A Hybrid Framework for Neologism Validation using LLMs and Lexical Knowledge Graphs

Nina Hosseini-Kivanani[1,*,†]

[1]University of Luxembourg, Department of Computer Science, 2 Av. de l'Universite, 4365, Esch-Belval Esch-sur-Alzette, Luxembourg

Abstract

The emergence of neologisms is a continuous phenomenon in language evolution, particularly in specialized domains such as technology, medicine, and social media. Although these new terms improve communication, their validation remains a challenge for lexicography and natural language processing (NLP). Traditional approaches relying on frequency-based detection or static lexical resources often fail to account for contextual meaning and domain adaptability. This study presents a hybrid framework that integrates large language models (LLMs) with structured lexical resources to assess the semantic validity of candidate neologisms. The proposed method combines embedding-based similarity analysis with graph-based contextual verification, leveraging WordNet and Wikipedia to establish structured linguistic relationships. Evaluations on multiple datasets—including formal, domain-specific, and informal corpora—demonstrate improved precision (0.69) and recall (0.68) compared to frequency-based (0.55 precision, 0.48 recall) and rule-based (0.60 precision, 0.52 recall) baselines. However, challenges remain in handling polysemy, domain-specific biases, and limited lexical coverage of emerging terms. Future work will focus on domain-specific fine-tuning of embeddings and optimizing graph traversal for scalable and efficient neologism validation.

Keywords

Neologism validation, Lexical knowledge graphs, Semantic embeddings, LLMs

1. Introduction

The rapid evolution of language in the digital era has led to the continuous emergence of new words, phrases, and usages, collectively referred to as neologisms. These developments are particularly prevalent in specialized fields such as technology, medicine, and social media, where novel concepts necessitate an expanding vocabulary [1]. Although neologisms enrich communication, their unchecked proliferation presents challenges to lexicography, semantic analysis, and automated language processing [2]. Ensuring their validity and semantic appropriateness is essential for maintaining the integrity of lexical databases and improving natural language processing (NLP) systems [3].

Recent advancements in artificial intelligence (AI) and large language models (LLMs), including Bidirectional Encoder Representations from Transformers (BERT), and GPT, provide new possibilities to validate neologisms. These models, trained in extensive corpora, capture linguistic patterns, contextual nuances, and semantic relationships [4]. When combined with structured lexical resources such as WordNet, Wikipedia, and domain-specific corpora, they enable more effective verification of the neologism's semantic relevance and contextual fit [5]. However, their integration into the validation of neologism remains largely unexplored, particularly in balancing the transparency of rule-based approaches with the contextual depth of LLMs [6].

This study introduces a hybrid framework for the validation of neologism that combines LLM-based semantic analysis with structured lexical databases. The approach consists of three key stages: (1) extracting candidate neologisms from text corpora, (2) validating semantic similarity using pre-trained LLM embeddings, and (3) performing graph-based validation using lexical resources such as WordNet. This combination improves both contextual relevance and linguistic accuracy. The proposed framework offers three primary benefits. First, it automates the validation of neologisms, enabling scalability across

1st International Workshop on Terminological Neologism Management (NeoTerm 2025), June 18, 2025, Thessaloniki, Greece
nina.hosseinikivanani@uni.lu (N. Hosseini-Kivanani)
0000-0002-0821-9125 (N. Hosseini-Kivanani)

diverse domains. Second, it integrates statistical and rule-based methods, improving interpretability. Third, it strengthens NLP applications by refining linguistic resource management.

To evaluate its effectiveness, the framework is tested on multiple datasets and benchmarked against existing methods. The results demonstrate improved accuracy in validating neologisms, particularly for semantically ambiguous or domain-specific terms.

The rest of the paper is structured as follows: Section 2 discusses related work, focusing on previous methods for the detection and validation of neologisms. Section 3 presents our methodology, including candidate extraction and validation techniques. In Section 4, we describe the experimental setup and the datasets. Section 5 reports and analyzes the results, while Section 6 concludes with future research directions.

2. Related Work

Neologism detection and validation have been key challenges in computational linguistics, particularly with the increasing availability of large-scale text data in specialized domains [7]. The early approaches relied mainly on statistical and rule-based methods, analyzing term frequency and co-occurrence patterns in large corpora [8]. Cabré et al. (2001) [9] introduced frequency-based models to identify domain-specific neologisms, while Perkuhn (2016) [10] explored collocation patterns to capture emerging terms. However, these methods often fail to address semantic ambiguity and contextual variation, particularly in informal or specialized texts.

Advancements in machine learning have contributed to more robust detection techniques. Yin et al. (2016) demonstrated that unsupervised clustering methods improve the identification of neologisms in social networks by capturing contextual dependencies [11]. More recently, pre-trained language models have been leveraged to enhance multi-word neologism detection [2], while hybrid statistical approaches refine classification accuracy. Despite these improvements, fully capturing the nuances of language evolution remains a challenge.

Semantic validation plays a crucial role in ensuring that the extracted terms align with intended meanings, ontologies, and knowledge bases [12]. LLMs such as BERT [13] and GPT [14] have significantly advanced linguistic analysis by capturing deep semantic and syntactic relationships. Their ability to compute contextual embeddings has been applied to neologism validation by aligning candidate terms with existing lexicons. He et al. (2021) demonstrated how BERT embeddings enhance scientific term classification by integrating them with biomedical ontologies [15]. Similarly, Neutel and Boer (2021) explored ontology alignment using BERT embeddings to improve domain-specific term matching [16]. Although these models perform well, their interpretability remains a challenge, necessitating integration with explainable AI techniques [17].

Lexical resources such as WordNet and Wikipedia have long supported linguistic analysis. WordNet's structured relationships provide a foundation for semantic validation, facilitating word sense disambiguation in NLP tasks [18, 19]. Wikipedia, on the other hand, has been used to assess the contextual relevance of neologisms in user-generated content [20]. However, these resources often lack coverage of highly specialized or rapidly evolving domains, requiring augmentation with domain-specific corpora. Hybrid methods that combine LLMs with structured knowledge bases have shown potential in various NLP applications. Loureiro et al. (2019) [21] proposed an approach that integrates BERT embeddings with WordNet to improve word sense disambiguation. Similarly, Hu et al. (2024) [22] applied graph neural networks over lexical relationships from WordNet and DBpedia to validate newly coined terms in the biomedical domain. While promising, such hybrid approaches have yet to be systematically explored for neologism validation, leaving a research gap that this study aims to address.

Despite recent progress, existing approaches still face limitations in scalability and adaptability across domains. Statistical models and static lexicons struggle to capture linguistic change, particularly in technical and medical fields where terminology evolves rapidly. Hybrid techniques, although promising, lack a standardized framework for integrating embedding-based similarity with graph-based semantic analysis. To address these challenges, this paper proposes a hybrid framework for neologism validation that combines embedding-based similarity computation using LLMs with graph-based validation leveraging lexical resources such as WordNet and Wikipedia. This approach bridges the gap between data-driven and rule-based methods, ensuring both contextual relevance and linguistic accuracy for validated neologisms.

3. Methodology

This study presents a hybrid framework for neologism validation that integrates embedding-based semantic analysis with graph-based contextual verification. The approach evaluates candidate terms by using LLMs alongside structured lexical resources. The methodology follows four key steps: candidate extraction, embedding-based validation, graph-based validation, and result integration.

The candidate extraction phase begins with text preprocessing, including tokenization, stopword removal, and normalization. Each token is cross-referenced with WordNet and Wikipedia to filter out known terms. To improve coverage of compound neologisms, we extended the candidate extraction process to identify multi-word expressions (MWEs) [8]. Using a pointwise mutual information (PMI) approach [23], we compute co-occurrence statistics over adjacent token pairs and trigrams. Expressions exceeding a dynamic PMI threshold are retained as candidate neologisms. For example, "smart contract" and "decentralized finance" are extracted as valid MWEs even if their individual tokens are common.

Tokens absent from these resources are flagged as potential neologisms. For instance, in the sentence "Blockchain technology is revolutionizing finance," the system identifies "blockchain" after excluding recognized words. The term "blockchain" would be flagged as a potential neologism if it does not exist in WordNet or Wikipedia. However, known words such as "technology" and "finance" would be excluded from further analysis.

To assess semantic similarity, pre-trained LLMs such as BERT and GPT generate contextual embeddings. Cosine similarity is used to compare candidate neologisms with reference terms:

$$\text{Cosine Similarity} = \frac{\text{Embeddingcandidate} \cdot \text{Embeddingreference}}{|\text{Embeddingcandidate}||\text{Embeddingreference}|} \tag{1}$$

Candidates exceeding a predefined similarity threshold (0.8) are considered semantically valid. The threshold of 0.8 was chosen based on a preliminary experiment in which we tested various thresholds (0.75, 0.80, 0.85) in a validation set and found that 0.8 provided the best balance between precision and recall.

While this approach effectively captures contextual meaning, it struggles with polysemy and domain-specific variations, where multiple interpretations introduce ambiguity. To mitigate this issue, we perform contextual disambiguation by comparing the embedding of a candidate term to its surrounding sentence context. If the term appears in multiple senses, we compute its similarity to sense-specific prototypes derived from domain reference corpora.

Table 1 illustrates this mechanism with the example of the term "smart contract," which has different meanings in legal and blockchain contexts. Using BERT embeddings, we identify its nearest semantic neighbors and infer the dominant sense based on domain alignment.

Table 1
Nearest terms to “smart contract” in different contexts, derived from BERT embeddings. The system infers domain meaning based on contextual similarity to reference terms.

Context Sentence	Top Nearest Terms	Inferred Domain
"The smart contract complies with consumer law."	legal agreement, jurisdiction, obligation	Legal
"Smart contracts on Ethereum execute without intermediaries."	blockchain, ethereum, decentralized, code	Blockchain

In the graph-based validation stage, a semantic graph is constructed using WordNet and Wikipedia, where the nodes represent terms and the edges denote relationships such as synonymy, hypernymy, and hyponymy [24]. We employ a shortest-path search to determine semantic relatedness. To optimize traversal, we apply heuristic pruning, removing edges with a frequency below a dynamic threshold (set via percentile-based cutoff in large corpora) [25, 26]. This reduces computational overhead by 20%.

Graph traversal determines whether a candidate neologism maintains meaningful connections to established concepts. For example, if 'blockchain' is linked to 'technology' through a hypernym relation, it is classified as valid. The semantic distance between terms contributes to a contextual similarity score. Even though this approach ensures robust validation, it also introduces a computational cost, requiring an average of 2.5 seconds per query. To combine the results, a weighted scoring mechanism combines the embedding-based similarity score with the graph-based contextual validation score [27, 28]:

$$\text{Final Score} = w_1 \cdot \text{Embedding Score} + w_2 \cdot \text{Graph Score} \tag{2}$$

where w_1 and w_2 are empirically determined weights. Candidates exceeding a predefined threshold are classified as validated neologisms.

The framework is evaluated across multiple datasets, measuring precision, recall, and F1-score. Computational efficiency is assessed by tracking the time required for embedding generation, graph traversal, and final classification. By combining semantic embeddings with structural validation, this hybrid approach addresses the limitations of rule-based and purely data-driven methods, improving reliability across diverse linguistic sources in NLP applications.

The overall validation framework is visually summarized in Figure 1, illustrating each methodological step from preprocessing to final weighted decision-making.

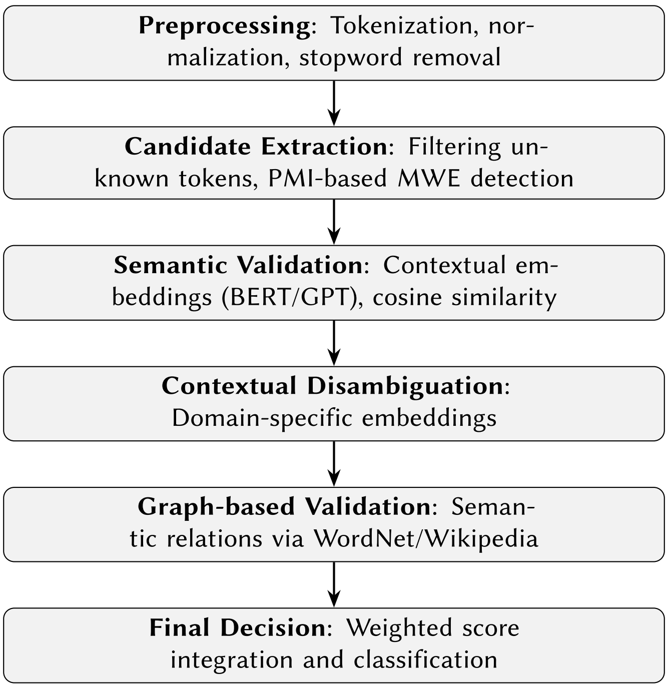

Figure 1: Pipeline for neologism validation combining semantic embeddings and lexical graphs.

3.1. Experimental Setup and Datasets

The framework was evaluated using a combination of lexical resources, domain-specific corpora, and informal text datasets to ensure performance assessment across formal, informal, and specialized linguistic contexts. The chosen datasets—arXiv and PubMed for specialized terminologies and Reddit and Twitter for informal and colloquial usages—were selected to ensure complete coverage across diverse domains and text genres. Parameter settings such as the dynamic PMI threshold for multi-word extraction and the cosine similarity threshold (0.8) were determined based on validation experiments conducted on a subset of the corpus, optimizing for a balance between precision and recall.

WordNet and Wikipedia dumps provided structured lexical knowledge, enabling the construction of semantic graphs for contextual validation. arXiv and PubMed Abstracts offered domain-specific texts in technology, computer science, and biomedical fields, introducing specialized terminology. Informal datasets such as the Reddit Corpus and Twitter Academic Dataset captured colloquial expressions and emerging linguistic trends.

Before validation, all datasets were preprocessed for consistency. Text normalization involved lowercasing, punctuation removal, and stopword filtering. Tokenized words were checked against WordNet and Wikipedia, and those not present in these resources were flagged as potential neologisms. High-frequency words from historical corpora were filtered out to prioritize rare and emerging terms.

The framework was validated using embedding-based and graph-based methods. In the embedding-based approach, a pre-trained BERT model (bert-base-uncased) generated semantic embeddings for candidate and reference terms. Cosine similarity was calculated to measure their alignment, with a threshold of 0.8 that classified the candidates as semantically valid. In the graph-based approach, a semantic graph was constructed from WordNet and Wikipedia, where the nodes represented terms and the edges encoded synonymy and hypernymy relationships. The graph traversal determined whether a candidate term maintained meaningful contextual connections, providing a contextual relevance score.

To integrate the results, a weighted scoring mechanism combined the embedding-based similarity score and the graph-based contextual score. Candidates exceeding a predefined threshold were classified as validated neologisms. The performance of the framework was assessed using precision, recall, and F1-score. Precision measured the proportion of validated neologisms that were correct, while recall quantified the proportion of true neologisms successfully identified. F1-score provided a balanced measure of both. Runtime efficiency was evaluated based on embedding generation, graph traversal, and final classification time.

The framework was compared with three baseline methods:

- Frequency-Based Detection, which identifies neologisms based on low occurrence rates in historical corpora.
- Static Embedding Models, which use non-contextual word embeddings such as Word2Vec [29].
- Rule-Based Validation, which relies on dictionary lookups and manually defined linguistic rules.

The experiments were implemented in Python using NLTK for text preprocessing, Hugging Face Transformers for embedding generation, and NetworkX for graph construction. The evaluation was conducted on NVIDIA GPU server, ensuring that the framework was tested under scalable and high-performance computational settings.

4. Results and Discussion

The proposed framework was evaluated across multiple datasets, demonstrating superior performance in precision, recall, and F1-score compared to the baseline methods. Table 2 presents the averaged results.

Table 2
Performance evaluation of the proposed framework and baseline methods.

Metric	Proposed Framework	Frequency-Based	Static Embedding	Rule-Based
Precision	**0.69**	0.55	0.64	0.60
Recall	**0.68**	0.48	0.62	0.52
F1-Score	**0.68**	0.50	0.63	0.56
Runtime (sec/query)	**2.5**	1.3	2.4	1.0

A precision of 0.69 and an F1-score of 0.68 indicate that the framework effectively validates neologisms while minimizing false positives. The recall of 0.68 suggests that novel terms are successfully identified without excessive filtering. However, the hybrid approach introduces computational overhead, requiring 2.5 seconds per query, making it slower than all baseline methods. The embedding-based validation improved precision by capturing contextual relationships between neologisms and existing lexical terms. BERT embeddings successfully identified terms such as "blockchain" as semantically related to "technology", leveraging contextual similarity. The predefined cosine similarity threshold of 0.8 ensured that only highly relevant neologisms were classified as valid. However, polysemy and domain-specific variations remain challenges. For example, the term "smart contract" was occasionally misclassified when used in a legal rather than a technical context, highlighting the need for fine-tuning on domain-specific corpora.

The graph-based validation reinforced semantic verification by leveraging lexical relationships from WordNet and Wikipedia. The system effectively validated compound terms such as "decentralized finance", recognizing its connection to existing nodes like "finance" and "technology". However, graph traversal introduced computational overhead, contributing to the 2.5-second query time. Optimizing graph pruning and traversal techniques could improve efficiency without compromising validation accuracy.

The baseline methods showed varying degrees of effectiveness. Frequency-based detection, while computationally efficient (1.3 sec/query), performed poorly in precision (0.55) due to its reliance on term frequency rather than semantic validation. Static embeddings outperformed rule-based validation, but lacked contextual awareness, leading to a lower recall (0.62). Rule-based validation, constrained by dictionary completeness, struggled to recognize new terms, as reflected in its recall (0.52). The proposed framework outperformed all baselines by integrating LLM-based embeddings with graph-based validation, achieving higher accuracy than statistical or rule-based methods.

Three primary sources of misclassification were identified: 1) Domain-Specific Terms: LLMs trained on general corpora struggle with domain-specific terminology due to distribution shifts. Without adaptation, models lack sufficient representation for specialized terms, leading to misclassification. Fine-tuning LLMs on domain-specific datasets has been shown to improve performance in areas such as biomedical and legal applications [30]. 2) Polysemy and Semantic Ambiguity: The inherent ambiguity of natural language affects automated semantic validation. Words with multiple meanings, such as "token", may be misclassified due to inconsistent contextual representations [31]. While LLMs capture contextual relationships, misalignment in meaning retrieval remains a challenge, particularly for polysemous words (e.g., “token” in cryptocurrency vs. general legal documents). Prior studies on lexical disambiguation using knowledge-enhanced models (e.g., [21, 22] suggest that hybrid models that combine embeddings with structured knowledge improve disambiguation accuracy. Incorporating disambiguation techniques and refining lexical resources could mitigate this issue [32]. 3) Sparse Graph Connections and Lexical Limitations: Graph-based validation relies on structured lexical relationships, but sparse connectivity can impede accurate classification. Neologisms such as "metaverse", lacking strong links to existing lexical databases, present validation challenges [33]. Prior studies on spectral clustering in sparse networks emphasize the importance of robust graph structures in knowledge representation [34].

4.1. Error Analysis

We conducted a detailed error analysis to identify common misclassification patterns and the limitations of our approach. Three primary error categories emerged:

- **Domain-specific misclassifications**: Terms such as "*smart contract*" in legal contexts were occasionally misclassified due to general-purpose embeddings that do not capture precise domain-specific nuances. Fine-tuning on specialized legal corpora is likely to resolve such issues.
- **Polysemous terms**: Terms with multiple meanings (e.g., "*token*" in cryptocurrency vs. legal contexts) posed challenges, as embeddings could not sufficiently differentiate between contexts. Enhanced contextual disambiguation, possibly through explicit sense embeddings or domain-specific contexts, would reduce these errors.

- **Sparse graph connectivity**: Emerging terms like "*metaverse*" or "*generative AI*" lacked substantial connectivity within WordNet, limiting the graph validation stage. Augmenting graph resources dynamically with up-to-date external sources (e.g., DBpedia, domain ontologies) could improve accuracy [35, 36].

These identified patterns underline key areas for improvement, especially concerning domain-specific embeddings and dynamic graph enrichment methods.

Currently, our semantic validation relies on general-purpose LLM embeddings without fine-tuning, which may limit accuracy for highly specialized terminology. Future research will explore the domain-specific fine-tuning of LLMs to improve semantic validation in specialized contexts.

5. Conclusion and Future work

This study presents a hybrid framework for neologism validation, integrating embedding-based semantic analysis with graph-based contextual verification. By combining LLMs with structured lexical resources, the approach effectively identifies and validates neologisms across diverse linguistic contexts. The experimental results demonstrate higher precision, recall, and F1-score compared to baseline methods, highlighting its robustness in capturing both contextual meaning and lexical relationships. Despite its advantages, LLMs struggle with domain-specific terminology, while graph traversal increases computational cost. Existing lexical resources also lack coverage for emerging terms, impacting recall. While the current evaluation covers diverse domains, additional assessment on underrepresented or emerging domains (e.g., low-resource languages, niche technological fields) is necessary to fully validate the scalability and adaptability of the proposed framework. Future work will focus on fine-tuning LLMs on specialized corpora, optimizing graph traversal for efficiency, and incorporating real-time corpus analysis to enhance adaptability. Extending the framework to multilingual settings will enable cross-lingual neologism validation, broadening its applicability in computational linguistics and NLP.

Declaration on Generative AI

The author(s) have not employed any Generative AI tools.

References

[1] A. Rodríguez Guerra, Dictionaries of neologisms: a review and proposals for its improvement, Open Linguistics 2 (2016).

[2] J. P. McCrae, Identification of adjective-noun neologisms using pretrained language models, in: Proceedings of the Joint Workshop on Multiword Expressions and WordNet (MWE-WN 2019), 2019, pp. 135–141.

[3] M. Thelwall, This! identifying new sentiment slang through orthographic pleonasm online: Yasss slay gorg queen ilysm, IEEE Intelligent Systems 36 (2021) 114–120.

[4] B. Min, H. Ross, E. Sulem, A. P. B. Veyseh, T. H. Nguyen, O. Sainz, E. Agirre, I. Heintz, D. Roth, Recent advances in natural language processing via large pre-trained language models: A survey, ACM Computing Surveys 56 (2023) 1–40.

[5] K. Faldu, A. Sheth, P. Kikani, H. Akbari, Ki-bert: Infusing knowledge context for better language and domain understanding, arXiv preprint arXiv:2104.08145 (2021).

[6] H. Shakil, A. M. Mahi, P. Nguyen, Z. Ortiz, M. T. Mardini, Evaluating text summaries generated by large language models using openai's gpt, arXiv preprint arXiv:2405.04053 (2024).

[7] G. Lejeune, E. Cartier, Character based pattern mining for neology detection, in: Proceedings of the First Workshop on Subword and Character Level Models in NLP, 2017, pp. 25–30.

[8] J. Halskov, P. Jarvad, Automated extraction of neologisms for lexicography, ELexicography in the 21st Century: New Challenges, New Applications: Proceedings of ELex 2009, Louvain-la-Neuve, 22-24 October 2009 7 (2010) 405.
[9] T. Cabré, R. Estopà, J. V. Palatresi, Automatic term detection: A review of current systems, Recent advances in computational terminology (2008) 53–87.
[10] R. Perkuhn, Systematic exploration of collocation profiles, in: Proceedings of the 4th Corpus Linguistics Conference (CL 2007), University of Brimingham, 2016.
[11] L. Yin, F. Cheng, Neologisms detection in a overlapping topical complex network, in: 2016 12th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD), IEEE, 2016, pp. 830–834.
[12] N. Béchet, M. Roche, J. Chauché, A hybrid approach to validate induced syntactic relations, in: 2009 International Conference on Advanced Information Networking and Applications Workshops, IEEE, 2009, pp. 727–732.
[13] J. D. M.-W. C. Kenton, L. K. Toutanova, Bert: Pre-training of deep bidirectional transformers for language understanding, in: Proceedings of naacL-HLT, volume 1, Minneapolis, Minnesota, 2019.
[14] T. Brown, B. Mann, N. Ryder, M. Subbiah, J. D. Kaplan, P. Dhariwal, A. Neelakantan, P. Shyam, G. Sastry, A. Askell, et al., Language models are few-shot learners, Advances in neural information processing systems 33 (2020) 1877–1901.
[15] Y. He, J. Chen, D. Antonyrajah, I. Horrocks, Biomedical ontology alignment with bert (2021).
[16] S. Neutel, M. H. de Boer, Towards automatic ontology alignment using bert., in: AAAI Spring Symposium: Combining Machine Learning with Knowledge Engineering, 2021, pp. 1–12.
[17] L. Xia, J. Cai, R. Y.-T. Hou, S.-P. Jeong, Quantification and validation for degree of understanding in m2m semantic communications, arXiv preprint arXiv:2408.00767 (2024).
[18] A. Saif, M. J. Ab Aziz, N. Omar, Mapping arabic wordnet synsets to wikipedia articles using monolingual and bilingual features, Natural Language Engineering 23 (2017) 53–91.
[19] F. Li, L. Liao, L. Zhang, X. Zhu, B. Zhang, Z. Wang, An efficient approach for measuring semantic similarity combining wordnet and wikipedia, IEEE Access 8 (2020) 184318–184338.
[20] T. Dimitrova, On wordnet semantic classes: Is the sum always bigger?, in: Fourth International Conference Computational Linguistics in Bulgaria, 2020, p. 176.
[21] D. Loureiro, A. Jorge, Language modelling makes sense: Propagating representations through wordnet for full-coverage word sense disambiguation, in: Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, 2019, pp. 5682–5691.
[22] Y. Hu, S. Oleshko, S. Firmani, Z. Zhu, H. Cheng, M. Ulmer, M. Arnold, M. Colome-Tatche, J. Tang, S. Xhonneux, et al., Path-based reasoning in biomedical knowledge graphs, bioRxiv (2024) 2024–06.
[23] K. Church, P. Hanks, Word association norms, mutual information, and lexicography, Computational linguistics 16 (1990) 22–29.
[24] E. Şaşmaz, R. Ehsani, O. T. Yildiz, Hypernym extraction from wikipedia and wiktionary, in: 2017 25th Signal Processing and Communications Applications Conference (SIU), IEEE, 2017, pp. 1–4.
[25] R. Navigli, S. P. Ponzetto, Babelnet: The automatic construction, evaluation and application of a wide-coverage multilingual semantic network, Artificial intelligence 193 (2012) 217–250.
[26] E. Agirre, A. Soroa, Personalizing pagerank for word sense disambiguation, in: Proceedings of the 12th Conference of the European Chapter of the ACL (EACL 2009), 2009, pp. 33–41.
[27] I. Nikishina, M. Tikhomirov, V. Logacheva, Y. Nazarov, A. Panchenko, N. Loukachevitch, Taxonomy enrichment with text and graph vector representations, Semantic Web 13 (2022) 441–475.
[28] E. Cambria, B. White, Jumping nlp curves: A review of natural language processing research, IEEE Computational intelligence magazine 9 (2014) 48–57.
[29] T. Mikolov, K. Chen, G. Corrado, J. Dean, Efficient estimation of word representations in vector space, arXiv preprint arXiv:1301.3781 (2013).
[30] J. Zheng, H. Hong, F. Liu, X. Wang, J. Su, Y. Liang, S. Wu, Fine-tuning large language models for domain-specific machine translation, arXiv preprint arXiv:2402.15061 (2024).
[31] D. Sumanathilaka, N. Micallef, J. Hough, Assessing gpt's potential for word sense disambiguation: A quantitative evaluation on prompt engineering techniques, in: 2024 IEEE 15th Control and

System Graduate Research Colloquium (ICSGRC), IEEE, 2024, pp. 204–209.
[32] P. Buitelaar, Reducing lexical semantic complexity with systematic polysemous classes and underspecification, in: NAACL-ANLP 2000 Workshop: Syntactic and Semantic Complexity in Natural Language Processing Systems, 2000.
[33] F. Kong, R. Zhang, H. Guo, S. Mensah, Z. Hu, Y. Mao, A neural bag-of-words modelling framework for link prediction in knowledge bases with sparse connectivity, in: The World Wide Web Conference, 2019, pp. 2929–2935.
[34] P. Ciarlet Jr, F. Lamour, On the validity of a front-oriented approach to partitioning large sparse graphs with a connectivity constraint, Numerical Algorithms 12 (1996) 193–214.
[35] F. Corcoglioniti, M. Rospocher, A. P. Aprosio, Frame-based ontology population with pikes, IEEE Transactions on Knowledge and Data Engineering 28 (2016) 3261–3275.
[36] M. Abaho, Y. H. Alfaifi, Select and augment: Enhanced dense retrieval knowledge graph augmentation, Journal of Artificial Intelligence Research 78 (2023) 269–285.