Automatic Construction of Knowledge Graphs from Text and Structured Data: A Preliminary Literature Review

Maraim Masoud
Insight SFI Research Centre for Data Analytics, Data Science Institute,
National University of Ireland Galway, Ireland

Bianca Pereira
Insight SFI Research Centre for Data Analytics, Data Science Institute,
National University of Ireland Galway, Ireland

John McCrae
Insight SFI Research Centre for Data Analytics, Data Science Institute,
National University of Ireland Galway, Ireland

Paul Buitelaar
Insight SFI Research Centre for Data Analytics, Data Science Institute,
National University of Ireland Galway, Ireland

Abstract

Knowledge graphs have been shown to be an important data structure for many applications, including chatbot development, data integration, and semantic search. In the enterprise domain, such graphs need to be constructed based on both structured (e.g. databases) and unstructured (e.g. textual) internal data sources; preferentially using automatic approaches due to the costs associated with manual construction of knowledge graphs. However, despite the growing body of research that leverages both structured and textual data sources in the context of automatic knowledge graph construction, the research community has centered on either one type of source or the other. In this paper, we conduct a preliminary literature review to investigate approaches that can be used for the integration of textual and structured data sources in the process of automatic knowledge graph construction. We highlight the solutions currently available for use within enterprises and point areas that would benefit from further research.

2012 ACM Subject Classification Information systems → Information extraction

Keywords and phrases Knowledge Graph Construction, Enterprise Knowledge Graph

Digital Object Identifier 10.4230/OASIcs.LDK.2021.19

Funding This work was supported by Science Foundation Ireland under grant SFI/12/RC/2289_P2 (Insight).

1 Introduction

The automatic construction of knowledge graphs from textual or structured data sources enables the generation of domain-specific enterprise knowledge graphs while decreasing the costs associated with manual generation of formal knowledge datasets [29]. The extraction from textual sources enables the representation of internal knowledge generated by employees and customers through the analysis of a domain of discourse, whereas the extraction from structured data sources (e.g. relational databases) enables the representation of enterprise information that is generated and applied in the provision of services and applications. Despite the potential benefits of both text and structured sources in the context of domain-specific enterprise knowledge graphs, the research community has focused on either one source of data or the other. In this paper, we perform a literature review that explores the

3rd Conference on Language, Data and Knowledge (LDK 2021).
Editors: Dagmar Gromann, Gilles Sérasset, Thierry Declerck, John P. McCrae, Jorge Gracia, Julia Bosque-Gil, Fernando Bobillo, and Barbara Heinisch; Article No. 19; pp. 19:1–19:9
OpenAccess Series in Informatics
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

combination of both types of data sources in the process of constructing domain-specific enterprise knowledge graph. As such construction tends to be automatically performed, the literature review investigates how both data sources can be integrated in the process of automatic knowledge graph construction.

First we highlight the initiatives promoting knowledge graph construction either from text or structured data sources in Section 2. Then, in Section 3, we present our methodology for analysis of the literature. In Section 4 we introduce the conceptual framework used to analyse the literature for domain-specific knowledge graph construction. Further, we present the results in Sections 5. We conclude the paper by highlighting the solutions available, lessons learned from this analysis, and pointing directions for future research.

2 Automatic Construction of Knowledge Graphs

Automatic knowledge graph construction has been the target of research challenges and initiatives in different research communities [2, 23, 31, 11].

Initiatives such as SemEval and OAEI focus only on the use of either textual or structured data sources. The SemEval Taxonomy Extraction Task [2], organised as part of SIGLEX/SIGSEM[1] conference in 2016, proposed the extraction of domain-specific taxonomic structures exclusively from textual data sources. Whereas the OAEI [11] organised as part of the International Semantic Web Conference (ISWC) since 2006, focuses only on the use of structured data sources where the goal is to create an integrated ontology based on the alignment between two (or more) already available ontologies or knowledge graphs.

In the intersection between text and structured data, we can find the TAC-KBP and NEEL challenges that aim at verifying if entities appearing in text are already represented in a structured knowledge base. The TAC-KBP Entity Linking task [23], a yearly event from 2009 to 2016 as part of TAC[2], aimed at taking advantage of a preexistent general domain structured data source and use the text as source of additional information to expand it. In a similar manner, the NEEL challenge [31], that run from 2013 to 2016 as part of the WWW Conference, focused on identifying if entities appearing in microblog messages (i.e. tweets) were already available in a general domain knowledge base.

Despite all the available initiatives in the automatic construction of knowledge graphs based on different sources, none of these initiatives explicitly focuses on both: (i) the aggregated use of text and structured data sources, and (ii) the generation of a domain-specific knowledge graph. Therefore, the goal of this paper is to analyse what are the available approaches in the literature that could be applied to the automatic construction of domain-specific enterprise knowledge graphs based on textual and structured data sources and what are the areas the require further research.

3 Methodology

The methodology used to select the literature followed four steps (Figure 1): (i) selection of seed papers, (ii) search, (iii) filtering, and (iv) analysis.

The selection of seed papers was performed based on a convenience sample, with a selection of survey and literature review papers known to the authors. A set of six seed papers was selected [5, 25, 26, 32, 33, 45].

[1] The Special Interest Group on the Lexicon (SIGLEX)/Special Interest Group on Computational Semantics (SIGSEM)

[2] Text Analysis Conference

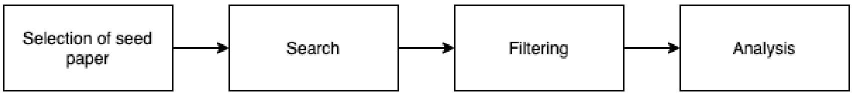

Figure 1 Methodology for literature review.

Table 1 Inclusion and exclusion criteria for selection of the literature.

Inclusion Criteria	Exclusion Criteria
- √ written in the English language. - √ published in conferences proceedings or in a journal. - √ the abstract and conclusion indicate the paper is in the topic of enriching the results of automatic extraction of knowledge graph with structured data.	- × only (semi-)manual approaches. - × no explanation of the method used. - × using only textual data sources.

Based on these seed papers, we expand our search to include also the literature in their list of references. Next, all the literature collected is filtered according to a set of inclusion and exclusion criteria (Table 1). Finally, all papers that passed the filtering criteria are analysed according to a series of dimensions of interest.

4 Dimensions Used for Analysis of the Literature

The analysis of the literature was performed based on four dimensions of interest: (i) point of integration, i.e. in which point the information coming from text and structured data sources was integrated during the process of automatic extraction of knowledge graphs; (ii) integration goal, i.e. if the goal is to expand the knowledge graph or validate its current content; (iii) format of the structured data, and (iv) format of the final knowledge graph, i.e. the type of knowledge graph expected as the output of the the extraction process using both text and structured data source.

Three **point of integration** were considered in our analysis. *Pre-construction* (Figure 2a) refers to the enrichment of the textual documents with information from the structured data source before they are provided as input to the knowledge graph construction algorithm. Integration that happens *during construction* (Figure 2b) assumes that both textual and structured data sources are not linked in advance and, instead, their linking will be performed during the process of automatic knowledge graph extraction. Last, *post-construction* (Figure 2c) stands for the connection between the output of the automatic extraction of knowledge graph from text and information coming from the structured data source.

The joint use of text and structured data sources can be used for two different **integration goals**: (i) knowledge graph completion, where data from both sources are combined to extend the knowledge graph, and (ii) knowledge graph validation, in which one data source is used to validate the information from the other data source.

Since enterprise environments work with heterogeneous types of data, the third dimension of interest relates to the **format of the structured data**. When analysing this dimension we are simplifying our categorisation by assuming that any data source that can be represented by an entity-relation diagram can be considered, or converted to, a graph.

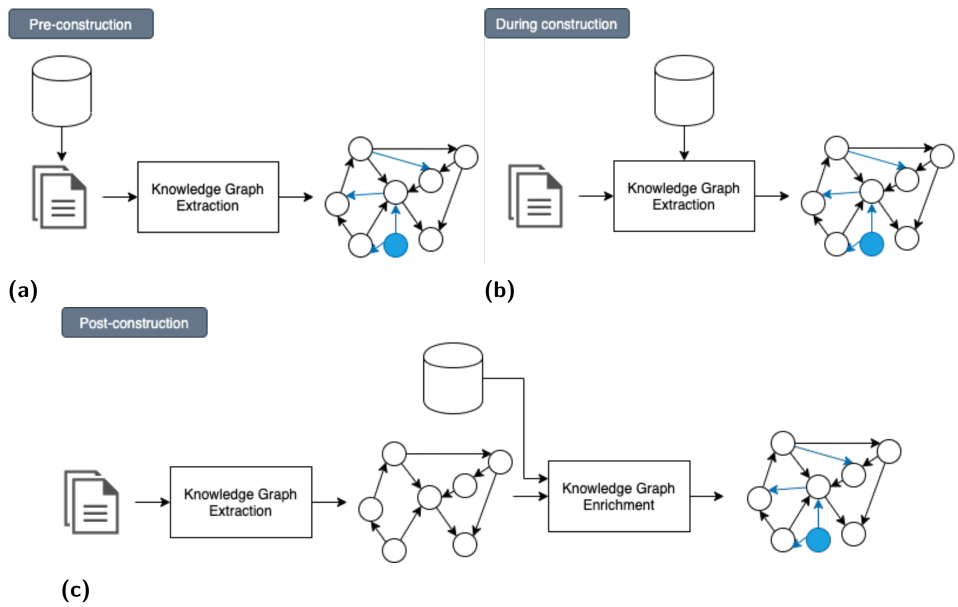

Figure 2 Integration of textual and structured data sources to enrich the automatic construction of knowledge graphs (a) pre-construction, (b) during construction, and (c) post-construction.

Last, given that there are multiple definitions of knowledge graph in the literature, we also analyse the **format of the output knowledge graph**. Four types of knowledge graph were considered: (i) term taxonomy, i.e. a knowledge graph only contains taxonomic relations and where vertices are terms, (ii) topic taxonomy, similar to term taxonomies but where vertices are represented by sets of terms, (iii) labeled graphs, in which relations receive a label representing a relation type (e.g. usedFor, preRequisite), and (iv) ontologies, where vertices and relations have a formal logical representation associated with them.

5 Results and Discussion

Based on the list of seed papers and their references, we start our analysis based on a set of 131 papers. From those, 97 papers were removed due to our exclusion criteria, resulting in one seed paper and further 34 papers analysed. Table 2 presents the categorisation of each paper and in the next sections we present the results of our analysis according to each of the dimensions chosen for literature review.

5.1 Point of Integration

The approaches used for **pre-construction** leverage the knowledge graph extraction algorithm in one of three ways: (i) by combining both textual and structured information into a single semantic space via the use of text embeddings and motifs [32], (ii) by generating a profile for each term using the metadata of the documents associated to the term (e.g. popularity of a document) [44], then using this profile to determine where the term fits in the structure of the final knowledge graph, and (iii) by using the structured data source to generate an initial graph that connects the different terms from text [42].

Regarding the integration **during construction**, the structured data source is provided as training data for the detection of relations between terms extracted from text. This detection is based on either: (i) relation classification, or (ii) relation prediction. In relation

Table 2 Papers categorised according to our dimensions of analysis.

Dimension of Analysis		Paper References
Point of integration	Pre-construction	[32, 42, 44]
	During construction	[39, 19, 12, 13, 41, 22, 40, 37, 16, 34, 28, 3, 14, 15]
	Post-construction	[7, 35, 9, 43, 46, 36, 17, 38, 8, 6, 20, 18, 1, 4, 10, 30, 27]
Integration goals	Completion	[21, 7, 39, 19, 12, 13, 41, 22, 40, 37, 16, 35, 9, 34, 28, 3, 14, 15, 43, 46, 36, 17, 38, 8, 6, 1, 4, 10, 24, 30, 27, 42, 44]
	Validation	[7, 39, 19, 12, 13, 41, 22, 40, 9, 14, 20, 18]
Format of structured data	Table	[32]
	Key-value pairs	[42, 44]
	Graph	[32, 7, 39, 19, 12, 13, 41, 22, 40, 37, 16, 35, 9, 34, 28, 3, 14, 15, 43, 46, 36, 17, 38, 8, 6, 20, 18, 1, 4, 10, 24, 30, 27]
Format of the output knowledge graph	Term taxonomy	–
	Topic taxonomy	[32, 42, 44]
	Labeled graph	[7, 39, 19, 12, 13, 41, 22, 40, 37, 16, 9, 34, 28, 3, 14, 15, 43, 46, 36, 17, 38, 8, 6, 20, 18, 1, 4, 10, 24, 30, 27]
	Ontology	[35]

classification, the types of relations existent in the structured data source are used to classify the relations existent between any two terms from text. In the reviewed literature, this is achieved by using neural networks ([3, 14, 15, 34]), or probabilistic methods ([28]). In contrast, in relation prediction, the goal is to identify what is the target term (or entity) to which any single term extracted should be related to. The literature focused on the use of neural networks using representations that are: (i) based on embeddings [39, 19, 12, 13, 41, 22, 40], or (ii) based on tensors [16, 37].

In **post-construction integration**, the knowledge graph built from text can be integrated with other knowledge graphs by: (i) graph alignment, where knowledge graphs are linked to each other but are still kept as separate entities [10, 17, 1, 35, 30, 7, 8, 46, 36, 6, 43, 8, 38], (ii) graph fusion, where the knowledge graphs have their terms and structures merged into a single knowledge graph [4, 9], or (iii) logical inference, where one knowledge graph is used to extract inference rules for expansion of the data in the other knowledge graph [27].

5.2 Integration Goals

We expect the literature to be heavily biased towards the goal of knowledge graph completion, i.e. extending the knowledge graph with information from both text and structured data sources. This is confirmed by our analysis where 33 papers out of 40 focused exclusively on this goal. The papers that focus only on the validation of knowledge graphs are limited to verifying the correctness of entities and relations but do not perform any additional step of correcting the detected errors [18, 20].

5.3 Format of the Structured Data

The literature explores the use of three different formats of structured data: (i) tables, (ii) key-value pairs , and (iii) graphs.

Tables and value-key pairs are provided as metadata to textual documents in pre-construction approaches either by the use of explicit links between entities in the structured data and documents that refer to those entities [32], or by the inference of these links via analysis of user interactions with both data sources [42, 44]. Graphs, on the other hand, are a dominant format in the analysed literature (Table 2), therefore enterprises wishing to integrate text and structured data for knowledge graph construction would have a higher availability of approaches if the structured data source is a graph-like structure.

5.4 Format of the Output Knowledge Graph

The literature analysed have a strong focus on labelled graphs, while the extraction of taxonomies and ontologies is underrepresented. This demonstrates that despite the amount of work on automatic extraction of taxonomies from text or ontology generation from structured data sources, it does not seem to be a common practice to integrate the two types of data sources when generating taxonomies or ontologies.

6 Conclusion

The goal of this paper is to provide knowledge on what is available in the literature for use by enterprises wishing to generate knowledge graphs based on their own internal data sources. For that, we present a preliminary literature review that investigates approaches used for the integration of textual and structured data sources in the process of automatic knowledge graph construction. Our analysis was based on: (i) point of integration (before, during or after knowledge graph construction), (ii) the goal for integrating sources, (iii) the format of the structured data source, and (iv) the structure of the constructed knowledge graph. Based on this analysis we conclude that, enterprises have a range of approaches available if aiming at the generation of a labelled knowledge graph that aggregates data from both textual and structured sources, where the structured data source used has a graph-like structure and the integration between textual and structured source is done only after a knowledge graph has been extracted from text (what we name post-construction integration). Meanwhile, the integration of data sources before they are used for automatic knowledge graph construction, as well as the use of tables or key-value pairs as structured data sources are still areas with possibility for further research.

7 Lessons Learned and Future Work

Many lessons can be drawn from this specific analysis in terms of our conceptual framework, survey analysis and findings.

Our categorization, while specific, has shown to be useful in classifying available approaches for constructing a domain-specific knowledge graph by; (i) categorizing similar approaches based on the selected dimensions, and (ii) displaying the patterns that influence the decision to adapt a specific variation of each dimension as discussed in the results section. The value of this classification is that it provides enterprises with a clear set of approaches for constructing a domain-specific knowledge graphs from structured and unstructured data sources. As a result, this survey is a step forward in understanding the possible solutions for generating domain-specific enterprise knowledge graph. It also assists enterprise practitioners who may prefer one approach over the another due to constrains in resource, time, and cost.

From the survey perspective, there is an opportunity to future investigate the research and the application of knowledge graph in enterprise domain. Future work will include expanding the survey to a systematic review with keyword-based seed papers. We envision this as a large-scale study that will examine the enterprise knowledge graph integration from different perspectives and demonstrate use cases from various application domains.

From the perspective of survey results, there are a range of options for generating knowledge graphs by aggregating structured and unstructured sources. According to our findings, integration using graph-like structure is a popular approach in comparison to tables and key-value pairs. The later formats are currently in the tentative stage as outlined in the results section. There is also an equal interest in integrating resources during or after the construction of the knowledge graph, as opposed to leveraging resources before the knowledge graph is created. In terms of integration goals, most research focuses on completing a knowledge graph, whereas only a few focus on using data sources to validate a populated knowledge graph. Finally, the majority of work is designed to generate a labelled graph as an output, with a few recent work focusing on topic knowledge graph for classification purposes.

References

1 Alessio Palmero Aprosio, Claudio Giuliano, and Alberto Lavelli. Automatic expansion of dbpedia exploiting wikipedia cross-language information. In *Extended Semantic Web Conference*, pages 397–411. Springer, 2013.

2 Georgeta Bordea, Els Lefever, and Paul Buitelaar. SemEval-2016 task 13: Taxonomy extraction evaluation (TExEval-2). In *Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016)*, pages 1081–1091. ACL, 2016.

3 Antoine Bordes, Nicolas Usunier, Alberto Garcia-Duran, Jason Weston, and Oksana Yakhnenko. Translating embeddings for modeling multi-relational data. In *NeurIPS*, pages 1–9, 2013.

4 Volha Bryl and Christian Bizer. Learning conflict resolution strategies for cross-language wikipedia data fusion. In *Proceedings of the 23rd International Conference on World Wide Web*, pages 1129–1134, 2014.

5 Hongyun Cai, Vincent W Zheng, and Kevin Chen-Chuan Chang. A comprehensive survey of graph embedding: Problems, techniques, and applications. *IEEE Transactions on Knowledge and Data Engineering*, 30(9):1616–1637, 2018.

6 Muhao Chen, Yingtao Tian, Kai-Wei Chang, Steven Skiena, and Carlo Zaniolo. Co-training embeddings of knowledge graphs and entity descriptions for cross-lingual entity alignment. In *IJCAI*, pages 3998–4004, 2018.

7 Muhao Chen, Yingtao Tian, Mohan Yang, and Carlo Zaniolo. Multilingual knowledge graph embeddings for cross-lingual knowledge alignment. In *IJCAI*, pages 1511–1517, 2017.

8 Gustavo de Assis Costa and José Maria Parente de Oliveira. Linguistic frames as support for entity alignment in knowledge graphs. In *Proceedings of the 20th International Conference on Information Integration and Web-based Applications & Services*, pages 226–229, 2018.

9 Xin Dong, Evgeniy Gabrilovich, Geremy Heitz, Wilko Horn, Ni Lao, Kevin Murphy, Thomas Strohmann, Shaohua Sun, and Wei Zhang. Knowledge vault: A web-scale approach to probabilistic knowledge fusion. In *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 601–610, 2014.

10 Arnab Dutta, Christian Meilicke, and Simone Paolo Ponzetto. A probabilistic approach for integrating heterogeneous knowledge sources. In *European Semantic Web Conference*, pages 286–301. Springer, 2014.

11 Jérôme Euzenat, Christian Meilicke, Heiner Stuckenschmidt, Pavel Shvaiko, and Cássia Trojahn. Ontology alignment evaluation initiative: Six years of experience. In Stefano Spaccapietra, editor, *Journal on Data Semantics XV*, pages 158–192. Springer Berlin Heidelberg, Berlin, Heidelberg, 2011.

12 Jun Feng, Minlie Huang, Yang Yang, and Xiaoyan Zhu. Gake: Graph aware knowledge embedding. In *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers*, pages 641–651, 2016.

13 Shu Guo, Quan Wang, Bin Wang, Lihong Wang, and Li Guo. Semantically smooth knowledge graph embedding. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 84–94, 2015.

14 Guoliang Ji, Shizhu He, Liheng Xu, Kang Liu, and Jun Zhao. Knowledge graph embedding via dynamic mapping matrix. In *Proceedings of the 53rd annual meeting of the ACL and the 7th international joint conference on natural language processing (volume 1: Long papers)*, pages 687–696, 2015.

15 Guoliang Ji, Kang Liu, Shizhu He, and Jun Zhao. Knowledge graph completion with adaptive sparse transfer matrix. In Dale Schuurmans and Michael P. Wellman, editors, *AAAI*, pages 985–991. AAAI Press, 2016.

16 Seyed Mehran Kazemi and David Poole. Simple embedding for link prediction in knowledge graphs. In *NeurIPS*, pages 4289–4300, 2018.

17 Chao Kong, Ming Gao, Chen Xu, Yunbin Fu, Weining Qian, and Aoying Zhou. EnAli: entity alignment across multiple heterogeneous data sources. *Frontiers of Computer Science*, 13(1):157–169, 2019.

18 Jens Lehmann, Daniel Gerber, Mohamed Morsey, and Axel-Cyrille Ngonga Ngomo. Defacto-deep fact validation. In *International semantic web conference*, pages 312–327. Springer, 2012.

19 Yankai Lin, Zhiyuan Liu, Maosong Sun, Yang Liu, and Xuan Zhu. Learning entity and relation embeddings for knowledge graph completion. In Blai Bonet and Sven Koenig, editors, *Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence, January 25-30, 2015, Austin, Texas, USA*, pages 2181–2187. AAAI Press, 2015.

20 Shuangyan Liu, M. d'Aquin, and E. Motta. Towards linked data fact validation through measuring consensus. In *LDQ@ESWC*, 2015.

21 Shuangyan Liu, Mathieu d'Aquin, and Enrico Motta. Measuring accuracy of triples in knowledge graphs. In *LDK*, pages 343–357. Springer, 2017.

22 Yuanfei Luo, Quan Wang, Bin Wang, and Li Guo. Context-dependent knowledge graph embedding. In *Proceedings of the 2015 Conference on EMNLP*, pages 1656–1661, 2015.

23 Paul McNamee and Hoa Trang Dang. Overview of the tac 2009 knowledge base population track. In *Text Analysis Conference (TAC)*, volume 17, pages 111–113, 2009.

24 Emir Muñoz, Aidan Hogan, and Alessandra Mileo. Triplifying wikipedia's tables. *LD4IE@ ISWC*, 2013.

25 Hoang Long Nguyen, Dang Thinh Vu, and Jason J Jung. Knowledge graph fusion for smart systems: A survey. *Information Fusion*, 61:56–70, 2020.

26 Heiko Paulheim. Knowledge graph refinement: A survey of approaches and evaluation methods. *Semantic web*, 8(3):489–508, 2017.

27 Heiko Paulheim and Simone Paolo Ponzetto. Extending dbpedia with wikipedia list pages. *NLP-DBPEDIA@ ISWC*, 13, 2013.

28 Boya Peng, Yejin Huh, Xiao Ling, and Michele Banko. Improving knowledge base construction from robust infobox extraction. In *Proceedings of NAACL HLT Conference, Volume 2 (Industry Papers)*, pages 138–148, 2019.

29 Bianca Pereira, C. Robin, Tobias Daudert, John P. McCrae, Pranab Mohanty, and P. Buitelaar. Taxonomy extraction for customer service knowledge base construction. In *SEMANTiCS*, 2019.

30 Dominique Ritze, Oliver Lehmberg, and Christian Bizer. Matching html tables to dbpedia. In *Proceedings of the 5th International Conference on Web Intelligence, Mining and Semantics*, pages 1–6, 2015.

31 G. Rizzo, B. Pereira, A. Varga, M. Van Erp, and A.E.C. Basave. Lessons learnt from the Named Entity rEcognition and Linking (NEEL) challengeÂ series. *Semantic Web*, 8(5):667–700, 2017.

32 Jingbo Shang, Xinyang Zhang, Liyuan Liu, Sha Li, and Jiawei Han. Nettaxo: Automated topic taxonomy construction from text-rich network. In *Proceedings of The Web Conference 2020*, pages 1908–1919, 2020.

33 Wei Shen, Jianyong Wang, and Jiawei Han. Entity linking with a knowledge base: Issues, techniques, and solutions. *IEEE Transactions on Knowledge and Data Engineering*, 27(2):443–460, 2014.

34 Richard Socher, Danqi Chen, Christopher D Manning, and Andrew Ng. Reasoning with neural tensor networks for knowledge base completion. In *NeurIPS*, pages 926–934. Citeseer, 2013.

35 Fabian M Suchanek, Serge Abiteboul, and Pierre Senellart. Paris: probabilistic alignment of relations, instances, and schema. *Proceedings of the VLDB Endowment*, 5(3):157–168, 2011.

36 Zequn Sun, Wei Hu, and Chengkai Li. Cross-lingual entity alignment via joint attribute-preserving embedding. In *International Semantic Web Conference*, pages 628–644. Springer, 2017.

37 Yi Tay, Anh Tuan Luu, Siu Cheung Hui, and Falk Brauer. Random semantic tensor ensemble for scalable knowledge graph link prediction. In *Proceedings of the Tenth ACM International Conference on Web Search and Data Mining*, pages 751–760, 2017.

38 Lucy Wang, Chandra Bhagavatula, Mark Neumann, Kyle Lo, Chris Wilhelm, and Waleed Ammar. Ontology alignment in the biomedical domain using entity definitions and context. In *Proceedings of the BioNLP 2018 workshop*, pages 47–55, 2018.

39 Zhen Wang, Jianwen Zhang, Jianlin Feng, and Zheng Chen. Knowledge graph embedding by translating on hyperplanes. In Carla E. Brodley and Peter Stone, editors, *AAAI*, pages 1112–1119, 2014.

40 Zhigang Wang, Juanzi Li, Zhiyuan Liu, and Jie Tang. Text-enhanced representation learning for knowledge graph. In *IJCAI*, pages 4–17, 2016.

41 Ruobing Xie, Zhiyuan Liu, and Maosong Sun. Representation learning of knowledge graphs with hierarchical types. In *IJCAI*, pages 2965–2971, 2016.

42 Xiaoxin Yin and Sarthak Shah. Building taxonomy of web search intents for name entity queries. In *Proceedings of the 19th International Conference on World Wide Web*, WWW '10, page 1001–1010. ACM, 2010.

43 Youmin Zhang, Li Liu, Shun Fu, and Fujin Zhong. Entity alignment across knowledge graphs based on representative relations selection. In *ICSAI*, pages 1056–1061. IEEE, 2018.

44 Yuchen Zhang, Amr Ahmed, Vanja Josifovski, and Alexander Smola. Taxonomy discovery for personalized recommendation. In *Proceedings of the 7th ACM international conference on Web search and data mining*, pages 243–252, 2014.

45 Xiaojuan Zhao, Yan Jia, Aiping Li, Rong Jiang, and Yichen Song. Multi-source knowledge fusion: a survey. *World Wide Web*, 23(4):2567–2592, 2020.

46 Hao Zhu, Ruobing Xie, Zhiyuan Liu, and Maosong Sun. Iterative entity alignment via joint knowledge embeddings. In *IJCAI*, volume 17, pages 4258–4264, 2017.