Journal on Artificial Intelligence
DOI:10.32604/jai.2021.018648
Article

Tech Science Press

Semantic Link Network Based Knowledge Graph Representation and Construction

Weiyu Guo[1,*], Ruixiang Jia[1] and Ying Zhang[2]

[1]The School of Computer Science and Technology, Qilu University of Technology (Shandong Academy of Sciences), Jinan, 250353, China

[2]Shandong Cable Interactive Service Co., Ltd., Jinan, 250100, China

[*]Corresponding Author: Weiyu Guo. Email: guo_weiyu@163.com

Received: 15 March 2021; Accepted: 26 April 2021

Abstract: A knowledge graph consists of a set of interconnected typed entities and their attributes, which shows a better performance to organize, manage and understand knowledge. However, because knowledge graphs contain a lot of knowledge triples, it is difficult to directly display to researchers. Semantic Link Network is an attempt, and it can deal with the construction, representation and reasoning of semantics naturally. Based on the Semantic Link Network, this paper explores the representation and construction of knowledge graph, and develops an academic knowledge graph prototype system to realize the representation, construction and visualization of knowledge graph.

Keywords: Knowledge graph; semantic link network; knowledge application

1 Introduction

With the rapid development of big data, information presents explosive growth. Massive heterogeneous and loosely-structured data is a serious challenge to the acquire and use information and knowledge effectively. Knowledge graph is a technology of knowledge representation and management, which emphasizes the ability of semantic retrieval.

In May 2012, Google proposed the knowledge graph, which can organize billions of knowledge triples in form of graph to improve search performance and knowledge discovery of search engine [1]. Knowledge graph describes various entities, concepts and the relationships between them. It can present knowledge to users systematically.

Knowledge graph provide a better way to organize, manage and understand the massive information on the Internet by a human-like cognitive method. Knowledge graph is a kind of relationship graph. Compared with simple text information, the representation of graph is more intuitive and easier to be understood. Knowledge graph has brought vitality to the Internet semantic search, and has great power in intelligent retrieval, big data analysis and decision support. It has become the infrastructure of knowledge-based intelligent services. Knowledge graph has become one of the hotspots and the core forces to develop artificial intelligence.

Google applied its technology to search engine to improve the quality of Google query results. Since then, it stimulated the enthusiasm of the Internet search engine companies and research institutions both foreign a domestic to build knowledge graphs, such as Microsoft's Probase [2], Baidu's "Zhixin", Sogou's "zhicube", etc. In foreign countries, large-scale knowledge graph had been constructed, including DBpedia [3], Yago [4], Nell, Freebase [5] and Google knowledge graph. These knowledge graphs contained millions of entities and billions of relationships. The importance of knowledge graph was also concerned by the government. In the new generation of artificial intelligence development plan, it focused on the core technologies of knowledge processing, deep search and visual interaction, and to realize the automatic acquisition of continuous incremental knowledge [6–7].

Knowledge graph containing billions of entities must be built, which includes concept recognition, entity discovery, attribute prediction, knowledge evolution modeling and relationship mining.

2 The Concept of Knowledge Graph

Knowledge graph, a structured knowledge base with semantic, is usually used to formally describe concepts and their relationships in the real world [8–10]. It is based on the triples of "Entity-Relationship-Entity", or "Entity-Attribute-Value". Entities are connected with each other through relationships, to form a network structure. Through knowledge graph, we can realize the semantic links among concepts, and support users to search by topic instead of string, so as to realize semantic retrieval. The search engines based on knowledge graph can feed back structured knowledge to users graphically. Users can accurately locate and acquire knowledge without browsing a large number of web pages.

Knowledge graph technology refers to an interdisciplinary research integrating cognitive computing, knowledge representation, knowledge reasoning, information retrieval and extraction, natural language processing and semantic web, data mining and machine learning. Knowledge graph research explores the theory and method of acquiring knowledge from Internet resources, as well as promotes the research of knowledge-driven language understanding. It can promote the transformation of information processing technology from information service to knowledge service [11–14].

3 Knowledge Graph Representation and Construction Based on Semantic Link Network

Knowledge graphs use RDF to describe and store the resources in the world in the form of triples [15–16]. RDF (Resource Description Framework) is a kind of resource description language. It is influenced by metadata standard, framework system, object-oriented language and so on. It is used to describe all kinds of network resources. Its appearance provides a standard data description framework to publish structured data on the web.

Triple is a general representation of knowledge graph G, that is, G = (E, R, S), where $E = \{e_1, e_2, .., e_n\}$ is the entity set in the knowledge base, which contains n different entities; $R = \{r_1, r_2, ..., r_m\}$ is the relationship set in the knowledge base, which contains m different relationships; $S \subseteq E \times R \times E$ represents the triple set in the knowledge base. The basic forms of triples mainly include entity1-relationship-entity2 and concept-attribute-value, that is, the concept refers to the conceptual representation of object formed in people's understanding of the world; entity is the specific thing in the objective world; attribute mainly refers to the attributes, characteristics and parameters that the objects may have; relationship describes the objective relationship among concepts, entities and events.

RDF is based on triple representation, which provides inter-operability between applications to exchange machine-understandable information on the Internet. However, because the knowledge graph contains a large number of entity-relationship triples, it cannot be directly represented. Semantic Link Network is an attempt in this area, and can deal with the construction, representation and reasoning of semantics better.

Semantic Link Network (SLN) is a model based on semantic links. It is a natural extension of the current Internet hyperlinks. Entities can be connected by some links with semantic to build a network. These links can be some predefined types: *cause-effective* link, *implication* link, *subtype* link, *similar-to* link, *instance* link, *sequential* link, *reference* link and so on. They can also defined by users. Semantic links provide the external semantics [17]. Entities are regarded as nodes, and the links are regarded as edges in Semantic link networks. Semantic nodes can be atomic nodes (entities) or complex nodes (a semantic link network)[18]. Reasoning, matching and visualization can be made based on the semantic links. SLN has some characters coincided with knowledge graph. So An attempt can be made to represent and construct knowledge graph based on semantic link network [19].

There are three ways to build knowledge graph: top-down way, bottom-up way and the combination of the two methods. The top-down approach is to first build the top-level relationship ontology, and then update the extracted entity matching to the top-level ontology [20]. In the bottom-up approach, the

categories, entities, attributes and relationships found in the extracted data are directly merged into the knowledge graph. We have proposed a new method of knowledge representation and construction of knowledge graph based on semantic link network. The process of representation and construction of knowledge graph based on Semantic link network can be shown as Fig. 1. The construction process is divided into four modules: knowledge acquisition, knowledge representation, knowledge storage and knowledge visualization.

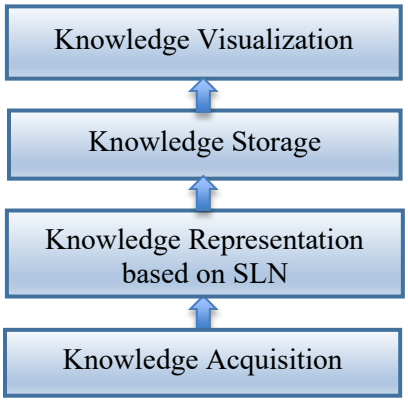

Figure 1: The process of knowledge graph based on SLN

1) Knowledge acquisition

Knowledge acquisition is a technology to extract structured elements such as entities, attributes and relationships from semi-structured or unstructured data. Data sources can be encyclopedic data (such as Wikipedia, Baidu Encyclopedia and Interactive Encyclopedia, etc.), search logs and Web pages of the World Wide Web. Through information extraction, the knowledge elements such as entity, relationship and attribute can be extracted from the original corpus. After disambiguation, cleaning and integration, the knowledge represented by triples can be obtained.

2) Knowledge graph representation based on semantic link network.

Knowledge graph can be regarded as a network structure of graph, in which nodes represent entities and edges represent relationships. We use the Semantic Link Network Builder as a tool to define, modify, verify and access the semantic links. SLN Builder can also transform a semantic link network definition into XML description for cross platform data exchange. SLN Builder consists of four parts: SLN definition, SLN verification, document verification and document storage [18].

By Semantic Link Network Builder interface, users can easily add and delete semantic links according to the relationship of knowledge entities. A document or a Web page can be handled as follows: the file is uploaded into SLN Builder, and then the semantic link tags can be constructed according to the user's understanding, for example, *cause-effective* link, *similar-to* link, *sequential* link and *reference* link. Matching and reasoning is supported by these Semantic links based on rules. Then, the data validation section is performed to check the correctness of semantic links and tags. Finally, the document storage function is performed to establish the XML description.

3) knowledge storage

The data storage part is for storing the structure of semantic link and document content. The defined semantic links are transformed into XML description. XML description includes two types: XML description of document content and XML description of semantic link. These two XML descriptions are stored in different files. The semantic link in the XML description of semantic link can be accessed through the ID number in the XML description of document content.

The data verification component includes semantic link verification and document verification. Because the semantic link is stored in the fixed data structure described above, if the user defines the

semantic link, it may produce illegal data structure. The purpose of semantic link verification is to check whether the semantic link is stored in the correct way. The document verification function mainly checks the correctness of the result document. It is mainly to check the following two errors: 1) check the mismatch of tags. 2) check the mistake between networks.

4) Knowledge visualization

Knowledge visualization uses SLN Browser to browse knowledge graph and perform reasoning and content recommendation in a convenient way for users to read. SLN Browser consists of five parts: matcher, reasoning rule set, reasoning engine, HTML converter and browser. Matcher provides an algorithm for matching between semantic links or views of semantic links. The reasoning rule set contains the semantic reason rules of SLN. Reasoning can be divided into two types: large-granularity reasoning based on SLN view matching algorithm and small-granularity reasoning based on semantic link reasoning rules. Reasoning rule sets mainly support small granularity reasoning. The HTML converter and browser convert the reasoning result of knowledge graph into web pages for common browsers to display.

With the help of reasoning mechanism, the SLN browser can recommend the entities semantically related to the current part to the user, and enable the user to foresee the content at the end of the semantic link network.

4 Application

As an application, we develop a prototype of academic data analysis based on knowledge graph, and tries to construct and express academic knowledge graph. The design goal of the system is to use the knowledge graph technology, such as knowledge extraction, semantic organization, entity association and relationship reasoning to construct academic knowledge graph. A knowledge graph can be built by standardizing scientific research entities, labeling scientific research entities and expressing multi-dimensional relationships for data resources such as institutions, scholars, documents and projects. It realizes the transformation and upgrading from literature retrieval to knowledge retrieval, and supports semantic retrieval based on scientific entity search and entity association navigation, other than the traditional keyword based retrieval method. The overall architecture of the prototype system is shown in Fig. 2.

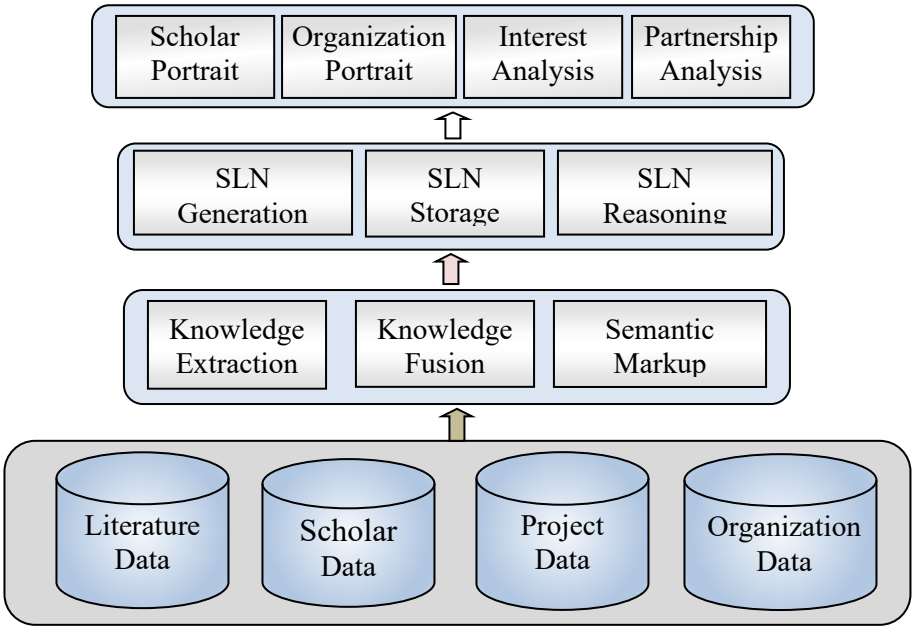

Figure 2: Architecture of the academic knowledge graph system

Academic data mainly includes literature data, personnel information, project information and organization information, which has the characteristics of large volume, multi-source, diversity and dynamic. Literature data includes books, papers, academic reports, patents, etc., which can be obtained from Baidu academic, CNKI, Wanfang and other academic knowledge bases. Project information includes vertical and horizontal project information, participant information, academic team, research scholars, research institutions and other data, which can also be obtained from office documents, or from the network by using web crawler. The data sources were very diverse.

Academic data includes unstructured data, semi-structured data and structured data. For different data sources, knowledge extraction methods are different.

For structured data from relational database, the extraction of entities, relationships and attributes is relatively simple. The query results can be mapped into triples when accessing the database. 1) The row of a table is used as an instance or resource; 2) If the column is a primary key or a foreign key, its value is marked as an entity; 3) If the column is a foreign key, its column name is marked as the relationship between the primary key and the foreign key in the same row; 4) If the column is not a foreign key, its column name is marked as an attribute of the primary key in the same row, and its column value is marked as the attribute value.

For semi-structured or unstructured data, words such as author, title, inventor and keywords are highlighted by parsing the documents or the Web pages. Then the entities, relationships and attributes contained in the page are extracted by manual annotation. For example, there is a "Author" relationship between a paper and the author, there is a "Cooperation" relationship between different authors. At the same time, there can be a "Similar" relationship between papers with the same keywords.

In addition to the traditional types of link, such as cause-effective link, sequential link and reference link, it also adds some special links such as *membership*, *publication*, *reference*, *participation* and *personnel rank*.

Through the knowledge extraction of academic data sources, a large number of entities, relationships and attributes are obtained, such as scholars, papers, projects, patents, organizations and other entities. Semantic link networks can be built by mark semantic tags and links. And, the results are stored. The reasoning mechanism based on semantic links can help to supplement the knowledge graph and show the hidden Associations. The complex semantic link network can be organized and browsed in the way of top-down and layer by layer decomposition. At the same time, it can carry out fuzzy query and extension based on nodes.

Based on the knowledge graph, information is integrated and can provide many intelligent services. Visualization service is one of them. Visualization service modules of scholar portrait, organization portrait, research interest analysis and cooperation relationship analysis are given in the system. It can provide more information about scholars and scientific research institutions, to evaluate their research achievements and introduce the progress and the development trends of scientific research.

1) Scholar portrait module. Scholar portrait refers to the full description of scholars' characteristics on the multi-dimensional analysis. This system mainly describes scholars from five aspects: basic information of scholars, academic papers, vertical projects, horizontal projects, patents. Scholar portrait analyzes and displays scholars in a visual way.

2) Organization portrait module: Organization portrait refers to the comprehensive analysis of the organization. It describes the organization structure, scholar form, research interest analysis, results analysis. It can display the scientific research progress and change trend of the organizations.

3) Interest analysis module: This module shows the research direction and interest of the individual scholars. It can be inferred through the keyword similarity analysis in statistical results of SLN. It can show the evolution process of the research direction and interest.

4) Partnership analysis module: This module shows the collaboration among all scholars in the organization. By counting the number of horizontal projects, vertical projects, papers, patents and other

research results that scholars have participated in, we can judge whether the cooperation between scholars is close or not. it can provide the possibility analysis of future cooperation among scholars, together with of the interest analysis results.

5 Conclusion

Knowledge Graph technology is a large-scale application of knowledge representation and knowledge base in the Internet environment. It shows the importance of knowledge in the intelligent system and is the basic knowledge resource to realize the intelligent system. We explore a method to represent and construct knowledge graphs based on Semantic Link Network. We also develop a prototype system of academic data knowledge graph. More efforts should be put into knowledge graph with semantics.

Funding Statement*:* The authors received no specific funding for this study.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

[1] A. Singhal, "Introducing the Knowledge Graph: things, not strings," Official Google Blog (2012) 1–8. [Online]. Available: https://www.blog.google/products/search/introducing-knowledge-graph-things-not/, 2012.

[2] C. Bizer, T. Heath and T. Berners-Lee, "Linked data-the story so far," *International Journal on Semantic Web & Information System*, vol. 5, no. 3, pp. 1–22, 2009.

[3] S. Auer, C. Bizer, G. Kobilarov, J. Lehmann, R. Cyganiak *et al.,* "Dbpedia: a nucleus for a web of open data," in *Proc. of the 6th Int. Semantic Web and 2nd Asian Conf. on Asian Semantic Web Conf.*, Spinger-Verlag, Berlin, Heidelberg, pp. 722–735, 2007.

[4] F. M. Suchanek, G. Kasneci and G. Weikum, "Yago: a core of semantic knowledge unifying wordnet and wikipedia," in *Proc. of the 16th Int. Conf. on World Wide Web*, New York, NY, USA, pp. 697–706, 2007.

[5] K. Bollacker, C. Evans, P. Paritosh and T. Sturge, "Freebase: A collaboratively created graph database for structuring human knowledge," [Online]. Available: http://ids.snu.ac.kr/w/images/9/98/ SC17.pdf, 2012.

[6] S. Fu, A. Lu and S. Yan, "Concept and application of knowledge graph," *Information and Communications Technology and Policy*, no. 5, pp. 10–13, 2018.

[7] Q. Liu, Y. Li, H. Duan, Y. Liu and Z. Qin, "Knowledge graph construction techniques," *Journal of Computer Research and Development*, vol. 53, no. 3, pp. 582–600, 2016.

[8] T. Hang, J. Feng and J. Lu, "Knowledge graph construction techniques: taxonomy, survey and future directions," *Computer Science*, vol. 48, no. 2, pp. 175–189, 2021.

[9] D. Li, Y. Zhang, D. Li and D. Lin, "Review of entity relation extraction methods," *Journal of Computer Research and Development*, vol. 57, no. 7, pp. 1424–448, 2020.

[10] S. Guan, X. Jin, Y. Wang and X. Cheng, "Shared embedding based neural networks for knowledge graph completion," in *Proc. of the 27th ACM Int. Conf. on Information and Knowledge Management*, New York, NY, USA, pp. 247–256, 2018.

[11] H. Huang, J. Yu, X. Liao and Y. Xi, "Review on knowledge graphs," *Computer Systems and Application*, vol. 28, no. 6, pp. 1–12, 2019.

[12] J. Zhou, X. Sun, X. Yu and X. Bian, "Knowledge graph and data application: intelligent recommendation," *Telecommunications Science*, vol. 35, no. 8, pp. 165–172, 2019.

[13] Z. Du, Z. Du and L. Wang, "Open knowledge graph representation learning based on neighbors and semantic affinity," *Journal of Computer Research and Development*, vol. 52, no. 12, pp. 2549–2561, 2019.

[14] S. Wang, Z. Du and X. Meng, "Research progress of large-scale knowledge graph completion technology," *Scientia Sinica Information*, vol. 50, no. 4, pp. 551–575, 2020.

[15] F. Du, Y. Chen and X. Du, "Survey of RDF query processing techniques," *Journal of Software*, vol. 24, no. 6, pp. 1222–1242, 2013.

[16] K. Kelloumenouer and Z. Kedad, "Discovering types in RDF datasets," in *Proc. of the 14th Int. Semantic Web*

Conf., Bethlehem, PA, USA, pp. 77–81, 2015.

[17] H. Zhuge and R. Jia, "Semantic link network builder and intelligent browser," *Concurrency and Computation: Practice and Experience*, vol. 16, no. 14, pp. 1585–1413, 2004.

[18] H. Zhuge, "Communities and emerging semantics in semantic link network: discovery and learning," *IEEE Transactions on Knowledge and Data Engineering*, vol. 21, no. 6, pp. 785–799, 2009.

[19] H. Zhuge, "Autonomous semantic link networking model for the knowledge grid," *Concurrency & Computation Practice & Experience*, vol. 19, no. 7, pp. 1065–1085, 2010.

[20] Y. Wang, S. Luo, Y. Yang and H. Zhang, "A survey on knowledge graph visualization," *Journal of Computer-Aided Design & Computer Graphics*, vol. 31, no. 10, pp. 1666–1676, 2019.