



## Ontology-Driven Text Classification and Data Mining: Beyond Keywords Toward Semantic Intelligence

Isaac Touza<sup>1,2\*</sup>, Gazissou Balama<sup>1,2</sup>, Warda Lazzar<sup>1,2</sup>, Kaladzavi Guidedi<sup>2,3</sup>, Kolyang<sup>2,4</sup>

<sup>1</sup> Department of Mathematics - Computer Science, Faculty of Sciences, University of Maroua, Maroua P.O. Box 814, Cameroon

<sup>2</sup> Laboratoire de Recherche en Informatique, University of Maroua, Maroua P.O. Box 46, Cameroon

<sup>3</sup> Department of Computer Science and Telecommunications, National Advanced School of Engineering, University of Maroua, Maroua P.O. Box 46, Cameroon

<sup>4</sup> Department of Computer Science, Higher Teacher's Training College, University of Maroua, Maroua P.O. Box 55, Cameroon

Corresponding Author Email: [isaac.touza@univ-maroua.cm](mailto:isaac.touza@univ-maroua.cm)

Copyright: ©2025 The authors. This article is published by IETA and is licensed under the CC BY 4.0 license (<http://creativecommons.org/licenses/by/4.0/>).

<https://doi.org/10.18280/ria.390301>

### ABSTRACT

**Received:** 7 May 2025

**Revised:** 29 May 2025

**Accepted:** 20 June 2025

**Available online:** 30 June 2025

#### Keywords:

*ontology, text classification, semantic enrichment, text mining, deep learning, knowledge representation, NLP, model*

The exponential increase of textual information on digital platforms exposes the shortcomings of conventional classification approaches, which often struggle to interpret meaning beyond surface-level keywords. This research explores the use of ontologies as an innovative approach to enhance semantic understanding in text classification. Ontologies serve as formal frameworks for representing domain knowledge, allowing systems to grasp complex conceptual relationships beyond simple statistical correlations. The paper provides a systematic review of ontology-based classification techniques, detailing their theoretical foundations, integration methods—from vector enrichment to deep learning architectures—and their effectiveness in fields like medicine and multilingual contexts. An empirical validation demonstrates that incorporating ontologies significantly improves classification performance, especially when combined with transformer-based models. Nonetheless, challenges such as scalability, multilingual support, and computational complexity remain. The study concludes with practical recommendations for implementation and suggests future research directions, including dynamic ontology learning, lightweight integration frameworks, and semantic alignment across languages. Ontology-driven classification emerges as a promising pathway toward more intelligent, interpretable, and domain-specific text analysis systems.

## 1. INTRODUCTION

With the exponential increase in daily textual data generated on the web, automating the classification of such information has become vital across various domains, including Natural Language Processing (NLP), content management, and information retrieval. Traditional methods, primarily based on statistical and machine learning techniques, often fall short in capturing the semantic richness inherent in text, restricting their accuracy and relevance [1-7]. To address this, integrating ontologies, a formal model of knowledge representation, offers new possibilities. Ontologies enable semantic structuring of information, allowing systems to better understand complex relationships between concepts, as demonstrated in many studies [8, 9]. This approach surpasses the limitations of classical methods by providing more sophisticated means to interpret and classify textual data [10]. Recent advances in data mining further enhance this integration, facilitating pattern extraction and the detection of implicit concept relationships [11, 12]. The goal of this paper is to review the current state of the art, highlighting how ontologies combined with data mining and machine learning improve text classification. The objective of this article is to

provide a comprehensive state of the art on the combined use of ontologies and data mining techniques for automatic text classification. We begin by presenting the theoretical foundations of ontologies and their role as semantic knowledge frameworks in NLP tasks. We then highlight the advantages of ontology-based approaches over traditional statistical or keyword-based methods. Next, we review the main strategies for integrating ontologies into text classification workflows, from vector-based representations to deep learning architectures such as transformers. This review is complemented by a critical analysis of existing research and an empirical validation using benchmark datasets. Finally, we discuss current limitations and outline future research directions, including dynamic ontology learning, cross-lingual alignment, and lightweight integration frameworks.

## 2. THEORETICAL FOUNDATIONS

The integration of ontologies with text classification requires a unified framework that connects semantic knowledge representation with statistical classification approaches. This section presents the foundational concepts of

both domains and establishes their interconnections, highlighting how ontological structures can enhance the feature space and semantic understanding in classification tasks.

## 2.1 Automatic text data classification

Let  $D = \{d_1, d_2, \dots, d_n\}$  represents the set of textual documents to be classified, where each document  $d_i$  is represented by a feature vector or variables  $D_i = \{x_1, x_2, \dots, x_m\}$ .

Each document is associated with a predefined label or category  $y$  selected from a set of predefined labels or categories  $Y = \{y_1, y_2, \dots, y_k\}$ .

Automatic classification aims to build a mathematical model or find a function  $f(d_i) : D \rightarrow Y$  that automatically assigns a label  $y_i$  to a given document  $d_i$  using machine learning methods. This function is learned from a training dataset  $(X_{train}, Y_{train})$ , where  $X_{train} = \{x_1, x_2, \dots, x_m\}$  represents the feature vectors of the training documents, and  $Y_{train} = \{y_1, y_2, \dots, y_m\}$  represents their corresponding labels [13].

In a literal sense, we define automatic text data classification as the process of automatically assigning a predefined label or category to an unstructured textual document based on features extracted from the text.

The diverse applications of automated text classification encompass many fields:

- In social media, it helps filter and categorize messages by topic or sentiment [6, 14, 15], aiding in the analysis of public opinions and emerging trends.
- In academic research [16, 17], it assists in organizing and analyzing articles, publications, and scholarly works, supporting the process of uncovering new insights.
- In corporate environments, it is used to classify emails, reports, and internal documents, promoting more efficient information management and better decision-making.

While automatic text classification provides advantages such as task automation, relevant information extraction, and improved productivity, it also presents challenges. Issues like language ambiguity, the necessity for high-quality labeled data, and algorithmic complexity can hinder performance. Proper annotation and effective document representation are essential for producing accurate and dependable results.

## 2.2 Ontologies

The term "ontology" originally comes from philosophy, where it pertains to the study or theory of existence, as noted by the Encyclopædia Universalis. In our context, however, we focus on computational ontologies.

Several scholars have offered definitions for this concept. The most well-known is by Gruber [8], who described an ontology as "an explicit specification of a conceptualization". Nonetheless, this definition was viewed as too broad by researchers like Smith and Welty [18]. Borst [19] expanded on this, defining an ontology as "a formal specification of a shared conceptualization", a characterization later refined by Studer et al. [20] as "a formal and explicit specification of a shared conceptualization". In the framework presented by Studer and colleagues, ontologies are depicted as a network of

interconnected concepts, linked through relationships and grounded in functions or axioms. They serve to model knowledge across various fields, often representing common understanding within specific domains, using a dedicated vocabulary for description.

Formally, an ontology ( $O$ ) has been represented in this study [19] as a tuple defined by Eq. (1):

$$O = (C, R, \tau) \quad (1)$$

where,

- $C$ : A set of concepts representing entities within the domain.
- $R$ : A set of relationships that define interactions or connections between concepts.
- $\tau$ : A function defined as Eq. (2):

$$\tau: C \rightarrow P(C \times R \times [0, 1]) \quad (2)$$

This function provides information about how conceptual terms are connected by returning finite sets of entries in  $C \times R \times [0, 1]$ . For each entry  $(c, r, w)$ ,  $c$  refers to a related concept term.  $r \in R$  indicates the type of relationship between the two conceptual terms (e.g., hypernymy, antonymy).  $w \in [0, 1]$  represents the weight of the relationship, specifying the degree of relatedness between the two conceptual terms, ranging from 0 (not related) to 1 (completely related).

In this study [21], the author enhanced this representation by defining ontology components as the tuple Eq. (3):

$$O = \langle C, H, R, A \rangle \quad (3)$$

where,

- $O$ : represents the ontology.
- $C$ : represents a set of classes (concepts).
- $H$ : represents a set of hierarchical links between concepts (taxonomic relationships).
- $R$ : represents a set of conceptual links (non-taxonomic relationships).
- $A$ : represents a set of rules and axioms.

In this study [22], the author represented ontologies using graph theory, modeling an ontology as a graph  $G = (V, E)$ , where,  $V$  is the set of nodes representing concepts.  $E$  is the set of edges representing relationships between the concepts.

From these definitions and formalizations, we conclude that an ontology is simply a collection of concepts, properties, and relationships that represent an explicit understanding of a particular domain. Its purpose is to formalize knowledge using precise terms and well-defined relationships, enabling a coherent and shared interpretation of information. This structured representation facilitates reasoning, information retrieval, and knowledge sharing across various applications.

## 2.3 Advantages of using ontologies in text classification

Ontologies have emerged as powerful tools for knowledge representation and management. In the field of text classification, their use offers several potential advantages that can enhance the performance of classification systems:

- **Structured Knowledge Representation:** Ontologies provide a structured framework for modeling domain-specific knowledge, offering rich semantic context for understanding textual data. This contextual information helps resolve

term ambiguities, capture nuanced relationships between concepts, reduce dimensionality [23], and improve the accuracy of text classification by correctly identifying the subject of a document [24].

- **Domain Expertise Integration:** Ontologies enable the integration of domain expertise into the classification process. Domain experts can contribute to the development and refinement of ontologies, ensuring that the classification model aligns with the terminology, relationships, and characteristics specific to the domain. This expertise can significantly enhance the performance of the classification system, especially in specialized fields where language use and semantics vary widely. For example, in the automatic classification of scientific articles in the medical field, ontologies can play a crucial role. Suppose we want to automatically classify articles based on their relevance to breast cancer research. A medical ontology specific to breast cancer can be developed in collaboration with medical experts. These experts can identify key concepts related to breast cancer, such as "tumor", "metastasis", "BRCA gene", and "treatment". They can also define relationships between these concepts, such as the "causes" relationship between the BRCA gene and breast cancer development. By using this ontology, textual documents like scientific articles can be enriched with semantic annotations. Relevant concepts identified in the ontology can be linked to specific terms found in the articles, creating a richer semantic representation of the documents.
- **Feature Space Enrichment:** Ontologies enable the enrichment of the feature space for text representation [25]. By leveraging ontological concepts and relationships, textual data can be represented in a more semantically meaningful way, capturing the inherent structure and semantics of the domain [26]. This enriched representation facilitates more efficient and accurate classification, particularly when dealing with complex or ambiguous text.

## 2.4 Data mining techniques for text classification

Data mining techniques for text classification involve methods and algorithms designed to extract relevant information from unstructured textual documents and categorize them into predefined categories. These techniques leverage textual features and patterns within the data to perform automatic classification. Below are some key techniques:

- **Bag-of-Words Techniques:** This approach represents textual documents as collections of words without considering their order or grammatical structure. Word occurrences or frequencies are used as features for classification [27]. Common algorithms include Naïve Bayes classifiers, decision trees, and Support Vector Machines (SVM).
- **Language Models:** Language models are statistical representations of word sequences in a text. They

estimate the probability of a word sequence or generate new text. Algorithms such as n-gram language models [28], Hidden Markov Models [29], and recurrent neural networks are commonly applied in text classification.

- **Word Embedding-Based Techniques:** Word embeddings are vector representations that capture the semantics of words. Models like Word2Vec, GloVe, and FastText are used to learn word embeddings from large text corpora. These embeddings are then employed as features for classification. Algorithms such as SVMs, neural networks, and random forests can be used alongside word embeddings [30-33].
- **Supervised Classification Methods:** Supervised methods such as SVMs, decision trees (e.g., C4.5), k-Nearest Neighbors (k-NN), neural networks, and random forests are widely used for text classification [34]. These methods train on labeled datasets to predict the categories of new documents.
- **Clustering Methods:** Clustering algorithms, such as hierarchical clustering, K-means, and DBSCAN, group similar documents based on their textual features. These methods identify clusters of documents that share similarities and help explore the underlying structure of data without requiring predefined category labels [2, 35, 36].
- **Transformer-Based Methods:** Transformer models, such as BERT, GPT-3, and T5, have revolutionized text classification by capturing bidirectional context and establishing complex semantic relationships between words. Unlike traditional methods, these models significantly enhance text understanding and enable more accurate classification, especially when fine-tuned on specific datasets [37-39].

These techniques and algorithms provide diverse approaches to text classification by leveraging textual features and patterns. The choice of technique and algorithm depends on the specific classification context and the characteristics of the textual data being processed.

## 2.5 Techniques for integrating ontologies in text classification

The integration of ontologies into text classification primarily involves five major approaches:

- **Ontology-Based Vector Representation:** This technique transforms documents and ontological concepts into vectors, enabling the use of vector similarity measures for classification [40-42]. Incorporating ontological knowledge significantly enhances accuracy by capturing term semantics.
- **Ontological Query Expansion:** This method enriches initial queries with related concepts derived from the ontology [43-45]. The expansion improves result relevance by establishing richer semantic links.
- **Contextual Knowledge Exploitation:** This approach integrates ontological concepts and relationships into document representation [46, 47]. Contextual enrichment strengthens the modeling of semantic interactions, thereby

enhancing classification precision.

- **Ontological Information Extraction:** This technique uses the ontology as a reference for concept-based annotation of documents [48, 49]. Ontological relationships and properties are utilized as additional features for classification.
- **Ontology Alignment:** This approach involves integrating multiple ontologies by identifying and leveraging conceptual correspondences [50]. Alignment facilitates interoperability and mutual enrichment of knowledge.

### 3. RELATED WORK

Numerous researchers have explored the integration of ontologies and data mining techniques to enhance automatic text classification. Below are notable contributions in this area:

Alipanah et al. [22] proposed a query expansion approach in distributed environments, leveraging ontology alignments to improve term relevance. However, the effectiveness of this method heavily depends on the accuracy of the ontology alignments. Similarly, an algorithm for domain-specific query expansion in sports was developed in this study [51], utilizing WordNet and domain ontologies to optimize document retrieval.

Camous et al. [52] used MeSH (Medical Subject Headings) to enhance the representation and sorting of MEDLINE documents. Hawalah [53] and Khabour et al. [15] introduced an improved architecture for classifying Arabic texts and a sentiment analysis approach utilizing ontologies and lexicons, respectively. These studies demonstrated significant improvements in classification due to the enrichment of semantic features.

Sivakami and Thangaraj [54] introduced COVOC, a solution that employs an ontology to extract relevant information related to the coronavirus. Wei et al. [55] proposed a model based on Resource Description Framework (RDF) for web document retrieval and classification. These works highlight the ability of ontologies to structure and enrich textual data.

Multilingual solutions have been explored in this study [56], combining automatic translators and ontologies for classification tasks. A hierarchical approach specific to the biomedical domain was presented in the study [57], integrating ontology alignments and cosine similarity scores to organize articles within hierarchies. Tao et al. [58] utilized a global ontology constructed from LCSH (Library of Congress Subject Headings) for topic generalization, while Bouchiha et al. [59] combined WordNet and SVMs to select and weight textual features. A four-step framework integrating deep learning and ontologies to enhance representation and classification of textual data was proposed in this study [60].

In this study [61], the authors presented a concept graph-based method enriched by ontologies for classifying medical documents. A supervised approach using pre-constructed ontologies to extract and enrich document features was discussed in this study [62].

A multi-class classification method exploiting deep learning and lexical ontologies was described in this study [63].

Li et al. [64] proposed a concept-based TextCNN model enriched with ontologies for predicting construction accidents, achieving high accuracy (88%) and AUC (0.92) but facing

domain-specific limitations.

Idrees et al. [65] introduced an enrichment multi-layer Arabic text classification model based on automated ontology learning, which achieved 97.92% accuracy in classification and 95% accuracy in ontology learning, though its multilingual applicability requires further exploration.

Giri and Deepak [66] developed a semantic ontology-infused deep learning model for disaster tweet classification, combining textual and image features to achieve an impressive F1 score of 98.416% on the Sri Lanka Floods dataset, despite challenges in computational complexity and multimodal data requirements.

Recent advancements in transformer-based ontology integration have significantly expanded the capabilities of text classification systems. These hybrid approaches enhance semantic understanding and context-aware prediction in a variety of domains, from healthcare and sentiment analysis to misinformation detection and low-resource languages. For instance, Hüsünbeyi and Scheffler [67] proposed an ontology-enhanced BERT model that significantly improved claim detection accuracy on the ClaimBuster and NewsClaims datasets by fusing OWL embeddings derived from ClaimsKG. In the context of under-resourced languages, Ali et al. [68] developed a sentiment analysis framework for Sindhi using DistilBERT augmented by a domain-specific ontology, achieving a notable increase in classification accuracy. Feng et al. [69] introduced OntologyRAG, which maps biomedical codes (e.g., SNOMED, ICD) via ontology-aware retrieval-augmented generation, showcasing the utility of ontology-guided LLMs for structured medical data interpretation.

In the medical recommendation domain, OntoMedRec integrates ontology embeddings within a transformer-based pipeline to recommend and classify treatments from clinical notes [70]. Meanwhile, this study [71] explored prompt-tuning enhanced by ontological cues to address few-shot classification challenges in low-resource settings. Similarly, Lee and Kim [72] proposed an ontology-based sentiment attribute classifier, demonstrating improved sentiment specificity in multilingual text processing. Cao et al. [73] applied ontology-enhanced LLMs to rare disease detection tasks, highlighting the benefits of structured semantic support in identifying niche biomedical entities.

Ouyang et al. [74] focused on fine-grained entity typing, enriched with ontological information, to improve type prediction accuracy in a zero-shot setting. Song et al. [75] introduced CoTel, a co-enhanced text labeling framework that combines rule-based ontology extraction and neural learning with ontology-enhanced loss prediction, significantly reducing labeling effort and time. Finally, Xiong et al. [76] developed a transformer-based approach that integrates ontology and entity type descriptions into the joint entity and relation extraction process, achieving improved performance on domain-specific datasets in the space science domain.

Additionally, Ngo et al. [77] demonstrated how integrating graph-based ontologies and transformers improves chemical disease relation classification in biomedical texts, offering superior performance to traditional deep learning models.

The contributions mentioned above demonstrate the potential of ontologies in enriching textual data, improving semantic understanding, and enhancing classification outcomes. Table 1 below provides a comparative analysis of these approaches, highlighting their application domains, employed techniques, achieved results, and noted limitation.

**Table 1.** Comparative analysis of ontology-based approaches for text classification

Ref.	Authors	Approach	Domain	Results	Limitations
[15]	Khabour et al.	Sentiment analysis with ontologies and lexicons	Sentiment analysis	Enrichment of semantic features	Complexity of lexical ontologies Dependence on the accuracy of ontology alignments
[22]	Alipanah et al.	Query expansion via ontology alignments	Distributed environments	Improved term relevance	Depends on quality and completeness of the domain ontology; fixed similarity threshold.
[51]	Chauhan et al.	Ontology-based semantic query expansion using concept similarity + synonym matching + threshold filtering	Document retrieval (sports)	Optimized document retrieval	Requires adaptation of MeSH for other domains
[52]	Camous et al.	Exploitation of MeSH for document sorting	Biomedical (MEDLINE)	Enhanced representation and sorting	Language-specific (Arabic)
[53]	Hawalrah	Architecture for Arabic text classification	Arabic texts	Significant improvement in classification	Limited to a specific domain (COVID-19)
[54]	Sivakami and Thangaraj	COVOC for extracting relevant information	Coronavirus	Effective information extraction	Complexity of RDF implementation
[55]	Wei et al.	RDF-based model for retrieval and classification	Web documents	Data structuring and enrichment	Dependence on the quality of automatic translations
[56]	Bentaallah	Multilingual solutions with translators and ontologies	Multilingual	Classification across multiple languages	Adaptability to other domains
[57]	Dollah and Aono	Hierarchical approach with ontology alignments	Biomedical	Organization of articles in hierarchies	Limited application to standard concepts
[58]	Tao et al.	Global ontology based on LCSH for generalization	Subject generalization	Improved subject structuring	Dependence on WordNet's coverage
[59]	Bouchiha et al.	WordNet and SVM for feature weighting	General text classification	Efficient selection of textual features	Complexity of multi-step framework
[60]	Nguyen et al.	Framework combining deep learning and ontologies	Data representation and classification	Improved classification	Complexity in graph construction
[61]	Shanavas et al.	Concept graphs enriched by ontologies	Medical documents	Effective document classification	Dependence on pre-constructed ontologies
[62]	Risch et al.	Supervised approach with feature extraction and enrichment	General	Precise feature extraction	Algorithmic complexity
[63]	Yelmen et al.	Multi-class classification with deep learning and lexical ontologies	General text classification	High-performance multi-class classification	Limited to construction-specific data and ontology scope
[64]	Li et al.	Ontology-based TextCNN for accident prediction	Construction industry	Achieved 88% accuracy and AUC of 0.92 for predicting construction accidents	Requires domain-specific adaptation; model performance may depend on dataset structure and richness
[65]	Idrees et al.	Enrichment multi-layer Arabic text classification model	Arabic text classification	95% in ontology learning	High computational complexity and dependence on multimodal data
[66]	Giri and Deepak	Semantic ontology-infused deep learning model for disaster tweet classification	Crisis response (tweets)	Achieved F1 score of 98.416% on the Sri Lanka Floods dataset	Dependency on ontology quality and coverage
[67]	Hüsünbeyi and Scheffler	BERT + OWL embeddings	Misinformation detection (Claim detection)	Improved accuracy and F1 on ClaimBuster/NewsClaims	Limited data, small sentiment ontology
[68]	Ali et al.	DistilBERT + custom ontology	Sentiment analysis in low-resource language (Sindhi)	Accuracy: 93% vs 82% baseline	

[69]	Feng et al.	OntologyRAG (LLM + retrieval + SNOMED/ICD ontology)	Biomedical code mapping	Improved mapping performance	Needs high compute for RAG inference
[70]	Tan et al.	OntoMedRec: logically-pretrained, model-agnostic ontology encoders	Medical recommendation system	Improved performance on full and few-shot EHR cases across multiple models	Depends on quality of medical ontologies; not end-to-end trainable alone
[71]	Ye et al.	Prompt tuning + ontological cues	Few-shot classification	Effective in low-resource scenarios	Prompt engineering complexity
[72]	Lee and Kim	Transformer + ontology-based attribute mapping	Sentiment attribute classification (multilingual)	Higher attribute-level accuracy	Difficulty scaling to new languages
[73]	Cao et al.	Ontology-guided LLM for entity/relation classification	Rare disease detection (biomedical)	F1 improvements in niche biomedical NER	Ontology incompleteness in rare disease domain
[74]	Ouyang et al.	Ontology enrichment (OnEFET)	Entity typing	Improved fine-grained type accuracy; outperforms zero-shot methods and rivals supervised ones	Requires curated and enriched ontologies for each domain
[75]	Song et al.	CoTel: hybrid approach with ontology + neural model	Semantic annotation (Text labeling)	Reduces time cost by 64.75% and labeling effort by 62.07%	Requires a high-quality ontology and well-tuned dual modules
[76]	Xiong et al.	Ontology + entity type descriptions integrated into PLM model	Joint entity and relation extraction (domain-specific)	+1.4% F1 score on relation extraction task (SSUIE-RE dataset, space science)	Performance validated on Chinese domain; generalization to other domains/languages not tested
[77]	Ngo et al.	Graph-enhanced Transformer	Biomedical – Chemical–Disease Relation Extraction	Outperformed standard DL models in relation classification	Requires detailed graph construction & domain-specific ontologies

#### 4. EMPIRICAL VALIDATION OF ONTOLOGY-BASED APPROACHES

To validate the theoretical advantages of ontology-based approaches, we conducted a comparative analysis using the OHSUMED corpus, a benchmark dataset for biomedical text classification. Three classification methods were evaluated: (1) a traditional bag-of-words model with SVM, (2) a deep learning approach using BERT, and (3) an ontology-enhanced BERT model integrating domain knowledge. The experiments employed 5-fold cross-validation, using 80% of the data for training and 20% for testing. Evaluation was based on precision, recall, and F1-score.

In the ontology-enhanced model, we incorporated MeSH (Medical Subject Headings) ontology to inject structured semantic information during both preprocessing and embedding stages. This biomedical ontology provided concept-level disambiguation and hierarchical semantic context that are not captured in purely lexical or contextual models. Table 2 below presents the comparative performance metrics of the three approaches:

**Table 2.** Comparative performance of classification approaches

Method	Precision	Recall	F1
<b>BOW + SVM</b>	71.4%	68.9%	70.1%
<b>BERT</b>	76.8%	75.3%	76.0%
<b>Ontology + BERT</b>	<b>83.5%</b>	<b>82.1%</b>	<b>83.2%</b>

The results clearly demonstrate that ontology-enhanced methods consistently outperform traditional and neural baselines across evaluation metrics. The improvement is particularly pronounced in the biomedical domain, where domain-specific ontology (MeSH) contributes structured knowledge that complements deep contextual embeddings. The 7.2% increase in F1-score over the vanilla BERT model confirms the added value of semantic enrichment, particularly in contexts where terminology is highly specialized and hierarchical.

These findings empirically validate the theoretical insights discussed in Section 2.3, emphasizing how ontology-based integration facilitates semantic disambiguation, improves generalization in domain-specific settings, and enhances the interpretability and robustness of classification outcomes.

## 5. RESULTS AND DISCUSSION

### 5.1 Comparative analysis of ontology-based approaches

Our comprehensive analysis of ontology-enhanced text classification methods reveals that ontologies are leveraged in distinct ways depending on the chosen approach, each influencing the architecture, learning capacity, and interpretability of the models. To clarify these differences, Table 3 summarizes how ontologies are utilized within five major classification strategies, and specifies the corresponding level of technical integration from shallow preprocessing to deep model fusion.

To better contextualize the performance improvements of

ontology-based classification approaches, we present in Table 4 the absolute F1-scores for each ontology integration approach compared to a standard BERT model. These results are derived from our own experimental findings.

**Table 3.** Overview of ontology usage and integration levels across classification approaches

Approach	Key Use of Ontologies	Integration Level
Ontology-Based Vector Representation	Direct semantic representation via mapped concepts	Input-level vector
Ontological Query Expansion	Lexical-semantic enrichment using ontology terms	Preprocessing-level
Contextual Knowledge Exploitation	Contextual enhancement using semantic links (e.g., is-a, part-of)	Graph/contextual fusion
Ontological Information Extraction	Annotation and labeling guided by ontology structure	Supervised labeling
Transformer-Based Ontology Integration	Direct fusion into transformer model	Model-level fusion

**Table 4.** Absolute F1-scores for ontology-based classification approaches compared to BERT baseline

Approach	Baseline BERT F1-score	Approach F1-score	Absolute Improvement
Ontology-Based Vector Representation	76.0%	80.2%	+4.2%
Ontological Query Expansion	76.0%	79.8%	+3.8%
Contextual Knowledge Exploitation	76.0%	81.7%	+5.7%
Ontological Information Extraction	76.0%	82.3%	+6.3%
Transformer-Based Ontology Integration	76.0%	83.2%	+7.2%

**Table 5.** Comparative analysis of ontology-based classification approaches

Approach	Computational Complexity	Domain Adaptability
Ontology-Based Vector Representation	Medium	High
Ontological Query Expansion	Low	Medium
Contextual Knowledge Exploitation	High	Medium
Ontological Information Extraction	High	Low
Transformer-Based Ontology Integration	Very High	Medium

The F1-score for transformer-based ontology integration (83.2%) closely aligns with our empirical validation results presented in Table 2, confirming the consistency between our

detailed experimental analysis and the broader comparative assessment presented here.

Building on this quantitative foundation, Table 5 presents a more comprehensive comparison of these five approaches across two key dimensions: (1) computational complexity, and (2) adaptability across domains.

Several key findings emerge from the comparative analysis of Table 4 and Table 5:

- **Performance-Complexity Tradeoff:** The data reveals a clear correlation between performance gains and computational demands, with transformer-based integration offering the highest F1-score improvement (+7.2%) but at very high computational cost.
- **Domain-Specific Suitability:** While ontology-based vector representation shows strong domain adaptability, ontological information extraction demonstrates more limited flexibility despite its superior accuracy gains.
- **Practical Implementation Considerations:** Query expansion emerges as the most computationally efficient approach, making it particularly suitable for resource-constrained environments where moderate performance gains (+3.8%) are acceptable.

These empirical results suggest that the optimal approach selection depends heavily on specific application requirements, particularly the available computational resources and the need for cross-domain generalization.

## 5.2 Identified gaps and challenges

The integration of ontologies into automatic text classification has marked a turning point in NLP. By enriching textual data with structured semantic knowledge, ontology-based approaches enable models to go beyond surface-level pattern recognition, offering deeper contextual understanding and improved interpretability. These methods have demonstrated notable improvements in classification accuracy, particularly in specialized domains such as biomedicine, where ontology like MeSH provide well-structured conceptual hierarchies. When combined with deep learning architectures such as BERT or transformers, ontologies enhance the ability to extract relevant features and disambiguate meanings in complex linguistic contexts.

However, despite these advances, the current landscape of ontology-enhanced classification presents several critical limitations that hinder its widespread adoption and scalability. One of the major challenges lies in scalability. Most existing systems struggle to scale effectively when confronted with large datasets or domains featuring rich and deep ontological structures. The computational overhead associated with integrating and reasoning over thousands of semantic concepts can become prohibitive, especially in real-time or resource-constrained environments.

Another pressing issue is domain dependence. Many ontology-based systems are built around domain-specific knowledge bases, which restrict their applicability to new or heterogeneous domains. Generalizing these approaches across different thematic areas requires either adaptable ontologies or cross-domain alignment strategies, which remain underdeveloped. This issue is further compounded by the lack of multilingual support in many ontological resources. Most are designed in a single language (often English), limiting their

utility in multilingual or cross-cultural applications where semantic equivalence is not always straightforward.

In addition, the field suffers from evaluation inconsistencies. There is no universally accepted framework for assessing ontology-based classifiers, making it difficult to compare different methods or replicate results. The absence of benchmark datasets and standardized metrics leads to fragmented evaluation practices, which impedes scientific progress. Furthermore, ontology construction itself remains a bottleneck. Building and maintaining high-quality ontologies is a labor-intensive process that demands expert knowledge, significant time investment, and specialized tools. This manual dependency makes it difficult to keep ontologies updated with emerging terminology and evolving knowledge domains.

Finally, integrating ontologies into complex machine learning pipelines introduces architectural and computational complexity. Injecting structured semantic knowledge into deep models whether via embeddings, attention mechanisms, or hybrid architectures can drastically increase system complexity, reduce efficiency, and complicate model interpretability and maintenance.

To overcome these limitations, several promising research directions emerge.

- First, automating ontology construction and enrichment using machine learning, text mining, or large language models could reduce dependence on manual curation and accelerate scalability.
- Second, incorporating multilingual and collaborative resources such as Wikidata, BabelNet, or ConceptNet may help improve semantic coverage and adaptability across languages and domains.
- Third, the development of hybrid approaches that fuse ontological reasoning with neural networks while minimizing complexity through dimensionality reduction or sparse attention could balance performance with efficiency.
- Lastly, the creation of open standards and collaborative platforms for ontology sharing, alignment, and evaluation would facilitate reproducibility, interoperability, and community-driven innovation.

## 6. CONCLUSION AND PERSPECTIVES

The use of ontologies in automatic text classification represents a meaningful step forward in NLP. By incorporating structured semantic knowledge into machine learning workflows, these methods go beyond the limits of traditional keyword-based or purely statistical approaches. They bring clearer interpretability, greater classification precision, and better performance in complex domains like healthcare, legal analysis, and multilingual content.

Our review, supported by experimental results, confirms that models enhanced with ontologies consistently outperform conventional techniques especially when domain-specific knowledge is well-organized and clearly defined. Ontologies such as MeSH, when integrated into deep learning models like BERT or other transformer-based systems, contribute not only to higher F1-scores but also to better contextual understanding and generalization.

However, these advantages come with real challenges.

Ontologies differ widely in quality, detail, and coverage, and creating or maintaining them often demands extensive manual effort. Integrating them into advanced machine learning pipelines can also increase the technical complexity and computational load. Still, the convergence of ontological knowledge with deep learning and automation paves the way for innovative, intelligent systems that are both scalable and adaptable.

Based on our findings, several practical guidelines can help researchers and developers. When choosing ontologies, it's critical to select ones that offer broad domain coverage, are actively maintained, and have the right level of granularity for the classification task. In environments with limited computational resources, ontology-based vectorization methods offer a simple and interpretable solution. For domain-specific use cases with rich ontological resources, leveraging contextual knowledge is key. And for high-performance applications in data-rich settings, combining ontologies with transformer architectures offers the most promising results.

Looking ahead, key areas of research need attention to make these systems more usable and scalable. Cross-lingual ontology alignment is crucial to support multilingual applications, as are zero-shot learning methods and the creation of multilingual benchmark datasets. There's also a need for modular, lightweight frameworks that allow for easier deployment in real-time or embedded environments. These should support concept selection, standardized APIs, and dynamic ontology management. Automating ontology construction, enrichment, and ongoing updates from real-world data streams will reduce human workload and improve relevance over time. Collaborative, open platforms can drive this evolution by enabling shared development and reuse of ontological resources.

## REFERENCES

- [1] Sebastiani, F. (2002). Machine learning in automated text categorization. *ACM Computing Surveys (CSUR)*, 34(1): 1-47. <https://doi.org/10.1145/505282.505283>
- [2] Manning, C.D., Raghavan, P., Schütze, H. (2008). *Introduction to Information Retrieval*. Cambridge University Press.
- [3] Yeh, A.S., Hirschman, L., Morgan, A.A. (2003). Evaluation of text data mining for database curation: Lessons learned from the KDD Challenge Cup. *Bioinformatics*, 19(Suppl. 1): i331-i339. <https://doi.org/10.48550/arXiv.cs/0308032>
- [4] Maron, M.E. (1961). Automatic indexing: An experimental inquiry. *Journal of the ACM (JACM)*, 8(3): 404-417. <https://doi.org/10.1145/321075.321084>
- [5] Cover, T., Hart, P. (1967). Nearest neighbor pattern classification. *IEEE Transactions on Information Theory*, 13(1): 21-27. <https://doi.org/10.1109/TIT.1967.1053964>
- [6] Hota, S., Pathak, S. (2018). KNN classifier based approach for multi-class sentiment analysis of Twitter data. *International Journal of Engineering and Technology*, 7(3): 1372-1375. <https://doi.org/10.14419/ijet.v7i3.12656>
- [7] Landauer, T.K., Foltz, P.W., Laham, D. (1998). An introduction to latent semantic analysis. *Discourse Processes*, 25(2-3): 259-284. <https://doi.org/10.1080/01638539809545028>



- [8] Gruber, T.R. (1993). A translation approach to portable ontology specifications. *Knowledge Acquisition*, 5(2): 199-220. <https://doi.org/10.1006/knac.1993.1008>
- [9] Guarino, N. (1998). *Formal Ontology in Information Systems: Proceedings of the First International Conference (FOIS'98)*, June 6-8, Trento, Italy (Vol. 46). IOS press.
- [10] Maedche, A., Staab, S. (2005). Ontology learning for the semantic web. *IEEE Intelligent Systems*, 16(2): 72-79. <https://doi.org/10.1109/5254.920602>
- [11] Fayyad, U., Piatetsky-Shapiro, G., Smyth, P. (1996). From data mining to knowledge discovery in databases. *AI Magazine*, 17(3): 37-37. <https://doi.org/10.1609/aimag.v17i3.1230>
- [12] Hotho, A., Staab, S., Stumme, G. (2003). Ontologies improve text document clustering. In *Third IEEE International Conference on Data Mining*, Melbourne, FL, USA, pp. 541-544. <https://doi.org/10.1109/ICDM.2003.1250972>
- [13] Aggarwal, C.C., Zhai, C. (2012). A survey of text classification algorithms. In *Mining Text Data*, pp. 163-222. [https://doi.org/10.1007/978-1-4614-3223-4\\_6](https://doi.org/10.1007/978-1-4614-3223-4_6)
- [14] Lanquillon, C. (2001). *Enhancing text classification to improve information filtering*. Doctoral dissertation. Otto-von-Guericke-Universität Magdeburg, Universitätsbibliothek.
- [15] Khabour, S.M., Al-Radaideh, Q.A., Mustafa, D. (2022). A new ontology-based method for Arabic sentiment analysis. *Big Data and Cognitive Computing*, 6(2): 48. <https://doi.org/10.3390/bdcc6020048>
- [16] Nédellec, C., Bossy, R., Chaix, E., Deléger, L. (2018). Text-mining and ontologies: New approaches to knowledge discovery of microbial diversity. In *Proceedings of the 4th International Microbial Diversity Conference*, Bari, Italy, pp. 221-227. <https://doi.org/10.48550/arXiv.1805.04107>
- [17] Langer, S., Neuhaus, F., Nürnberger, A. (2024). CEAR: Automatic construction of a knowledge graph of chemical entities and roles from scientific literature. <https://doi.org/10.48550/arXiv.2407.21708>
- [18] Smith, B., Welty, C. (2001). Ontology: Towards a new synthesis. *Formal Ontology in Information Systems*, 10(3): 3-9. <https://doi.org/10.1145/505168.505201>
- [19] Borst, W.N. (1997). *Construction of engineering ontologies for knowledge sharing and reuse*. Ph.D. thesis, University of Twente. Centre for Telematics and Information Technology (CTIT). <https://research.utwente.nl/en/publications/construction-of-engineering-ontologies-for-knowledge-sharing-and->
- [20] Studer, R., Benjamins, V.R., Fensel, D. (1998). *Knowledge engineering: Principles and methods*. *Data & Knowledge Engineering*, 25(1-2): 161-197. [https://doi.org/10.1016/S0169-023X\(97\)00056-6](https://doi.org/10.1016/S0169-023X(97)00056-6)
- [21] Zouaq, A. (2011). An overview of shallow and deep natural language processing for ontology learning. In *Ontology Learning and Knowledge Discovery Using the Web: Challenges and Recent Advances*. Hershey, PA: IGI Global Scientific Publishing, pp. 16-37. <https://doi.org/10.4018/978-1-60960-625-1.ch002>
- [22] Alipanah, N., Parveen, P., Menezes, S., Khan, L., Seida, S.B., Thuraisingham, B. (2010). Ontology-driven query expansion methods to facilitate federated queries. In *2010 IEEE International Conference on Service-Oriented Computing and Applications (SOCA)*, Perth, WA, Australia, pp. 1-8. <https://doi.org/10.1109/SOCA.2010.5707141>
- [23] Oleiwi, S.S. (2015). Enhanced ontology-based text classification algorithm for structurally organized documents.
- [24] Wijewickrema, C.M. (2015). Impact of an ontology for automatic text classification. *Annals of Library and Information Studies (ALIS)*, 61(4): 263-272. <http://op.niscpr.res.in/index.php/ALIS/article/viewFile/4163/191>
- [25] Wu, S.H., Tsai, R.T.H., Hsu, W.L. (2003). Text categorization using automatically acquired domain ontology. In *Proceedings of the Sixth International Workshop on Information Retrieval with Asian Languages*, pp. 138-145. <https://aclanthology.org/W03-1118.pdf>
- [26] Bloehdorn, S., Hotho, A. (2004). Text classification by boosting weak learners based on terms and concepts. In *Fourth IEEE International Conference on Data Mining (ICDM'04)*, Brighton, UK, pp. 331-334. <https://doi.org/10.1109/ICDM.2004.10077>
- [27] Zhang, Y., Jin, R., Zhou, Z.H. (2010). Understanding bag-of-words model: A statistical framework. *International Journal of Machine Learning and Cybernetics*, 1: 43-52. <https://doi.org/10.1007/s13042-010-0001-0>
- [28] Cavnar, W.B., Trenkle, J.M. (1994). N-gram-based text categorization. In *Proceedings of SDAIR-94, 3rd Annual Symposium on Document Analysis and Information Retrieval*, 161175: 14. <https://dsac13-2019.github.io/materials/CavnarTrenkle.pdf>
- [29] Rabiner, L.R. (2002). A tutorial on hidden Markov models and selected applications in speech recognition. *Proceedings of the IEEE*, 77(2): 257-286. <https://doi.org/10.1109/5.18626>
- [30] Mikolov, T., Chen, K., Corrado, G., Dean, J. (2013). Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*. <https://doi.org/10.48550/arXiv.1301.3781>
- [31] Pennington, J., Socher, R., Manning, C.D. (2014). Glove: Global vectors for word representation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, Doha, Qatar, pp. 1532-1543. <https://aclanthology.org/D14-1162.pdf>
- [32] Joulin, A., Grave, E., Bojanowski, P., Mikolov, T. (2017). Bag of tricks for efficient text classification. *arXiv preprint arXiv:1607.01759*. <https://doi.org/10.48550/arXiv.1607.01759>
- [33] Zhang, Y., Wallace, B. (2015). A sensitivity analysis of (and Practitioners' Guide to) convolutional neural networks for sentence classification. *arXiv preprint arXiv:1510.03820*. <https://doi.org/10.48550/arXiv.1510.03820>
- [34] McCallum, A., Nigam, K. (1998). A comparison of event models for naive bayes text classification. In *Proceedings of the Workshop on Learning for Text Categorization*, pp. 41-48. <http://yangli-feasibility.com/home/classes/lfd2022fall/media/aaaiws98.pdf>
- [35] Jain, A.K. (2010). Data clustering: 50 years beyond K-means. *Pattern Recognition Letters*, 31(8): 651-666. <https://doi.org/10.1016/j.patrec.2009.09.011>
- [36] Ester, M., Kriegel, H.P., Sander, J., Xu, X.W. (1996). A density-based algorithm for discovering clusters in large

- spatial databases with noise. In Proceedings of the 2nd International Conference on Knowledge Discovery and Data Mining (KDD), pp. 226-231.
- [37] Devlin, J., Chang, M.W., Lee, K., Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. In Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers), pp. 4171-4186. <https://doi.org/10.18653/v1/N19-1423>
- [38] Brown, T., Mann, B., Ryder, N., Subbiah, M., et al. (2020). Language models are few-shot learners. Advances in Neural Information Processing Systems 33 (NeurIPS 2020), pp. 1877-1901. <https://proceedings.neurips.cc/paper/2020/hash/1457c0d6bfc4967418bfb8ac142f64a-Abstract.html>.
- [39] Raffel, C., Shazeer, N., Roberts, A., Lee, K., et al. (2020). Exploring the limits of transfer learning with a unified text-to-text transformer. Journal of Machine Learning Research, 21(140): 1-67. <https://www.jmlr.org/papers/v21/20-074.html>.
- [40] Wróbel, K., Wielgosz, M., Smywiński-Pohl, A., Pietron, M. (2016). Comparison of SVM and ontology-based text classification methods. In 15th International Conference on Artificial Intelligence and Soft Computing (ICAISC 2016), Zakopane, Poland, pp. 667-680. [https://doi.org/10.1007/978-3-319-39378-0\\_57](https://doi.org/10.1007/978-3-319-39378-0_57)
- [41] Kastrati, Z., Imran, A.S., Yayilgan, S.Y. (2015). An improved concept vector space model for ontology based classification. In 2015 11th International Conference on Signal-Image Technology & Internet-Based Systems (SITIS), Bangkok, Thailand, pp. 240-245. <https://doi.org/10.1109/SITIS.2015.102>
- [42] Ngo, V.M., Cao, T.H. (2010). Ontology-based query expansion with latently related named entities for semantic text search. In Advances in Intelligent Information and Database Systems, pp. 41-52. [https://doi.org/10.1007/978-3-642-12090-9\\_4](https://doi.org/10.1007/978-3-642-12090-9_4)
- [43] Raza, M.A., Ali, M., Pasha, M., Ali, M. (2022). An improved semantic query expansion approach using incremental user tag profile for efficient information retrieval. VFAST Transactions on Software Engineering, 10(3): 1-9. <https://doi.org/10.21015/vtse.v10i3.1136>
- [44] Kumar, R., Sharma, S.C. (2023). Hybrid optimization and ontology-based semantic model for efficient text-based information retrieval. The Journal of Supercomputing, 79(2): 2251-2280. <https://doi.org/10.1007/s11227-022-04708-9>
- [45] Wan, J., Wang, W.C., Yi, J.K., Chu, C., Song, K. (2012). Query expansion approach based on ontology and local context analysis. Research Journal of Applied Sciences, Engineering and Technology, 4(16): 2839-2843.
- [46] Alagha, I. (2022). Leveraging knowledge-based features with multilevel attention mechanisms for short arabic text classification. IEEE Access, 10: 51908-51921. <https://doi.org/10.1109/ACCESS.2022.3175306>
- [47] Lee, Y.H., Hu, P.J.H., Tsao, W.J., Li, L. (2021). Use of a domain-specific ontology to support automated document categorization at the concept level: Method development and evaluation. Expert Systems with Applications, 174: 114681. <https://doi.org/10.1016/j.eswa.2021.114681>
- [48] El Khettari, O., Nishida, N., Liu, S.S., Munne, R.F., et al. (2024). Mention-Agnostic information extraction for ontological annotation of biomedical articles. In Proceedings of the 23rd Workshop on Biomedical Natural Language Processing, pp. 457-473. <https://doi.org/10.18653/v1/2024.bionlp-1.37>
- [49] Cutting-Decelle, A.F., Digeon, A., Young, R. I., Barraud, J.L., Lamboley, P. (2018). Extraction of technical information from normative documents using automated methods based on ontologies: Application to the ISO 15531 MANDATE standard - Methodology and first results. arXiv preprint arXiv:1806.02242. <https://doi.org/10.48550/arXiv.1806.02242>
- [50] Jurisch, M., Igler, B. (2018). RDF2Vec-based classification of ontology alignment changes. arXiv preprint arXiv:1805.09145. <https://doi.org/10.48550/arXiv.1805.09145>
- [51] Chauhan, R., Goudar, R., Rathore, R., Singh, P., Rao, S. (2012). Ontology based automatic query expansion for semantic information retrieval in sports domain. In International Conference on Eco-friendly Computing and Communication Systems, Kochi, India, pp. 422-433. [https://doi.org/10.1007/978-3-642-32112-2\\_49](https://doi.org/10.1007/978-3-642-32112-2_49)
- [52] Camous, F., Blott, S., Smeaton, A.F. (2007). Ontology-based MEDLINE document classification. In International Conference on Bioinformatics Research and Development, Berlin, Germany, pp 439-452. [https://doi.org/10.1007/978-3-540-71233-6\\_34](https://doi.org/10.1007/978-3-540-71233-6_34)
- [53] Hawalah, A. (2019). Semantic ontology-based approach to enhance arabic text classification. Big Data and Cognitive Computing, 3(4): 53. <https://doi.org/10.3390/bdcc3040053>
- [54] Sivakami, M., Thangaraj, M. (2021). Ontology based text classifier for information extraction from coronavirus literature. Trends in Sciences, 18(24): 47. <https://doi.org/10.48048/tis.2021.47>
- [55] Wei, G.Y., Wu, G.X., Gu, Y.Y., Ling, Y. (2008). An ontology based approach for Chinese web texts classification. Information Technology Journal, 7: 796-801.
- [56] Bentaallah, M.A. (2015). Utilisation des ontologies dans la catégorisation de textes multilingues. Doctoral dissertation. Université de Sidi Bel Abbès-Djillali Liabes.
- [57] Dollah, R.B., Aono, M. (2011). Ontology-based approach for classifying biomedical text abstracts. International Journal of Data Engineering, 2(1): 1-15.
- [58] Tao, X.H., Delaney, P., Li, Y.F. (2021). Text categorisation on semantic analysis for document categorisation using a world knowledge ontology. IEEE Intelligent Informatics Bulletin, 21(1): 13-24. <http://comp.hkbu.edu.hk/~cib/2021/Article1.pdf>.
- [59] Bouchiha, D., Bouziane, A., Doumi, N. (2023). Ontology based feature selection and weighting for text classification using machine learning. Journal of Information Technology and Computing, 4(1): 1-14. <https://doi.org/10.48185/jitc.v4i1.612>
- [60] Nguyen, T.T.S., Do, P.M.T., Nguyen, T.T., Quan, T.T. (2023). Transforming data with ontology and word embedding for an efficient classification framework. EAI Endorsed Transactions on Industrial Networks and Intelligent Systems, 10(2): 2.
- [61] Shanavas, N., Wang, H., Lin, Z., Hawe, G. (2020). Ontology-based enriched concept graphs for medical document classification. Information Sciences, 525: 172-

181. <https://doi.org/10.1016/j.ins.2020.03.006>
- [62] Risch, J.C., Petit, J., Rousseaux, F. (2016). Ontology-based supervised text classification in a big data and real time environment.
- [63] Yelmen, I., Gunes, A., Zontul, M. (2023). Multi-class document classification using lexical ontology-based deep learning. *Applied Sciences*, 13(10): 6139. <https://doi.org/10.3390/app13106139>
- [64] Li, X., Shu, Q., Kong, C., Wang, J., et al. (2025). An intelligent system for classifying patient complaints using machine learning and natural language processing: Development and validation study. *Journal of Medical Internet Research*, 27: e55721. <https://preprints.jmir.org/preprint/55721>.
- [65] Idrees, A.M., Al-Solami, A.L.M. (2024). An enrichment multi-layer Arabic text classification model based on siblings patterns extraction. *Neural Computing and Applications*, 36(14): 8221-8234. <https://doi.org/10.1007/s00521-023-09405-z>
- [66] Giri, K.S.V., Deepak, G. (2024). A semantic ontology infused deep learning model for disaster tweet classification. *Multimedia Tools and Applications*, 83(22): 62257-62285. <https://doi.org/10.1007/s11042-023-16840-6>
- [67] Hüsünbeyi, Z.M., Scheffler, T. (2024). Ontology enhanced claim detection. *arXiv preprint arXiv:2402.12282*. <https://doi.org/10.48550/arXiv.2402.12282>
- [68] Ali, A., Ghaffar, M., Somroo, S.S., Sanjrani, A.A., Ali, T., Jalbani, T. (2025). Ontology based Semantic Analysis framework in Sindhi Language. *VFAST Transactions on Software Engineering*, 13(1): 193-206. <https://doi.org/10.21015/vtse.v13i1.2080>
- [69] Feng, H., Yin, Y., Reynares, E., Nanavati, J. (2025). OntologyRAG: Better and faster biomedical code mapping with retrieval-augmented generation (RAG) leveraging ontology knowledge graphs and large language models. *arXiv preprint arXiv:2502.18992*. <https://doi.org/10.48550/arXiv.2502.18992>
- [70] Tan, W., Wang, W., Zhou, X., Buntine, W., Bingham, G., Yin, H. (2024). OntoMedRec: Logically-pretrained model-agnostic ontology encoders for medication recommendation. *World Wide Web*, 27(3): 28. <https://doi.org/10.1007/s11280-024-01268-1>
- [71] Ye, H., Zhang, N., Deng, S., Chen, X., et al. (2022). Ontology-enhanced prompt-tuning for few-shot learning. In *WWW '22: Proceedings of the ACM Web Conference 2022, Virtual Event, Lyon, France*, pp. 778-787. <https://doi.org/10.1145/3485447.3511921>
- [72] Lee, S.J., Kim, H.K. (2025). Ontology-Based Sentiment Attribute Classification and Sentiment Analysis. *Journal of KIISE: Software and Applications*, KoreaScience, pp. 23-32. <https://www.earticle.net/Article/A464412>
- [73] Cao, L., Sun, J., Cross, A. (2024). AutoRD: A framework for rare disease entity and relation extraction using ontology-guided LLMs. *arXiv preprint arXiv:2403.00953*. <https://doi.org/10.48550/arXiv.2403.00953>
- [74] Ouyang, S., Huang, J., Pillai, P., Zhang, Y., Zhang, Y., Han, J. (2024). Ontology enrichment for effective fine-grained entity typing. In *KDD '24: Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, New York, NY, United States, pp. 2318-2327. <https://doi.org/10.1145/3637528.3671857>
- [75] Song, M.H., Zhang, L., Yuan, M., Li, Z., Song, Q., Liu, Y., Zheng, G. (2023). Cotel: Ontology-neural co-enhanced text labeling. In *WWW '23: Proceedings of the ACM Web Conference 2023, Austin, TX, USA*, pp. 1897-1906. <https://doi.org/10.1145/3543507.3583533>
- [76] Xiong, X., Wang, C., Liu, Y., Li, S. (2023). Enhancing ontology knowledge for domain-specific joint entity and relation extraction. In *22nd China National Conference on Chinese Computational Linguistics*, Harbin, China, pp. 237-252. [https://doi.org/10.1007/978-981-99-6207-5\\_15](https://doi.org/10.1007/978-981-99-6207-5_15)
- [77] Ngo, N.H., Nguyen, A.D., Thi, Q.T.P., Dang, T.H. (2024). Integrating Graph and transformer-based models for enhanced chemical-disease relation extraction in document-level contexts. In *13th International Symposium on Information and Communication Technology*, Danang, Vietnam, pp. 174-187. [https://doi.org/10.1007/978-981-96-4285-4\\_15](https://doi.org/10.1007/978-981-96-4285-4_15)