



VIII ISKO BRASIL 2025



Uso da Sumarização Automática de Textos na Representação da Informação e do Conhecimento

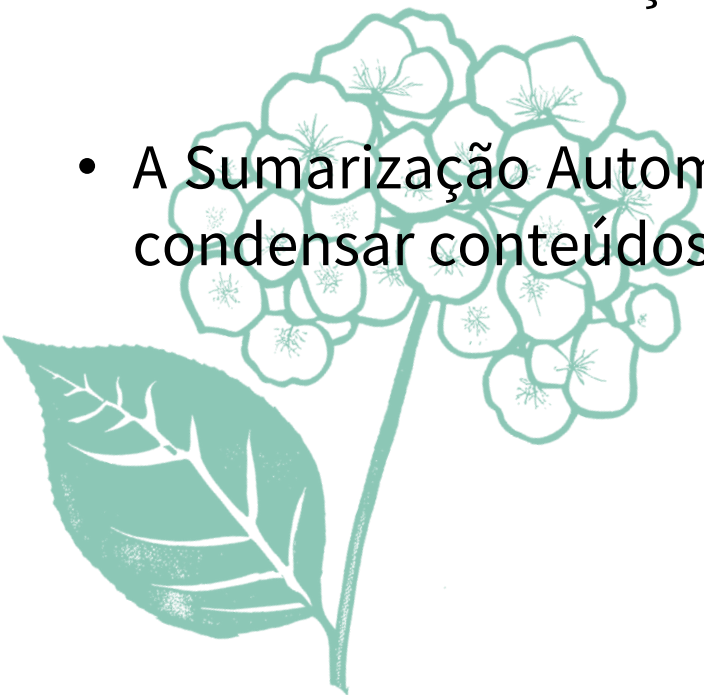
João Matheus Nascimento Rodrigues*
Valdirene Pereira da Conceição*



Arquivar São Luís*; mateusnasc33@gmail.com
Universidade Federal do Maranhão*; valdirene.pc@ufma.br



- **Crescimento exponencial da informação digital** demanda métodos mais eficazes de organização, acesso e recuperação da informação.
- Tecnologias como o Processamento de Linguagem Natural (PLN) têm sido aplicadas na Ciência da Informação para melhorar esses processos.
- A Sumarização Automática de Textos (SAT) **surge como uma solução** promissora para condensar conteúdos e destacar informações relevantes.

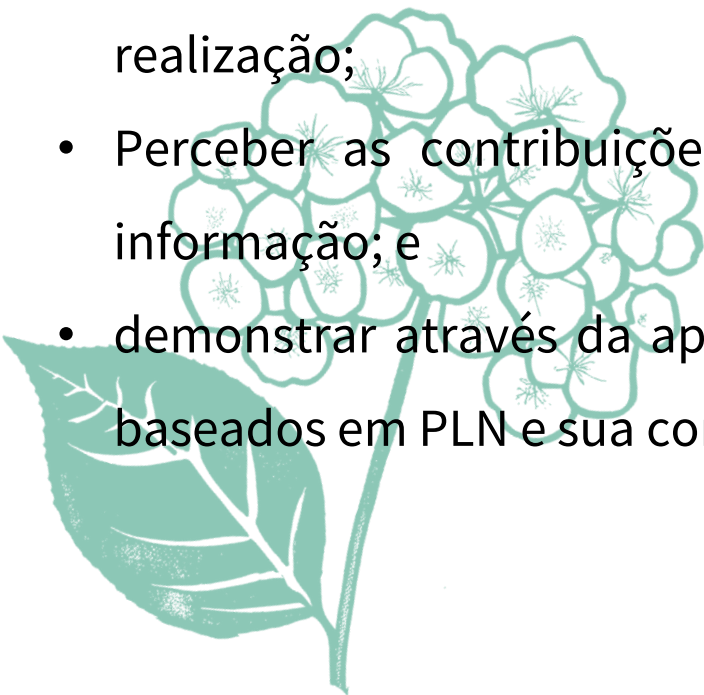


PROBLEMA

Como a Sumarização Automática de Textos pode contribuir para a representação temática da informação em ambientes informacionais?

OBJETIVOS

- Conhecer o processo de sumarização automática de textos e as ferramentas disponíveis para a sua realização;
- Perceber as contribuições da SAT para a representação temática, organização e recuperação da informação; e
- demonstrar através da aplicação da sumarização automática de textos, o funcionamento de sistemas baseados em PLN e sua contribuição para a Organização da Informação e do Conhecimento.



- Tipo de Pesquisa: aplicada, exploratória e qualitativa, voltada à Sumarização Automática de Textos em ambientes informacionais.
- Utiliza pesquisa bibliográfica e documental com fontes acadêmicas para fundamentação teórica.
- Foram testados dois sumarizadores extrativos, um de software livre via web e outro comercial para desktop, aplicados ao corpus, composto por cinco artigos científicos sobre Covid-19.

(Marconi; Lakatos, 2003; Prodanov e Freitas, 2014; Gil, 2010).



Fundamentos: Sumarização Automática de Textos na Representação da Informação



VIII ISKO BRASIL 2025



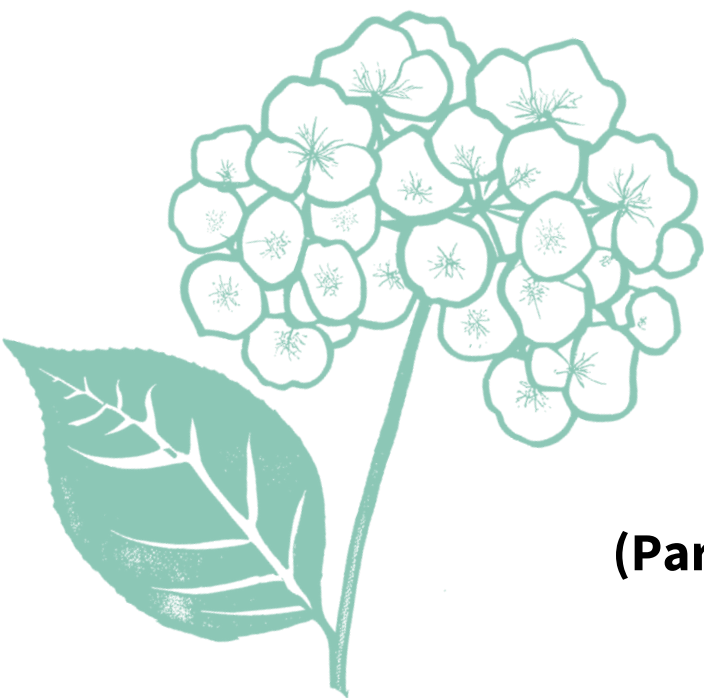
A SAT é um processo computacional que condensa automaticamente um ou mais textos, simulando a produção humana de resumos e facilitando o acesso rápido à informações.

Textos de entrada são submetidos por processos de análise, transformação e síntese, culminando em um texto de saída mais sucinto, condensado e que apresente as principais informações extraídas da estrutura de um ou mais documentos.

(Pardo, 2008).



Utilizada na indexação e recuperação da informação, a SAT destaca conteúdos temáticos relevantes e se aproxima dos processos tradicionais de representação da informação.



A SAT enfrenta desafios relacionados à extração de termos significativos, mas apresenta grande potencial para otimizar o tratamento informacional, especialmente diante do volume crescente de dados digitais.

(Pardo, 2008).



Sumarização Automática de Textos: uso exploratório na Representação da Informação

- **Tipos de sumários (Pardo, 2008):**
 - Informativo, indicativo e crítico (função);
 - Genéricos ou focados (audiência);
 - Extrato ou abstract (formação).
- **Abordagens da SAT:**
 - Superficial: baseada em estatísticas simples (ex.: palavras-chave);
 - Profunda: usa conhecimento linguístico (ex.: análise semântico-discursiva);
 - Híbrida: combina as duas anteriores.
- **Classificações do processo:**
 - Monodocumento: resume um único texto;
 - Multidocumento: integra conteúdos de vários textos.
- **Etapas do processo:**
 - Análise → Transformação → Síntese.

SUMARIZADOR	CATEGORIA	ENDEREÇO WEB
<u>Esummarizer</u>	Proprietário/Comercial	www.esummarizer.com
FreeSummarizer	Proprietário	www.freesummarizer.com/
Gist Summarizer	Acadêmico	www.gistsummarizer.com
Intellexer Summarizer NE	Proprietário/Comercial	www.intellexer.com/summarizer-ne
Linguakit	Livre/Gratuito	www.linguakit.com
Neural Summarizer	Acadêmico	www.neuralsummarizer.com
Resoomer	Livre/Gratuito	www.resoomer.com
SMMRY	Livre/Gratuito	www.smmry.com
SummarizeThis™	Livre/Gratuito	www.summarizethis.com
Summary Generator	Livre/Gratuito	www.summarygenerator.com
Text Compactor	Livre/Gratuito	www.textcompactor.com
Text Summarizer	API/Livre/Gratuito	www.textsummarizer.com
Turbine Text	Livre/Gratuito	www.turbinetext.com
ChatGPT	Proprietário/Comercial	chat.openai.com
TLDR This	Livre/Gratuito	www.tldrthis.com
Paperpile Summarizer	Proprietário/Comercial	www.paperpile.com/summarizer
Scholarcy	Proprietário/Comercial	www.scholarcy.com
<u>Prepostseo Summarizer</u>	Livre/Gratuito	www.prepostseo.com/summarizer



Aplicações da Sumarização Automática de Textos na Representação da Informação



VIII ISKO BRASIL 2025



- **CrITÉrios de escolha dos softwares:**

- Sumarizadores extrativos, por serem mais comuns.
- Um de software livre (Turbine Text) e outro comercial (Intellexer Summarizer).
- Preferência por ferramentas pouco exploradas academicamente.

- **Corpus utilizado:**

- Cinco artigos científicos sobre Covid-19, obtidos no Portal da Capes.
- CritÉrios: qualidade, rigor acadÊmico e temÁtica atual.

- **Procedimentos adotados:**

- Textos convertidos em .txt para compatibilidade com as ferramentas.
- Remoção de resumos e metadados para não interferir na sumarização.

- **Objetivo da aplicação:**

- Avaliar o desempenho dos SSA na representação temática.
- Observar como lidam com novas terminologias e estruturas acadêmicas.

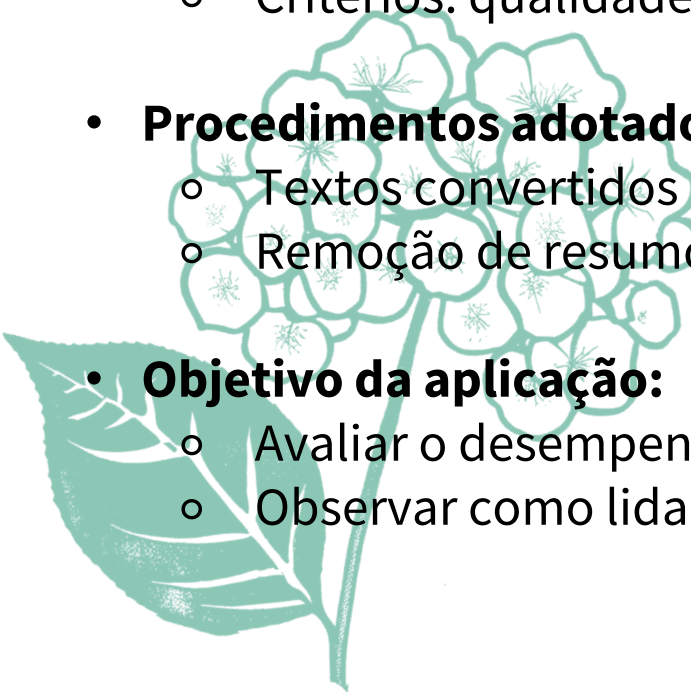


Ilustração processo de sumarização no sistema: TURBINE TEXT



VIII ISKO BRASIL 2025



Gerador de Texto Automático - x +

turbinetext.com/pt/resumo

Português ▾ Resumir Pouco... ▾

→

Digite seu texto aqui...

Aqui aparecerá o texto resumido.

Resumir texto

Acesso gratuito 0 caracteres | 0 palavras

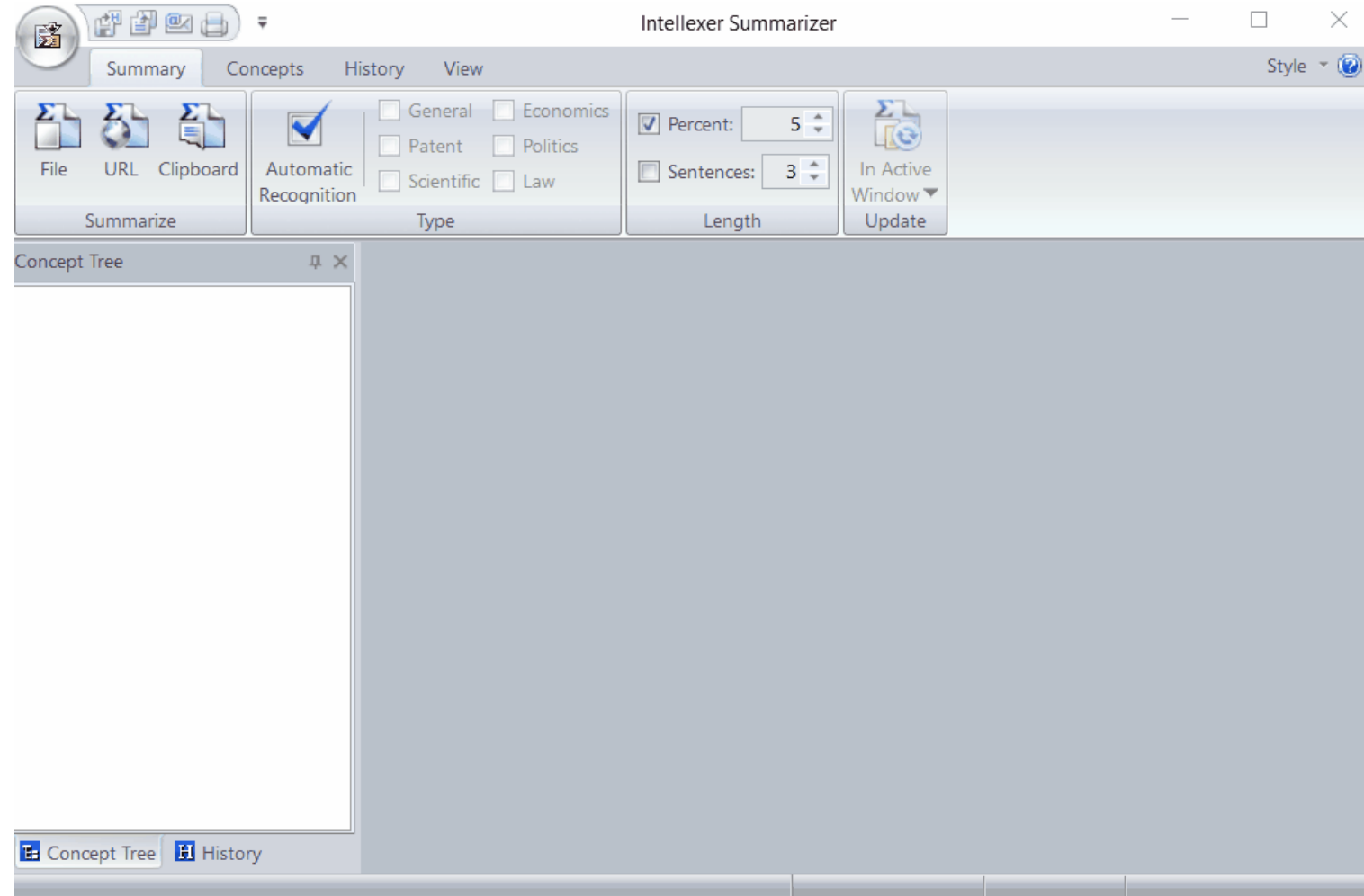
0 caracteres | 0 palavras



Ilustração processo de sumarização no sistema: INTELLEXER SUMMARIZER



VIII ISKO BRASIL 2025



Resumo original vs Resumo Gerado Automaticamente

Resumo Original	Turbine Text	Interllex Summarizer
Objetivo: Esta comunicação tem como objetivo discutir as Fake News no cenário brasileiro de COVID-19. Desenvolvimento: no período entre 29 de janeiro e 31 de março de 2020, foi realizada uma busca no banco do Ministério da Saúde brasileiro e foram identificadas 70 Fake News sobre o COVID-19, sendo: 40 informações relacionadas aos discursos de autoridades na saúde, 17 sobre terapêutica, nove com medidas de prevenção, duas referentes aos prognósticos da doença e duas de vacinação. Conclusão: a literatura sobre o tema pode ser escassa, no entanto, observa-se que a velocidade de produção de Fake News é relevante, especialmente ao pressupor o impacto social e a capacidade de circulação destas notícias.	Anos mais tarde (2012), o MERS-Cov, outro tipo de coronavírus, provocou surto na Arábia Saudita, propagando o contágio por meio de perdigotos e deficiência na higiene, especialmente das mãos, dentre outros hábitos hodiernos da cultura dos cuidados, para além das lacunas nas políticas públicas, quando 27 países foram atingidos. Em tempos de avanços tecnológicos, estas notícias falsas são veiculadas nas redes sociais, de forma rápida e multiplicada entre a população, que, em linguagem metafórica, pode-se entender como um vírus que contamina a comunicação e promove ações e comportamentos contrários às orientações das autoridades técnicas no campo da saúde. Estas, com termos técnicos, próprios dos centros de pesquisa, precisam ser decodificadas à população para melhor entendimento, o que remete à aplicação da técnica da comunicação denominada de AIDA — Atenção, Interesse, Desejo e Atitude — utilizada pelos jornalistas para a imprensa social. Elas atraem aos que as disseminam quando há interesse, mas possuem sobrevida curta, bem como fazem os operadores das comunicações até que outro assunto seja mais interessante para a indústria e/ou comércio das informações, visando aos leitores na formação de opinião pública e as redes sociais como seus consumidores.	A busca das notícias Fake News ocorreu no banco de dados do Ministério da Saúde, no cenário da pandemia de COVID-19, no período de 29 de janeiro a 31 de março de 2020, quando foram identificados 70 registros. Refletir sobre as Fake News na contemporaneidade é pensar nas publicações com base nas evidências científicas. Isto se articula com a intencionalidade que a indústria da informação e seus operadores aplicam às matérias jornalísticas, no intuito do efeito de real, no sentido de poder fazer e crer, o que se faz ver. Por um lado, não se pode negar que a tipificação não foi apresentada à análise, o que permite lacunas nesta comunicação; por outro lado, instiga o aprofundamento das tipificações para análise em estudos futuros. Ademais, destaca-se que a literatura brasileira é escassa sobre a pandemia de COVID-19 e a velocidade desta produção do conhecimento vai de encontro com a produção das Fake News. Não obstante, enfatiza-se a necessidade de a população conhecer o site do Ministério da Saúde brasileiro, o qual aponta as Fake News, para que ocorra educação em saúde com informações corretas e seguras.



Avaliação dos resumos originais vs resumos sumarizados

Critérios de Avaliação Utilizados

- Baseados em Dias da Silva et al. (2007), Lancaster (2004), Dang (2004, 2005).
- Critérios:
 - Clareza e concisão.
 - Fidelidade ao conteúdo original.
 - Abrangência.
 - Objetividade.
 - Gramática e ortografia.

Classificação dos Resumos

- Ótimo: Atende plenamente aos critérios de avaliação.
- Bom: Atende à maioria dos critérios, com pequenos problemas.
- Regular: Apresenta falhas relevantes em um ou mais critérios.

Quadro de Avaliação

Critérios de Avaliação	Resumo Original	Turbine Text	Interllex Summarizer
Clareza e Consisão	Ótimo	Bom	Regular
Fidelidade ao conteúdo original	Ótimo	Bom	Regular
Abrangência	Ótimo	Bom	Regular
Objetividade	Ótimo	Bom	Regular
Gramática e Ortografia	Ótimo	Bom	Bom

O uso de ferramentas de Sumarização Automática de Textos (SAT) **contribui para agilizar a organização e representação temática da informação**, especialmente no processo de indexação.

Essas ferramentas, quando integradas ao Processamento de Linguagem Natural (PLN), podem melhorar significativamente a eficiência em ambientes informacionais.

No entanto, é necessário aprofundar pesquisas teóricas e práticas para fortalecer o uso dessas tecnologias, além de capacitar profissionais para sua aplicação adequada.

Obrigado!!!



- ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS. **NBR 6028**. Informação e documentação: resumos: apresentação. Rio de Janeiro, 2021.
- BRASCHER, Marisa; CAFÉ, Lígia. Organização da Informação ou Organização do Conhecimento? In: Encontro Nacional de Pesquisa em Ciência da Informação, 9., 2008, São Paulo. **Anais[...]**. São Paulo: Ancib, 2014. p. 1 - 14. Disponível em: <http://enancib.ibict.br/index.php/enancib/ixenancib/paper/view/3016>. Acesso em: 2 jan. 2025.
- CARDOSO, Paula Christina Figueira. **Exploração de métodos de sumarização automática multidocumento com base em conhecimento semântico-discursivo**. 2014. Tese (Doutorado em Ciências de Computação e Matemática Computacional) - Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos, 2014. doi:10.11606/T.55.2014.tde-16032015-161912. Acesso em: 03 jan. 2025.
- CARDOSO, Paula Christina Figueira. **Exploração de métodos de sumarização automática multidocumento com base em conhecimento semântico-discursivo**. 2014. Tese (Doutorado em Ciências de Computação e Matemática Computacional) - Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos, 2014. doi:10.11606/T.55.2014.tde-16032015-161912. Acesso em: 03 jan. 2025.
- CARMO, Juliana Rabelo do; CONCEIÇÃO, Valdirene Pereira da. Processamento da linguagem natural do domínio musical: do sentido à gestão terminológica no ambiente Etermos. **Informação & Informação**, Londrina, n. 3, v. 23, p.314-341, set./dez. 2018. Disponível em: <http://www.uel.br/revistas/uel/index.php/informacao/article/view/29622/pdf>. Acesso em: 22 dez. 2024.
- DANG, Hoa Trang. Overview of DUC 2005. In: **Proceedings of the Document Understanding Conference**, 2005. Disponível em: <https://duc.nist.gov/pubs/2005papers/OVERVIEW05.pdf>. Acesso em: 15 jan. 2025.
- DIAS DA SILVA, Bento Carlos et al. **Introdução ao Processamento das Línguas Naturais e Algumas Aplicações**. São Carlos: NILC - ICMC-USP, 2007. 121 p. (Série de Relatórios do Núcleo Interinstitucional de Linguística Computacional). Disponível em: <http://conteudo.icmc.usp.br/pessoas/taspardo/NILCTR0710-DiasDaSilvaEtAl.pdf>. Acesso em: 25 jan. 2025.
- GIL, Antonio Carlos. **Como elaborar projetos de pesquisa**. 5. ed. São Paulo: Atlas, 2010. 183p.
- LANCASTER, F. W. **Indexação e Resumos**: teoria e prática. 2. ed. Brasília: Briquet de Lemos/livros, 2004.
- MARCONI, Marina de Andrade; LAKATOS, Eva Maria. Fundamentos de metodologia científica. São Paulo: Atlas, 2003.
- MARTINS, Camilla Brandel. **Introdução à sumarização automática**. 2001. Disponível em: <https://sites.icmc.usp.br/taspardo/RTDC00201-CMartinsEtAl.pdf>. Acesso em: 19 dez. 2024.

- PARDO, Thiago Alexandre Salgueiro. **Sumarização Automática**: Principais Conceitos e Sistemas para o Português Brasileiro. São Carlos: NILC - ICMC-USP, 2008. 15 p. (Série de Relatórios do Núcleo Interinstitucional de Linguística Computacional). Disponível em: <http://conteudo.icmc.usp.br/pessoas/taspardo/NILCTR0804-Pardo.pdf>. Acesso em: 22 dez. 2024.
- PRODANOV, Cleber Cristiano; FREITAS, Ernani Cesar de. **Metodologia do trabalho científico**: métodos e técnicas da pesquisa e do trabalho acadêmico. 2. ed. Novo Hamburgo: Feevale, 2013. Disponível em: <http://www.feevale.br/Comum/midias/8807f05a-14d0-4d5b-b1ad-1538f3aef538/E-book%20Metodologia%20do%20Trabalho%20Cientifico.pdf>. Acesso em: 22 jan. 2025.
- RINO, Lúcia Helena Machado; PARDO, Thiago Alexandre Salgueiro. A Sumarização Automática de Textos: principais características e metodologias. In: **Anais do XXIII Congresso da Sociedade Brasileira de Computação, Vol. VIII: III Jornada de Minicursos de Inteligência Artificial (III MCIA)**, 2003, p. 203-245. Campinas-SP.
- SANTOS, Ângelo Filipe da Silva dos. **Sumarização automática de texto**. 2012. 76 f. Dissertação (Mestrado) - Curso de Mestrado em Engenharia Informática, Universidade da Beira Interior, Covilhã, 2012. Disponível em: <https://ubibliorum.ubi.pt/handle/10400.6/3738>. Acesso em: 20 jan. 2025.
- SANTOS, Ângelo Filipe da Silva dos. **Sumarização automática de texto**. 2012. 76 f. Dissertação (Mestrado) - Curso de Mestrado em Engenharia Informática, Universidade da Beira Interior, Covilhã, 2012. Disponível em: <https://ubibliorum.ubi.pt/handle/10400.6/3738>. Acesso em: 20 dez. 2025.
- SOUZA, Osvaldo de et al. Um método de sumarização automática de textos através de dados estatísticos e Processamento de Linguagem Natural. **Informação & Sociedade: Estudos**, v. 27, n. 3, 24 dez. 2017. Disponível em: <https://periodicos.ufpb.br/ojs2/index.php/ies/article/view/32571>. Acesso em: 09 jan. 2025.

