



O PROCESSAMENTO AUTOMÁTICO DE LÍNGUAS NATURAIS: PLN E A CIÊNCIA DA INFORMAÇÃO

The automatic processing of natural languages: PLN and information science

João Matheus Nascimento Rodrigues, Universidade Federal do Maranhão -
mateusnasc_@live.com

Valdirene Pereira da Conceicao, Universidade Federal do Maranhão -
cvaldireneufma@gmail.com

Eixo 4: Ciência da Informação: diálogos e conexões

INTRODUÇÃO

O mundo está profundamente marcado por mudanças de ordem informacional e tecnológica. As Tecnologias de Informação e Comunicação (TICs) mudaram e estão mudando de forma significativa o modo de fazer e viver do homem. A cada dia surgem novas tecnologias, ferramentas e maneiras de realizar processos e atividades, frutos da vasta produção informacional e da expansão em larga escala do mundo digital.

Desta forma, a tamanha produção informacional aliada à expansão em larga escala do mundo digital, provocam uma produção desenfreada de informações, o que gera cada vez mais a necessidade de instrumentos, ferramentas e métodos que garantam o tratamento dessas informações para seu posterior acesso e uso.

Instrumentos e ferramentas que, desde sua concepção, até os dias atuais, foram evoluindo para dar conta do número cada vez mais crescente de informações, assim como, da necessidade cada vez mais ágil por parte dos usuários, de acesso a essas informações produzidas e disponíveis em instituições e organizações.

A título de exemplo, temos os códigos de catalogação e classificação, que tiveram e desempenham um papel muito importante no tratamento da informação. E são utilizados em processos de tratamento e organização, das informações disponíveis em acervos físicos de bibliotecas, arquivos e museus, para possibilitar a organização, recuperação e acesso da informação.

Processos que, com a introdução do computador em meados dos anos 1950, sofreram significativas mudanças e evoluções, além de serem potencializados, com o advento das TICs e o emprego de softwares e hardwares, no sentido de dar suporte no desenvolvimento e realização de atividades de representação, organização e recuperação da informação, não apenas em ambientes tradicionais ou físicos, mas



também agora em ambientes digitais. Assim, tornou-se um grande desafio conseguir acompanhar esse novo momento e desenvolver instrumentos que pudessem tratar essas informações em formato tradicional e ao mesmo tempo, em formato digital.

Brascher e Café (2008, p. 5), definem organização da informação como “[...] um processo que envolve a descrição física e de conteúdo dos objetos informacionais.” com o intuito de “[...] arranjá-los sistematicamente em coleções, nesse caso, temos a organização da informação em bibliotecas, museus, arquivos, tanto tradicionais quando eletrônicos.” (BRASCHER; CAFÉ, 2008, p. 6).

Assim, é possível conceber a organização da informação, e a necessidade de novos métodos e técnicas de processamento de dados/informação que auxiliem e acompanhem os avanços tecnológicos para permitir o acesso e uso às informações disponíveis nos diversos tipos de suportes.

Nesse sentido, tendo como pressuposto a interdisciplinaridade presente na CI, estudos que envolvem aplicações de técnicas e métodos de outros domínios do conhecimento tem sido cada vez mais comum. Citamos o Processamento Automático de Línguas Naturais (PLN), vertente da Inteligência Artificial, que se constitui a partir de contribuições teórico-metodológicas das Ciências da Computação, da Ciência da Informação, Linguística e outras áreas do conhecimento, e que estuda a possibilidade de computadores simular à compreensão e o entendimento humano no que se refere ao tratamento da língua natural.

Tendo em vista esse cenário, o presente trabalho, objetiva através da literatura, conhecer, tal como, caracterizar o PLN, e as ferramentas oriundas desse tipo de tecnologia, a fim de perceber as contribuições do PLN para atividades oriundas à Ciência da Informação.

Moresi (2003) afirma que a pesquisa científica é uma atividade que tem por objetivo aproximar a teoria da realidade, a partir do relacionamento estabelecido entre aquela, e os dados desta. Nesse sentido, afim de sustentar metodologicamente o presente estudo, este estudo, é caracterizado como uma pesquisa de natureza aplicada, exploratória, em que utiliza como procedimentos técnicos e meios de investigação a pesquisa bibliográfica e documental.

Entende-se por pesquisa aplicada quando esta “objetiva gerar conhecimentos para aplicação prática dirigidos à solução de problemas específicos” (PRODANOV; FREITAS, 2013, p. 51), tendo em vista a viabilidade da aplicação prática dos



resultados obtidos para contribuir em determinada realidade. Em que, no presente contexto, objetiva-se gerar conhecimentos no âmbito do PLN.

A pesquisa exploratória por sua vez, proporciona maior familiaridade com o problema e assunto em que está sendo desenvolvida a pesquisa, pela sua natureza de “sondagem”. Além de que “Seu planejamento é, portanto, bastante flexível, de modo que possibilite a consideração dos mais variados aspectos relativos ao fato estudado.”, (GIL, 2010, p. 41), o que possibilita o desenvolvimento dos estudos sob diversos ângulos e aspectos.

Enquanto que a pesquisa bibliográfica “[...] abrange toda bibliografia já tornada pública em relação ao tema de estudo, desde publicações avulsas, boletins, jornais, revistas, livros, pesquisas, monografias, teses, material cartográfico etc. [...]” (MARCONI; LAKATOS, 2003, p. 181), e a “A pesquisa documental assemelha-se muito à pesquisa bibliográfica. A diferença essencial entre ambas está na natureza das fontes.” (GIL, 2010, p. 45). Assim, no que diz respeito ao desenvolvimento e sustentação teórica desta pesquisa, utilizou-se artigos científicos, teses, dissertações, monografias, livros e afins, e quanto à pesquisa documental, faz uso de relatórios técnicos científicos, slides de apresentações em eventos científicos e outros materiais desta natureza.

PROCESSAMENTO AUTOMÁTICO DE LÍNGUAS NATURAIS: DEMARCAÇÕES INICIAIS

Introduzidos na cultura ocidental no início dos anos 40, os computadores têm se mostrado grandes aliados no desenvolvimento de atividades que, até outrora, eram realizadas exclusivamente pelo homem. Desde então nota-se, grandes avanços no desenvolvimento de tecnologias digitais que possam substituir o homem em tarefas arriscadas, mecânicas, sistemáticas, repetitivas e enfadonhas, além de estarem mais presentes em nossa sociedade, sobretudo com o avanço da internet. (DIAS DA SILVA, 2006).

Transpondo o contexto para qual foram desenvolvidas, – e a necessidade cada vez maior de uma comunicação mais efetiva entre homem e máquina, superando as linguagens de programação e interfaces gráficas, desenvolvidas para facilitar sobremaneira a comunicação entre usuário e computador, – surgem novas



possibilidades que sustentam essa nova realidade em que ambos estão inseridos atualmente.

Entretanto, estabelecer meios que possibilitem a comunicação entre homem e máquina, se configura um desafio de difícil realização, tendo em vista as diferentes linguagens utilizadas entre esses sujeitos. “Os computadores normalmente estão aptos a compreenderem instruções escritas em linguagens computacionais – tais como Java, C++, Python, Perl –, mas possuem muita dificuldade para entender comandos simples escritos em uma linguagem humana.” (EVERS, 2013, p. 86). Em contrapartida, temos a complexidade da língua natural utilizada pelos humanos, imbuída de facetas e formalismos linguísticos. Desafios que tomam proporções maiores, em se tratando de criar mecanismos que possibilitem a compreensão e entendimento por parte da máquina, das línguas naturais.

A esse respeito, Sousa (2015, p. 15), afirma que

Pela forma como é constituído, o computador é incapaz de compreender comandos ambíguos, como se caracteriza a comunicação humana, sendo trabalho do PLN organizar a língua em modelos exatos que possam ser inteligíveis por máquinas.

Nasce assim, nesse contexto de interação homem e máquina, o Processamento Automático de Línguas Naturais, (doravante PLN). É compreendido como “[...] à área de pesquisa que se dedica a investigar, propor e desenvolver formalismos, modelos, técnicas, métodos e sistemas computacionais que têm a língua natural como objeto primário.”. (NUNES, 2008, p. 3).

O PLN é um domínio de pesquisa amplo e controverso, de natureza inter/multidisciplinar, tendo em vista a pluralidade de objetivos e interesses dos pesquisadores que desenvolvem estudos nesse domínio. Embora seja uma vertente da Inteligência Artificial, que ajuda computadores a entender, interpretar e manipular a linguagem humana, tanto escrita quanto falada, compreende técnicas e estudos conceituais de diversas áreas do conhecimento, e constitui-se a partir de contribuições teórico-metodológicas das Ciências da Computação, da Ciência da Informação, Estatística, Filosofia, Linguística, Linguística Computacional, Lógica, Matemática, Psicologia e outras. (LADEIRA; ALVARENGA, 2012; SOUSA, 2015).

Sustentando assim, a perspectiva inter/multidisciplinar dos estudos em PLN, mas levando em consideração as possibilidades que os Sistemas de Processamento Automático de Línguas Naturais (SPLN) podem oferecer, é comum encontrar na

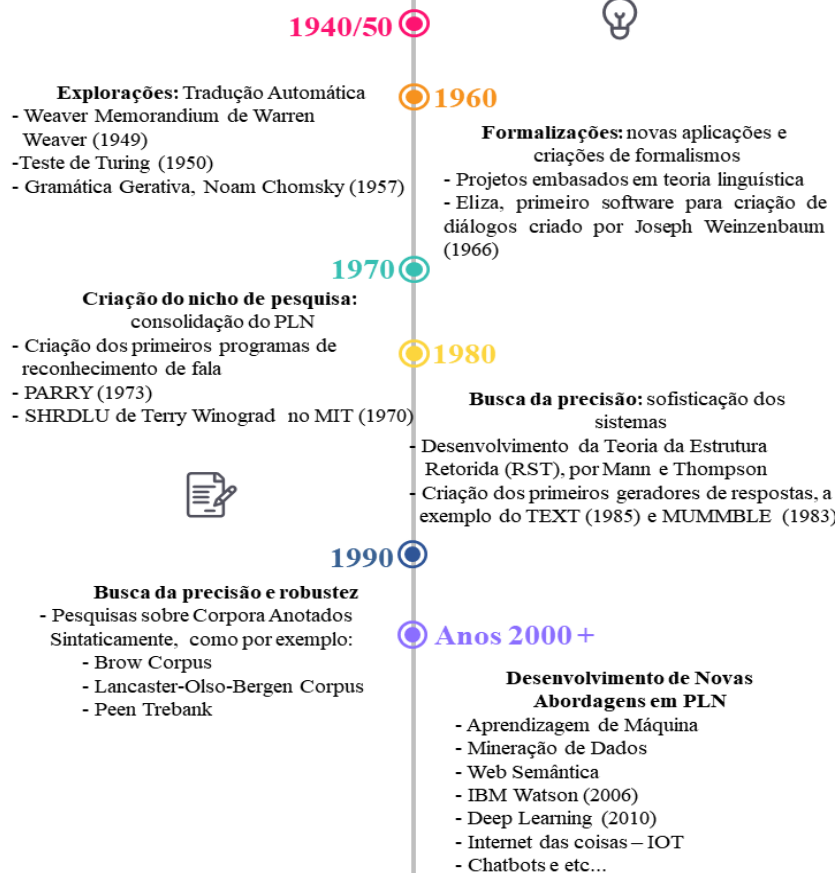


literatura diversas formas de defini-lo e designá-lo. Entretanto, todas as definições integram a noção de armazenamento em computador e manipulação de dados linguísticos. “Em outras palavras, o PLN pode ser definido como a habilidade de um computador em processar a mesma linguagem que os humanos usam no dia a dia”. (ROSA, 2011, p. 137).

Em face desses fatos, apresentar uma narrativa que tenha como cerne a evolução histórica do PLN, torna-se uma tarefa difícil, quando consideramos as diversas abordagens apresentadas pela literatura, no sentido de métodos e estratégias utilizadas. Entretanto, tem-se como marco inicial dos estudos em PLN, datada no início da década de 40, a Tradução Automática (TA), “[...] considerada pela maioria dos autores o marco inicial do uso do computador para a investigação das línguas naturais, [o que] permite também apresentar uma síntese da evolução dos estudos nesse campo.” (DIAS DA SILVA *et al*, 2007, p. 5).

Logo tendo em vista, uma melhor visualização dos marcos que estiveram em torno da evolução do PLN, a Figura 1, apresenta de forma sucinta os principais eventos históricos que amparam os avanços na área.

Figura 1 – Evolução dos estudos em PLN.



Fonte: adaptado de Dias da Silva *et al.*, (2007); Vieira e Lopes, (2010); Rosa (2011); Pardo (2008).

Nesse sentido, com o desenvolvimento dos estudos em PLN ao longo do tempo, surgem novos recursos e ferramentas linguísticas computacionais, agregando com maior frequência os conhecimentos linguísticos, em especial o léxico, a gramática e a semântica, potencializando as possibilidades de interação entre homem e máquina, quanto ao uso de línguas naturais e artificiais. Estudos que foram impulsionados sobremaneira após o surgimento da internet e as várias possibilidades de aplicações em Línguas Naturais.

PLN NA CIÊNCIA DA INFORMAÇÃO

Em Ciência da Informação, o PLN tem sido estudado em uma concepção teórico-metodológica, entretanto, em virtude das possibilidades apresentadas nos avanços dos estudos nessa área, têm ganhado espaço em uma perspectiva prática, em especial no campo da “[...] Indexação e Recuperação da Informação, pela possibilidade de softwares baseados nesse modelo propiciarem a extração de termos



com maior precisão semântica para recuperação da informação em sistemas de buscas automatizados [...]” (CARMO; CONCEIÇÃO, 2018, p. 317), potencializando as tarefas automatizadas com o emprego do PLN.

Nesse sentido, percebe-se um considerável destaque dessas técnicas de Processamento Automático de Línguas Naturais no contexto da Organização, Representação e Recuperação da Informação, tendo o emprego dessas ferramentas, voltado para Sistemas de Recuperação da Informação; Análise de Informações/Dados, e outros. (LADEIRA; ALVARENGA, 2012).

Desse modo, Barbosa e Kobashi (2017), salientam para as possibilidades de uso da visualização de dados no cenário da recuperação da informação, a partir das funções de extroversão e descoberta.

Enquanto Rampão e Tsunoda (2019) apontam para a mineração de dados aplicada a tarefas de classificação e associação e descoberta de padrões através da análise de grandes conjuntos de dados, no contexto específico de bases de dados jurídicas.

Vieira e Lima (2001) apontam para aplicações de PLN voltadas para SRI. De acordo com os autores, a RI é uma área que se envolve com a recuperação de documentos relevantes, dado um determinado tema, e não necessariamente com a recuperação de informação específica ou com a obtenção de respostas a dadas perguntas.

Assim, em relação ao uso do PLN em face da Recuperação da Informação, Fernerda (2003), o aponta como possível solução aos problemas referentes a RI, – levando em consideração a natureza do PLN –, pela simples observação de que, tanto os documentos como as expressões de busca, utilizados para recuperá-los em sistemas automatizados, são objetos linguísticos. Nhacuongue (2020, *online*) corrobora nessa perspectiva, ao propor “[...] estratégias de recuperação da informação baseadas no PLN, para extrair relações semânticas da WordNet.Pt, e utilizá-las na representação de documentos e de expressões de busca dos usuários”.

Ainda nesse âmbito, Câmara Júnior (2013, p. 74), fala que “A área de RI pode ser vista, sob alguns aspectos, como uma aplicação de sucesso de PLN. O crescimento rápido, desordenado e acachapante da Internet só foi possível por causa de motores de busca livres, disponíveis e efetivos, a maioria desenvolvida com técnicas de PLN”, sucesso este relacionado, a possibilidade de esses sistemas de RI,



baseados em modelos de PLN, realizarem a manipulação dos termos de entrada no sistema, traduzindo-os para a linguagem utilizada pelo sistema, e recuperar os documentos que sejam satisfatórios à pesquisa realizada pelo usuário.

Nunes (2008) destaca ainda que com o crescimento da Web, as aplicações em SRI ganharam bastante destaque, tendo o sistema de busca do Google, um dos mais conhecidos representantes. O grande destaque, para ela, está na possibilidade de hoje, o usuário ser livre para formalizar suas consultas em língua natural, em contraste, às linguagens rígidas de busca utilizadas anteriormente (a exemplo do SQL). Possibilidade esta, que também se faz presente em SRIs utilizados em Unidades de Informações, a exemplo de Bibliotecas, Museus e Arquivos.

Assim, em módulos de busca em Sistemas de Recuperação da Informação, as consultas realizadas de maneira erradas ou inexatas são processadas através de aplicações de PLN, que transformam os termos de busca por termos próximos ou semelhantes do correto, possibilitando uma recuperação mais precisa. (LIMA; NUNES; VIEIRA, 2007).

Transcendendo esse cenário, em torno da recuperação de informações e documentos, existem aplicações direcionadas a processos de organização e representação da informação. É o caso dos sistemas de extração e mineração de informação/dados ou mineração de textos, que podem ser aplicados no intuito de auxiliar a execução de atividades de indexação; construção de vocabulários controlados; sumarização de informação, destinado à interpretação de conteúdo dos documentos; e análise de conteúdo. (LADEIRA, 2010; LADEIRA; ALVARENGA, 2012). Tal como, o uso de ferramentas de PLN para a construção dos *Simple Knowledge Organization System* (SKOS).

Pardo (2008) corrobora com essa perspectiva, ao destacar que, além dos sistemas baseados em perspectivas de extração e mineração de dados, existem também sistemas baseados em processos que permitem a condensação de textos, com a finalidade de possibilitar a compreensão do conteúdo mais importante existente no texto, a exemplo das ferramentas de Sumarização Automática.

Em contrapartida, à luz das possibilidades apresentadas pelos fundamentos teóricos e metodológicos expostos na subseção anterior, é possível vislumbrar o uso dessas aplicações em outros ramos da Ciência da Informação, em que, podemos citar a possibilidade de uso de Chat Boots e Assistentes Virtuais em Sistemas de



Referência Virtual; Aplicações de PLN em motores de busca de Bases de Dados e Repositórios Institucionais; e o uso de api's baseados em modelos de PLN voltados para tradução automática, integrados também em bases de dados.

Nesse contexto, a fim de caracterizar as pesquisas e tendências em PLN na área de Ciência da Informação, o Quadro 1, apresenta dados em formato de referência bibliográfica, de teses e dissertações desenvolvidas na área de Ciência da Informação, que possuem como foco de pesquisa vertentes do Processamento Automático de Linguagem Natural. O levantamento dos dados disponibilizados no quadro a seguir, foram possíveis através de buscas realizadas na Biblioteca Digital Brasileira de Teses e Dissertações – BDTD.

Quadro 1 – Pesquisa em PLN na CI.

REFERÊNCIA
CÂMARA JÚNIOR, Auto Tavares da. Indexação automática de acórdãos por meio de processamento de linguagem natural. 2007. Dissertação (Mestrado em Ciência da Informação) – Universidade de Brasília, Brasília, 2007.
BRUZINGA, Graciane Silva. Indexação automática de documentos textuais: proposta de critérios essenciais. 2009. Dissertação (Mestrado em Ciência da Informação) – Universidade Federal de Minas Gerais, Belo Horizonte, 2010.
NASCIMENTO, Gustavo Diniz. Dos sintagmas nominais aos descritores documentais: estudo de caso na indexação de teses e dissertações da área de Direito. 2015. Dissertação (Mestrado em Ciência da Informação) – Universidade Federal de Pernambuco, Recife, 2015.
CELERINO, Victor Galvão. Proposta de normalização dos sintagmas nominais em termos para indexação automática. 2018. Dissertação (Mestrado em Ciência da Informação) – Universidade Federal de Pernambuco, Recife, 2018.
SOUZA, Renato Rocha. Uma proposta de metodologia para escolha automática de descritores utilizando sintagmas nominais. 2005. Tese (Doutorado em Ciência da Informação) – Universidade Federal de Minas Gerais, Belo Horizonte, 2005.
MAIA, Luiz Claudio Gomes. Uso de sintagmas nominais na classificação automática de documentos eletrônicos. 2008. Tese (Doutorado em Ciência da Informação) – Universidade Federal de Minas Gerais, Belo Horizonte, 2008.
LADEIRA, Ana Paula. Processamento de linguagem natural: caracterização da produção científica dos pesquisadores brasileiros. 2010. Tese (Doutorado em Ciência da Informação) – Universidade Federal de Minas Gerais, Belo Horizonte, 2010.
MARTINS, Agnaldo Lopes. O uso do sintagma nominal na recuperação de documentos: proposta de um mecanismo automático para classificação temática de textos digitais. Tese (Doutorado em Ciência da Informação) – Universidade Federal de Minas Gerais, Belo Horizonte, 2012.
MESQUITA, Luiz Antonio Lopes. Sintagmas nominais na indexação automática: uma análise estrutural da distribuição de termos relevantes em teses de doutorado da UFMG. 2012. Tese (Doutorado em Ciência da Informação) – Universidade Federal de Minas Gerais, Belo Horizonte, 2012.
CAMARA JUNIOR, Auto Tavares da. Processamento de linguagem natural para indexação automática semântico-ontológica. 2013. Tese (Doutorado em Ciência da Informação) – Universidade de Brasília, Brasília, 2013.



REFERÊNCIA

SILVA, Tiago José da. **Indexação automática por meio da extração e seleção de sintagmas nominais em textos em língua portuguesa**. 2014. Dissertação (Mestrado em Ciência da Informação) – Universidade Federal de Pernambuco, Recife, 2014.

Fonte: elaborado pelos autores (2022).

CONSIDERAÇÕES FINAIS

A princípio, evidenciamos a importância da ótica da interdisciplinaridade, pois compreendemos a necessidade da colaboração de diferentes campos do conhecimento, em prol de uma mesma causa. Dessa forma, compreende-se os ganhos para a área da Ciência da Informação, a partir das relações estabelecidas apoiadas pelas possibilidades de uso e aplicação de teorias e metodologias de outras áreas do conhecimento, em atividades oriundas à CI.

Nesse bojo, após as exposições realizadas anteriormente, são animadoras as múltiplas possibilidades do emprego de sistemas baseados em modelos de PLN na Ciência da Informação, levando em consideração as várias frentes que cada uma das áreas se dispõe a investigar e a natureza inter/multidisciplinar em que estão.

Entretanto, após o resultado e sistematização das buscas, percebemos ainda uma predominância das pesquisas quanto às temáticas, pois grande parte delas estão voltadas para estudos entorno de processos cujo objetivo é contribuir para o Tratamento Temático da Informação.

A partir do estudo, podemos nos dar conta de que algumas dessas aplicações de PLN já estão presentes no âmbito da Organização, Representação e Recuperação da Informação em ambientes relacionados à área da Biblioteconomia e Ciência da Informação, a exemplo da Mineração de Textos, Sistemas Inteligentes de Recuperação da Informação, Sintagmas Nominais e outros.

Entretanto, indicamos que mais estudos nessa área possam ser desenvolvidos no sentido de dar maior sustentação e bases teóricas para o desenvolvimento de técnicas de representação e organização da informação, para permitir uma aproximação mais sólida e concreta entre as áreas em destaque.

Acreditamos que este trabalho contribua com os fundamentos teóricos-metodológicos na Representação e Organização da Informação, com o intuito de estimular novos estudos sobre aplicação do PLN no processo de indexação e no



desenvolvimento de serviços e produtos inerentes à Biblioteconomia e Ciência da Informação.

Como trabalhos futuros, recomendamos investigações sobre os seguintes temas: o uso de ferramentas de PLN voltadas para indexação de imagens, através de tecnologias de OCR; estudos sobre o reconhecimento de entidades nomeadas; uso de *chatbots* em serviços de referência virtuais; integração de sistemas de conversação texto-fala em bases de dados e repositórios instrucionais, com o intuito de assegurar a acessibilidade à deficientes visuais e outros.

REFERÊNCIAS

BARBOSA, Eduardo Caetano; KOBASHI, Nair Yumiko. Extroversão e descoberta: visualização de dados no auxílio a buscas e recuperação de informações. **Revista Brasileira de Biblioteconomia e Documentação**, v. 13, p. 115-120, 2017. Disponível em: <http://hdl.handle.net/20.500.11959/brapci/1157>. Acesso em: 05 out. 2020.

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 jul. 2019.

CAMARA JUNIOR, Auto Tavares da. **Processamento de linguagem natural para indexação automática semântico-ontológica**. 2013. 181 f. Tese (Doutorado em Ciência da Informação) - Universidade de Brasília, Brasília, Df, 2013.

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 abr. 2019.

DIAS DA SILVA, Bento Carlos Dias da. O estudo Lingüístico-Computacional da Linguagem. **Letras de Hoje**, Porto Alegre, v. 41, n. 2, p. 103-138, jun. 2006. Disponível em: <https://revistaseletronicas.pucrs.br/ojs/index.php/fale/article/view/597>. Acesso em: 29 jun. 2020.



EVERS, Aline. Processamento de Língua Natural. *In*: EVERS, Aline. **Processamento de língua natural e níveis de proficiência do português: um estudo de produções textuais do exame CELPE-ras**. 2013. 174 f. Dissertação (Mestrado em Letras) – Universidade Federal do Rio Grande do Sul, Porto Alegre, 2013. Disponível em: <http://www.ufrgs.br/acervocelpebras/arquivos/textos-publicados/evers-aline>. Acesso em: 26 jul. 2020.

FERNEDA, Edberto. **Recuperação de informação**: análise sobre a contribuição da ciência da computação para a ciência da informação. 2003. 147 f. Tese (Doutorado em Ciência da Informação) – Escola de Comunicação e Artes, Universidade de São Paulo, São Paulo, 2003. Disponível em: <https://www.teses.usp.br/teses/disponiveis/27/27143/tde-15032004-130230/pt-br.php>. Acesso em 26 maio 2020.

GIL, Antonio Carlos. **Como elaborar projetos de pesquisa**. 5. ed. São Paulo: Atlas, 2010. 183p.

LADEIRA, Ana Paula. **Processamento de linguagem natural**: caracterização da produção científica dos pesquisadores brasileiros. 2010. 262 f. Tese (Doutorado em Ciência da Informação) - Universidade Federal de Minas Gerais, Belo Horizonte, 2010. Disponível em: <https://repositorio.ufmg.br/handle/1843/ECID-8B3Q6C>. Acesso em: 21 jul. 2020.

LADEIRA, Ana Paula; ALVARENGA, Lídia. Processamento de linguagem natural: em busca de evidências temáticas nas publicações nacionais e contemporâneas. *In*: ENCONTRO NACIONAL DE PESQUISA E PÓS-GRADUAÇÃO EM CIÊNCIA DA INFORMAÇÃO, 10, 2009, João Pessoa. **Anais [...]**. João Pessoa: Ancib, 2009. Disponível em: <http://repositorios.questoesemrede.uff.br/repositorios/bitstream/handle/123456789/69/GT%20%20Txt%2019-%20LADEIRA%2c%20Ana%20Paula%3b%20ALVARENGA%2c%20L%3c%addia.pdf?sequence=1>. Acesso em: 21 jul. 2020.

LIMA, Vera Lúcia Strube de; NUNES, Maria das Graças Volpe; VIEIRA, Renata. Desafios do processamento de línguas naturais. *In*: XXVII CONGRESSO DA SOCIEDADE BRASILEIRA DE COMPUTAÇÃO, 2007. p. 2202–2216. **Anais [...]**. Disponível em: <https://www.inf.pucrs.br/linatural/Recursos/Desafios.pdf>. Acesso em: 23 jul. 2020. linguagem. 2019. Disponível em: https://grad.letras.ufmg.br/arquivos/monitoria/ApostilaConceitos%20b%3c%a1sicos_Aula1.p df. Acesso em 15 jul. 2020.

MARCONI, Marina de Andrade; LAKATOS, Eva Maria. **Fundamentos de metodologia científica**. São Paulo: Atlas, 2003.



MORESI, Eduardo. **Metodologia da Pesquisa**. Brasília: UCB, 2003. Disponível em: <http://www.inf.ufes.br/~pdcosta/ensino/2010-2-metodologia-de-pesquisa/MetodologiaPesquisa-Moresi2003.pdf>. Acesso em 22 set. 2020

NHACUONGUE, Januário Albino; DUTRA, Moisés Lima. A terminologia em Sistemas de Recuperação da Informação baseada na WORDNET.PT. **Informação & Sociedade: Estudos**, v. 30, n. 2, 26 maio 2020. Disponível em: <https://periodicos.ufpb.br/ojs2/index.php/ies/article/view/50756>. Acesso em: 21 set. 2020.

NUNES, Maria das Graças Volpe. **Processamento de línguas naturais**: para quê e para quem. Notas Didáticas ICMC-USP, n. 73. São Carlos, 2008.

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 abr. 2019.

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 set. 2020.

RAMPÃO, Talita de Souza; TSUNODA, Denise Fukumi. Mineração de dados em bases jurídicas: um estudo de caso. **Revista Brasileira de Educação em Ciência da Informação**, v. 6, n. Especial, p. 61-76, 2019. Disponível em: <http://hdl.handle.net/20.500.11959/brapci/114081>. Acesso em: 05 out. 2020.

ROSA, João Luís Garcia. Processamento de Línguas Naturais. *In*: ROSA, João Luís Garcia. **Fundamentos da Inteligência Artificial**. Rio de Janeiro: Ltc, 2011. P. 136-169. Disponível em: <http://walderson.com/2011-2/IA/FIA.pdf>. Acesso em 02 jun. 2020.

SOUSA, Aline Rocha de. **Processamento automático de línguas naturais**: um estudo sobre a localização do IBM Watson™ para o português do Brasil. 2015. 76 f., Trabalho de Conclusão de Curso (Bacharelado em Línguas Estrangeiras Aplicadas)



– Universidade de Brasília, Brasília, 2015. Disponível em: https://bdm.unb.br/bitstream/10483/18730/1/2015_AlineRochaDeSousa_tcc.pdf. Acesso em 28 jun. 2020.

VIEIRA, Renata; LIMA, Vera Lúcia Strube de. Linguística computacional: princípios e aplicações. *In*: ENCONTRO NACIONAL DE INTELIGÊNCIA ARTIFICIAL, 2001. **Anais** [...]. Disponível em: <https://www.inf.pucrs.br/linatural/Recursos/jaia-2001.pdf>. Acesso em: 20 jul. 2020.

VIEIRA, Renata; LOPES, Lucelene. Processamento de linguagem natural e o tratamento computacional de linguagens científicas. *In*: PERNA, Cristina Lopes; DELGADO, Heloísa Koch; FINATTO, Maria José (Org.). **Linguagens especializadas em corpora**: modos de dizer e interfaces de pesquisa. Porto Alegre: Edipucrs, 2010. p. 183-201. Disponível em: <http://bibliodigital.unijui.edu.br:8080/xmlui/handle/123456789/1496>. Acesso em: 22 jul. 2020.