruciais da Ciência da Informação, que é o de propiciar acesso informacional aos mais diversos públicos.

As formas de organização dos documentos podem ser limitantes e restritivas, visto que apresentam “um” ponto de vista e não “o” único, pois são elaboradas a partir de objetivos e normas previamente estipuladas. Em virtude das linguagens documentárias serem elaboradas para uma área específica e se referirem às áreas de especialidade, elas tornam os documentos inacessíveis intelectualmente para as classes sociais mais populares, que contam com vocabulário menos especializado.

Levando em consideração que o público dos museus, de modo geral, é bastante heterogêneo e as obras que comportam possibilitam múltiplas interpretações, intenta-se reduzir o cerceamento representativo propiciado pelas linguagens documentárias, por meio de uma ampliação vocabular, pela integração de novos termos equivalentes, genéricos, específicos e associativos. Tal ampliação deve ser realizada por intermédio de fontes que amparem pesquisas de termos.

Uma adequada articulação de termos e conceitos voltados às camadas menos especializadas seria imprescindível para agregar novos expectadores, potencialmente interessados em adquirir conhecimentos da área artística, como é o caso dos públicos das grandes exposições, que nem sempre são artistas, historiadores, escultores, críticos ou outros.

A preocupação principal, portanto, é a efetiva comunicação do acervo museológico e bibliográfico dos museus com os diversos públicos. Recorre-se, para tanto, ao trabalho de identificação, adequação e inserção de termos para ampliar possibilidades de transmissão da informação, tentando com isso preservar a memória e contribuir para a dinâmica cultural.

2. PROPOSTA DE APERFEIÇOAMENTO DOCUMENTÁRIO PARA OBRAS DO ACERVO MUSEOLÓGICO

O acervo museológico, composto de obras, objetos e demais artigos de interesse para exposições são agrupados sobre determinados critérios e, de modo descritivo, exibem algumas informações como título da obra, nome do artista, período de elaboração da obra, entre outras.

Diferentes são as formas de descrição e apresentação das obras nos museus, porém, embora custoso e exaustivo, propõe-se como complemento que um trabalho terminológico de orientação semasiológica seja implantado. Tal trabalho procederia do termo ao conceito, cujo produto seria uma lista alfabética de termos ou expressões acompanhados por seus respectivos significados, para cada obra do acervo permanente do museu, respeitando os limites de cada campo, como demonstrado no exemplo abaixo:

Figura 1: A origem da Via-Láctea, de Tintoretto



Ficha Descritiva

Artista: Jacopo Tintoretto

Dados Biográficos: Veneza – Itália, 1518 – 1594.

Título: A origem da Via-Láctea

Tipo de obra: Pintura

Técnica: Óleo sobre tela

Dimensões: 124,5 x 165 cm

Estilo Artístico: Maneirismo

Data de realização da obra:

1577 - 1578

Na biblioteca teria uma - Lista Informativa sobre as Obras do Museu:

A Origem da Via-Láctea: A composição dessa tela diz respeito a uma imagem desequilibrada pela remoção de faixa na parte inferior...

Jacopo Tintoretto foi o mais prolífico pintor de Veneza no fim do século 16. No início de sua carreira, ele lutou para conseguir o reconhecimento...

3. PROPOSTA DE APERFEIÇOAMENTO DOCUMENTÁRIO PARA OBRAS DO ACERVO BIBLIOGRÁFICO

Na Biblioteca, a atribuição de termos aos documentos artísticos (neste caso, impressos) é feita após análise dos materiais e a consulta ao tesauro devidamente atualizado. A terminologia seria utilizada para analisar contextos, a fim de propor remissivas e mais relações entre termos, em linguagem menos específica.

Tomando-se como exemplo o assunto “Maneirismo” e já devidamente aplicada a metodologia terminológica, teríamos as seguintes figuras, abaixo relacionadas:

Figura 2: Apresentação gráfica das relações do “Maneirismo”



Figura 3: Apresentação gráfica tesauroística das relações do “Maneirismo”

MANEIRISMO

UP Arte Maneirista
 UP Estilo Maneirista
 UP Maniérisme
 UP Manierismo
 UP Manierista
 UP Maniériste

TG ARTE EUROPEIA

TE MANEIRISMO ITALIANO
 TE MANEIRISMO DOS PAÍSES BAIXOS

TR RENASCENÇA
 TR INQUIETAÇÃO NEURÓTICA
 TR DISTORÇÃO DAS FORMAS
 TR TENSÕES E OPOSTOS
 TR CONTRADIÇÕES
 TR PARADOXO
 TR VIRTUOSIDADE EXCESSIVA
 TR IMITAÇÃO SUPERFICIAL DO CLÁSSICO
 TR DESPROPORÇÃO FIGURATIVA
 TR ÊNFASE NA SUBJETIVIDADE

TR JACOPO PONTORMO
TR FIORENTINO ROSSO
TR DOMENICO BECCAFUMI
TR PARMIGIANINO
TR GIORGIO VASARI
TR AGNOLO BRONZINO
TR EL GRECO
TR JACOPO TINTIRETTO
TR BARTHOLOMAEUS SPRANGER
TR HANS VON ASCHEN
TR GIULIO ROMANO
TR ANDREA PALLADIO

4. CONSIDERAÇÕES FINAIS

A ideia de apresentar as obras do acervo museológico remetendo a informações complementares na biblioteca e de aumentar o número de termos do acervo bibliográfico para proporcionar ampla representatividade dos documentos artísticos centra-se na preocupação de “torná-las acessíveis” a um maior número de pessoas, objetivando desfazer fronteiras socioculturais tradicionalmente construídas.

Porém, trabalhar de forma isolada implica a perda de referências setoriais do museu, pois impossibilita a integração dos fluxos de trabalho, impedindo também a padronização das formas de apresentação das obras aos públicos e o realinhamento dos objetivos das instituições.

307

REFERÊNCIAS

Smit, Johanna Wilhelmina. 2010. O arquivo de museu e a informação. In: SEMINÁRIO INTERNACIONAL ARQUIVOS DE MUSEUS E PESQUISA, 1. São Paulo. *Anais...* MAC USP, p. 84 – 92.

Ethics and epistemology of classification

Jens-Erik Mai
University of Copenhagen
jenserikmai.info

Knowledge organization theory and practice have in recent years been pushed in a more relativistic orientation. While traditional, mainstream knowledge organization thinking was grounded in naïve realism, contemporary scholarship is oriented toward the social and cultural aspects that determine how classifications and other knowledge organization structures are made. This orientation toward the social and cultural has reconceptualized classification theory and demanded a new epistemological basis.

In this talk I will first outline this new epistemological basis for contemporary classification scholarship. I will then argue that what has been missing in this reconceptualization of classification research toward what we can call a ‘relativistic’ approach is a sound ethical foundation. Whereas much contemporary classification research ends with calls to editors of classifications to make their epistemological positions clear, I will argue that that alone is not enough – they also need to be held accountable to their ethics and moral positions.

After providing the conceptual basis for the epistemological and ethical foundation for contemporary classification research, I will conclude the talk by providing a framework for a practice-based ethics of classification.

Knowledge organization for its own sake: relationships between theory and applications

Claudio Gnoli (Università di Pavia)
gnoli@aib.it

The domain of knowledge organization (KO) consists of several layers: KO theory looks for the best principles to arrange the parts of human knowledge into general, consistent views; KO systems provide structures and symbols to do it; KO representation records these as formalized, nowadays digital, data; KO applications make use of all this for the particular needs of various information services. It is important that each of these layers be in sound relationships with the other ones. I have elsewhere discussed relationships between theory and systems [Gnoli 2011a] and between systems and representation [Gnoli 2011b]. In this occasion I intend to focus on relationships between theory (together with systems and representations) and applications. It is often said that KO should be driven by the needs of its users, so that user surveys can be a major source for KO theory. Supporters of domain analysis believe that each community needs a different organization of knowledge, so their approach also starts from the final layer to determine the shape of the previous ones. I would like to propose an approach that is in some way opposite: KO may identify theoretical principles, and produce certain systems, that are not in compliance with any particular need, beside the intellectual satisfaction of understanding the world and our knowledge of it. Such desire has existed among men in all times, although it is now often forgotten due to the dominance of technology and pragmatics in modern life. One could argue that KO for its own sake be a useless luxury that we cannot afford. However, experience with research teaches that many major innovations originated from research that was purely theoretical. None can know which will ever be the applications of discovering the Higgs boson, despite scientists feel that understanding it is a central aim of contemporary research. Research on graphene, a form of carbon atom lattice, originally had not any application purpose, but was later found extremely productive and their authors were awarded the Nobel Prize for it in 2010. In the same way, KO theory, and systems reflecting it, should be allowed to develop freely without necessarily having any particular application in sight. As witnessed by Ernst Mayr [1981], KO systems that best reflect the properties of reality as it is often turn out to be also the most efficient ones for the purposes of information retrieval.

Trabalhos premiados

Os seguintes trabalhos premiados serão publicados na revista *Knowledge Organization*:

Relações semânticas em sistemas de organização do conhecimento

Marisa Bräscher Basílio Medeiros

A contribuição da Bibliometria para os estudos metateóricos e de Análise de Domínio

Renata Cristina Gutierres Castanha / Maria Cláudia Cabrini Grácio

Desafios para a classificação de canções populares

Rodrigo de Santis/ Rosali Fernandez de Souza