Constructing Low-Redundant and High-Accuracy Knowledge Graphs for Education

Wentao Li, Huachi Zhou, Junnan Dong, Qinggang Zhang, Qing Li, George Baciu, Jiannong Cao, and Xiao Huang

Department of Computing, Hong Kong Polytechnic University, Hong Kong
{wentao0406.li,huachi.zhou}@connect.polyu.hk
{hanson.dong,qinggangg.zhang}@connect.polyu.hk
{qing-prof.li,csgeorge,jiannong.cao,xiao.huang}@polyu.edu.hk

Abstract. Motivated by the successful applications of commonsense knowledge graphs (KGs) and encyclopedia KGs, many KG-based applications have been developed in education, such as course content visualization and learning path/material recommendations. While KGs for education are often constructed manually, attempts have been made to leverage machine learning algorithms to extract triples from teaching materials. However, education-related KGs learned by existing algorithms contain significant amounts of redundancy and noise. It is because the entities and relations in teaching materials are often instructional, abstract, and implicit, while textbooks often contain detailed explanations, examples, and illustrations. Off-the-shelf KG construction algorithms are designed for concrete entities. To this end, we propose an effective framework to construct low-redundant and high-accuracy KGs for education. First, we design an ontology that is tailored for education. By choosing related Wikidata items, we construct an instructional entity set. We avoid using traditional methods such as named-entity recognition to extract entities from textbooks, aiming to reduce redundancy. Then, we add subtopic relations among our selected instructional entities based on the corresponding hierarchy in Wikidata, and form a backbone. Second, we design a machine reading comprehension model with pre-defined questions to extract other types of relations, such as equivalent to, applied to, and inventor of. Third, we apply active KG error detection to further refine the KG with minimal human effort. In the experiments, we take the artificial intelligence domain as an example and demonstrate the effectiveness of the proposed framework. Our KG achieves an accuracy of around 80% scored by domain experts.

Keywords: Educational knowledge graphs · Knowledge graph construction.

1 Introduction

Many encyclopedia knowledge graphs (KGs) [15], commonsense KGs [5], and KGs for medical science[1] have been developed, with a wide range of applica-

[1] https://bioportal.bioontology.org

tions such as search and recommendations [26]. Motivated by this, KG-based applications have been developed in education, such as course content visualization [13, 18], learning path/material recommendations [20], and university course management [3]. For example, course content KGs could illustrate concepts and knowledge in a hierarchical and systematic way [18].

While KGs for education are often constructed manually [20, 18], a few attempts have been made to leverage machine learning algorithms to extract triples from teaching materials [6]. Studies on machine-learning-based KG construction for education could be divided into two classes. The former one aims to construct KGs to represent course concepts [7]. The latter targets to learn KGs to organize multimedia learning resources [11, 22]. Existing studies usually employ query languages and web scrapes to construct KGs from structured data sources such as knowledge bases and HTML web pages. As for unstructured data sources like textbooks, traditional pipelines containing named entity recognition (NER) and relation extraction (RE) modules are a common solution for most KG construction tasks [4, 2].

However, educational KGs developed by existing machine-learning-based algorithms often contain significant numbers of redundant relations, and erroneous triples. These issues reduce their usability in real-world teaching and learning scenarios. First, *low-accuracy*. Different from real-world or common-sense KGs, entities and relations in KGs for education are more abstract and hard to represent and extract. Even for human readers, it is difficult to distinguish the exact borders of scientific terms compared with a person's or organization's name. Inconsistent spellings in domain papers and textbooks, further interrogated the problems of entity recognition. While many existing pipelines finish this work in a sequential manner, the errors produced by the NER stage will be magnified in the RE stage, leading to noisy outputs. In addition, it is often quite difficult to infer from the text description the relationships between educational topics, such as their hierarchy. This can result in noisy or inaccurate relations in the final KGs. Second, *redundancy*. Existing methods often try to extract as many triples as possible from available data sources. While frameworks generate a large number of triples with a certain level of redundancy, the KG refinement module is often an ignored part of previous research. Without an effective way to filter the results, a large portion of the information in KGs is trivial for educational purposes. Students need to spare extra effort to distinguish valuable information when exploring the KGs. Furthermore, it remains difficult to filter out incorrect or redundant triples in KGs, given that annotations are expensive.

To solve these problems, we propose to leverage reliable resources to construct KGs for education. Starting from the initial entity set, we extend the KGs with reliable data sources like Computer Science Ontology [19] and Wikidata [24]. Then our framework takes the entities and relations from the first step as the concise backbone and then effectively expands the KG with unstructured data sources, i.e., textbooks and Wikipedia, without introducing erroneous triples into KGs. Specifically, we designed a targeted machine reading comprehension (MRC) method to extract concepts with pre-defined question patterns. By focusing on particular relations in each step, e.g., *subtopic_of*, we automatically generate questions like 'What are subtopics of Natural Language Processing?' and extract credible answers, i.e., candidate tail concepts of the topic entity *Natural_Language_Processing* with respect to relation *subtopic_of*. We avoid employing traditional methods like NER, which will lead to noise in the final KG. After that, we filter the noisy triples by applying active error detection methods. By effectively utilizing the backbone of KG and tailored ontology, our framework achieve ideal accuracy and redundancy on the final output.

To this end, we propose a novel course KG construction framework for education guided by a standard ontology. Specifically, entity attributes and relations are well defined as criteria for computer science to avoid ambiguity. Along with the protective confines, we effectively learn from concisely structured data for building a KG backbone, as well as the abundant text corpus as unstructured data to expand it. Moreover, we apply a tailored KG error detection method with minimal human effort to further actively refine the final output by singling out suspicious triples. Extensive experiments and sufficient visualization are included in this paper to show the robustness of our constructed KG. Our contributions are summarized as follows.

- We propose an effective framework to construct low-redundant and high-accuracy KGs for education.
- To reduce the redundancy, we learn a backbone based on related Wikidata items and hierarchy, and avoid using named-entity recognition.
- To improve the accuracy, we design a machine reading comprehension task with pre-defined questions to extract relations.
- We take the artificial intelligence domain as an example and empirically evaluate the effectiveness of the proposed framework.

2 Methodology

Now we introduce the pipeline of the proposed framework as shown in Figure 1, which is composed of four components. (i) The first component is ontology design, which is tailored for the education in the artificial intelligence domain. (ii) The second component aims to build the backbone of KG from reliable structured data, including hierarchical relationships. (iii) The third component designs MRC tasks and leverages the pre-trained language model to extract triples from massive real-world unstructured resources. (iv) In the fourth component, we finalize the KG by adopting an active learning based KG refinement model to remove redundant triples. In each component, we invite several AI-oriented experts to apply human inspection to the acquired triples to ensure the accuracy of the KGs for education.

2.1 Ontology Design

The objective of KGs construction is to extract a set of triples (h, r, t), composed of a head entity h, a relation r, and a tail entity t from external resources. We

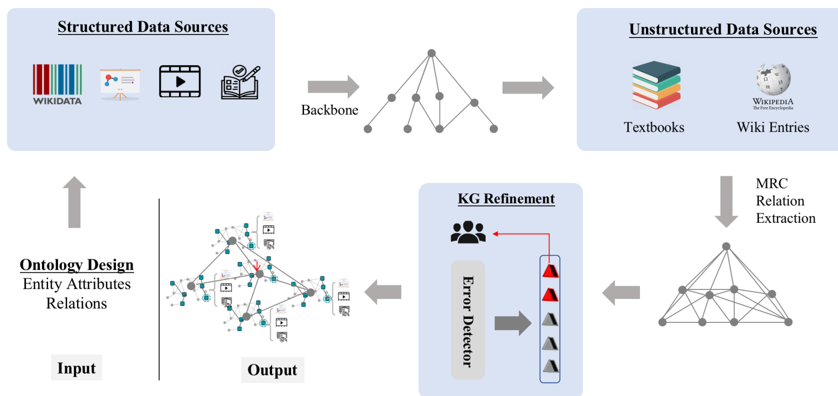

Fig. 1. A framework for constructing high-accuracy and low-redundancy KGs.

believe that the ontology design should be in line with the motivation facilitating the minimization of errors and redundancy in the KGs. To reduce potential errors and redundancy, we design the ontology with the capability to uniquely identify each entity. Each entity is associated with six attributes: Type, Description, Wikidata ID, Wikipedia Link, Tutorial Videos and Books. The definitions and detailed explanations of entities and relations are shown in Table 1.

Entity Attributes Definition	
Type	Topics or people
Description	Short texts explain the entity
Wikidata ID	Wikidata knowledge ID (if any)
Wikipedia Link	Link to Wikipedia (if any)
Tutorial Videos	Link to recommended videos (if any)
Tutorial Books	Link to recommended books (if any)
Relation types Definition	
Subtopic_of	A is a subtopic of B
Equivalent_to	A is an alias of B
Applied_to	A is applied to B
Invented_by	A is proposed/invented by B

Table 1. The definitions of entities and relations in KGs.

2.2 Backbone with Entity Set and Hierarchical Relations

After defining the entities and relations, we focus on the backbone construction of the target KG. Most real-world available knowledge could be roughly categorized into *structured* and *unstructured* forms. Empirically, structured data have been preprocessed by labor engineering while in a format directly accessible by machines, so the extracted triples are less prone to noise. It could satisfy our requirements for accuracy and low redundancy compared with those extracted from unstructured data. Therefore, we will shed light on how we start with reliable structured data to obtain initial triples first.

Instructional Entity Set and Hierarchical Relations The backbone of KG establishes connections between a bunch of relevant core concepts in the specific domain, which is able to picture a concise and organized taxonomy for the domain. As we are constructing KG for abstract instructional entities, it is essential to first obtain a set of high-quality entities and relations as a backbone at the beginning to assist later construction processes. In this step, we focus on subtopic relations between the backbone entities set as an essential basis. To build the hierarchical backbone from the top, we collect high-level entities and select them from textbook glossaries, Wikidata, and the Computer Science Ontology, a large-scale ontology for taxonomy in computer science generated by the Klink-2 algorithm [16].

To add low-level entities to the backbone, we continue focusing on the predefined relation subtopic and refer to a large open data source Wikidata, a structured version of Wikipedia to collect instructional entities. Wikidata allows many editors to collaboratively update entries and store it in a structured manner. Therefore, the extracted knowledge has relatively higher confidence in maintaining a high-level accuracy and low-redundancy. We assign the entity name in the obtained backbone of KG as the subject of the query and use SPARQL to retrieve entities that satisfy the subtopic relation. Only the valuable triples matching the ontology design are left to conduct entity alignment and linked to the original backbone KG. The combination of extracted triples between CSO[19] and Wikidata the two reliable resources completes the backbone structure construction and the probability of introduced errors will be further reduced.

Multimedia Entity Attributes We enrich the backbone by adding information to the entity's attributes, such as the video and book links. The linked e-learning resources provide comprehensive, relevant tutorial videos and teaching books for the target entity. They are beneficial for students to understand the background knowledge when exploring the KGs. The resource selection combines both domain experts' recommendations and search engine querying. For search engine querying, we define the main search keyword as the entity's name in the obtained backbone KGs. To prevent incorporating unrelated resources with the same name, we define sophisticated rules to filter erroneous resources. The ranking of candidate videos and books takes the total number of views and average ratings into consideration. Finally, we select top-1 high-quality tutorial videos and books as the target entity attributes.

2.3 Completing Knowledge Graph with Unstructured Data

After the KG backbone construction, we aim to extend the scale of KG by adding more relations from unstructured data such as texts and web pages. Currently, there are two mainstream methods to adopt. One is based on the open-source IE packages, and the other utilizes fine-tuned machine reading comprehension (MRC) tasks. The most popular models in IE packages are designed for specific relation extractions and often require a large amount of training corpus for a new target domain. The training data often requires expensive engineering expertise to train an applicable model. To be free of this labor, we pay attention to how we extract triples based on fine-tuned MRC tasks.

We consider two text sources: Wikipedia entries, and classical textbooks, which are relatively reliable and high-quality. The core task is to add multiple trustworthy relations, except for subtopics from different publications and open-source knowledge bases to the KG backbone.

Relation	Question	Answer
Applied_to	Where is e1 applied to?	e2
Subtopic_of	What is the subtopic of e3?	e4
Inventor_of	Who is the inventor of e5?	e6

Table 2. Question and answer examples used in MRC.

Relation Extraction as Machine Reading Comprehension Task The general relation extraction procedure can be divided into two key stages: entity recognition (NER) and relation extraction (RE). In the NER stage, to push the model to identify the specialized terms in the sentences, high volume annotated data are necessary to train deep learning models while simple rule-based extraction methods have insufficient generalization capacity to extract meaningful facts. Motivated by the aforementioned problems, we propose a novel deep learning based approach by virtue of pre-trained models to identify new entities.

As many studies point out, large-scale pre-trained language models already contain a specific prerequisite knowledge about low-level semantics and factoid commonsense to produce high-quality results on many downstream tasks [17]. Pre-trained language models learn to extract generic features with unsupervised learning from large-scale unlabeled data. Therefore, we only need to fine-tune it with a few labeled data without substantial architecture modifications and training corpus. We apply it to automatically identify entities and extract relations from texts by creating MRC tasks. Given a clean text and questions with a well-defined format in advance, MRC tasks push the chosen model to retrieve answers from the texts [12].

In this way, we can transform the relation extraction task into finding the answer to a formulated question from a given text. The question input to the model is related to a particular relation and head/tail entity. Then the returned answer will be the text span of the most likely tail/head entity in the given text.

The whole method consists of the following steps as shown in the Algorithm 1. First, we use several manually defined template questions tailored to extract triples for specific relations. The example questions are shown in Table 2.3. The placeholder will be replaced with the entity's name. Then, the long texts are segmented according to the maximum input length of the MRC model, and overlaps between segments are required to prevent splitting the possible answer span into two segments. For each text from segmented texts, we then feed each generated question in order along with text to the MRC model to get multiple candidate answers, which are sorted by the confidence scores. Last, we keep only one answer with the highest confidence scores, which represents the wanted tail entity.

Algorithm 1: MRC for Relation Extraction

Input: $inputText$; $questionTemplate$; $headEntity$;
Output: $tailEntity$
initialization;
$textList \leftarrow splitLongText(inputText, maxlength)$;
$questionList \leftarrow generateQuestion(headEntity, questionTemplate)$;
for t *in* $textList$ **do**
  for q *in* $questionList$ **do**
    $ans, ansScore \leftarrow MRCmodel(t, q)$;
    if $ansScore \geq scoreThreshold$ **then**
      $ansList = ansList + (ans, ansScore)$
$ansList.SortbyScore()$;
$tailEntity \leftarrow ansList.Pop()$;
return $tailEntity$

Next, we improve the quality of the returned triples by excluding results with low confidence scores and those not meeting the requirements (e.g., too long word length or containing special symbols). The answer spans may consist of juxtaposed nouns, which include multiple potential entities. We further truncate this juxtaposed noun into several entities. Finally, the extracted head/tail entities and relations are linked to the KG backbone or knowledge bases obtained in the previous step.

2.4 Active Learning Based Knowledge Graph Refinement

In this section, we introduce a tailored KG refinement model to automatically single out suspicious and redundant triples. Early studies of KG refinement mainly rely on rule-based methods. Typically, AMIE [10] introduces a rule mining model under the Open World Assumption and uses an altered metric to measure the degree of each data instance being true, and AMIE+ [9] extends this method to large-scale KGs. Nonetheless, they are all limited by the difficulty of obtaining sufficient and correct rules. Given correct rules, those methods can

Algorithm 2: An Active Learning Strategy for KG Refinement

Input: a KG for pre-training $\mathcal{G}_p$ and a noisy KG for predicting $\mathcal{G}_t$
Output: the noisy set $\mathcal{S}_n$ and the clean set $\mathcal{S}_c$

$\mathcal{T}_l, \mathcal{T}_u \leftarrow \texttt{SPLIT}(\mathcal{G}_p, \tau)$
$\mathcal{S}_n \leftarrow \varnothing$; $\mathcal{S}_c \leftarrow \varnothing$
train a detector $\mathcal{D}$ based on the annotated triple set $\mathcal{T}_l$
$N \leftarrow$ the maximum number of iterations
$m \leftarrow 0$
while $m < N$ **do**
  $T \leftarrow \texttt{queryStrategy}(\mathcal{T}_u)$
  get labels of T
  $\mathcal{T}_l' \leftarrow \mathcal{T}_l \cup T$; $\mathcal{T}_u' \leftarrow \mathcal{T}_u - T$
  retrain the detector $\mathcal{D}$ based on $\mathcal{T}_l'$
  $m \leftarrow m + 1$
foreach $t \in \mathcal{G}_t$ **do**
  $\hat{l} \leftarrow$ predict the label of t according to $\mathcal{D}$
  if $\hat{l} == -1$ **then** //noisy set
    $\mathcal{S}_n \leftarrow$ add t into $\mathcal{S}_n$
  else //clean set
    $\mathcal{S}_c \leftarrow$ add t into $\mathcal{S}_c$
return $\mathcal{S}_n$ *and* $\mathcal{S}_c$

spot erroneous triples that violate the rules, but if they are not able to detect errors that are not included in these rules, then some potential errors with complex patterns may escape detection.

Thus, we tend to leverage external labeled data to learn an effective error detector that classifies triples in KGs into *noisy* and *clean* sets. Large-scale labeled data could theoretically boost the detection ability of the proposed detector. However, gathering plenty of labeled data requires a lot of manual work. In order to decrease the number of labels while maintaining the detection model's reliability, we integrate active learning (AL) techniques into the training process.

The pseudocode of our proposed AL-based error detection method, i.e. ALED, is presented in Algorithm 2. Concretely, we first divide the pretraining KG $\mathcal{G}_p$ into a sizable unlabeled triple set $\mathcal{T}_u$ and a tiny labeled set $\mathcal{T}_l$. Then, using the labeled data $\mathcal{T}_l$, it trains a detector $\mathcal{D}$. In order to get desirable detection performance, ALED iteratively chooses a set T of data from the unlabeled pool U to retrain the detector $\mathcal{D}$, using a different query method for each iteration. A well-trained detector $\mathcal{D}$ is created after N iterations. Next, ALED employs the detector $\mathcal{D}$ to anticipate the label of each triple t in the noisy KG $\mathcal{G}_t$. If t is erroneous, it will be classified into *noisy* set, i.e. $\mathcal{S}_n$. Otherwise, it will be added into the *clean* set, i.e. $\mathcal{S}_c$.

3 Experiments

3.1 Human Evaluation and Scoring

To verify the quality of the KG, we first randomly sampled 50 triples from the KG and employed 3 experts and PhD students in the AI domain to score the results, with 0 being considered incorrect and 1 being considered correct. Our KG received a accuracy rate from the three scorers by 84%, 78%, and 82%, respectively. The Table 3 shows the comparison results for KG generated by different methods, especially rule-based methods [21] and machine learning methods [25], with the same inputs. Therefore, we can draw the conclusion that, when compared to other existing construction methods, our framework is able to produce results with higher accuracy.

Table 3. Evaluation Scores of KGs by different construction methods.

	Our KG	KG by Rules	KG by ML Models
Annotator 1	84%	24%	24%
Annotator 2	78%	16%	40%
Annotator 3	82%	36%	52%

We then examined the coverage of domains in our KG by hiring 3 PhD students to score the taxonomic relationships and comprehensiveness based on a set of expert-annotated hierarchical relations, which contains 135 key domain relationships as a gold standard. The results show that our KG has 56 perfect matches with the expert annotated KG and 46 half matches (only matches with head entity or tail entity) within 135 key domain relationships, which guarantees satisfactory coverage in the computer science domain.

3.2 Visualization and Case Study

As we designed this KG with the ability to visualize connections between concepts, we then visualized a part of the results of our KG for a case study. As shown in Figure 2, a visualization example for partial relations in KG is provided. When choosing artificial intelligence as the focused topic of the graph, our KG clearly pictures its relationship with other subtopics. It also pinpoints the relationship between the application of one topic to other domains. With the educational ontology design, the KG does not contain too much redundant information, and the visualization of the connections between entities is more intuitive. E-learning resources are included in attributes, such as recommended textbooks, which are very helpful for students to understand or learn new concepts. For clearer illustration, we further present the local structure of one subgraph in Figure 3 centered on one subtopic.

With our KGs, students can discover potential connections between concepts through direct interaction or search queries, which is often difficult to accomplish

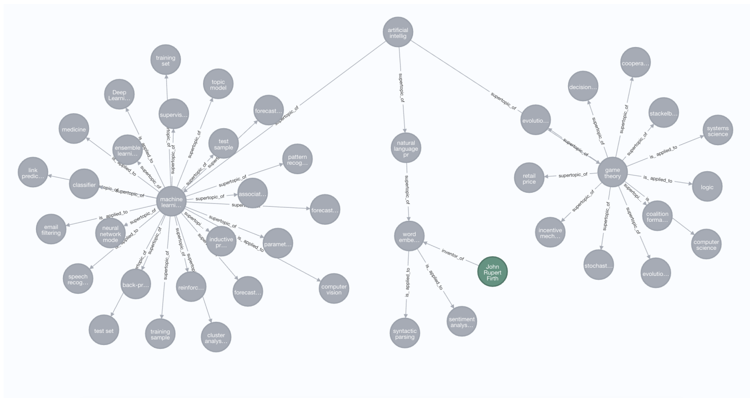

Fig. 2. Overall visualization of KG for certain subtopics.

Fig. 3. Local structure of one subgraph.

through literal reading and traditional classroom lecturing. Compared to many large KGs, it maintains the clarity of taxonomy without providing overly complex information for beginners. It concentrates on the necessary relationships and other learning resources.

4 Related Work

Although high-quality KGs for education purposes often require significant amounts of effort from experts, considerable research has focused on constructing KGs with deep learning methods. We divide the previous research work into two classes as follows [1].

Course concept knowledge graphs. This line of research aims to extract educational concepts from textbooks and other resources to represent relations between different concepts in courses. This provides visualization for abstract concepts in a domain, which helps students to have a better understanding of certain domains and establish clearer connections. Dessì et al. extracted large-scale KGs from published papers in the CS domain with relation extraction models but without emphasis on educational purposes [8]. Qin et al. studied how KGs can be used in teaching a specific subject in computer science [18]. They used web crawlers to obtain raw data from websites and then extracted entities and relations using trained machine-learning modules.

Prerequisite relations are important in teaching and learning. Prerequisite relations mining focuses on finding prerequisite relations from course descriptions and related materials for course learning. For example, which courses or knowledge should be acquired first in order to understand a new topic or take a new course. Liang et al. explored data-driven methods to recover prerequisite relations for different university courses [14]. Recently, Sun et al. proposed ConLearn, a contextual-knowledge-aware approach to tackle this task [23]. Their work utilizes pre-trained language models to generate contextual information and graph neural networks to process the features. This mainly provides convenience for teachers to plan and prepare curriculum design. Building KGs focused on prerequisite relations concentrates more on the learning schedule and course plan.

Multimedia learning material knowledge graphs. Several studies focus on integrating e-learning resources into KGs and linking different concepts with relations between courses. Recently, Dang et al. constructed KGs for MOOCs with a focus on education using Wikipedia data [7]. Li et al. proposed a system named MEduKG to integrate multi-modal information into KGs for education [11]. The previous one concentrates on resource organization based on existing MOOC data into KGs. The latter one concentrates more on constructing KG from multimedia learning resources like slides, textbooks, and recordings. They used NER and RE components to construct KGs but from multi-media data resources. Aliyu et al. constructed KG for university course management based on structured data [3]. They used KGs to store and manage data for university courses, like instructors, teaching semesters, etc. It can be used to store teaching arrangement information for universities.

Errors may be accumulated during the NER process. Existing educational KGs constructed by machine learning algorithms often contain noises and redundant triples, which make these KGs impractical to some extent. Our framework employs tailored designs to learn low-redundant and high-accuracy KGs.

5 Conclusions and Future Work

In this paper, we proposed a framework to construct KGs for education. Unlike other KG construction methods, we aim to build a KG with high accuracy and low redundancy. Starting from the ontology to build the backbone, the framework then leverages reliable data sources with tailored methods. Moreover, we apply active KG error detection to refine the KG with minimal human effort. Our case study and visualization demonstrate how this framework produces KG for education with high-quality and integrated learning resources, enabling the final KG to be a useful aid for education. We also designed a machine reading comprehension method to extract relations from unstructured data. This method can be applied for relation extraction in various domains without a large amount of training data. The experimental results show that it performs well for handling practical real-world data. In the future, we will focus on the following aspects: Firstly, further expanding the size of the KG by introducing more relations and entity types for education, while keeping high accuracy and low redundancy. Secondly, explore algorithms to refine the KGs more effectively and measure the quality of the outputs. Thirdly, apply the KGs to educational downstream tasks, such as question and answer systems, learning course planning, etc.

References

1. Abu-Salih, B.: Domain-specific knowledge graphs: A survey. Journal of Network and Computer Applications **185** (2021)
2. Agichtein, E., Gravano, L.: Snowball: Extracting relations from large plain-text collections. In: Proceedings of the fifth ACM conference on Digital libraries. pp. 85–94 (2000)
3. Aliyu, I., Kana, A., Aliyu, S.: Development of knowledge graph for university courses management. International Journal of Education and Management Engineering **2** (2020)
4. Batista, D.S., Martins, B., Silva, M.J.: Semi-supervised bootstrapping of relationship extractors with distributional semantics. In: Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing. pp. 499–504 (2015)
5. Bosselut, A., Rashkin, H., Sap, M., Malaviya, C., Celikyilmaz, A., Choi, Y.: Comet: Commonsense transformers for automatic knowledge graph construction. In: Annual Meeting of the Association for Computational Linguistics. pp. 4762–4779 (2019)
6. Chen, P., Lu, Y., Zheng, V.W., Chen, X., Yang, B.: KnowEdu: A system to construct knowledge graph for education. IEEE Access **6**, 31553–31563 (2018)
7. Dang, F.R., Tang, J.T., Pang, K.Y., Wang, T., Li, S.S., Li, X.: Constructing an educational knowledge graph with concepts linked to wikipedia. Journal of Computer Science and Technology **36**(5), 1200–1211 (2021)
8. Dessì, D., Osborne, F., Reforgiato Recupero, D., Buscaldi, D., Motta, E., Sack, H.: Ai-kg: an automatically generated knowledge graph of artificial intelligence. In: International Semantic Web Conference. pp. 127–143. Springer (2020)
9. Galárraga, L., Teflioudi, C., Hose, K., Suchanek, F.M.: Fast rule mining in ontological knowledge bases with amie. The International Journal on Very Large Data Bases **24**(6), 707–730 (2015)
10. Galárraga, L.A., Teflioudi, C., Hose, K., Suchanek, F.: Amie: association rule mining under incomplete evidence in ontological knowledge bases. In: International World Wide Web Conference. pp. 413–422 (2013)
11. Li, N., Shen, Q., Song, R., Chi, Y., Xu, H.: Medukg: A deep-learning-based approach for multi-modal educational knowledge graph construction. Information **13**(2) (2022)

12. Li, X., Yin, F., Sun, Z., Li, X., Yuan, A., Chai, D., Zhou, M., Li, J.: Entity-relation extraction as multi-turn question answering. In: Annual Meeting of the Association for Computational Linguistics. pp. 1340–1350 (2019)
13. Li, Y., Zhao, J., Yang, L., Zhang, Y.: Construction, visualization and application of knowledge graph of computer science major. In: International Conference on Big Data and Education. pp. 43–47 (2019)
14. Liang, C., Ye, J., Wu, Z., Pursel, B., Giles, C.L.: Recovering concept prerequisite relations from university course dependencies. In: AAAI Conference on Artificial Intelligence (2017)
15. Mahdisoltani, F., Biega, J., Suchanek, F.M.: YAGO3: A knowledge base from multilingual wikipedias. In: Conference on Innovative Data Systems Research (2015)
16. Osborne, F., Motta, E.: Klink-2: integrating multiple web sources to generate semantic topic networks. In: International Semantic Web Conference. pp. 408–424. Springer (2015)
17. Petroni, F., Rocktäschel, T., Lewis, P., Bakhtin, A., Wu, Y., Miller, A.H., Riedel, S.: Language models as knowledge bases? In: Conference on Empirical Methods in Natural Language Processing. pp. 2463–2473 (2019)
18. Qin, Y., Cao, H., Xue, L.: Research and application of knowledge graph in teaching: Take the database course as an example. In: Journal of Physics: Conference Series. vol. 1607 (2020)
19. Salatino, A.A., Thanapalasingam, T., Mannocci, A., Osborne, F., Motta, E.: The computer science ontology: a large-scale taxonomy of research areas. In: International Semantic Web Conference. pp. 187–205 (2018)
20. Shi, D., Wang, T., Xing, H., Xu, H.: A learning path recommendation model based on a multidimensional knowledge graph framework for e-learning. Knowledge-Based Systems **195**, 105618 (2020)
21. Stewart, M., Liu, W.: Seq2kg: an end-to-end neural model for domain agnostic knowledge graph (not text graph) construction from text. In: Proceedings of the International Conference on Principles of Knowledge Representation and Reasoning. vol. 17, pp. 748–757 (2020)
22. Su, Y., Zhang, Y.: Automatic construction of subject knowledge graph based on educational big data. In: International Conference on Big Data and Education. pp. 30–36 (2020)
23. Sun, H., Li, Y., Zhang, Y.: Conlearn: Contextual-knowledge-aware concept prerequisite relation learning with graph neural network. In: SIAM International Conference on Data Mining. pp. 118–126 (2022)
24. Vrandečić, D., Krötzsch, M.: Wikidata: a free collaborative knowledgebase. Communications of the ACM **57**(10), 78–85 (2014)
25. Wadden, D., Wennberg, U., Luan, Y., Hajishirzi, H.: Entity, relation, and event extraction with contextualized span representations. arXiv preprint arXiv:1909.03546 (2019)
26. Wang, H., Zhao, M., Xie, X., Li, W., Guo, M.: Knowledge graph convolutional networks for recommender systems. In: International World Wide Web Conference. pp. 3307–3313 (2019)