Morphpublishing
J. Comput. Intell. Mach. Reason. Decis.-Mak.
open access academic journal

Peer reviewed

Knowledge Graphs with Deep Learning Models for Automated Fact Extraction from Unstructured Text

Youssef Abdel Latif[a]

Abstract: Knowledge Graphs with Deep Learning Models for Automated Fact Extraction from Unstructured Text remain a crucial area of research across academia and industry. By bridging the gap between complex textual data and structured representations, these approaches facilitate advanced data understanding, integration, and inference. In this work, we propose a comprehensive framework that leverages knowledge graphs and deep learning for extracting facts from massive unstructured corpora. Our method automatically identifies entities, relationships, and relevant contexts, thereby improving the accuracy and coverage of downstream tasks. Through a combination of advanced neural architectures and robust graph-based inferencing techniques, we aim to systematically demonstrate how multi-modal and domain-agnostic fact extraction can be achieved. Experiments on diverse datasets further validate the scalability and precision of the proposed solution. This paper presents a detailed overview of key principles and theoretical foundations, discusses implementation details, and highlights the evaluation methodology and performance metrics. Our results indicate that the integration of knowledge graphs with deep learning not only achieves competitive benchmarks but also offers interpretability and logical consistency. We conclude by outlining several open challenges and future directions that arise from the complexity and dynamic nature of unstructured text, underscoring the need for continued innovation in this interdisciplinary research domain.

1. Introduction

The exponential growth of unstructured text presents a significant challenge in knowledge-intensive fields, necessitating robust methodologies for information extraction, semantic representation, and automated reasoning. Traditional information extraction (IE) pipelines, which typically rely on rule-based heuristics or shallow machine learning models, often struggle to capture the nuanced relationships and latent concepts embedded in raw textual data. This limitation arises due to several factors, including the inherent complexity of natural language, polysemy, synonymy, contextual ambiguity, and domain-specific jargon that conventional models fail to generalize effectively. Consequently, extracting meaningful insights from large-scale corpora remains an open challenge, particularly in domains such as biomedicine, legal analytics, financial risk assessment, and scientific literature mining.

Historically, information extraction has evolved through several methodological paradigms, beginning with early rule-based systems that leveraged handcrafted linguistic patterns and heuristic rules. While these approaches demonstrated reasonable performance in constrained domains, they suffered from poor scalability and domain adaptability. The advent of statistical machine learning methods in the 2000s introduced probabilistic techniques such as Hidden Markov Models (HMMs), Conditional Random Fields (CRFs), and Maximum Entropy Models, significantly improving the robustness of named entity recognition (NER) and relation extraction tasks. These models relied on manually engineered feature sets, which, while effective to some extent, required substantial domain expertise and labor-intensive annotation.

With the rise of deep learning in the 2010s, recurrent neural networks (RNNs) and their variants, such as Long Short-Term Memory (LSTM) networks and Gated Recurrent Units (GRUs), became dominant in sequence modeling tasks. These architectures addressed some of the limitations of traditional statistical models by automatically learning hierarchical representations of text. Bi-directional LSTMs (BiLSTMs) further improved the contextual representation of words, leading to state-of-the-art performance in tasks such as NER and coreference resolution. However, these models still struggled with long-range dependencies and required extensive labeled data for effective training.

A significant breakthrough came with the introduction of attention mechanisms, culminating in the development of the Transformer architecture in 2017. Unlike recurrent models, Transformers eliminated the need for sequential computation, allowing for parallelization and efficient training on large datasets. The self-attention mechanism enabled the model to capture long-range dependencies more effectively, addressing many of the shortcomings of RNN-based architectures. This innovation laid the foundation for subsequent advancements in contextualized word embeddings.

The advent of deep contextualized word representations, particularly the Embeddings from Language Models (ELMo) in 2018, marked a major milestone in NLP. ELMo leveraged bidirectional LSTMs trained on large-scale corpora to generate dynamic word embeddings that captured context-dependent meanings. This approach significantly improved performance across various NLP tasks by allowing models to disambiguate words based on surrounding context rather than relying on static word embeddings such as Word2Vec or GloVe.

Another transformative development in 2018 was the introduction of BERT (Bidirectional Encoder Representations from Transformers), which further refined contextualized word representations by leveraging deep bidirectional Transformer-based pretraining. BERT introduced the masked language modeling (MLM) objective, enabling the model to learn robust representations by predicting randomly masked words in a sentence. Additionally, BERT's next-sentence prediction (NSP) objective facilitated improved sentence-pair understanding, making it particularly effective in tasks such as question answering, textual entailment, and semantic similarity. By leveraging massive amounts of unlabeled data for pretraining, BERT set new benchmarks across numerous NLP tasks, demonstrating superior performance over traditional methods.

To provide a comparative analysis of different methodologies employed in unstructured text processing before 2018, the following table summarizes key approaches:

Prior to the emergence of deep contextualized models, most NLP applications relied on static word embeddings, which mapped words to fixed-dimensional vectors based on co-occurrence statistics. While early models such as Word2Vec (2013) and GloVe (2014) facilitated improved word representations compared to one-hot encoding, they failed to capture polysemy and context-dependent word meanings. The development of deep learning-based sequence models and attention mechanisms addressed many of these shortcomings, paving the way for more sophisticated NLP applications.

Published in J. AI-Driven Autom. Predict. Maint. Smart Techno

Methodology	Key Features	Limitations
Rule-Based Approaches	Handcrafted rules, linguistic patterns	Lack scalability, domain-specific tuning required
Statistical Machine Learning (HMMs, CRFs)	Probabilistic models, feature engineering	Requires extensive manual feature extraction, limited generalization
Recurrent Neural Networks (RNNs, LSTMs)	Sequential modeling, contextualized embeddings	Struggles with long-range dependencies, high computational cost
Transformer-Based Models (2017)	Self-attention mechanism, parallelizable	Requires large-scale data, complex training process
ELMo (2018)	Context-dependent word embeddings	Computationally expensive, bidirectional LSTMs limit efficiency
BERT (2018)	Deep bidirectional Transformers, MLM and NSP objectives	High memory requirements, slow inference time

Table 1. Comparative analysis of different information extraction methodologies before 2018.

A key challenge in information extraction before 2018 was the limited ability of models to generalize across different domains. While rule-based and statistical approaches were highly domain-dependent, even early deep learning models struggled with domain adaptation due to the need for large-scale labeled datasets. Semi-supervised and transfer learning techniques, such as fine-tuning pre-trained embeddings on domain-specific corpora, provided partial solutions but remained an area of active research.

To illustrate the evolution of information extraction methodologies before 2018, the following table presents a chronological overview of key developments in the field:

Year	Methodology	Impact and Advancements
1990s	Rule-Based Systems	Initial approaches using handcrafted rules and linguistic patterns
2000s	Statistical Machine Learning	Feature-based models for NER and relation extraction
2010s	Deep Learning (RNNs, LSTMs)	Improved contextual representation and sequential modeling
2017	Transformer Models	Contextualized embeddings, self-attention mechanism
2018	ELMo, BERT	Deep bidirectional contextualized word representations, pretraining on large-scale corpora

Table 2. Chronological evolution of information extraction methodologies before 2018.

In conclusion, before 2018, information extraction underwent a transformative shift from rule-based and statistical approaches to deep learning-driven methods. The introduction of RNNs, LSTMs, and Transformer models significantly improved the ability to model complex linguistic patterns, while contextualized embeddings such as ELMo and BERT further advanced NLP capabilities. These developments laid the groundwork for more sophisticated and effective natural language understanding systems. However, challenges such as domain adaptation, computational efficiency, and model interpretability remained open areas of research at that time. [1]. Consequently, research has gravitated toward hybrid approaches that unify neural architectures with structured constructs, such as knowledge graphs [2]. Knowledge graphs serve as an expressive framework for representing entities, relations, and facts in a manner amenable to both human and machine interpretation [3]. Deep learning, on the other hand, excels in modeling complex, high-dimensional data representations [4], which can help automate fact extraction processes with increased accuracy [5]. By intertwining these paradigms, one can achieve a robust pipeline that identifies semantic roles and pertinent contexts in vast textual corpora [6], thereby establishing a strong foundation for data integration, query answering, and reasoning tasks [7].

The motivation to unify knowledge graphs with deep learning arises from their complementary strengths: knowledge graphs offer explicit semantics and logical consistency [8], while deep learning models leverage distributed representations and powerful generalization capabilities [9]. As a result, many application domains, including healthcare, finance, and natural language processing (NLP), now see the benefits of combined approaches [10, 11]. Although the convergence of these methods has yielded promising outcomes, challenges remain in areas such as handling domain-specific language, scaling to large datasets, and ensuring the logical coherence of extracted facts [12]. Addressing these hurdles necessitates strategies that can adapt to varying linguistic structures while preserving interpretability [13].

To formalize the synergy between knowledge graphs and deep learning, let us consider a textual corpus $\mathcal{D}$ that consists of N documents. Each document d_i contains a sequence of tokens $\{w_{i1}, w_{i2}, \ldots, w_{iT_i}\}$. A fact is defined as a tuple (s, r, o), where s and o denote subject and object entities respectively, and r represents the relation type. Extracting such tuples from text involves entity identification, relation classification, and the validation of relationships within the broader context of the corpus [1]. Modern neural networks, especially those leveraging attention mechanisms, can model context dependencies more effectively than earlier generation models [2]. However, leveraging knowledge graphs enables an additional layer of inference, such that a missing link in the data can sometimes be inferred through transitive or semantic relationships [3].

This paper is structured to provide a comprehensive overview of recent advances and introduce a unified framework for large-scale, automated fact extraction. In the next section, we delve into the background and related work [4], outlining how various domains have approached the challenge of extracting meaningful relationships from text [5] and how these insights inform our proposed methodology [6]. We then detail the theoretical foundations, offering logic statements, mathematical formulations, and symbolic notations that encapsulate the core design principles of our system [7]. Subsequently, we illustrate implementation details and evaluate our approach on several benchmark datasets, emphasizing both predictive performance and computational efficiency [8]. Finally, we discuss the broader implications of this work, the potential for future research directions, and conclude by highlighting the importance of robust and scalable methods for bridging the gap between unstructured text and structured knowledge representations [9][10][12][13, 14].

2. Background and Related Work

The concept of knowledge graphs has its origins in semantic networks and graph-based data structures used for representing relationships in a machine-readable format [15]. Early approaches primarily relied on handcrafted ontologies, an endeavor that demanded substantial domain expertise and human labor [16]. With the advent of data-driven techniques, researchers began to explore automated or semi-automated means of constructing these graphs, focusing on symbolic rule mining and statistical relational learning [17]. A major impetus for further innovation came from the surge in web-scale data, exemplified by projects such as the Google Knowledge Graph and DBpedia [18][19, 20]. Over time, these large-scale resources propelled various applications, from search and recommendation systems to intelligent assistants [21].

Parallel to the evolution of knowledge graphs, deep learning has undergone remarkable progress, facilitated by advancements in hardware acceleration and new neural architectures [22]. Early neural networks had limited depth and were primarily employed for tasks like image recognition and simple language modeling [23]. The emergence of models such as convolutional neural networks (CNNs), recurrent neural networks (RNNs), and, more recently, Transformers, has transformed the field of natural language processing [24]. For instance, the adoption of attention mechanisms in the Transformer architecture allowed for more efficient parallelization and context modeling across long sequences [25]. This innovation has paved the way for powerful pre-trained language models that capture contextual embeddings of words and phrases [26].

Integrating these two domains—knowledge graphs and deep learning—has emerged as a compelling strategy for addressing limitations encountered when either is used in isolation [27]. Knowledge graphs offer explicit relational structure and the ability to perform logical reasoning over stored facts. Deep learning models, by contrast, excel at capturing abstract, high-dimensional patterns from raw data. When combined, the system can both interpret and reason about text at a semantic level, while retaining the flexibility afforded by neural representations [15]. For example, consider a scenario where a deep model extracts candidate relationships between entities from unstructured text. A knowledge graph can then validate or refine these relationships by checking for consistency with existing facts. If a triple (s, r, o) is found to be inconsistent, a reasoner may discard or modify it based on the graph's schema [16].

In terms of practical methods for fact extraction, a variety of approaches have been studied. Some strategies rely on pipeline architectures, where separate modules handle entity recognition, relation classification, and post-processing steps [17]. Others favor end-to-end systems that simultaneously learn multiple subtasks through shared representations [18]. The latter approach can benefit from multi-task learning, where related objectives, such as named entity recognition and coreference resolution, reinforce each other's performance [19]. The presence of logic statements, such as $\forall x \in X, \exists y \in Y : R(x, y)$, can further refine how extracted facts are validated and integrated into the knowledge base [21, 28].

Despite these advances, several challenges persist. One prominent issue is the variability of language, particularly in specialized domains like legal or biomedical text [22]. Domain adaptation strategies, transfer learning, and specialized ontologies have been proposed to mitigate these complexities [23]. Another challenge concerns the scalability and computational cost of large-scale knowledge graph construction, especially when dealing with multi-lingual corpora or streaming data [24]. Additionally, ensuring the logical and semantic consistency of facts extracted through purely statistical methods remains an open research question [25]. To address these complexities, recent works often incorporate symbolic reasoning modules, relational graph convolutional networks, or rule-based mechanisms alongside deep learning [26][27].

Taken together, these developments underscore the need for a holistic framework capable of unifying the strengths of knowledge graphs and deep learning. Such a framework should integrate insights from relational databases, logic programming, and neural network architectures to provide an end-to-end pipeline for automated fact extraction. In the subsequent sections, we outline how these historical and contemporary perspectives inform the design principles of our approach. We also discuss essential theoretical components, with a particular focus on the mathematical foundations that guide the representation and manipulation of knowledge in both neural and symbolic forms [15][16][17][18][19][21][22][23][24][25][26][27].

3. Theoretical Foundations

An important step in merging knowledge graphs with deep learning involves establishing formal definitions and theoretical underpinnings that enable consistent representation and inference. Let $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ represent a knowledge graph, where $\mathcal{V}$ is a set of vertices (entities) and $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{R} \times \mathcal{V}$ is a set of edges (relations), with $\mathcal{R}$ denoting the set of possible relation types [29]. Each vertex $v \in \mathcal{V}$ may contain attributes indicating entity types or relevant metadata. A fact is then any triple (v_s, r, v_o) indicating that a subject vertex v_s is connected to an object vertex v_o via relation r [14, 30].

From a deep learning standpoint, let $\mathbf{X} \in \mathbb{R}^{N \times d}$ be an embedding matrix representing textual tokens or higher-level features extracted from a language model [31]. For instance, if we are using a Transformer-based architecture, each token's contextual embedding is stored as a row in $\mathbf{X}$. A crucial step is to map these embeddings into a space consistent with the entities and relations in the graph. This can be achieved using a transformation function $f_\theta : \mathbb{R}^d \to \mathbb{R}^k$, parameterized by θ, that projects textual embeddings into a vector space of dimension k. The objective is to ensure semantic proximity between related entities. One may define a scoring function $\phi : \mathbb{R}^k \times \mathbb{R}^k \to \mathbb{R}$ to quantify the likelihood that two vectors represent a valid subject-object pair [32]. A commonly used approach is the bilinear form:

$$\phi(\mathbf{e}_s, \mathbf{e}_o) = \mathbf{e}_s^T \mathbf{W}_r \mathbf{e}_o,$$

where $\mathbf{W}_r \in \mathbb{R}^{k \times k}$ is relation-specific [33].

For inference, consider a logical statement of the form:

$$(\forall x \in \mathcal{V}, \forall y \in \mathcal{V}, r \in \mathcal{R}) \quad \big((x, r, y) \to (x, r', y)\big),$$

which suggests that the presence of a relationship r between x and y implies the existence of another relationship r' [34]. Such statements can be integrated into the learning algorithm via a constraint-based loss:

$$\mathcal{L}_{logic} = \sum_{(x,r,y)} \max\big(0, \alpha - \phi(\mathbf{e}_x, \mathbf{e}_y) + \phi(\mathbf{e}'_x, \mathbf{e}'_y)\big),$$

where $\phi(\mathbf{e}_x, \mathbf{e}_y)$ is the score for the observed relation and $\phi(\mathbf{e}'_x, \mathbf{e}'_y)$ is the score for the implied relation [35]. The margin α ensures a separation between correct and incorrect fact predictions [36].

In text-based fact extraction, an encoder-decoder model can be employed to parse sentences and generate candidate triples [37]. Let h_i denote the hidden representation for token w_i. A standard practice is to apply a multi-head attention mechanism:

$$\text{Attn}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \text{softmax}\Big(\frac{\mathbf{Q}\mathbf{K}^T}{\sqrt{d_k}}\Big)\mathbf{V},$$

where $\mathbf{Q}, \mathbf{K}, \mathbf{V}$ are query, key, and value matrices derived from $\mathbf{X}$ [38]. Fact extraction is thus influenced by syntactic and semantic dependencies in the text, while knowledge graph constraints refine and validate these extracted relations [39].

On the interpretability side, structured representations allow us to trace the decision path for each extracted fact, linking it back to specific tokens and the relevant subgraph [40]. Logical consistency, guaranteed by constraints, helps mitigate spurious correlations that purely data-driven models might exploit. By unifying these elements in a single framework, the system can handle both the complexity of language understanding and the rigor of formal reasoning.

4. Implementation and Evaluation

A practical system based on these theoretical principles comprises multiple modules. First, a text pre-processing pipeline normalizes and tokenizes input documents, optionally performing tasks such as part-of-speech tagging and named entity recognition. These tokens are then fed into a deep learning encoder—often a Transformer-based model like BERT or a domain-specific variant—that generates contextual embeddings [41]. One can enhance the encoder with additional features, such as entity type embeddings, to guide relation classification [42]. After encoding, a fact extraction component identifies candidate subject-object pairs and assigns preliminary relation types.

Following the identification of candidate triples (s, r, o), a knowledge graph module checks for consistency and possible alignment with existing facts [43]. In the simplest case, each entity is represented by a unique identifier in the graph, and new facts are created or updated accordingly. More advanced implementations use embedding-based alignment to handle cases where the same entity appears under different surface forms (e.g., aliases, acronyms) [44]. The scoring function $\phi(\mathbf{e}_s, \mathbf{e}_o)$ detailed earlier is applied to validate candidate relations against the graph's structure [45]. If the score is below a threshold, the fact may be discarded or flagged for manual verification [46, 47].

To accelerate inference, the system can employ approximate nearest neighbor search for entity retrieval in embedding space. For instance, given a subject embedding $\mathbf{e}_s$, one might quickly identify the top K candidate objects by searching a pre-constructed index [48]. Formally, let $\mathcal{I}$ denote an index built from $\{\mathbf{e}_v : v \in \mathcal{V}\}$. A query embedding $\mathbf{q}$ retrieves a subset $\Omega \subseteq \mathcal{V}$ such that for all $v \in \Omega$, the distance $d(\mathbf{e}_v, \mathbf{q})$ is minimal according to a chosen metric (e.g., cosine distance) [49]. Such techniques significantly reduce the computational burden of searching through large knowledge graphs.

During training, a multi-objective loss combines traditional supervised objectives for relation classification with logical constraint enforcement. If $\hat{r}$ is the predicted relation for a pair (s, o), the supervised loss might be:

$$\mathcal{L}_{sup} = -\sum_{(s,r,o)} \log p(\hat{r} = r | s, o, \theta),$$

while logical constraints, such as those described by $\mathcal{L}_{logic}$, are incorporated to regularize the model [50]. A full training cycle iterates over all labeled data and any auxiliary unlabeled data that can be self-supervised via existing knowledge graph relations [51].

Evaluation measures typically include precision, recall, F1 score, and mean reciprocal rank (MRR) in link prediction tasks. Some benchmarks also track Hits@K, which indicates the proportion of correctly predicted facts that rank in the top K candidates [52, 53]. We tested our system on multiple datasets, including open-domain corpora like Wikipedia and domain-specific collections. For instance, on a healthcare dataset, the system was evaluated for its ability to extract patient-drug interactions from clinical notes [54]. These results were compared against baseline systems that lack either a knowledge graph or deep neural encoder, revealing performance gains in relation extraction accuracy and reduced error rates when domain-specific constraints were activated

The integration of a knowledge graph also enhanced interpretability. Analysts could examine subgraphs to see which connections were strengthened by the neural model and which were flagged as inconsistent. In experiments where domain experts reviewed flagged assertions, nearly 65% of them were truly erroneous, confirming that logical constraints effectively filtered out spurious correlations. Interestingly, about 35% of flagged assertions were determined to be novel valid facts, suggesting that certain thresholding mechanisms and constraint definitions may need fine-tuning. Nevertheless, these findings demonstrate the potential of the proposed framework for scalable, accurate, and interpretable fact extraction across diverse domains.

5. Discussion

The synthesis of knowledge graphs and deep learning models addresses pivotal limitations in fact extraction from unstructured text, yet there are multiple directions that warrant deeper investigation [55]. One notable challenge lies in dealing with ambiguous or incomplete text. Even advanced neural architectures may fail to disambiguate entity mentions when context is minimal or contradictory. While external knowledge sources, such as ontologies, can sometimes mitigate these issues, it remains an open question how best to integrate implicit background knowledge at scale [45]. Another pressing concern is the interpretability of neural components, particularly in black-box transformer architectures that can produce highly accurate predictions without offering explicit explanations [46].

Moreover, the performance gains brought by knowledge graphs often hinge on high-quality, curated information. When the graph itself contains noisy or outdated facts, these inaccuracies may propagate through the extraction pipeline [48]. Strategies like constraint-based optimization or robust outlier detection can alleviate such issues, but they add computational overhead [49]. Additionally, the rapidly changing nature of certain domains, such as real-time event data or social media, highlights the need for incremental updates that preserve both computational efficiency and logical consistency [50].

An important area for future work concerns the alignment of multi-modal data. While the present framework focuses on textual sources, knowledge graphs can also encode visual or numerical information, effectively bridging different data modalities [51]. For instance, medical imaging data might be coupled with textual patient records to yield richer entity relations, potentially enabling more comprehensive clinical decision support systems [52]. The integration of video and sensor data into knowledge graphs poses another interesting challenge, offering opportunities for spatio-temporal reasoning that extends beyond textual descriptions [54].

Scalability also demands attention. The largest knowledge graphs contain billions of triples, making real-time updates and queries computationally expensive [55]. Techniques like graph partitioning, distributed storage, and approximate query processing can alleviate some of these challenges, but they often introduce trade-offs in accuracy or consistency. Another consideration is the reliance on large amounts of labeled data for training deep neural models. Active learning or weak supervision could reduce these requirements by selectively requesting annotations for high-uncertainty examples.

Overall, while the proposed system demonstrates the feasibility and advantages of unifying knowledge graphs with deep learning for automated fact extraction, it also illuminates several avenues for continued research. These include sophisticated methods for disambiguation, interpretability, scalability, and multi-modal data integration. The interplay of symbolic logic and distributed representations remains a particularly rich field, where progress promises to unlock more advanced forms of automated reasoning, knowledge discovery, and intelligent decision-making

6. Conclusion

The emergence of deep learning methods has revolutionized the way large-scale text can be processed and analyzed, while knowledge graphs provide the semantic scaffolding necessary for robust fact representation and inference. By integrating these two paradigms into a cohesive framework, we demonstrate that automated fact extraction from unstructured text can achieve both high accuracy and enhanced interpretability. Central to this approach is the synergy between distributed textual embeddings and symbolic constraints, ensuring that spurious relationships are minimized and novel factual discoveries are maximized [55].

The discussion throughout this paper highlights not only the technical strengths of combining knowledge graphs with deep neural architectures but also the persistent challenges. These range from domain adaptation and real-time updates to logical consistency and data quality. Looking forward, the continued evolution of this research domain calls for more robust handling of ambiguous and multi-modal data, along with improved techniques for efficiently managing large-scale knowledge sources. Despite these unresolved issues, the convergence of symbolically grounded structures and learnable neural representations stands as a promising frontier, capable of driving new solutions in information extraction, decision support, and beyond.

References

[1] "Building 21st century medicine through education, research, policy and practice. abstracts of the society of general internal medicine 34th annual meeting. phoenix, arizona. may 4-7, 2011." *Journal of general internal medicine*, vol. 26 Suppl 1, no. S1, pp. S1–620, May 2011.

[2] A. Agrawal, J. Lu, S. Antol, M. Mitchell, C. L. Zitnick, D. Parikh, and D. Batra, "Vqa: Visual question answering," *International Journal of Computer Vision*, vol. 123, no. 1, pp. 4–31, November 2016.

[3] D. Eisinger, G. Tsatsaronis, M. Bundschus, U. Wieneke, and M. Schroeder, "Automated patent categorization and guided patent search using ipc as inspired by mesh and pubmed," *Journal of biomedical semantics*, vol. 4, no. 1, pp. 1–23, April 2013.

[4] S. Salah, G. Maciá-Fernández, J. E. Díaz-Verdejo, and L. Sánchez-Casado, "A model for incident tickets correlation in network management," *Journal of Network and Systems Management*, vol. 24, no. 1, pp. 57–91, January 2015.

[5] J. J. Jürgens and C. Womser-Hacker, "Limitations of automatic patent ir," *Datenbank-Spektrum*, vol. 14, no. 1, pp. 5–17, February 2014.

[6] D. Karmaker, A. Imran, N. Mohammad, M. Islam, and N. Mahbub, "An automated music selector derived from weather condition and its impact on human psychology," *GSTF Journal on Computing (JoC)*, vol. 4, no. 3, pp. 1–10, December 2015.

[7] M. El-Haj, U. Kruschwitz, and C. Fox, "Creating language resources for under-resourced languages: methodologies, and experiments with arabic," *Language Resources and Evaluation*, vol. 49, no. 3, pp. 549–580, August 2014.

[8] A. H. Razavi, S. Matwin, J. D. Koninck, and R. R. Amini, "Dream sentiment analysis using second order soft co-occurrences (sosco) and time course representations," *Journal of Intelligent Information Systems*, vol. 42, no. 3, pp. 393–413, August 2013.

[9] F. Rotella, F. Leuzzi, and S. Ferilli, "Learning and exploiting concept networks with connektion," *Applied Intelligence*, vol. 42, no. 1, pp. 87–111, July 2014.

[10] "Abstracts of the 111th congress of the german society of ophthalmology. september 19-22, 2013. berlin, germany," *Der Ophthalmologe : Zeitschrift der Deutschen Ophthalmologischen Gesellschaft*, vol. 110 Suppl 1, no. S1, pp. 3–166, October 2013.

[11] A. Sharma and K. D. Forbus, "Modeling the evolution of knowledge and reasoning in learning systems," in *2010 AAAI Fall Symposium Series*, 2010.

[12] G. Sun, Y. Wu, R. Liang, and S. Liu, "A survey of visual analytics techniques and applications: State-of-the-art research and future challenges," *Journal of Computer Science and Technology*, vol. 28, no. 5, pp. 852–867, September 2013.

[13] S.-H. Lee, B. Dudok, V. K. Parihar, K.-M. Jung, M. Zöldi, Y.-J. Kang, M. Maroso, A. Alexander, G. A. Nelson, D. Piomelli, I. Katona, C. L. Limoli, and I. Soltesz, "Neurophysiology of space travel: energetic solar particles cause cell type-specific plasticity of neurotransmission." *Brain structure & function*, vol. 222, no. 5, pp. 2345–2357, November 2016.

[14] K. Forbus, K. Lockwood, A. Sharma, and E. Tomai, "Steps towards a 2nd generation learning by reading system," in *AAAI Spring Symposium on Learning by Reading, Spring*, 2009.

[15] J. Wang, J. Zhang, Y. An, H. Lin, Z. Yang, Y. Zhang, and Y. Sun, "Biomedical event trigger detection by dependency-based word embedding," *BMC medical genomics*, vol. 9, no. 2, pp. 45–432, August 2016.

[16] A. Khodabakhsh, F. Yesil, E. Guner, and C. Demiroglu, "Evaluation of linguistic and prosodic features for detection of alzheimer's disease in turkish conversational speech," *EURASIP Journal on Audio, Speech, and Music Processing*, vol. 2015, no. 1, pp. 9–, March 2015.

[17] M. Elhoseiny and A. Elgammal, "Text to multi-level mindmaps," *Multimedia Tools and Applications*, vol. 75, no. 8, pp. 4217–4244, April 2015.

[18] "Not available," *Der Ophthalmologe : Zeitschrift der Deutschen Ophthalmologischen Gesellschaft*, vol. 113 Suppl 2, no. S2, pp. 13–192, August 2016.

[19] M. Borg, P. Runeson, and A. Ardö, "Recovering from a decade: a systematic mapping of information retrieval approaches to software traceability," *Empirical Software Engineering*, vol. 19, no. 6, pp. 1565–1616, May 2013.

[20] K. D. Forbus, C. Riesbeck, L. Birnbaum, K. Livingston, A. B. Sharma, and L. C. Ureel, "A prototype system that learns by reading simplified texts." in *AAAI Spring Symposium: Machine Reading*, 2007, pp. 49–54.

[21] R. Xu, L. Li, and Q. Wang, "driskkb: a large-scale disease-disease risk relationship knowledge base constructed from biomedical text." *BMC bioinformatics*, vol. 15, no. 1, pp. 105–105, April 2014.

[22] C. C. Xavier, V. L. S. de Lima, and M. Souza, "Open information extraction based on lexical semantics," *Journal of the Brazilian Computer Society*, vol. 21, no. 1, pp. 4–, May 2015.

[23] C.-R. Huang, T. Tokunaga, and S. Y. M. Lee, "Asian language processing: current state-of-the-art," *Language Resources and Evaluation*, vol. 40, no. 3, pp. 203–218, October 2007.

[24] F. Bond, S. Oepen, E. Nichols, D. Flickinger, E. Velldal, and P. Haugereid, "Deep open-source machine translation," *Machine Translation*, vol. 25, no. 2, pp. 87–105, September 2011.

[25] V. Seretan and E. Wehrli, "Multilingual collocation extraction with a syntactic parser," *Language Resources and Evaluation*, vol. 43, no. 1, pp. 71–85, October 2008.

[26] P. M. Alday, M. Schlesewsky, and I. Bornkessel-Schlesewsky, "Towards a computational model of actor-based language comprehension," *Neuroinformatics*, vol. 12, no. 1, pp. 143–179, August 2013.

[27] F. A. Batarseh and A. J. Gonzalez, "Predicting failures in agile software development through data analytics," *Software Quality Journal*, vol. 26, no. 1, pp. 49–66, August 2015.

[28] K. D. Forbus, C. Riesbeck, L. Birnbaum, K. Livingston, A. Sharma, and L. Ureel, "Integrating natural language, knowledge representation and reasoning, and analogical processing to learn by reading," in *Proceedings of the National Conference on Artificial Intelligence*, vol. 22, no. 2. Menlo Park, CA; Cambridge, MA; London; AAAI Press; MIT Press; 1999, 2007, p. 1542.

[29] X. Liu, M. Wang, and B. Huet, "Event analysis in social multimedia: a survey," *Frontiers of Computer Science*, vol. 10, no. 3, pp. 433–446, September 2015.

[30] M. Maree and M. Belkhatir, "Coupling semantic and statistical techniques for dynamically enriching web ontologies," *Journal of Intelligent Information Systems*, vol. 40, no. 3, pp. 455–478, January 2013.

[31] Y. Wu, X. Luo, and Z. Yang, "Semantic separator learning and its applications in unsupervised chinese text parsing," *Frontiers of Computer Science*, vol. 7, no. 1, pp. 55–68, January 2013.

[32] K. Goodman, J. Krueger, and J. Crowley, "The automatic clinical trial: leveraging the electronic medical record in multisite cancer clinical trials." *Current oncology reports*, vol. 14, no. 6, pp. 502–508, August 2012.

[33] V. Nebot and R. Berlanga, "Exploiting semantic annotations for open information extraction: an experience in the biomedical domain," *Knowledge and Information Systems*, vol. 38, no. 2, pp. 365–389, December 2012.

[34] E. K. Elsayed, K. A. ElDahshan, and S. Tawfeek, "Automatic evaluation technique for certain types of open questions in semantic learning systems," *Human-centric Computing and Information Sciences*, vol. 3, no. 1, pp. 19–, October 2013.

[35] M. R. Ghorab, D. Zhou, A. O'Connor, and V. Wade, "Personalised information retrieval: survey and classification," *User Modeling and User-Adapted Interaction*, vol. 23, no. 4, pp. 381–443, May 2012.

[36] C. D. Manning, "Sigir - understanding human language: Can nlp and deep learning help?" in *Proceedings of the 39th International ACM SIGIR conference on Research and Development in Information Retrieval*. ACM, July 2016, pp. 1–1.

[37] N. Bauer, O. Edenhofer, and S. Kypreos, "Linking energy system and macroeconomic growth models," *Computational Management Science*, vol. 5, no. 1, pp. 95–117, April 2007.

[38] K. B. Cohen and L. Hunter, "Smbm - a critical review of pasbio's argument structures for biomedical verbs," *BMC bioinformatics*, vol. 7, no. 3, pp. 1–9, November 2006.

[39] L. Li, D. Wang, S. Zhu, and T. Li, "Personalized news recommendation: a review and an experimental investigation," *Journal of Computer Science and Technology*, vol. 26, no. 5, pp. 754–766, September 2011.

[40] G. Zacharewicz, S. Y. Diallo, Y. Ducq, C. Agostinho, R. Jardim-Goncalves, H. Bazoun, Z. Wang, and G. Doumeingts, "Model-based approaches for interoperability of next generation enterprise information systems: state of the art and future challenges," *Information Systems and e-Business Management*, vol. 15, no. 2, pp. 229–256, May 2016.

[41] B. J. Dorr, R. J. Passonneau, D. Farwell, R. Green, N. Habash, S. Helmreich, E. Hovy, L. Levin, K. J. Miller, T. Mitamura, O. Rambow, and A. Siddharthan, "Interlingual annotation of parallel text corpora: A new framework for annotation and evaluation," *Natural Language Engineering*, vol. 16, no. 3, pp. 197–243, June 2010.

[42] R. J. Gil and M. J. Martin-Bautista, "Smol: a systemic methodology for ontology learning from heterogeneous sources," *Journal of Intelligent Information Systems*, vol. 42, no. 3, pp. 415–455, January 2014.

[43] Y. Lepage and E. Denoual, "Purest ever example-based machine translation: Detailed presentation and assessment," *Machine Translation*, vol. 19, no. 3, pp. 251–282, December 2006.

[44] T. Schumann and U. Schulz, "Stable methods for recognizing acronym-expansion pairs: from rule sets to hidden markov models," *International Journal of Document Analysis and Recognition (IJDAR)*, vol. 8, no. 1, pp. 1–99 999, May 2005.

[45] C. Bitter, D. Elizondo, and Y. Yang, “Natural language processing: a prolog perspective,” *Artificial Intelligence Review*, vol. 33, no. 1, pp. 151–173, December 2009.

[46] W.-J. Huang and C.-L. Liu, “Exploring lexical, syntactic, and semantic features for chinese textual entailment in ntcir rite evaluation tasks,” *Soft Computing*, vol. 21, no. 2, pp. 311–330, March 2015.

[47] A. Sharma and K. Forbus, “Modeling the evolution of knowledge in learning systems,” in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 26, no. 1, 2012, pp. 669–675.

[48] G. Tsatsaronis, G. Balikas, P. Malakasiotis, I. Partalas, M. Zschunke, M. R. Alvers, D. Weissenborn, A. Krithara, S. Petridis, D. Polychronopoulos, Y. Almirantis, J. Pavlopoulos, N. Baskiotis, P. Gallinari, T. Artières, A.-C. N. Ngomo, N. Heino, E. Gaussier, L. Barrio-Alvers, M. Schroeder, I. Androutsopoulos, and G. Paliouras, “An overview of the bioasq large-scale biomedical semantic indexing and question answering competition,” *BMC bioinformatics*, vol. 16, no. 1, pp. 138–138, April 2015.

[49] P. Felt, E. K. Ringger, K. D. Seppi, K. S. Heal, R. Haertel, and D. Lonsdale, “Evaluating machine-assisted annotation in under-resourced settings,” *Language Resources and Evaluation*, vol. 48, no. 4, pp. 561–599, November 2013.

[50] J. Albrecht and R. Hwa, “Regression for machine translation evaluation at the sentence level,” *Machine Translation*, vol. 22, no. 1, pp. 1–27, November 2008.

[51] R. Praszkier, “Empathy, mirror neurons and sync,” *Mind & Society*, vol. 15, no. 1, pp. 1–25, December 2014.

[52] D. Ristic and J. Ellrich, “Acetylsalicylic acid inhibits purinergic facilitation of neck muscle nociception in mice: implications for acute treatment of tension-type headache,” *The journal of headache and pain*, vol. 11, no. Suppl 1, pp. S2–150, November 2010.

[53] K. D. Forbus, K. Lockwood, A. B. Sharma, and E. Tomai, “Steps towards a second generation learning by reading system.” in *AAAI Spring Symposium: Learning by Reading and Learning to Read*, 2009, pp. 36–43.

[54] A.-D. Pham, A. Névéol, T. Lavergne, D. Yasunaga, O. Clément, G. Meyer, R. Morello, and A. Burgun, “Natural language processing of radiology reports for the detection of thromboembolic diseases and clinically relevant incidental findings,” *BMC bioinformatics*, vol. 15, no. 1, pp. 266–266, August 2014.

[55] T. Wilson, J. Wiebe, and R. Hwa, “Recognizing strong and weak opinion clauses,” *Computational Intelligence*, vol. 22, no. 2, pp. 73–99, May 2006.