

Um mapeamento sistemático sobre técnicas, abordagens e avaliações de desambiguação de sentido de palavra

Clovis Holanda do Nascimento, Vinicius Cardoso Garcia, Ricardo de Andrade Araújo.

Resumo: máximo 200 palavras

Este artigo apresenta um Mapeamento Sistemático da literatura sobre técnicas, abordagens e avaliações de Desambiguação de Sentidos de Palavras. O objetivo é fornecer uma visão geral dos estudos empíricos e teóricos publicados que lidam com ambiguidade. Aplicamos um método de revisão sistemática, para pesquisar e selecionar artigos publicados, extrair e sintetizar dados dos artigos selecionados que relataram estudos sobre desambiguação de palavras. Nossa busca resultou em mais de 13.000 artigos, dos quais selecionamos 107 artigos publicados entre 2012 e 2023. O problema de desambiguação foi definido em 1940 e é considerado como um problema de Inteligência Artificial completo pela literatura científica, sua relevância é crucial para a qualidade de varias tarefas de Processamento de Linguagem Natural, tradução automática de idiomas, sumarização de textos, análise de sentimentos entre outros. Nossa pesquisa revelou que apesar dos avanços com técnicas emergentes como o BERT, métodos mais antigos como o clássico LESK ainda é utilizado, encontramos mais de 100 métodos para resolver o problema de ambiguidade, mais de 30 formas para avaliar estes métodos, e ainda as limitações existentes nas abordagens, idiomas estudados, países envolvidos, objetos de estudos, domínios entre outros aspectos são publicados neste mapeamento.

Palavras chave: Mapeamento Sistemático, Desambiguação de Sentidos de Palavras, Inteligência Artificial, Processamento de Linguagem Natural.

Introdução

Devido à vasta quantidade de conteúdo textual que cresce exponencialmente na internet e nos meios digitais, a compreensão destes textos por computadores se mostra como uma poderosa ferramenta para ajudar as pessoas a aproveitar este universo de informações, pois a máquina faz o “papel” do ser humano, processando e compreendendo os textos, porém de forma mais rápida e em grande volume de dados (L'HEUREUX et al. 2017; OKUL, AKSU, AYDIN, 2019).

No entanto, apesar do ser humano não ser capaz de processar informações na velocidade e na quantidade de dados que um computador pode, ele consegue interpretar com facilidade a ambiguidade das palavras presente nas línguas naturais, com isso, obter uma compreensão contextual sintática e semântica, que ainda é um problema complexo para os computadores, observemos, a palavra banco é polissêmica, sendo assim, pode ter muitos sentidos dependendo do seu contexto, por exemplo, banco da praça, banco financeiro, banco de dados entre outros (Navigli 2018).

Para resolver o problema da ambiguidade nas palavras polissêmicas nos idiomas surgiu a área de pesquisa denominada Word Sense Disambiguation (WSD), que é a tarefa de identificar o sentido pretendido de uma palavra de maneira computacional com base no contexto em que ela esta inserida, essa área de pesquisa vem sendo estudada desde os primórdios do Processamento de Linguagem Natural (PLN) (Weaver 1955; Agirre e Edmonds 2006; Navigli 2009).

Atualmente, entre as tecnologias para WSD, destaca-se o modelo BERT (Pre-training of Deep Bidirectional Transformers for Language Understanding) que alcançou desempenhos de ponta em Natural Language Understanding (NLU). No entanto, os modelos existentes ainda não conseguem capturar informações específicas do contexto, levando a um desempenho ruim ao lidar com a ambiguidade generalizada da linguagem causada pelas palavras polissêmicas, embora o BERT seja uma linha de base forte, estudos mostram que o modelo não converge para sua capacidade total e seu esquema de treinamento ainda apresenta oportunidades de aprimoramento (Duan et al. 2021; Liu et al. 2019; Yang et al. 2019).

Através deste Mapeamento Sistemático da literatura (MS) foram encontradas evidências que indicam que a resolução da ambiguidade é um problema ainda em aberto desde 1940. Visto que, para resolver o problema de ambiguidade é necessário superar vários obstáculos inerentes às linguagens naturais, tais como (sinonímia, homônimos, parônimos, hiperônimo, hiponímia, neologismo entre outros) e que podem ocorrer conjuntamente, e os trabalhos encontrados não foram concebidos para superar todas estas questões simultaneamente, ficando as outras sem tratamento (Dhungana, Shakya 2015; Sajini, Kallimani 2017; Farahani, Janfada, Bidgoli 2020; Sousa, Milios, Berton 2020).

O restante deste trabalho está organizado da seguinte forma. Na seção “Trabalhos Relacionados” são apresentados trabalhos relacionados ao tema. Na seção “Mapeamento Sistemático”, é abordado o método adotado e o protocolo de pesquisa. A seção “Resultados e Discussão” apresenta as análises e resultados. A seção “Metanálise” detalha a metanálise realizada nos estudos selecionados. Na seção “Ameaças à validade”, são detalhadas as ameaças à validade encontradas. Por fim, na seção “Conclusão”, é apresentada a conclusão.

Trabalhos Relacionados

Nesta seção são expostos os principais trabalhos relacionados a esta pesquisa, consideramos trabalhos relacionados, aqueles que fazem a compilação de vários estudos primários sobre WSD. Até o momento da conclusão da investigação sobre revisões da literatura voltadas para WSD, apenas foram encontrados surveys desatualizados e nenhum trabalho usando um protocolo de MS para pesquisar sobre técnicas, abordagens e avaliações WSD foi descoberto.

Os mapeamentos e revisões sistemáticas tem menor risco de ter vieses durante o seu planejamento e execução, devido seu rigor científico, tem um processo controlado e rigoroso, garantindo que a mesma seja passível de auditoria, repetível e imparcial aos interesses dos envolvidos. Além de permite identificar com mais precisão lacunas na pesquisa atual e, como consequência, identificar possíveis linhas de pesquisa para investigação futura avaliação (SCANNAVINO et al. 2017).

Os trabalhos que mais se assemelham a este estudo estão listados abaixo, que apresentam alguns tópicos de pesquisa sobre WSD:

(Camacho-Collados, Pilehvar 2018) Apresentam uma pesquisa contendo uma abrangente visão geral da ampla gama de técnicas dos dois principais ramos de representação de sentidos (não supervisionado e baseado em conhecimento), a pesquisa cobre formas de avaliação, aplicações para este tipo de representação, e fornece uma análise de quatro de seus aspectos importantes: interoperabilidade, granularidade de sentido, adaptabilidade a diferentes domínios e composicionalidade. Segundo a análise dos autores, algumas abordagens são mais adequadas para certas aplicações ou domínios, sem qualquer conclusão geral clara, essas questões abertas certamente ainda são relevantes e encorajam pesquisas futuras sobre representações distribuídas de significado, com muitas áreas a serem exploradas.

(Popov 2018) Esta pesquisa apresenta uma visão geral do uso de redes neurais artificiais para a tarefa de WSD. Mais especificamente, examina os avanços nos modelos de linguagem neural nos últimos anos que resultaram em métodos para a representação distribuída eficaz de unidades linguísticas, representações - embeddings de palavras, embeddings de contexto, embeddings de sentido. Também foi examinada uma série de aplicações bem-sucedidas de redes neurais recorrentes para a tarefa de desambiguação.

(Aliwy, Taher 2019) Neste estudo, os algoritmos DLS foram classificados em três categorias, como técnicas baseadas no conhecimento, supervisionadas e não supervisionadas. Cada categoria foi estudada em detalhes com explicação de quase todos os algoritmos em cada categoria. Portanto, os exemplos de trabalho para cada método foram tomados com a linguagem utilizada, os corpora utilizados e outros fatores. Os benefícios e as desvantagens de cada método foram registrados. Algumas dessas técnicas têm limitações em algumas situações, portanto este trabalho irá auxiliar os pesquisadores no campo do processamento de linguagem natural, para selecionar os algoritmos adequados para resolver seu problema específico em DLS. A novidade do trabalho pode ser vista na comparação das obras utilizadas

e dos algoritmos utilizados. Deste trabalho, concluiu-se que (i) alguns métodos fornecem alta precisão para a linguagem, mas baixo para outro, (ii) o tamanho do conjunto de dados usado afeta o desempenho do algoritmo usado, (iii) algumas dessas abordagens podem ser executadas rapidamente, mas com limitação da precisão e (iv) a maioria dessas abordagens são implementadas para muitos idiomas com sucesso.

(Bhattacharjee et al. 2020) Este artigo apresenta uma pesquisa sobre as abordagens de DLS nos principais métodos de PLN, comparando diferentes abordagens para DLS em algoritmos supervisionados, não supervisionados e baseados em conhecimento. Este artigo também tem como objetivo fornecer uma análise de lacunas em sistemas DLS pesquisados, comparando os pontos fortes e fracos de vários sistemas pesquisados e sua precisão. Com base nas descobertas, uma futura abordagem híbrida que sinergia métodos baseados em regras e métodos baseados em aprendizado de máquina são contemplados. Os resultados desta pesquisa são previstos por meio de uma pesquisa em andamento no algoritmo Meta-Search baseado em WSD sob a alçada do Centro de Desenvolvimento de Computação Avançada (C-DAC) para um sistema baseado em PLN inteligente para detectar o sentido real das consultas de pesquisa e fornecer classificação semântica de manchetes de notícias e fragmentos contendo palavras ambíguas.

(Pavithra, Mandal 2021) Neste trabalho os autores desenvolveram um estudo comparativo de três abordagens, a saber, algoritmo LESK, técnicas de incorporação, e técnicas de Rede Neural baseadas no texto coletado de livros de histórias infantis (StoryWeaver). Descreve o histórico das técnicas e as metodologias para construir cada um dos modelos apresentados. Foi explorada uma abordagem que combina a rede neural Bi-LSTM com o Knowledge Graph para prever o significado contextual das palavras. O estudo mostra que a precisão da abordagem combinada é 80,34 abordagens. Este estudo comparativo mostra que uma combinação híbrida de abordagem baseada no conhecimento e aprendizagem profunda pode produzir melhores resultados. A precisão dos modelos de aprendizado profundo pode ser aprimorada ainda mais com o aprimoramento das técnicas de incorporação.

Mapeamento Sistemático

Nesta seção é apresentado o MS sobre trabalhos relacionados a WSD, este trabalho é classificado como estudo secundário, pois é uma revisão de estudos primários. Um MS consiste em um protocolo sistemático para investigar e selecionar estudos de uma área de pesquisa, a fim de extrair informações e mapear os resultados para prover uma visão ampla de um tópico de pesquisa. Resultados de um MS auxiliam na identificação de lacunas da área, capazes de propor pesquisas futuras e prover um guia para posicionar adequadamente novas atividades de pesquisa (Kitchenham & Charters 2007; Petersen et al. 2008; Kitchenham et al. 2011). Por se basear em um protocolo de pesquisa, pode ser reproduzido por outros pesquisadores (Kitchenham 2004).

O presente estudo baseou-se no protocolo proposto por (Kitchenham 2004) e tem como objetivo o desenvolvimento de um MS com o propósito de identificar, caracterizar, classificar e meta-analisar trabalhos científicos de forma a caracterizar o uso de técnicas, abordagens, formas de avaliação e limitações das técnicas WSD. A definição das questões de pesquisa, fonte de dados, estratégia de busca, critérios de seleção e extração de dados serão descritos nas seções a seguir.

Questões de pesquisa

As perguntas de pesquisa foram desenvolvidas com o objetivo de apresentar uma visão geral da área, destacando os principais aspectos dos estudos primários (Kitchenham 2004; Petersen, Vakkalanka, & Kuzniarz 2015). Para este estudo, as questões de pesquisa tentam fornecer uma visão específica sobre os aspectos relevantes de WSD.

As seguintes perguntas de pesquisa específicas foram usadas para orientar a extração de dados, análise, síntese e apresentação dos resultados:

- RQ1:
Quais as técnicas utilizadas para desambiguar?
- RQ2:
Quais as formas utilizadas para avaliar as técnicas de desambiguações?
- RQ3:
Quais as limitações das técnicas utilizadas para desambiguar?
- RQ4:
Quais os objetos de estudos mais ativos na área de desambiguação?
- RQ5:
Quais os idiomas dos datasets mais ativos na área de desambiguação?
- RQ6:
Quais os domínios mais ativos na área de desambiguação?
- RQ7:
Quais as abordagens utilizadas na área de desambiguação?
- RQ8:
Quanto trabalhos foram publicados nos anos selecionados?
- RQ9:
Quais os países que publicaram trabalhos nessa área?
- RQ10:
Quanto trabalhos foram publicados nos sites de buscas?

Fontes de dados e estratégia de pesquisa

As fontes de dados que mais retornaram estudos usando a string de buscas foram escolhidas, principalmente o Scopus, por contemplar a maior quantidade de trabalhos da literatura científica, com mais de 21.950 *journals* e 120 mil conferências (Scopus 2019), foram definidas quatro fontes de informação para efetuar as buscas:

Fonte de Informação	Endereço online	Tipo
Scopus	www.scopus.com	Motor de busca
IEEE Xplore	ieeexplore.ieee.org	Base bibliográfica
ACM Digital Library	dl.acm.org	Híbrida
Engineering Village	www.engineeringvillage.com	Motor de busca

A figura 1 mostra todo o ciclo de vida do processo de busca e seleção dos estudos primários que resultou em 107 trabalhos.

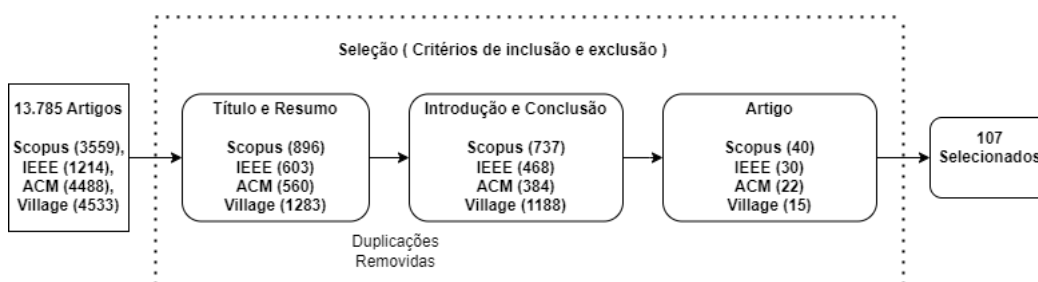


Figura 1: Processo de Busca e seleção dos artigos.

Strings de Busca:

Durante o estudo preliminar, em buscas ad-hoc, foi observado que existiam varias dúvidas sobre quais são as técnicas de desambiguação, como são avaliadas, e as limitações dessas técnicas. Diante disto, cerca de 100 artigos sobre desambiguação foram lidos com o intuito de coletar o maior numero possível de palavras chaves para montar a string de busca, resultando nas as seguintes palavras chave:

Disambiguation, Ambiguity, Ambiguous, polysemy, word sense disambiguation, WSD, sense embedding, Natural language processing, NLP.

Após a coleta das palavras chaves foram elaboradas as strings de buscas para cada fonte de dados, passando por vários testes e refinamentos no intuito de resultar na maior quantidade de estudos relacionados.

String Scopus:

TITLE-ABS-KEY (disambiguation) OR TITLE-ABS-KEY (ambiguity) OR TITLE-ABS-KEY (ambiguous) OR TITLE-ABS-KEY (polysemy) OR TITLE-ABS-KEY ("word sense disambiguation") OR TITLE-ABS-KEY (wsd) OR TITLE-ABS-KEY ("sense embedding") AND TITLE-ABS-KEY ("natural language processing") OR TITLE-ABS-KEY (nlp) AND (PUBYEAR > 2011)

String IEEE:

(((((Disambiguation OR Ambiguity OR Ambiguous OR polysemy OR word sense disambiguation OR WSD OR sense embedding) AND (Natural language processing OR NLP)))))

String ACM:

[Title: disambiguation] OR [Title: ambiguity] OR [Title: ambiguous] OR [Title: polysemy] OR [Title: "word sense disambiguation"] OR [Title: wsd] OR [[Title: "sense embedding"] AND [Title: "natural language processing"]]] OR [Title: pln] AND [Publication Date: (01/01/2012 TO *)]

String Engineering Village:

((((Disambiguation OR Ambiguity OR Ambiguous OR polysemy OR word sense disambiguation OR WSD OR sense embedding) AND (Natural language processing OR NLP)) WN KY) + (2023 OR 2022 OR 2021 OR 2020 OR 2019 OR 2018 OR 2017 OR 2016 OR 2015 OR 2014 OR 2013 OR 2012)

Critério de seleção

Cada estudo retornado através das buscas foi avaliado em relação a sua relevância quanto a esse MS. Assim, todos passaram por critérios de inclusão e exclusão.

Critérios de Inclusão:

- Escopo principal técnica de desambiguação
- Escopo principal avaliação de técnica de desambiguação
- Apresentar limitação do trabalho
- Apresentar informações do dataset
- Idioma Inglês
- A partir de 2012
- Estudo com mais de cinco páginas

Critérios de Exclusão:

- Não conseguir ter acesso ao estudo
- O estudo não é um estudo primário
- Estudo sem resumo
- Artigo duplicado
- O estudo ser uma versão antiga de um estudo selecionado
- Não esta escrito em inglês

Extração de dados

Para responder as 10 (dez) questões de pesquisas definidas, foram necessárias as extração dos dados dos estudos primários selecionados, sendo utilizada uma planilha contendo:

- Título do Artigo
- Abordagem usada
- Objeto de estudo
- Idioma do dataset
- Domínio
- Métodos usados
- Formas de Avaliações dos métodos
- Limitações do método
- Ano
- Site do engenho
- País
- Autores

Resultados e Discussão

Nesta seção serão apresentados os resultados dos estudos selecionados, respondendo às questões de pesquisa apresentadas anteriormente. Como foi dito, ao final do processo de seleção 107 estudos foram selecionados. Os dados foram analisados descritivamente por meio de frequências absolutas e percentuais e os resultados foram elaborados por meio de tabelas e gráficos.

A Tabela 1 e figura 2 apresentam os principais métodos encontrados nos estudos primários (RQ1), foram utilizados 120 métodos nos 107 artigos selecionados, os mais frequentes foram: similarity algorithms (citado em 25,2% dos artigos), graphs (22,4%), ad hoc (15,0%), LESK (11,2%). Os percentuais relativos aos 120 métodos utilizados, os mais frequentes foram: similarity algorithms (22,5% dos methods), graphs (20,0%) e (27,5%) correspondeu a diversos outros métodos, e que estão listados no Apêndice A.

Tabela 1 – Métodos mais frequentes nos 107 artigos selecionados

Método	N (%) ⁽¹⁾	N (%) ⁽²⁾
Similarity algorithms	27 (22,5%)	27 (25,2%)
Graphs	24 (20,0%)	24 (22,4%)
ad hoc	16 (13,3%)	16 (15,0%)
LESK	12 (10,0%)	12 (11,2%)
BERT	8 (6,7%)	8 (7,5%)
Outros	33 (27,5%)	33 (30,8%)
TOTAL	120 (100,0%)	120 (112,1%)

(1) Percentuais obtidos com base no total das frequências. No caso 120 Métodos.

(2) Percentuais obtidos com base nos 107 artigos selecionados na pesquisa.

Considerando a possibilidade de mais de um método por artigo a soma das frequências pode ser superior ao total correspondente aos 107 artigos.

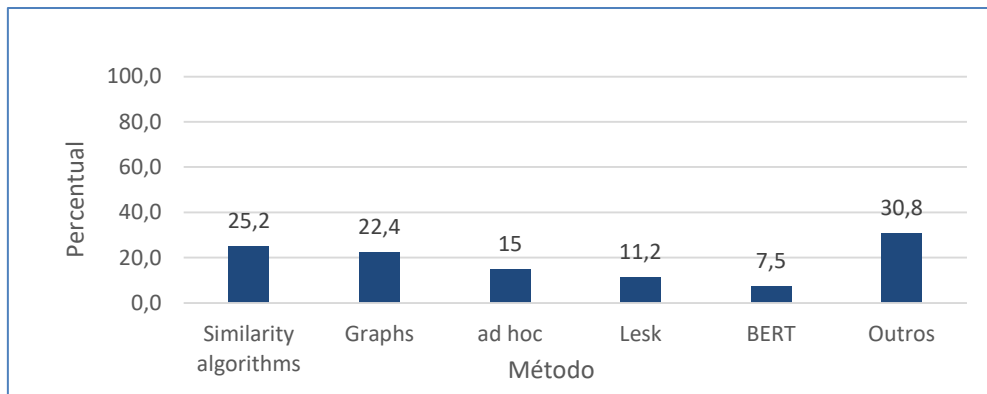


Figura 2 – Métodos mais frequentes nos 107 artigos selecionados

Em resposta a pergunta 1, RQ1-Quais as técnicas utilizadas para desambiguar textos, sendo as principais:-.

- 27 Similarity algorithms: (Saha, Mou, Mittra 2019; Zhang et al. 2018; Li, Cui, Wang 2020; AL-Qawasmeh, AL-Smadi, Fraihat 2016; Tuan et al. 2020; Laatar, Aloulou, Belguith 2020; Sarika, Sharma 2016; Sabbir, Jimenoyepes, Kavuluru 2017; Ding 2021; Khabsa, Treeratpituk, Giles 2015; Sun et al. 2020; Seifollahi, Shajari 2019; Kwon, Oh, Ko 2021; Dongsuk et al. 2018; Orkphol, Yang 2019; Aouicha, Taieb, Marai 2018; Asgari-Bidhendi et al. 2021; Gogoi, Baruah, Nath 2021; Rais, Lachkar 2016; Yin et al. 2019; Pereira. Silva. smin 2014; Oele, Noord 2017; Wiedemann et al., 2019; Bouarroudj, Boufaida, Bellatreche 2022; Bhingardive et al. 2015; Dutta, Borgohain 2021; Rouhizadeh, Shamsfard, Rouhizadeh 2020).
- 24 Graphs: (Abderrahim, Abderrahim 2022; Asgari-Bidhendi et al. 2021; Emami 2019; Arab, Jahromi, Fakhrahmad 2016; Neuman, Assaf, Cohen 2013; Zhang, Yan, Zheng 2019; Duque et al. 2018; Qiang et al. 2022; Bouarroudj, Boufaida, Bellatreche 2022; Pershina, He, Grishman 2015; Cohen, Zhong, Li 2022; Rademaker, Chalub, Freitas 2017; Tripodi, Pelillo 2017; Zhang et al. 2018; Ganea, Hofmann 2017; Hermansson et al. 2013; Vretinaris et al. 2021; Sun et al. 2020; Guerrieri et al. 2016; Barrena, Soroa, Agirre 2021; Vascon et al. 2021; Kwon, Oh, Ko 2021; Dongsuk et al. 2018; Aouicha, Taieb, Marai 2018).
- 16 ad hoc: (Tahat, Ahmad 2013; Li et al. 2015; Diekmann, Tratt 2019; Martínez, Berzal, Cubero 2021; Duran et al. 2013; Zhang et al. 2018; Veres 2018; Suzuki et al. 2018; Li. Cui. Wang 2020; Cabezudo, Pardo 2017; Nóbrega, Pardo 2014; Diamantini et al. 2015; Seifollahi, Shajari 2019; Aouicha, Taieb, Marai 2018; Duan et al. 2021; Asgari-Bidhendi et al. 2021).
- 12 LESK: (Petroliro 2016; Kumar et al. 2020; Basile, Caputo, Semeraro 2014; Ayetiran Agbele 2016; Sharma, Joshi 2019; Oele, Noord 2017; Cabezudo, Pardo, 2017; Nóbrega, Pardo 2014; Kumar, Siddappa 2021; Basuki et al. 2019; Kharate, Patil 2021; Tripathi et al. 2020).
- 8 BERT: (Al-Hajj, Jarrar 2022; Saeidi 2021; Yin et al. 2019; Wiedemann et al. 2019; Tan, Na 2019; Ionita et al. 2019; Duan et al. 2021; Qiang et al. 2022).

A Tabela 2 e figura 3 apresentam as formas de avaliações dos métodos (RQ2), foram registrada a presença de 36 tipos de avaliação, sendo as mais prevalentes: compared with other approaches (72,0%), F-Score (F1) (41,1%), Precision (37,4%) e Accuracy (34,6%) e os demais tipos de avaliação variaram de 6,5% a 28,0%, a demais formas de avaliações podem ser consultadas no Apêndice A.

Tabela 2 – Formas de avaliações mais frequentes nos 107 artigos selecionados

Avaliação	N (%) ⁽¹⁾	N (%) ⁽²⁾
Compared with other approaches	77 (2,1%)	77 (72,0%)
F-Score(F1)	44 (1,2%)	44 (41,1%)
Precision	40 (1,1%)	40 (37,4%)
Accuracy	37 (1,0%)	37 (34,6%)
Recall	30 (0,8%)	30 (28,0%)
Senseval2	15 (0,4%)	15 (14,0%)
Senseval3	13 (0,4%)	13 (12,1%)
SemEval2007	10 (0,3%)	10 (9,3%)
SemEval2013	8 (0,2%)	8 (7,5%)
SemEval2015	7 (0,2%)	7 (6,5%)

(1) Percentuais obtidos com base no total das frequências. No caso 36 formas de avaliações.

(2) Percentuais obtidos com base nos 107 artigos selecionados na pesquisa. Considerando a possibilidade de mais de uma forma de avaliação por artigo a soma das frequências pode ser superior ao total correspondente aos 107 artigos.

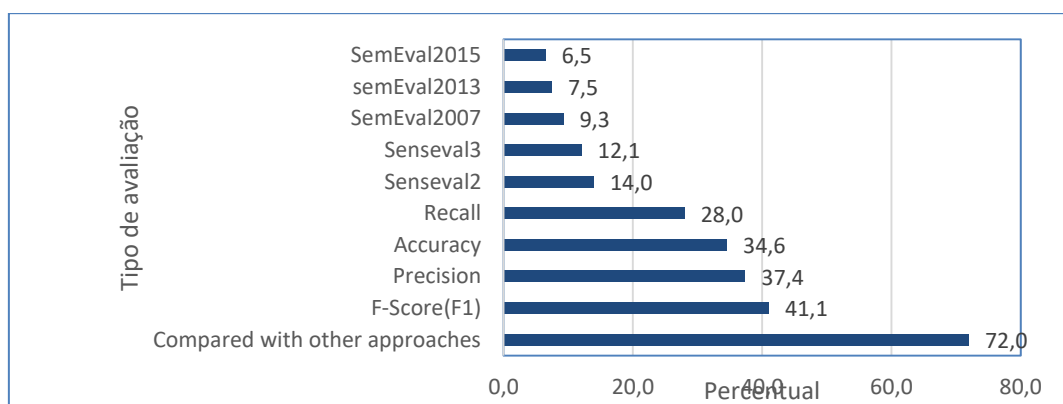


Figura 3 – Formas de avaliações mais frequentes nos 107 artigos selecionados

Na tabela 3 e figura 4 apresentam as limitações dos artigos (RQ3), os 4 tipos de limitações mais citadas tiveram frequências que variaram de 3 a 5 cada e a maioria (82,2%) se for com base nas 90 limitações e 69,2% se for com base nos 107 artigos foi agregada em outras limitações, e que estão listadas no Apêndice A.

Um ponto a destacar é alta quantidade de problemas encontrados, ou seja, as limitações das abordagens nos artigos, e que não são intrínsecos das línguas naturais, mas da própria abordagem, isso dá indícios que indica a dificuldade de resolver o problema WSD.

Tabela 3 – Tipos de limitações mais frequentes nos 107 artigos selecionados

Limitação	N (%) ⁽¹⁾	N (%) ⁽²⁾
Works only nouns	5 (5,6%)	5 (4,7%)
It's not fully automatic	4 (4,4%)	4 (3,7%)
Works only verbs	4 (4,4%)	4 (3,7%)
Depends on availability and human capacity	3 (3,3%)	3 (2,8%)
Other	74 (82,2%)	74 (69,2%)

(1) Percentuais obtidos com base no total das frequências. No caso 90 limitações.

(2) Percentuais obtidos com base nos 107 artigos selecionados na pesquisa.

Considerando a possibilidade de mais de uma limitação por artigo a soma das frequências pode ser superior ao total correspondente aos 107 artigos.

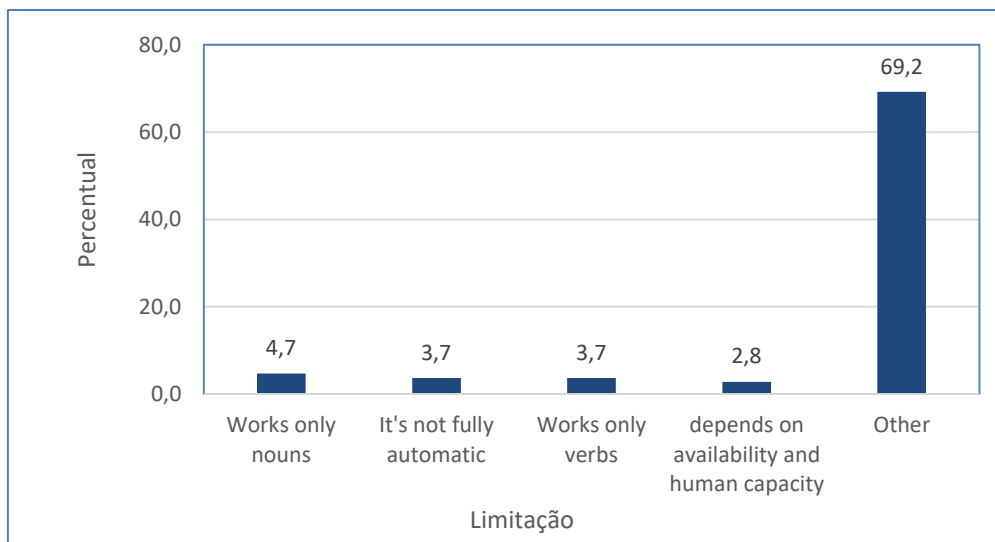


Figura 4 – Tipos de limitações mais frequentes nos 107 artigos selecionados

A tabela 4 e figura 5 informam que mais da metade (56,1%) dos artigos teve texts como objeto de estudo (RQ4), seguido de 15,0% como entity e 11,2% como outros objetos, e que estão listados no Apêndice A.

Tabela 4 – Objetos de estudos mais frequentes nos 107 artigos selecionados

Objetos do estudo	N (%) ⁽¹⁾
Texts	60 (56,1%)
Entity	16 (15,0%)
Author name	8 (7,5%)
Documents	6 (5,6%)
Person Name	5 (4,7%)
Other	12 (11,2%)

(1) Percentuais obtidos com base nos 107 artigos que foram selecionados.

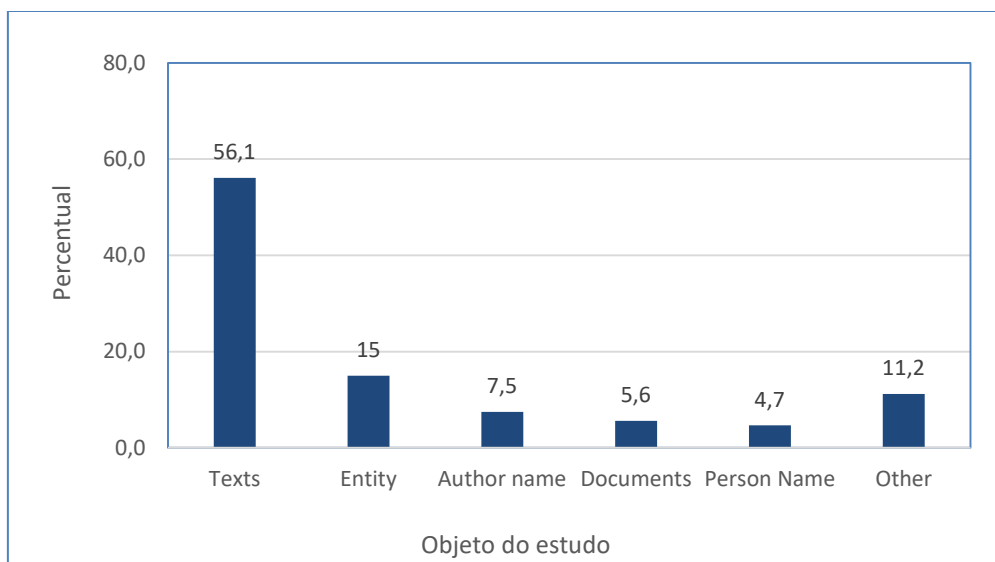


Figura 5 – Objetos de estudos mais frequentes nos 107 artigos selecionados

Na tabela 5 e figura 6 apresentam os idiomas utilizados nos datasets (RQ5), a maioria (64,5%) dos datasets dos artigos foram em English, seguido de 13,1% publicado em several idiomas, 7,5% em outros e os percentuais dos demais idiomas listados variaram de 1,9% a 4,7%, para saber o restante dos idiomas consulte o Apêndice A.

Tabela 5 – Idiomas dos datasets mais frequentes nos 107 artigos selecionados

Idioma	N (%) ⁽¹⁾
English	69 (64,5%)
Several	14 (13,1%)
Arabic	5 (4,7%)
Hindi	5 (4,7%)
Portuguese	4 (3,7%)
Chinese	2 (1,9%)
Others	8 (7,5%)

(1) Percentuais obtidos com base nos 107 artigos que foram selecionados.

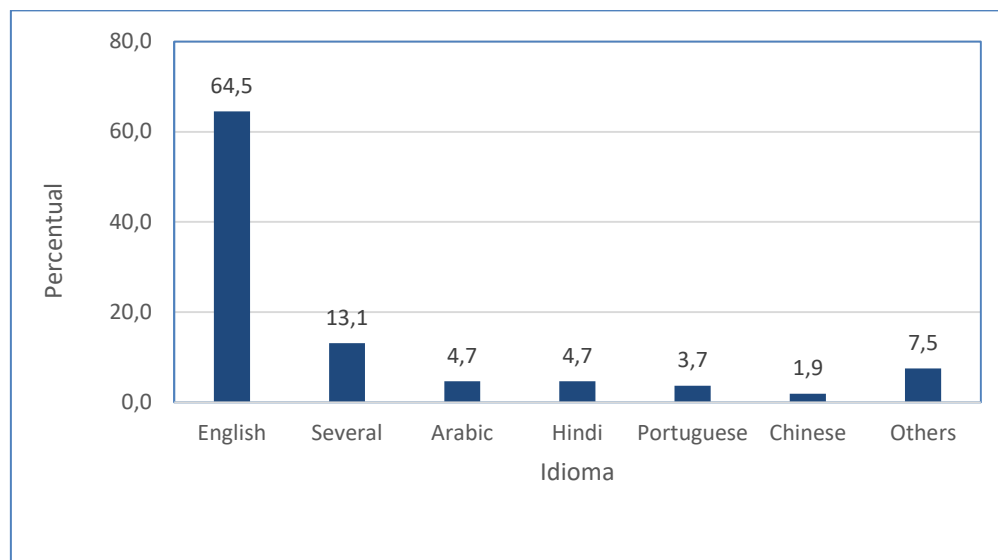


Figura 6 – Idiomas mais frequentes nos 107 artigos selecionados

Na tabela 6 e figura 7 apresentam os domínios, a maioria (80,4%) dos domínios (RQ6) correspondeu a several domínios, seguido de 11,2% no domínio library e as frequências dos outros 4 tipos listados variaram de um (0,9%) a 5,6%.

Tabela 6 – Domínio mais frequentes nos 107 artigos selecionados

Domínio	N (%) ⁽¹⁾
Several	86 (80,4%)
Library	12 (11,2%)
Biomedical	6 (5,6%)
Instructions	1 (0,9%)
News	1 (0,9%)
Psychological data	1 (0,9%)

(1) Percentuais obtidos com base nos 107 artigos que foram selecionados.

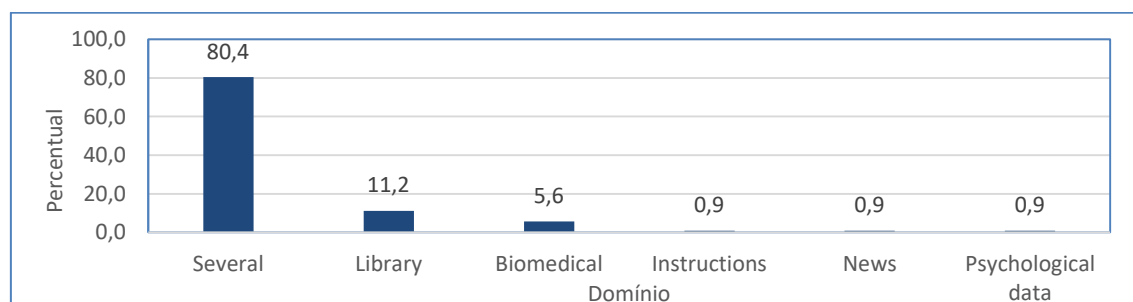


Figura 7 – Domínios mais frequentes nos 107 artigos selecionados

A tabela 7 e figura 8 enfatizam que: os tipos de abordagens (RQ7) mais frequentes foram: supervised (58,9%) e not supervised (22,4%) e os demais tipos tiveram percentuais que variaram de 0,9% a 13,1%, a abordagem que foi denominada como other, que corresponde a 1 artigo com (0,9%), foi chamada assim porque o autor não deixou claro a abordagem que usou.

Tabela 7 – Abordagens mais frequentes nos 107 artigos selecionados

Abordagem	N (%) ⁽¹⁾
Supervised	63 (58,9%)
Not Supervised	24 (22,4%)
knowledge-based	14 (13,1%)
semi-supervised	3 (2,8%)
Hybrid	2 (1,9%)
Other	1 (0,9%)

(1) Percentuais obtidos com base nos 107 artigos que foram selecionados.

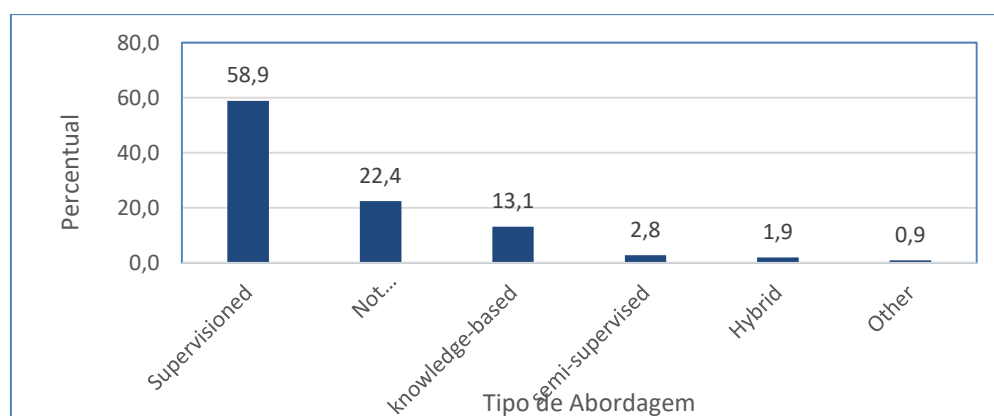


Figura 8 – Abordagens mais frequentes nos 107 artigos selecionados

Na tabela 8 e figura 9 apresentam os anos (RQ8), os anos com maiores frequências de artigos corresponderam a 2021 (22,4%) e 2019 (16,8%) e os percentuais dos demais anos analisados variaram de 3,7% a 9,3%. No caso do ano de 2022, a baixa quantidade de publicações pode ter sido afetada pelo filtro que foi aplicado em junho de 2022.

Tabela 8 – Publicações por ano nos 107 artigos selecionados

Anos	N (%) ⁽¹⁾
2012	4 (3,7%)
2013	7 (6,5%)
2014	5 (4,7%)
2015	6 (5,6%)
2016	8 (7,5%)
2017	8 (7,5%)
2018	12 (5,9%)
2019	18 (16,8%)
2020	10 (9,3%)
2021	24 (22,4%)
2022	5 (4,7%)

(1) Percentuais obtidos com base nos 107 artigos que foram selecionados.

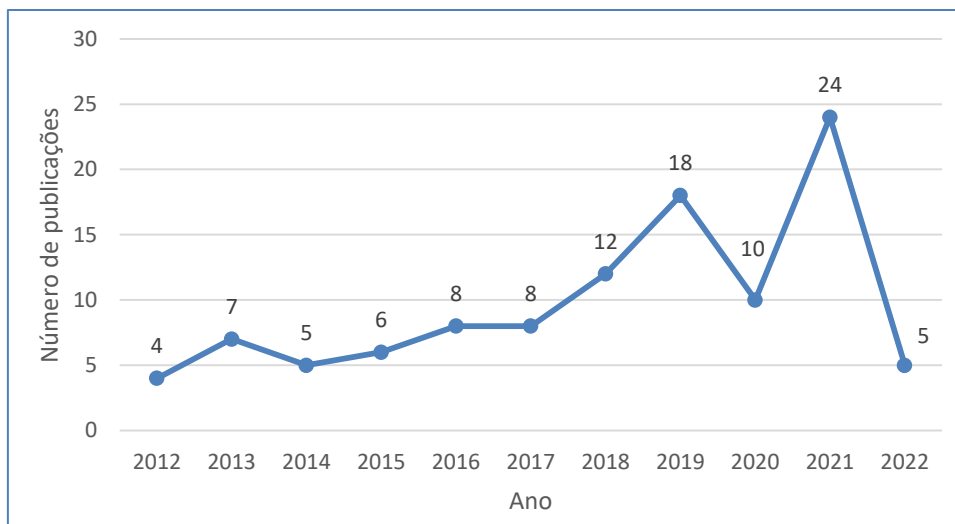


Figura 9 – Publicações por ano nos 107 artigos selecionados

A tabela 9 e figura 10 apresentam os países (RQ9) que publicaram nos 107 artigos, os três mais frequentes foram China (18,7%), EUA (10,3%) e Italy (9,3%), 29,0% correspondeu a outros países que estão listados no apêndice A, e os demais relacionados tiveram percentuais que variaram de 2,8% a 8,4%.

Tabela 9 – Países mais frequentes nos 107 artigos selecionados

País	N (%) ⁽¹⁾
China	20 (18,7%)
EUA	11 (10,3%)
Italy	10 (9,3%)
India	9 (8,4%)
Brazil	8 (7,5%)
Iran	5 (4,7%)
Saudi Arabia	4 (3,7%)
Canada	3 (2,8%)
South Korea	3 (2,8%)
Spain	3 (2,8%)
Outros	31 (29,0%)

(1) Percentuais obtidos com base nos 107 artigos que foram selecionados.

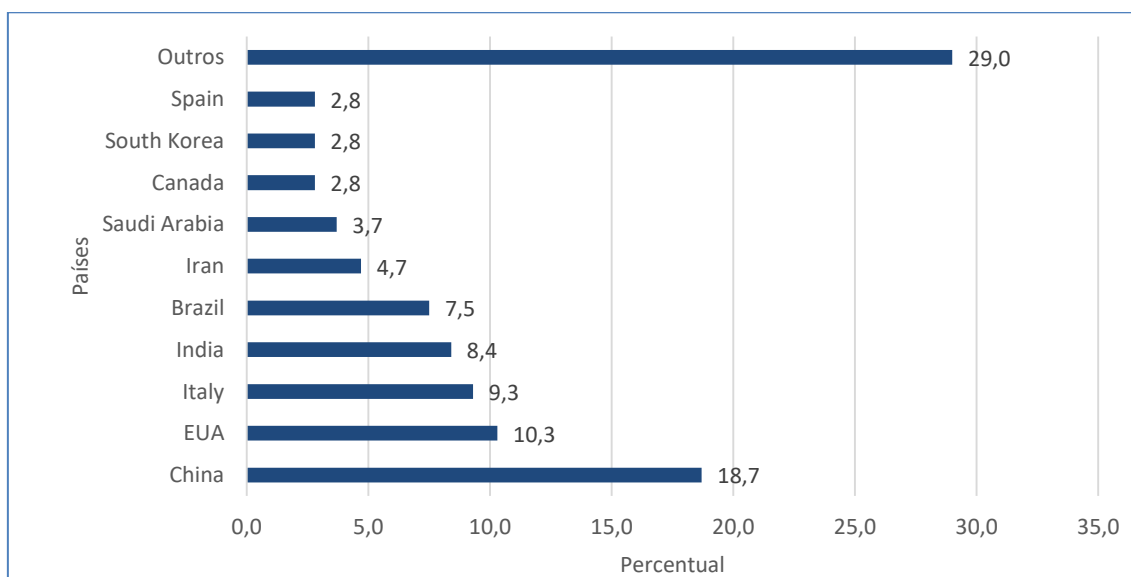


Figura 10 – Países mais frequentes nos 107 artigos selecionados

A Tabela 10 e figura 11 mostram os sites de buscas (RQ10), o site mais frequente foi Scopus com 37,4% do total dos 107 artigos selecionados, o menor valor (14,0%) correspondeu ao site Village e os percentuais dos outros dois sites relacionados variaram de 20,6% a 28,0%.

Tabela 10 – Frequência dos artigos nos repositórios

Sites / Engenho	N (%) ⁽¹⁾
Scopus	40 (37,4%)
IEEE	30 (28,0%)
ACM	22 (20,6%)
Village	15 (14,0%)

(1) Percentuais obtidos com base nos 107 artigos que foram selecionados.

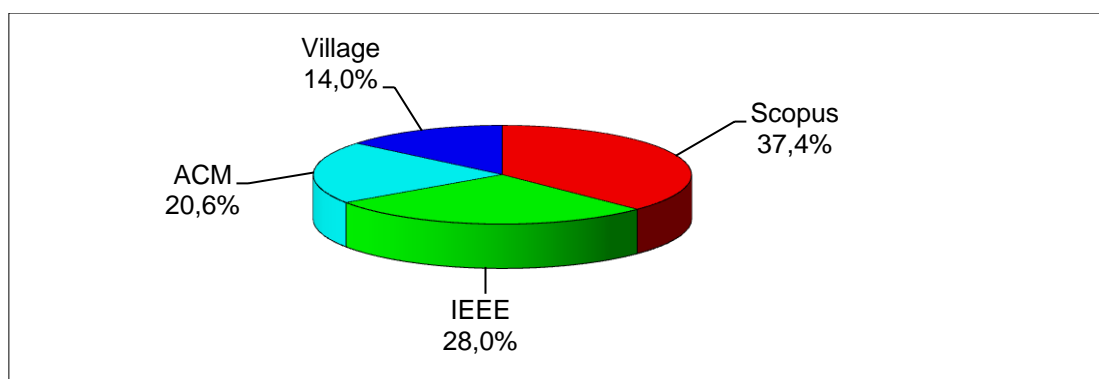


Figura 11 – Frequência dos artigos nos repositórios

Metanálise

A metanálise é uma análise estatística usada para interpretar os resultados de diferentes estudos individuais e independentes, geralmente extraídos de trabalhos publicados, com a aplicação de uma ou mais técnicas estatísticas, e com o objetivo de integrá-los, combinando e resumindo seus resultados, para sintetizar as suas conclusões ou mesmo extrair uma nova conclusão, além de possibilitar a inclusão de futuros estudos que venham a ser publicados (Monteiro 2010).

Sendo assim, pode-se afirmar que a metanálise é um estudo observacional da evidência e que se fundamenta na aplicação do método estatístico a um estudo de revisão sistemática, integrando dois ou mais estudos primários (Santos, Cunha 2013). Uma de suas vantagens é elevar a objetividade das revisões sistemáticas de literatura, minimizando possíveis vieses e aumentando a quantidade de estudos analisados (Filho et al. 2014). Portanto, a metanálise possibilita uma estimativa imparcial com aumento da precisão.

Os resultados de uma metanálise são representados por um gráfico do tipo *Forest plot*. A vantagem dos *Forest plots* é sumarizar todas as informações sobre a eficácia dos métodos avaliados, acurácia dos métodos dos algoritmos, no nosso caso, e a contribuição de cada estudo para a análise (Berwanger et al. 2007). Foi utilizado neste trabalho o software gratuito R Studio versão 2023.06.1 para gerar o gráfico *Forest plot* (Team 2014).

No *Forest plot*, o eixo vertical determina a ausência de efeito (Risco Relativo ou RR $\frac{1}{2}$ 1,0), apontando que não existem diferenças significativas entre os grupos quanto ao efeito das intervenções estudadas. De outro lado, o eixo horizontal determina valores de risco relativo (razão de taxas). Valores deslocados à esquerda do eixo vertical indicam que os achados favorecem a intervenção de comparação (controle); os valores à direita favorecem o tratamento em estudo. O status do controle ou tratamento é relativo ao estudo, pois o gráfico também pode ser usado com dois tratamentos, nosso caso, e o deslocamento para a esquerda ou para a direita indicará a intervenção preferida (algoritmo de desambiguação) (Medina & Pailaquilén 2010).

Para cada estudo, o ponto central representa a estimativa pontual do efeito e o comprimento da linha representa seu Intervalo de Confiança (IC). Corresponde ao intervalo de valores dentro do qual se assume que, com 95% de significância, se encontra o verdadeiro valor do efeito, permitindo também estabelecer direção e precisão. Linhas mais curtas dos Intervalos de Confiança indicam maior precisão dos resultados, melhorando sua validade. Quando esta linha toca o eixo central, há indicação de que mais estudos e mais amostras populacionais são necessários para alcançar a significância estatística da possível diferença entre as intervenções, (Medina & Pailaquilén 2010).

Outras variáveis presentes no Forest Plot são, I^2 de Higgins e Thompson, que representa a heterogeneidade entre os estudos, esta medida pode variar entre 0% e 100%, outra maneira de verificar a existência em metanálise é o teste Q de Cochran, a variável p ou p-value, apresentado pelo X^2 representa a probabilidade de ocorrer um determinado evento (Higgins et al. 2002; Rodrigues 2010).

O Coeficiente de Correlação de Kendall (representado pela letra grega τ), é uma variável que representa uma medida não-paramétrica que avalia a associação entre duas variáveis, sendo interpretado como uma medida de concordância entre dois conjuntos de classificações relativas a um conjunto de objetos de estudo. O valor negativo indica associação negativa, quanto mais alta a magnitude, mais forte a associação (Higgins et al. 2002; Rodrigues 2010).

Os dados de vários estudos só podem ser combinados na metanálise se tiverem características semelhantes (tipos de participantes, intervenções e medidas de resultados), (Medina & Pailaquilén 2010), sendo assim, após exame rigoroso nos 107 artigos selecionados, dois estudos foram elegíveis para a metanálise, (Qiang et al. 2022) e (Yin et al. 2019), apresentaram em comuns a medida dos resultados dos algoritmos de desambiguação, no caso a acurácia, além disso, comparam os resultados dos mesmos algoritmos, o Graph e o BERT.

De cada artigo foi coletados a quantidade de dados (amostra) analisados (representados no gráfico pela variável Total), como o total de erros das predições feitas por cada algoritmo aplicado (representados no gráfico pela variável Evento). O variável Evento conta as saídas com efeitos adversos, portanto para este trabalho ela representa o número de previsões erradas feitas por cada algoritmo avaliado.

Analisando o gráfico Forest plot deste MS na Figura 12, foi observado que, o IC do (Yin 2019) não tocou o RR, revelando que há diferença estatística favorecendo o BERT. Em contrapartida, por ter uma comprovação relevante, o estudo de (Qiang 2022) teve uma maior influencia nos resultados desta metanálise, pois seus pesos (aleatório = 50% e fixo = 99,7%) superaram os valores dos pesos do trabalho de (Yin 2019) (aleatório = 50% e fixo = 0,3%). Apesar de (Qiang 2022) favorecer o Graph e (Yin 2019) favorecer O BERT, o diamante ultrapassa o eixo central, indicando que não há diferença estatística significativa, para o favorecimento de um dos dois algoritmos. Na Tabela 11 é apresentado o resumo desta metanálise dos algoritmos avaliados.

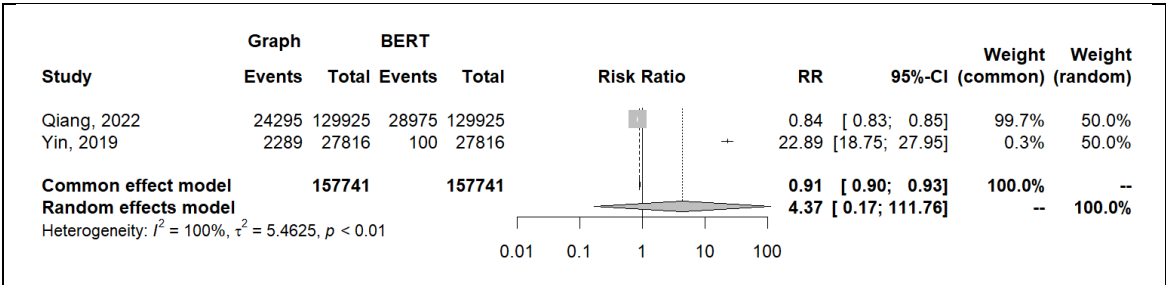


Figura 12 – Graph x BERT

Tabela 11 – Resumo da metanálise.

Artigos Avaliados	Algoritmos Avaliados	Heterogeneidade (I^2 %)	χ^2 p-value	Algoritmo vencedor (com diferença estatística significativa)
Qiang, 2022	Graph e BERT	Alta (100%)	p < 0,01	Não Houve
Yin, 2019				

Fonte: Elaborada pelo Autor.

Ameaças a Validade

As ameaças à validade podem restringir a habilidade de interpretar e/ou descrever resultados dos dados obtidos. Sendo assim, não há como desconsiderar as seguintes ameaças encontradas nesse estudo.

Validade de Construção: A *string* de busca e as questões de pesquisa usadas podem não contemplar toda a área WSD. Para mitigar essa ameaça, tentou-se elaborar uma *string* mais abrangente possível, quanto aos termos que pudessem ser usados na área, utilizando vários sinônimos. Estes termos foram identificados e refinados, efetuando a leitura de mais de 100 artigos sobre desambiguação.

Validade Interna:

Viés de Seleção: Os artigos foram incluídos ou excluídos de acordo com julgamento dos próprios pesquisadores. Consequentemente, alguns estudos podem ter sido categorizados incorretamente. Para mitigar estas ameaças, o conhecimento do assunto e a seleção minuciosa foram importantes para diminuir o viés.

Viés de Classificação: Alguns artigos selecionados não deixaram claro qual técnica foi utilizada/desenvolvida para efetuar a desambiguação, sendo assim, estes dados foram classificados como algoritmos ad hoc.

Validade Externa: Apesar da Scopus ser um motor de busca que consegue obter a maior quantidade de trabalhos da literatura científica, com mais de 21.950 *journals* e 120 mil conferências (Scopus 2019), não é possível alegar que os resultados desse mapeamento sistemático contemplam toda a área da WSD. Outra limitação foi não ter acesso a todos os artigos. No entanto, este trabalho apresentou evidências das principais técnicas utilizadas para desambiguação, identificando lacunas a serem exploradas e servindo como guia para futuros trabalhos nesta linha.

Conclusão

Neste trabalho, foi realizado um MS quantitativo, com o objetivo de conhecer os estudos empíricos e teóricos publicados que lidam com WSD publicados entre 2012 e 2023, para identificar, selecionar, extrair e caracterizar os dados seguindo o protocolo de pesquisa definido, e após a leitura de mais de 13.000 artigos, teve como resultado a seleção de 107 trabalhos, que apresentou o panorama das técnicas, abordagens, formas de avaliação das técnicas e outras informações relevantes.

Como resultados, valem destacar, os principais métodos foram Similarity algorithms com 27 estudos (25,2%), Graphs, com 24 (22,4%), ad hoc, com 16 (15,0%), LESK, com 12 (11,2%) e BERT, com 8 (7,5%). Em relação as formas de avaliação dos métodos, as mais utilizadas foram Compared with other approaches, com 77 (72,0%), F-Score(F1), com 44 (41,1%), Precision, com 40 (37,4%), Accuracy, com 37 (34,6%) e Recall, com 30 (28,0%). No contexto das limitações dos métodos, aparecem Works only nouns, com 5 (4,7%), It's not fully automatic, com 4 (3,7%), Works only verbs, com 4 (3,7%) e Depends on availability and human capacity, com 3 (2,8%).

Na metanálise, dois estudos foram elegíveis para análise, (Qiang et al. 2022) e (Yin et al. 2019), considerando os algoritmos Graph e BERT (Pre-training of Deep Bidirectional Transformers for Language Understanding), onde tivemos alta heterogeneidade, não apresentando diferença estatística significativa, sendo assim, não teve algoritmo vencedor.

Objetivo foi alcançado, ao analisar as publicações em ordem cronológica ao longo dos anos, foi perceptível a evolução das soluções para mitigar o problema de ambiguidade das linguagens naturais, no entanto, as limitações encontradas nos trabalhos revelam indícios que ainda a um longo caminho para percorrer, no intuito de preencher estas lacunas, trabalhos futuros devem focar nestes problemas, melhorando a qualidade dos datasets, ampliando os pré-processamentos, e desenvolvendo algoritmos que trate simultaneamente os problemas que dificultam a WSD.

Acreditamos que este trabalho é importante para a academia e desenvolvedores de aplicações PLN em geral, pois apresenta a trajetória e tendências. Além disso, pode oferecer mais uma forma de encontrar as melhores soluções para o seu problema no tratamento da ambiguidade nas linguagens naturais.

Referências

1. ABDERRAHIM, M. A.; ABDERRAHIM, M. E.-A. Arabic word sense disambiguation for information retrieval. *Transactions on Asian and Low-Resource Language Information Processing*, ACM New York, NY, v. 21, n. 4, p. 1–19, 2022.
2. Agirre, E. and Edmonds, P.G. (2006) *Word Sense Disambiguation: Algorithms and Applications*, Vol. 33, Springer Science & Business Media, Netherlands.
3. AL-HAJJ, M.; JARRAR, M. Arabglossbert: Fine-tuning bert on context-gloss pairs for wsd. *arXiv preprint arXiv:2205.09685*, 2022.
4. AL-QAWASMEH, O.; AL-SMADI, M.; FRAIHAT, N. Arabic named entity disambiguation using linked open data. In: IEEE. *2016 7th International Conference on Information and Communication Systems (ICICS)*. [S.l.], 2016. p. 333–338.
5. ALIWY, A. H.; TAHER, H. A. Word sense disambiguation: Survey study. *Journal of Computer Science*, v. 15, n. 7, p. 1004–1011, 2019.
6. AOUICHA, M. B.; TAIEB, M. A. H.; MARAI, H. I. Wordnet and wiktionary-based approach for word sense disambiguation. In: SPRINGER. *Transactions on Computational Collective Intelligence XXIX*. [S.l.], 2018. p. 123–143.
7. ARAB, M.; JAHROMI, M. Z.; FAKHRAHMAD, S. M. A graph-based approach to word sense disambiguation. an unsupervised method based on semantic relatedness. In: IEEE. *2016 24th Iranian Conference on Electrical Engineering (ICEE)*. [S.l.], 2016. p. 250–255.
8. ASGARI-BIDHENDI, M.; JANFADA, B.; HAVANGI, A.; HOSSAYNI, S. A.; MINAEIBIDGOLI, B. An unsupervised language-independent entity disambiguation method and its evaluation on the english and persian languages. *arXiv preprint arXiv:2102.00395*, 2021.
9. AYETIRAN, E. F.; AGBELE, K. An optimized lesk-based algorithm for word sense disambiguation. *Open Computer Science, Sciendo*, v. 8, n. 1, p. 165–172, 2016.
10. BASILE, P.; CAPUTO, A.; SEMERARO, G. An enhanced lesk word sense disambiguation algorithm through a distributional semantic model. In: *Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers*. [S.l.: s.n.], 2014. p. 1591–1600.

11. BASUKI, S.; KHOLIMI, A. S.; MINARNO, A. E.; SUMADI, F. D. S.; EFFENDY, M. R. A. Word sense disambiguation (wsd) for indonesian homograph word meaning determination by lesk algorithm application. In: IEEE. *2019 12th International Conference on Information & Communication Technology and System (ICTS)*. [S.l.], 2019. p. 8–15.
12. BARRENA, A.; SOROA, A.; AGIRRE, E. Towards zero-shot cross-lingual named entity disambiguation. *Expert Systems with Applications*, Elsevier, v. 184, p. 115542, 2021.
13. BERWANGER, O. et al. Como avaliar criticamente revisões sistemáticas e metanálises. *Ver Bras Ter Intensiva*, SciELO Brasil, v. 19, n. 4, p. 475–80, 2007.
14. BHATTACHARJEE, K.; SHIVAKARTHIK, S.; MEHTA, S.; KUMAR, A.; PHATANGARE, S.; PAWAR, K.; UKARANDE, S.; WANKHEDE, D.; VERMA, D. Survey and gap analysis of word sense disambiguation approaches on unstructured texts. In: IEEE. *2020 International Conference on Electronics and Sustainable Communication Systems (ICESC)*. [S.l.], 2020. p. 323–327.
15. BHINGARDIVE, S.; SINGH, D.; RUDRAMURTHY, V.; REDKAR, H.; BHATTACHARYYA, P. Unsupervised most frequent sense detection using word embeddings. In: *Proceedings of the 2015 conference of the North American Chapter of the Association for Computational Linguistics: Human language technologies*. [S.l.: s.n.], 2015. p. 1238–1243.
16. BOUARROUDJ, W.; BOUFAIDA, Z.; BELLATRECHE, L. Named entity disambiguation in short texts over knowledge graphs. *Knowledge and Information Systems*, Springer, v. 64, n. 2, p. 325–351, 2022.
17. CABEZUDO, M. A. S.; PARDO, T. A. Exploring classical and linguistically enriched knowledge-based methods for sense disambiguation of verbs in brazilian portuguese news texts. *Sociedad Española para el Procesamiento del Lenguaje Natural*, 2017.
18. CAMACHO-COLLADOS, J.; PILEHVAR, M. T. From word to sense embeddings: A survey on vector representations of meaning. *Journal of Artificial Intelligence Research*, v. 63, p. 743–788, 2018.
19. COHEN, F. S.; ZHONG, Z.; LI, C. Semantic graph for word disambiguation in machine translation. *Multimedia Tools and Applications*, Springer, v. 81, n. 30, p. 43485–43502, 2022.
20. DHUNGANA, U. R.; SHAKYA, S. Hypernymy in wordnet, its role in wsd, and its limitations. In: IEEE. *2015 7th International Conference on Computational Intelligence, Communication Systems and Networks*. [S.l.], 2015. p. 15–19.
21. DIEKMANN, L.; TRATT, L. Default disambiguation for online parsers. In: *Proceedings of the 12th ACM SIGPLAN International Conference on Software Language Engineering*. [S.l.: s.n.], 2019. p. 88–99.
22. DING, Y. News article name disambiguation model based on reinforcement learning. In: IEEE. *2021 International Conference on Artificial Intelligence, Big Data and Algorithms (CAIBDA)*. [S.l.], 2021. p. 122–127.
23. DIAMANTINI, C.; MIRCOLI, A.; POTENA, D.; STORTI, E. Semantic disambiguation in a social information discovery system. In: IEEE. *2015 International Conference on Collaboration Technologies and Systems (CTS)*. [S.l.], 2015. p. 326–333.
24. DONGSUK, O.; KWON, S.; KIM, K.; KO, Y. Word sense disambiguation based on word similarity calculation using word vector representation from a knowledge-based graph. In: *Proceedings of the 27th international conference on computational linguistics*. [S.l.: s.n.], 2018. p. 2704–2714.

25. DUAN, R.; YANG, X.; WANG, Q.; ZHAO, Y. A parameter-adaptive convolution neural network for capturing the context-specific information in natural language understanding. In: IEEE. *2021 2nd International Conference on Electronics, Communications and Information Technology (CECIT)*. [S.l.], 2021. p. 111–121.
26. DUQUE, A.; STEVENSON, M.; MARTINEZ-ROMO, J.; ARAUJO, L. Co-occurrence graphs for word sense disambiguation in the biomedical domain. *Artificial intelligence in medicine*, Elsevier, v. 87, p. 9–19, 2018.
27. DURAN, M. S.; SCARTON, C.; ALUÍSIO, S.; RAMISCH, C. Identifying pronominal verbs: Towards automatic disambiguation of the clitic 'se' in portuguese. In: *Proceedings of the 9th Workshop on Multiword Expressions*. [S.l.: s.n.], 2013. p. 93–100.
28. DUTTA, A.; BORGOHAIN, S. K. Verb sense disambiguation by measuring semantic relatedness between verb and surrounding terms of context. *International Journal of Advanced Computer Science and Applications*, Science and Information (SAI) Organization Limited, v. 12, n. 2, 2021.
29. EMAMI, H. A graph-based approach to person name disambiguation in web. *ACM Transactions on Management Information Systems (TMIS)*, ACM New York, NY, USA, v. 10, n. 2, p. 1–25, 2019.
30. FARAHANI, Y. V.; JANFADA, B.; BIDGOLI, B. M. A review of algorithms, datasets, and criteria in word sense disambiguation with a view to its use in islamic texts. In: IEEE. *2020 8th Iranian Joint Congress on Fuzzy and intelligent Systems (CFIS)*. [S.l.], 2020. p. 172–179.
31. FILHO, D. B. F. et al. O que é, para que serve e como se faz uma meta-análise? *Teoria & Pesquisa: Revista de Ciência Política*, v. 23, n. 2, 2014.
32. GANEA, O.-E.; HOFMANN, T. Deep joint entity disambiguation with local neural attention. *arXiv preprint arXiv:1704.04920*, 2017.
33. GOGOI, A.; BARUAH, N.; NATH, L. J. Assamese word sense disambiguation using cuckoo search algorithm. *Procedia Computer Science*, Elsevier, v. 189, p. 142–147, 2021.
34. GUERRIERI, A.; RAHIMIAN, F.; GIRDZIJAUSKAS, S.; MONTRESOR, A. Tovel: Distributed graph clustering for word sense disambiguation. In: IEEE. *2016 IEEE 16th International Conference on Data Mining Workshops (ICDMW)*. [S.l.], 2016. p. 623–630.
35. HERMANSSON, L.; KEROLA, T.; JOHANSSON, F.; JETHAVA, V.; DUBHASHI, D. Entity disambiguation in anonymized graphs using graph kernels. In: *Proceedings of the 22nd ACM international conference on Information & Knowledge Management*. [S.l.: s.n.], 2013. p. 1037–1046.
36. HIGGINS, J.; THOMPSON, S.; DEEKS, J.; ALTMAN, D. Statistical heterogeneity in systematic reviews of clinical trials: a critical appraisal of guidelines and practice. *Journal of health services research & policy*, SAGE Publications Sage UK: London, England, v. 7, n. 1, p. 51–61, 2002.
37. IONITA, M.; KASHNITSKY, Y.; KRIGE, K.; LARIN, V.; LOGVINENKO, D.; ATANASOV, A. Resolving gendered ambiguous pronouns with bert. *arXiv preprint arXiv:1906.01161*, 2019.

38. KHABSA, M.; TREERATPITUK, P.; GILES, C. L. Online person name disambiguation with constraints. In: *Proceedings of the 15th acm/ieee-cs joint conference on digital libraries*. [S.l.: s.n.], 2015. p. 37–46.
39. KHARATE, N. G.; PATIL, V. H. Word sense disambiguation for marathi language using wordnet and the lesk approach. In: SPRINGER. *Proceeding of First Doctoral Symposium on Natural Computing Research: DSNCR 2020*. [S.l.], 2021. p. 45–54.
40. Kitchenham, B. (2004). Procedures for performing systematic reviews. Keele, UK, Keele University, 33(2004), 1–26.
41. Kitchenham, B. & Charters, S. (2007). Guidelines for performing systematic literature reviews in software engineering. Technical report EBSE-2007-01. School of Computer Science and Mathematics, Keele University.
42. Kitchenham, B.A.; Brereton, P.; Li, Z.; Budgen, D.; Burn, A (2011) Repeatability of systematic literature reviews. 15th Annual Conference on Evaluation Assessment in Software Engineering (EASE 2011), p. 46-55.
43. KWON, S.; OH, D.; KO, Y. Word sense disambiguation based on context selection using knowledge-based word similarity. *Information Processing & Management*, Elsevier, v. 58, n. 4, p. 102551, 2021.
44. KUMAR, B. M.; SIDDAPPA, M. Kannada word sense disambiguation using semantic relations. In: IOP PUBLISHING. *Journal of Physics: Conference Series*. [S.l.], 2021. v. 1767, n. 1, p. 012025.
45. KUMAR, M.; MUKHERJEE, P.; HENDRE, M.; GODSE, M.; CHAKRABORTY, B. Adapted lesk algorithm based word sense disambiguation using the context information. *International Journal of Advanced Computer Science and Applications*, Science and Information (SAI) Organization Limited, v. 11, n. 3, p. 254–260, 2020.
46. LAATAR, R.; ALOULOU, C.; BELGUITH, L. H. Disambiguating arabic words according to their historical appearance in the document based on recurrent neural networks. *ACM Transactions on Asian and Low-Resource Language Information Processing (TALLIP)*, ACM New York, NY, USA, v. 19, n. 6, p. 1–16, 2020.
47. L'HEUREUX, A.; GROLINGER, K.; ELYAMANY, H. F.; CAPRETZ, M. A. Machine learning with big data: Challenges and approaches. *Ieee Access*, IEEE, v. 5, p. 7776–7797, 2017.
48. LI, H.; CUI, Y.; WANG, T. An effective approach for automatic author name disambiguation based on multiple strategies. In: *Proceedings of the 3rd International Conference on Computer Science and Software Engineering*. [S.l.: s.n.], 2020. p. 169–175.
49. LI, Z.; ZHANG, X.; HUANG, H.; XIE, Q.; ZHU, J.; ZHOU, X. Addressing instance ambiguity in web harvesting. In: *Proceedings of the 18th International Workshop on Web and Databases*. [S.l.: s.n.], 2015. p. 6–12.
50. LIU, Y.; OTT, M.; GOYAL, N.; DU, J.; JOSHI, M.; CHEN, D.; LEVY, O.; LEWIS, M.; ZETTLEMOYER, L.; STOYANOV, V. Roberta: A robustly optimized bert pretraining approach. *arXiv preprint arXiv:1907.11692*, 2019.
51. MARTÍNEZ, V.; BERZAL, F.; CUBERO, J.-C. Disambiguation of semantic relations using evidence aggregation according to a sense inventory. *IEEE Transactions on Knowledge and Data Engineering*, IEEE, v. 33, n. 7, p. 2875–2887, 2021.

52. Medina, E. U., & Pailaquilen, R. M. B. (2010). Systematic review and its relationship with evidence-based practice in health. *Revista Latino-Americana de Enfermagem*, 18(4), 1–8.
53. MONTEIRO, R. N. M. *Metodologias de meta-análise aplicadas nas Ciências da Saúde*, Universidade da Beira Interior, 2010.
54. NEUMAN, Y.; ASSAF, D.; COHEN, Y. A cognitively motivated word sense induction algorithm. In: IEEE. *2013 IEEE Symposium on Computational Intelligence, Cognitive Algorithms, Mind, and Brain (CCMB)*. [S.l.], 2013. p. 66–72.
55. NÓBREGA, F. A. A.; PARDO, T. A. S. General purpose word sense disambiguation methods for nouns in portuguese. In: SPRINGER. *International Conference on Computational Processing of the Portuguese Language*. [S.l.], 2014. p. 94–101.
56. OELE, D.; NOORD, G. V. Distributional lesk: Effective knowledge-based word sense disambiguation. In: *IWCS 2017—12th International Conference on Computational Semantics—Short papers*. [S.l.: s.n.], 2017.
57. OKUL, Ş.; AKSU, D.; AYDIN, M. A. Applications of deep learning and big data technologies. In: IEEE. *2019 16th International Multi-Conference on Systems, Signals & Devices (SSD)*. [S.l.], 2019. p. 550–553.
58. ORKPOL, K.; YANG, W. Word sense disambiguation using cosine similarity collaborates with word2vec and wordnet. *Future Internet*, MDPI, v. 11, n. 5, p. 114, 2019.
59. PAVITHRA, C.; MANDAL, S. An overview of relevant literature on different approaches to word sense disambiguation. In: IEEE. *2021 Fourth International Conference on Electrical, Computer and Communication Technologies (ICECCT)*. [S.l.], 2021. p. 1–6.
60. PEREIRA, D. A.; SILVA, E. E. B. da; ESMIN, A. A. Disambiguating publication venue titles using association rules. In: IEEE. *IEEE/ACM Joint Conference on Digital Libraries*. [S.l.], 2014. p. 77–86.
61. PERSHINA, M.; HE, Y.; GRISHMAN, R. Personalized page rank for named entity disambiguation. In: *Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. [S.l.: s.n.], 2015. p. 238–243.
62. Petersen, K., Vakkalanka, S., & Kuzniarz, L. (2015). Guidelines for conducting systematic mapping studies in software engineering: An update. *Information and Software Technology*, 64, 1–18. doi:10.1016/j.infsof.2015.03.007
63. Petersen, K.; Feldt, R.; Mujtaba, S.; Mattsson, M. (2008) Systematic Mapping Studies in Software Engineering. *12nd International Conference on Evaluation and Assessment in Software Engineering (EASE 2008)*, p. 1-10.
64. PETROLITO, T. A language-independent lesk based approach to word sense disambiguation. In: *Proceedings of the 8th Global WordNet Conference (GWC)*. [S.l.:s.n.], 2016. p. 275–281.
65. POPOV, A. Neural network models for word sense disambiguation: an overview. *Cybernetics and information technologies*, v. 18, n. 1, p. 139–151, 2018.
66. QIANG, C.; LI, X.; MA, X.; HAN, W.; YAO, Y. Entity disambiguation method based on graph attention networks. In: IEEE. *2022 14th International Conference on Measuring Technology and Mechatronics Automation (ICMTMA)*. [S.l.], 2022. p. 912–919.

67. RADEMAKER, A.; CHALUB, F.; FREITAS, C. Two corpus based experiments with the portuguese and english wordnets. In: *LDK Workshops*. [S.l.: s.n.], 2017. p. 134–145.
68. RAIS, M.; LACHKAR, A. Biomedical word sense disambiguation context-based: Improvement of senserelate method. In: IEEE. *2016 International Conference on Information Technology for Organizations Development (IT4OD)*. [S.l.], 2016. p. 1–6.
69. R. Navigli, "Natural language understanding: Instructions for (presente and future) use," in *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence, IJCAI*, 2018, pp. 5697-5702.
70. Roberto Navigli. 2009. Word Sense Disambiguation: A Survey. *ACM Computing Surveys (CSUR)*, 41(2):10.
71. RODRIGUES, C. L. Metanálise: um guia prático. 2010.
72. SABBIR, A.; JIMENO-YEPES, A.; KAVULURU, R. Knowledge-based biomedical word sense disambiguation with neural concept embeddings. In: IEEE. *2017 IEEE 17th International Conference on Bioinformatics and Bioengineering (BIBE)*. [S.l.], 2017. p. 163–170.
73. SAEIDI, M. Contextbert: Contextual graph representation learning in text disambiguation. 2021.
74. SAHA, P. K.; MOU, A. D.; MITTRA, T. A bangla word sense disambiguation technique using minimum edit distance algorithm and cosine distance. In: IEEE. *2019 23rd International Computer Science and Engineering Conference (ICSEC)*. [S.l.], 2019. p. 1–6.
75. SANTOS, E. J. F. D.; CUNHA, M. Interpretação crítica dos resultados estatísticos de uma meta-análise: Estratégias metodológicas. *Millenium*, Instituto Politécnico de Viseu, n. 44, p.85–89, 2013.
76. SARIKA; SHARMA, D. K. Hindi word sense disambiguation using cosine similarity. In: SPRINGER. *Proceedings of International Conference on ICT for Sustainable Development: ICT4SD 2015 Volume 2*. [S.l.], 2016. p. 801–808.
77. SAJINI, G.; KALLIMANI, J. S. Recognition and disambiguation of polysemy words in entered hindi documents. In: IEEE. *2017 International Conference on Electrical, Electronics, Communication, Computer, and Optimization Techniques (ICEECCOT)*. [S.l.], 2017. p. 1–8.
78. SCANNAVINO, K. R. F.; NAKAGAWA, E. Y.; FABBRI, S. C. P. F.; FERRARI, F. C. Revisão sistemática da literatura em engenharia de software: teoria e prática. 2017.
79. SCOPUS. *Scopus - Elsevier Database*. 2019. Disponível em: <<https://www.scopus.com>>.
80. SEIFOLLAHI, S.; SHAJARI, M. Word sense disambiguation application in sentimento analysis of news headlines: an applied approach to forex market prediction. *Journal of Intelligent Information Systems*, Springer, v. 52, p. 57–83, 2019.
81. SHARMA, P.; JOSHI, N. Design and development of a knowledge based approach for word sense disambiguation by using wordnet for hindi. *International Journal of Innovative Technology and Exploring Engineering*, v. 8, n. 3, 2019.

82. SOUSA, S.; MILIOS, E.; BERTON, L. Word sense disambiguation: an evaluation study of semi-supervised approaches with word embeddings. In: IEEE. *2020 International Joint Conference on Neural Networks (IJCNN)*. [S.l.], 2020. p. 1–8.
83. SUN, Q.; PENG, H.; LI, J.; WANG, S.; DONG, X.; ZHAO, L.; PHILIP, S. Y.; HE, L. Pairwise learning for name disambiguation in large-scale heterogeneous academic networks. In: IEEE. *2020 IEEE International Conference on Data Mining (ICDM)*. [S.l.], 2020. p. 511–520.
84. SUZUKI, R.; KOMIYA, K.; ASAHARA, M.; SASAKI, M.; SHINNOU, H. All-words word sense disambiguation using concept embeddings. In: *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*. [S.l.: s.n.], 2018.
85. TAHAT, S.; AHMAD, K. A method on lexical disambiguation in distributed heterogeneous autonomous database. In: IEEE. *2013 International Conference on Research and Innovation in Information Systems (ICRIIS)*. [S.l.], 2013. p. 330–335.
86. TAN, S.-S.; NA, J.-C. Positional attention-based frame identification with bert: A deep learning approach to target disambiguation and semantic frame selection. *arXiv preprint arXiv:1910.14549*, 2019.
87. TEAM, R. C. *R: A Language and Environment for Statistical Computing*. 2014. Disponível em: <<https://www.r-project.org/>>.
88. TRIPATHI, P.; MUKHERJEE, P.; HENDRE, M.; GODSE, M.; CHAKRABORTY, B. Word sense disambiguation in hindi language using score based modified lesk algorithm. *International Journal of Computing and Digital Systems*, University of Bahrain, v. 10, p. 2–20, 2020.
89. TRIPODI, R.; PELILLO, M. A game-theoretic approach to word sense disambiguation. *Computational Linguistics*, MIT Press One Rogers Street, Cambridge, MA 02142-1209, USA journals-info . . . , v. 43, n. 1, p. 31–70, 2017.
90. TUAN, K.-W.; LIN, Y.-C.; CHANG, J. S.; LEE, K.-L.; CHEN, L.-K. Consenses: Disambiguating content word groups based on knowledge base and definition embedding. In: IEEE. *2020 International Conference on Technologies and Applications of Artificial Intelligence (TAAI)*. [S.l.], 2020. p. 260–265.
91. VASCON, S.; ASLAN, S.; BIGAGLIA, G.; GIUDICE, L.; PELILLO, M. Transductive visual verb sense disambiguation. In: *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*. [S.l.: s.n.], 2021. p. 3050–3059.
92. VERES, C. Task specific disambiguation in a web page: Towards symbiotic annotation. In: *Proceedings of the 8th International Conference on Web Intelligence, Mining and Semantics*. [S.l.: s.n.], 2018. p. 1–6.
93. VRETINARIS, A.; LEI, C.; EFTHYMIOU, V.; QIN, X.; ÖZCAN, F. Medical entity disambiguation using graph neural networks. In: *Proceedings of the 2021 international conference on management of data*. [S.l.: s.n.], 2021. p. 2310–2318.
94. Warren Weaver. 1955. Translation. Machine translation of languages, 14:15-23.
95. WIEDEMANN, G.; REMUS, S.; CHAWLA, A.; BIEMANN, C. Does bert make any sense? interpretable word sense disambiguation with contextualized embeddings. *arXiv preprint arXiv:1909.10430*, 2019.
96. YANG, Z.; DAI, Z.; YANG, Y.; CARBONELL, J.; SALAKHUTDINOV, R. R.; LE, Q. V. Xlnet: Generalized autoregressive pretraining for language understanding. *Advances in neural information processing systems*, v. 32, 2019.

97. YIN, X.; HUANG, Y.; ZHOU, B.; LI, A.; LAN, L.; JIA, Y. Deep entity linking via eliminating semantic ambiguity with bert. *IEEE Access*, IEEE, v. 7, p. 169434–169445, 2019.
98. ZHANG, K.; ZHU, Y.; GAO, W.; XING, Y.; ZHOU, J. An approach for named entity disambiguation with knowledge graph. In: IEEE. *2018 International Conference on Audio, Language and Image Processing (ICALIP)*. [S.l.], 2018. p. 138–143.
99. ZHANG, W.; YAN, Z.; ZHENG, Y. Author name disambiguation using graph node embedding method. In: IEEE. *2019 IEEE 23rd international conference on computer supported cooperative work in design (CSCWD)*. [S.l.], 2019. p. 410–415.
100. ZHANG, Y.; ZHANG, F.; YAO, P.; TANG, J. Name disambiguation in aminer: Clustering, maintenance, and human in the loop. In: *Proceedings of the 24th ACM SIGKDD international conference on knowledge discovery & data mining*. [S.l.: s.n.], 2018. p. 1002–1011.

Apêndice A

Este apêndice lista o restante dos dados das tabelas classificados como others dos 107 trabalhos selecionados, e demais achados durante a pesquisa.

Na tabela 1 lista os 33 métodos restantes dos 107 artigos selecionados (RQ1), e que foram classificados como others, bem como os demais métodos e tecnologias subjacentes para WSD, encontrados durante a leitura de mais de 13.000 trabalhos do MS. Os 33 artigos estão com asterisco no final da palavra para diferenciar dos demais na tabela.

Tabela 1 – Métodos e tecnologias subjacentes para WSD.

Métodos		
Algorithm of Mihalcea	FastTex	PageRank*
Alkie	Firefly algorithm	Part of speech*
Ant Colony Optimization	Fuzzy logic	Patterns
algorithms (ACO)*	Game-theoretic*	Perceptron
Associative memories	Gated Recurrent Units-GRU	Probability of association
Attention mechanism*	Generative Adversarial	Quadratic programming (QP)
Autoencoder	Networks (GAN)	Quantum Probability
Bayesian Classification	Generative model	Theory(QPT)*
Bidirectional Beam Search	Genetic Algorithm (GA)*	Random Forest
with Gold path (BiBSG)	Glossary*	Random walk algorithm
BiLSTM	GlossBERT	Reinforcement learning
Biobert	Glove	Reservoir computing (RC)*
BioCBERT	Harmony Search Algorithm	RoBERTa*
BLSTM	(HSA)	Rules*
Case Based Reasoning	Hidden Semantic Information	Semantic similarity*
CBOW	Extractor (HSIE)	Sense incorporation learning
Cconditional probability	Imply	algorithm
Centrality scoring	Incremental algorithm for	Sense2Vec
Classifier*	mining	SenseRelate
Cluster*	Kernel	Seq2Seq*
Conceptual Density (CD)	K-means	Skip-gram
Conditional Random Fields	Knearest neighbors (KNN)	SLAND (Self-training Lazy
(CRFs)	Label Propagation	Associative Name
Context-gloss augmentation	Latent Dirichlet Allocation	Disambiguator)
Convolutional latent semantic	(LDA)*	Sparse Associative Memories
model (CLSM)	Latent Semantic Analysis	(SAM)
Convolutional neural network	(LSA)*	Subsume
Co-occurrence	Latent semantic space	Supervised term expansion
Crowdsourcing	Levenshtein distance*	Support Vector
Cuckoo search algorithm	Limited performance on some	Machines (SVM)*
Decision Tree (DT)	datasets	Surrounding Word Sense
Deep bi-directional ESN	Linear discriminant analysis	Model (SWSM)
(DBiESN) *	Linked open data	T5 Encoder
Deep generative adversarial	Local Neighborhood	Temporal logic*
networks	Long Short-Term Memory	Topic modeling*
Deep Learning*	(LSTM)*	Topic-specific Language*
Diffusion	Markov Logic Network	Model (TLM method)
Dirichlet Process Gaussian	mBERT	Top-k similarity
Mixture Model (DPGMM)	mDistilBERT	Transformer*
Discrete Crow Search	Naive Bayes*	Triplet Networks (TNs)
Algorithm (DCSA)	Nearest cluster model	Unsupervised term expansion
Doc2Vec	Needs large vocabularies for	Variational Semantic Memory
Domain knowledge*	best results	(VSM)
Echo State Network (ESN)*	Neural Attention Module	Vector Space Models*
ELMo	(NAM)	Word2Vec*
Embeddings	Non-negative matrix	WorNet
Encoder-decoder	factorization (NMF)	Yarowsky algorithm
Evolutionary algorithm*	Optimum Path Forest (OPF)*	

Os dados foram classificados como others porque foram menos usados em relação aos dados que aparecem nas tabelas do texto principal deste MS, e foram colocados neste Apêndice para não gerar um número de páginas excessivas no texto principal.

Na tabela 2 lista as demais formas de avaliações dos métodos (RQ2) que não aparecem no texto principal:

Tabela 2 – Demais formas usadas para avaliar sistemas WSD.

Formas de avaliações
Ad hoc Attempt Babelfy Coverage DBLP Computer Science Bibliography DBpedia Error rate Execution Time FrenchSemEval General Language Understanding Evaluation (GLUE) K Metric OMSTI SemCor SemEval-2010 Success rate

Na tabela 3 lista as limitações dos artigos (RQ3) que não aparecem na tabela do texto principal:

Tabela 3 – Limitações dos trabalhos.

Limitações dos trabalhos
Ability to capture latent and fine-grained semantics is limited Adaptability performance decreases on larger scale datasets Annotation errors in some datasets Association of ambiguous identifiers with their entities Attention captures only adjacent nodes Cannot identify all types of entities Classification errors may occur Complex rules Considers only the context of the word Depends on complex vector representations. Depends on preprocessing to use another language Difficulty in semantic understanding of multiple demonstrative pronouns Do not eliminate the ambiguity of part of speech other than nouns Does not accept embedding phrases Does not act in a specific domain Does not handle semantic structure does not handle syntactic structure Does not work online with data on the web Errors caused by metonymy Existence of noise without result Fails if polysemy is high Failure attribution of sense may occur Failure to choose the meaning of the word may occur Failure, in some cases, to identify the entity. Gender bias may occur Is unable to effectively resolve situations complex It cannot resolve all correspondence issues It does not automatically decide which word in the sentences is strongly related to the target word It does not consider the context It does not solve all hierarchical grouping problems.

It does not treat interrogative words
 It doesn't perform great on all datasets.
 It is unable to handle sparse nameset
 It needs a lot of data to ensure model performance
 Its performance depends on the quality of the dictionaries.
 Limited applicability due to poor performance in some cases
 Limited word coverage of the lexical semantic network limits accuracy
 Low accuracy
 Low coverage dictionaries limit the accuracy of the method
 Low performance
 low performance coarse-grained corpora
 Method performance depends on language brightness vectors
 Model can get confused and suggest more than one word
 Model is subject to identification failures.
 Need quality annotations to return correct results
 Needs a parallel multilingual corpus
 No high efficacy in co-occurrence of terms within attributes
 Not a wsd system of all words
 Not applied to other datasets
 Not effective for recognizing sentences with complex semantics
 Not the same performance in all languages
 Occurrences of entities that are not correctly disambiguated
 Occurrences of false positives and false negatives
 Occurrences of wrong predictions
 Only in short text entities
 Over-fitting
 Performance decreases when no data is reported in the training set, along with the test step set.
 Performance is limited by language links from a dataset
 Performance is limited by the quality of the knowledge base
 Performance is limited by the sense coverage of the dataset
 Performance is restricted due to data spacing issue
 Pre-processing inefficiency
 Presence of unrelated words
 Results in average accuracy
 Slow running algorithms
 Some predictions are not reliable enough.
 Some types of entities had relatively lower accuracy results
 The algorithm used is not very fast
 The best choices of sense depend on the size of the corpus and the target word
 The method fails on existence of noisy data
 The method fails to detect some error situation.
 The quality of the disambiguation depends on the quality of the instances.
 The results are no better than the best performing algorithms.
 The results with verbs is not good
 The selected sense does not match the context.
 The size of the context window influences the result
 The use of different supervised signals did not bring significant improvements
 There are cases where you cannot determine the sense of the word
 There are some cases where the models fail and I don't infer the correct sense
 There are some lexical gaps that lead to choosing a generic sense
 Topic modeling needed
 WordNet only considers nouns and verbs
 Works only adjectives and verbs
 Works only nouns, verbs, and adjectives.
 Wrongly disambiguated words can propagate to other words

A Tabela 4 lista os objetos de estudo (RQ4) denominados como others, que não aparecem no texto principal:

Tabela 4 - Objetos de estudos

Objetos de Estudos
Document representation
Named entity
Nouns
Phrase
Schema
Venue of Publication
Words

Na tabela 5 lista os idiomas utilizados nos datasets (RQ5) classificados como others:

Tabela 5 – Idioma dos datasets

Idioma dos Datasets
Assamese
Bangla
Cannada
France
Indonesian
Japanese
Marathi
Persian

Na tabela 9 lista os paises que publicaram os artigos (RQ9) classificados como others.

Tabela 9 – Paises que publicaram trabalhos

Países
Algeria
Bangladesh
Bulgaria
France
Germany
Indonesia
Israel
Japan
Jordan
Lebanon
Malaysia
Mexico
Morocco
Netherlands
Nigeria
Norway
Poland
Portugal
Singapore
Sweden
Switzerland
Taiwan
Tunisia
United Kingdom