PRASEMap: A Probabilistic Reasoning and Semantic Embedding based Knowledge Graph Alignment System

Zhiyuan Qi∗
Tencent Jarvis Lab
Shenzhen, China
qizhyuan@gmail.com

Ziheng Zhang∗
Tencent Jarvis Lab
Shenzhen, China
zihengzhang@tencent.com

Xi Chen
Tencent Platform and Content Group
Shenzhen, China
jasonxchen@tencent.com

Jiaoyan Chen
University of Oxford
Oxford, UK
jiaoyan.chen@cs.ox.ac.uk

Yefeng Zheng∗
Tencent Jarvis Lab
Shenzhen, China
yefengzheng@tencent.com

1 INTRODUCTION

Knowledge Graph (KG) alignment attempts to discover equivalent entities, relations, and attributes (i.e., mappings) between two KGs. The existing approaches utilize either reasoning-based or semantic embedding-based techniques, but few studies explore their combination. In this demonstration, we present PRASEMap, an unsupervised KG alignment system that iteratively computes the Mappings with both Probabilistic Reasoning (PR) and Semantic Embedding (SE) techniques. PRASEMap can support various embedding-based KG alignment approaches as the SE module, and it also enables easy human computer interaction that additionally provides an option for users to feed the mapping annotations back to the system for better results. The demonstration showcases these features via a stand-alone Web application with user friendly interfaces. The demo is available at https://prasemap.qizhy.com.

CCS CONCEPTS
• Computing methodologies → Knowledge representation and reasoning.

KEYWORDS
knowledge graph, knowledge graph alignment, probability reasoning, embedding

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

ACM Reference Format:
Zhiyuan Qi, Ziheng Zhang, Jiaoyan Chen, Xi Chen, and Yefeng Zheng. 2021. PRASEMap: A Probabilistic Reasoning and Semantic Embedding based Knowledge Graph Alignment System. In Proceedings of the 30th ACM International Conference on Information and Knowledge Management (CIKM '21), November 1–5, 2021, Virtual Event, QLD, Australia. ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3459637.3481972

Note that the whole system can be performed in two modes, i.e., an automatic mode without any seed mappings or human intervention and a semi-automatic mode accepting the human annotated mappings for augmentation with an annotation interface specifically designed to identify entity mappings. To summarize, the Web application PRASEMap allows users to test the widely-used
2 SYSTEM DESIGN

We consider the knowledge graph alignment that aims at discovering the entity mappings, i.e., equivalent entities referring to the same real-world object, between two KGs $\mathcal{KG}_1$ and $\mathcal{KG}_2$. This section first presents the architecture and pipeline of PRASEMap and then briefly introduces the two core modules — Probabilistic Reasoning (PR) and Semantic Embedding (SE).

2.1 Architecture and Pipeline

As shown in Figure 1, PRASEMap consists of the data preprocessing module (denoted as PRE module), the PR module, and the SE module. These modules work in an iterative pipeline whose steps are described as follow.

[Step 1. Initialization] The PRE module parses the two KG files that are either pre-stored in the system or uploaded by the user, and constructs two KG objects that contain standardized relations and attributes. Note that for cross-lingual KGs, the PRE module can call the translation API to translate the text into the same language for subsequent PR and SE modules.

[Step 2. Running PR Module] The PR module receives the KG objects, and iteratively discovers cross-KG literal mappings, entity mappings, and relation mappings. After several self-iterations, it outputs the entity mappings (cf. Section 2.2).

[Step 3. Running SE Module] The SE module uses the entity mappings from the PR module as seeds for training, and calculates entity embeddings and more entity mappings as the output (cf. Section 2.3).

[Step 4. Rerunning PR Module] The PR module is performed again with the entity embeddings and mappings from the SE module. Different from Step 2, entity embeddings are utilized to enable more accurate probabilistic reasoning, and the received entity mappings are utilized to disclose more potential mappings.

Given a task, Step 1 and Step 2 are executed once, while Step 3 and Step 4 can be repeated for several times (which can be configured by the user) to fully explore the benefits of the two modules. This is the automatic mode of PRASEMap relying on no seed mapping or human interaction. In the semi-automatic mode, the PR module selects and presents some uncertain mappings to the user and receives the manual annotations, based on which it further discovers more mappings and deletes some wrong mappings.

2.2 Probabilistic Reasoning Module

A famous reasoning-based alignment system termed PARIS [5] is adopted and upgraded as the PR module. PARIS was originally implemented in Java. To be compatible with many embedding-based KG alignment algorithms implemented in Python, we re-implement PARIS using C++ and create its Python bindings using the PyBind11 library, which makes PRASEMap much more effective than our original PRASE that is implemented in Python and evaluated in [4].

Briefly, the original PARIS first computes the functionality of the KG relations and attributes. Given a head entity, the functionality can be used to determine whether two tail entities or literals across two KGs are identical. Next, PARIS discovers initial literal mappings and entity mappings by lexical matching, and further iteratively updates the probabilities of these mappings as well as the subsumption relationships between relations and attributes by reasoning with the functionality. PARIS finally outputs entity mappings along with other mappings after several iterations when the mappings converge. Please refer to [5] for more details.

In Step 2, the PR module uses the original PARIS while in Step 4, the PR module uses the upgraded PARIS that has several new features. First, it can exploit the semantic embeddings output by the SE module to adjust the estimated probability of entity mappings during its reasoning process, so as to improve the recall and the precision of the resultant mappings. Specifically, the new formulation of computing the equivalent probability of two entities $e_1 \in \mathcal{KG}_1$ and $e_2 \in \mathcal{KG}_2$ is given by $Pr_{new}(e_1 \equiv e_2) := \alpha Pr_{old}(e_1 \equiv e_2) + (1 - \alpha) \text{sim}(e_1, e_2)$ with $\alpha \in (0, 1)$, where the first term corresponds to the original probability estimation formulation of PARIS; the second term measures the (cosine) similarity of entity embeddings used to adjust $Pr_{old}$. We use the Eigen library to manipulate embeddings in the process of reasoning. Second, the upgraded PARIS directly adopts those high-quality entity mappings calculated by the SE module into its initial entity mapping set for further exploring more potential mappings during its self-iterations. Please refer to [4] for detailed explanations of the PR module.

Considering the difficulty in aligning two complex KGs, manual labeling often plays an important role in the real-world applications [16]. To further leverage manual labeling, the PR module enables optional human interaction through which those uncertain mappings are sent to users for feedback. Specifically, the PR module randomly selects a small number of the mappings with low scores for labeling. The users can mark a score of 0 (i.e., false) or 1 (i.e., true) for each entity mapping. Note that PRASEMap regards the labels from user feedback as ground truth and will not update these labeled mappings’ scores during the subsequent iterations.

---

1. [http://webdam.inria.fr/paris/](http://webdam.inria.fr/paris/)
This section provides a demonstration towards PRASEMap and a brief evaluation of the human annotation during the iterations.

3 DEMONSTRATION

3.1 Web Application

For better human-computer interaction, we construct a stand-alone Web application with multiple key features and visualization interfaces. Specifically, the back end is developed based on SpringBoot, Mybatis, and Redis. The front end is developed using Vue.js and antv/g6 — a graph visualization engine. The system itself stores some widely-used KG alignment benchmarks from popular general-purpose KGs, such as DBpedia and Wikidata, as well as an industry benchmark from Tencent’s medical KGs [10]. The Web application also allows users to upload their own KG files for KG alignment and visualization. The resultant mappings can be downloaded. We next introduce three important interfaces and their usage.

[Task Running and Mapping Visualization] The user registers at the Web application with his/her email, and then receives a unique task ID for his/her own KG alignment task on the Web application. After login at the Web application, the user submits a KG alignment task, and the server enqueues the task on a work queue to maintain the workload. When PRASEMap starts to run for a task, some logging information is presented for the user to monitor the running process including the running stage, the time cost, the training loss, and the number of discovered mappings. To illustrate the alignment result, the Web application can visualize each mapping, where the corresponding one-hop neighbourhoods of the two associated entities of the mapping are presented. With the neighbourhoods, the user can better understand why entities are matched, and meanwhile the Web application can present more mappings between their neighbouring entities, which can further justify why the two entities are matched. Figure 4 shows the visualization of an example mapping between “Florin_Tene” of DBpedia-EN (blue-circled entities) and “Florin_Tene” of DBpedia-DE (pink-circled entities), where the blue/pink lines represent the relations within each single KG, and the orange lines represent all the entity mappings. Furthermore, the user can click any individual entity with the mouse to check the corresponding attributes that are also helpful to explain a mapping.

[Manual Labeling] The Web application enables the user to return their feedback to PRASEMap for improvement. PRASEMap selects a number of entity mappings with low confidence and presents them to the user in a table, named as User Feedback Form in Figure 3. For...
3.2 Evaluation of Human Annotation

The performance of the automatic mode of PRASEMap has been verified in the performance evaluation of PRASE [4]. This work complements the evaluation on its semi-automatic mode, i.e., the effectiveness of human annotation feedback.

We present the results of PRASEMap on Tencent’s industry benchmark MED-BBK-9K [10] using MultiKE as the SE module, with 10, 50 and 100 human annotations fed back to iteration. Note that we use undiscovered ground truth mappings as human feedback for convenience in evaluation. The results are shown in Table 1, where P, R and F1 denote precision, recall, and F1-score, respectively. It can be observed that, compared with PARIS, both the precision and the recall of PRASEMap significantly increase, leading to a higher F1-score. Besides, the performance of PRASEMap benefits from the human labeling as the P/R/F1 increases as the number of annotations increases. This finding also sheds light on the importance and effectiveness of human interaction in KG alignment. Please refer to [4] for more comprehensive evaluation about PRASEMap.

![Figure 3: The Interface for Manual Mapping Labeling.](image)

![Figure 4: Neighbourhood Visualization.](image)

Table 1: Experimental results on MED-BBK-9K [10].

| Model          | P (%) | R (%) | F1 (%) |
|----------------|-------|-------|--------|
| PARIS          | 80.588| 37.110| 50.818 |
| PRASEMap       | 83.833| 61.973| 71.265 |
| PRASEMap w/ 10 Annotations | 84.186| 62.115| 71.486 |
| PRASEMap w/ 50 Annotations  | 84.402| 62.606| 71.889 |
| PRASEMap w/ 100 Annotations | 84.760| 63.130| 72.363 |

4 CONCLUSION AND DISCUSSION

In this demonstration, we present a KG alignment system termed PRASEMap that leverages both probabilistic reasoning and semantic embedding for the automatic mapping discovery. The system is implemented with a Web application with friendly input, output, and visualization interfaces as well as different settings for fully configurable PRASEMap. It also supports human-computer interaction that utilizes human feedback for better results.

The Tencent team is actively developing new features, such as involvement of ontology [8], and meanwhile maintaining PRASEMap for higher reliability, better performance, and industry deployment. For future work, we will utilize PRASEMap as a primary component in the construction of large-scale medical KGS. Moreover, the manual labeling interface in PRASEMap will be integrated into the annotation platform to assist the human annotators and improve their productivity in various tasks beyond mapping annotation, such as KG correction and entity canonicalization [1]. We also plan to explore more learning-based strategies for the selection of undiscovered mappings to reduce the annotation workload.

ACKNOWLEDGMENTS

This work was funded by National Key R&D Program of China (2018YFC000702). We also acknowledged the kind assistance from Shengzhang Lai and Chun Zhang in developing and maintaining the Web application.
REFERENCES

[1] Jiaoyan Chen, Xi Chen, Ian Horrocks, Erik B. Myklebust, and Ernesto Jimenez-Ruiz. 2020. Correcting Knowledge Base Assertions. In Proceedings of The Web Conference 2020 (Taipei, Taiwan) (WWW ’20). Association for Computing Machinery, New York, NY, USA, 1537–1547. https://doi.org/10.1145/3366423.3380226

[2] Muhao Chen, Yingtao Tian, Mohan Yang, and Carlo Zaniolo. 2017. Multilingual Knowledge Graph Embeddings for Cross-lingual Knowledge Alignment. In Proceedings of the 26th International Joint Conference on Artificial Intelligence. Melbourne, Australia, 1511–1517.

[3] Ernesto Jiménez-Ruiz and Bernardo Cuenca Grau. 2011. LogMap: Logic-based and Scalable Ontology Matching. In Proceedings of the 10th International Semantic Web Conference. Bonn, Germany, 273–288.

[4] Zhiyuan Qi, Zheng Zhang, Jiaoyan Chen, Xi Chen, Yuejia Xiang, Ningyu Zhang, and Yefeng Zheng. 2021. Unsupervised Knowledge Graph Alignment by Probabilistic Reasoning and Semantic Embedding. In Proceedings of the 30th International Joint Conference on Artificial Intelligence. Montreal, Canada.

[5] Fabian M. Suchanek, Serge Abiteboul, and Pierre Senellart. 2012. PARIS: Probabilistic Alignment of Relations, Instances, and Schema. In Proceedings of the 38th International Conference on Very Large Databases. Istanbul, Turkey, 157–168.

[6] Zequn Sun, Qingheng Zhang, Wei Hu, Chengming Wang, Muhao Chen, Farahnaz Akrami, and Chengkai Li. 2020. A Benchmarking Study of Embedding-based Entity Alignment for Knowledge Graphs. In Proceedings of the 46th International Conference on Very Large Databases. Tokyo, Japan, 2326–2349.

[7] Zhichun Wang, Qingsong Lv, Xiaohan Lan, and Yu Zhang. 2018. Cross-lingual Knowledge Graph Alignment via Graph Convolutional Networks. In Proceedings of the Conference on Empirical Methods in Natural Language Processing. Brussels, Belgium, 349–357.

[8] Yuejia Xiang, Zheng Zhang, Jiaoyan Chen, Xi Chen, Zhenxi Lin, and Yefeng Zheng. 2021. OntoEA: Ontology-guided Entity Alignment via Joint Knowledge Graph Embedding. In Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021. Association for Computational Linguistics, Online, 1117–1128. https://doi.org/10.18653/v1/2021.findings-acl.96

[9] Qingheng Zhang, Zequn Sun, Wei Hu, Muhao Chen, Lingbing Guo, and Yuzhong Qu. 2019. Multi-view Knowledge Graph Embedding for Entity Alignment. In Proceedings of the 28th International Joint Conference on Artificial Intelligence. Macao, China, 5429–5435.

[10] Zheng Zhang, Jiaoyan Chen, Xi Chen, Hualuo Liu, Yuejia Xiang, Bo Liu, and Yefeng Zheng. 2020. An Industry Evaluation of Embedding-based Entity Alignment. In Proceedings of the 28th International Conference on Computational Linguistics. Barcelona, Spain, 179–189.