Unified Workbench for Knowledge Graph Management

Ryutaro Ichise[1,2], Natthawut Kertkeidkachorn[2,1], Lihua Zhao[2,1], and Esrat Farjana Rupu[1]

[1] National Institute of Informatics, Tokyo 101-8430, Japan
{ichise,esrat_farjana}@nii.ac.jp
[2] National Institute of Advanced Industrial Science and Technology, Tokyo 135-0064, Japan
{n.kertkeidkachorn, lihua.zhao}@aist.go.jp

Abstract. Knowledge Graphs become essential knowledge resources for AI applications. Nevertheless, massive real-world knowledge is produced every day. Ignoring new knowledge greatly affects the outcomes of an application due to missing or inaccurate knowledge. To deal with the new knowledge and its life cycle, the holistic framework for curating and manipulating knowledge is needed. In this study, we present our solution, namely Unified Workbench for Knowledge Graph Management (UWKGM), that unifies several technologies to deal with knowledge curation and manipulation in a knowledge graph. Also, We demonstrate some example cases of this framework in use.

1 Introduction

Many AI related tasks, e.g. question and answering systems, entity resolution systems, and information retrieval systems, widely use knowledge graphs (KGs) as knowledge resources. Consequently, KGs are increasingly in demand. However, curating and manipulating KGs requires huge efforts. Based on the analysis on KGs, we found that there are three major problems: 1) Adding New Knowledge, 2) Erroneous Knowledge Injection and 3) Inadequate Knowledge. Adding new knowledge is the most important issue because new knowledge, by nature, is produced every day; in consequence, it is beyond the human effort to deal with. Indeed, the entities and relations contained in KGs are usually dynamically added rather than static. Hence, allowing a new entity or a relation to be added to KGs is required. Erroneous knowledge injection is another problem that degrades the quality of KGs. This problem involves the injection of erroneous knowledge into KG. It is impossible to completely eliminate erroneous knowledge in KGs because erroneous knowledge may be included by accident, e.g. users attempt to create new knowledge or an automated KGs population. Using low-quality KGs may trigger the malfunction on AI applications. Verification and validation on KGs should be considered. Inadequate knowledge is the result of incomplete knowledge resources. It is impossible to prepare all the necessary information in advance. Reasoning necessary information at the time becomes a key to perform actual AI applications on incomplete KGs. Completing missing knowledge is therefore also a key for manipulating KGs.

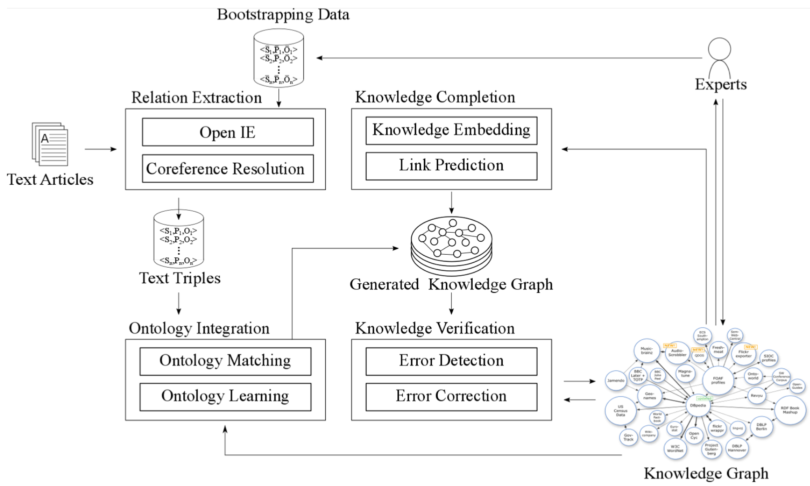

Fig. 1. UWKGM architecture

As discussed those three major problems on KGs above, the holistic framework for curating and manipulating knowledge is needed. In this paper, we therefore introduce our on-going framework, Unified Workbench for Knowledge Graph Management (UWKGM), which integrates many technologies in order to address three problems above. Also, use-cases and achievements of UWKGM are demonstrated.

2 Unified Workbench for Knowledge Graph Management

UWKGM consists of four main components: 1) Relation Extraction (RE), 2) Ontology Integration (OI), 3) Knowledge Verification (KV) and 4) Knowledge Completion (KC), as shown in Fig. 1. RE and OI aim to solve the adding new knowledge problem. KV is to deal with the erroneous knowledge injection problem. KC copes with the inadequate knowledge problem. The details of components and their implementation are as follows.

Relation Extraction is a component to retrieve relationships between entities as triples from unstructured data, particularly text. There are two modules: OpenIE and Coreference Resolution. Open Information Extraction (OpenIE) finds and extracts the relation between two entities as a triplet, while Coreference Resolution unifies surface forms of an entity into one representation. In addition, the framework allows two modules to employ bootstrapping data curated by domain experts in order to enhance results when creating triples. Currently, Stanford OpenIE and Coreference Resolution are implemented by Stanford Open Information Extraction [1]. Our configuration follows T2KG [3].

Ontology Integration is a component to serialize text triples to RDF standard triples and also integrate RDF triples to existing KGs. There are two modules: ontology learning and ontology matching. Ontology learning is to learn and populate (extend) the ontology, while ontology matching is to reuse classes, attributes, relations (T-Box) and individuals (A-Box) in existing KGs. When the ontology used in the KGs is inapplicable, the module is capable of modifying the ontology to accommodate the new triples. This component relies on both T2KG [3] and FITON [6]. The preliminary results of UWKGM on the relation extraction together with the ontology integration was reported in T2KG framework [3]

As presented in T2KG [3], RE and OI can solve the adding new knowledge problem. Concretely, not only non-existing facts in KGs can be discovered but also a new entity and a new property (as a relation) can be populated.

Knowledge Verification is to verify and validate RDF triples before publishing to KGs. Two modules in this component are error detection and error correction. The error detection module is to detect erroneous triples by using constraints in the ontology or analyzing the patterns of triples, such as value range. The error correction module is to correct erroneous triples by finding the most suitable replacement entities. The error detection module utilizes the studies [5, 4], while the error correction module is implemented based on FIXRVE [4]. As shown in the studies [5, 4], the erroneous triples are detected and corrected; in consequence, the erroneous knowledge injection can be avoid and fixed.

Knowledge Completion is to learn the embedding representations and perform inference over the existing KGs to discover missing knowledge. In this component, there are two modules: knowledge embedding and link prediction. the knowledge embedding is to learn entities and relations in KGs as vector representations in the low-dimensional space. Such representations can be used in many applications, e.g. fact-checking. The link prediction module is to predict the missing relationship between entities in the KGs. We implement these two modules by TorusE [2]. In TorusE [2], the missing knowledge is discovered and KGs is enriched; as a result, the inadequate knowledge problem is alleviated.

3 Use Cases and Demonstration

To demonstrate the capability of our UWKGM framework, four main use cases are presented as follows. Also, some demonstration videos and supplementary materials are available at http://ri-www.nii.ac.jp/UWKGM/ .

KG Construction: The first UWKGM use-case focuses on transforming text data extracted from any domain into the KGs. In this use case, RE creates triples from texts. After that, OI integrates those text triples into KGs. A successful example of this specific UWKGM use-case is described in T2KG [3]. To date, we have used T2KG to transform more than 100,000 unstructured text articles into text triples and have integrated those triples into KGs using the ontology learning module.

KG Population: The second UWKGM use-case deals with populating the new knowledge to the KG. There are two minor use-cases: 1) external resource-based population, and 2) internal resource-based population. In external resource-based population, the ontology matching module in OI is used to merge new knowledge to the KG. One example of the external resource-based population is T2KG, in which KG triples are integrated under DBpedia ontology [3]. In an internal resource-based population, KC is employed to predict new knowledge from the KG. A recent specific use-case of this type is TorusE, in which KG triples are used to predict new relationships between entities [2].

KG Revision: The third use-case demonstrate how to alleviate errors by revising the KG. In UWKGM, KV is the key module for handling such use-cases. Currently, a well-performed example of this use-case is FIXRVE [4], in which a pre-defined ontology is used to find and to resolve incorrect triples by using entity profiles.

KG Embedding: The fourth use-case is to learn embedding representations of entities and relations in KGs. In the knowledge embedding module of KC, the embedding representations are learned by the translation-based model, TorusE [2]. We provide RESTful API to retrieve vector representations of the entities and the relations in KGs for using other applications, e.g. link prediction in the study [2].

4 Conclusion

In this paper, we discussed the problem on curating and manipulating KGs. Based on our analysis on the problem, we then proposed our on-going framework, called Unified Workbench for Knowledge Graph Management (UWKGM), and illustrated its use-cases and achievements. We are currently planning to release the framework to the public in the future. Our current demo video is available on http://ri-www.nii.ac.jp/UWKGM/.

Acknowledgment

This work was partially supported by the New Energy and Industrial Technology Development Organization (NEDO).

References

1. Angeli, G., Premkumar, M.J.J., Manning, C.D.: Leveraging linguistic structure for open domain information extraction. In: Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics. pp. 344–354. ACL (2015)
2. Ebisu, T., Ichise, R.: TorusE: Knowledge graph embedding on a lie group. In: Proceedings of the 32nd AAAI Conference on Artificial Intelligence. AAAI (2018)
3. Kertkeidkachorn, N., Ichise, R.: An automatic knowledge graph creation framework from natural language text. IEICE Transactions on Information and Systems **101-D**(1), 90–98 (2018)
4. Lertvittayakumjorn, P., Kertkeidkachorn, N., Ichise, R.: Resolving range violations in DBpedia. In: Proceedings of the 7th Joint International Semantic Technology Conference. pp. 121–137. Springer (2017)
5. Rahoman, M.M., Ichise, R.: Automatic erroneous data detection over type-annotated linked data. IEICE Transactions on Information and Systems **E99-D**(4), 969–978 (2016)
6. Zhao, L., Ichise, R.: Ontology integration for linked data. Journal on Data Semantics **3**(4), 237–254 (2014)