electronics

MDPI

Article

Application of Knowledge Graph in Water Conservancy Education Resource Organization under the Background of Big Data

Yangrui Yang *, Yaping Zhu and Pengpeng Jian

School of Information Engineering, North China University of Water Resources and Electric Power, Zhengzhou 450018, China

* Correspondence: yangyangrui@ncwu.edu.cn; Tel.: +86-19937023308

Abstract: The key to improving the readability and usage of educational resources is their orderly arrangement and integration. Knowledge graphs, which are a large-scale form of knowledge engineering, are an effective tool for managing and organizing educational resources. The water conservancy's educational big data is separated into three tiers of objectives–courses–knowledge units based on the connotation level of self-directed learning. Combined with the idea of Outcome-based Education(OBE), the goal-oriented knowledge graph structure of water conservancy disciplines and graph creation method is proposed. The focus is the error accumulation problem brought about by the traditional relational extraction method of Named Entity Recognition based on rules or sequence labeling. We first complete this objective, and then the relationship classification is performed according to the water conservancy disciplines entity and relations joint extraction (WDERJE) model, on which the prompt mechanism design is based. Think of the entity-relationship extraction task as a sequence-to-sequence generation task, and take the structured extraction language to unify the coding entity extraction and relationship extraction structures. The evaluation results of the WDERJE model show that the F_0.5 value of each entity extraction is above 0.76, and the cumulative extraction relationship triple is nearly 180,000. The graph fully optimizes the organization and management of water conservancy education resources and effectively improves the readability and utilization rate of water conservancy teaching resources.

Keywords: knowledge graph; knowledge engineering; water conservancy educational big data; relationship extraction; prompt mechanism

check for updates

Citation: Yang, Y.; Zhu, Y.; Jian, P. Application of Knowledge Graph in Water Conservancy Education Resource Organization under the Background of Big Data. *Electronics* **2022**, *11*, 3913. https://doi.org/10.3390/electronics11233913

Academic Editor: George A. Tsihrintzis

Received: 12 October 2022
Accepted: 24 November 2022
Published: 26 November 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Both UN-Water's focus on global water issues and UNESCO's emphasis on education for sustainable development require us to take a fresh look at Water Conservancy discipline education today and in the future. In addition, China's "14th Five-Year Plan for Water Science and Technology Innovation" also proposes to lead the new phase of high-quality Water Conservancy development with a high level of scientific and technological innovation support [1]. Without a doubt, science and technology innovation requires a solid system of disciplinary knowledge as a foundation. One of the most crucial approaches to helping water conservancy professionals create a solid disciplinary knowledge system is to organize and integrate the knowledge of water conservancy disciplines using the knowledge graph [2]. When using water conservancy education materials for self-directed learning, students should be aware of three crucial concepts. They need to: (1) Clearly position their own development goals (Why learn). For example, water affairs, engineering, hydrology and water resources, water conservancy, and hydropower engineering, among many others, are among the subfields of water conservation. Students need a clear direction for their studies in order to learn and practice with purpose. (2) Understand the courses that need to be taken in this direction and how these courses are related to each other (How to learn).

Different directions have different emphases. For instance, the direction of agriculture water conservation focuses on courses related to Soil and Farming as well as Water-saving Irrigation. The direction of Water Conservancy and Hydropower Engineering focuses on Water Conservancy Engineering Geology and Water Conservancy and Hydropower Planning and other courses. Moreover, many courses in the same direction seem to be independent of one another, but some courses have a strict sequential relationship between them. For example, Engineering Graphics is a prerequisite course for Water Conservancy Engineering Cartography. When learning Water Conservancy Engineering Cartography, there must be a reserve of Engineering Graphics knowledge in the mind of the learner. The learning effect of the prerequisite course will directly affect the learning efficiency and effectiveness of subsequent courses. Students can only develop a complete learning plan after they understand the logical connections between the courses. (3) Effectively organize and summarize specific knowledge units (What learn). There are many unstructured and trivial knowledge units in the course, such as knowledge points and course experiments, while there are also some orderly knowledge units between related courses, which need to be sorted out and summarized to understand what we have learned.

The rational use of water conservancy education resources is necessary to address the three issues mentioned above. However, a variety of their effective uses are significantly impacted by the water education resources' multi-source, heterogeneous, and redundant nature. The knowledge graph, as we are all aware, is essential for organizing and integrating subject knowledge [3]. But knowledge needs to be organized in a reasonable structure, and simply integrating knowledge is not a good way to help students sort out their knowledge. According to two experienced water conservancy professors, a goal-driven learning approach can effectively improve students' learning outcomes and enable them to become higher-quality and more professional water conservancy talents. Based on the above research, inspired by OBE [4], in this paper, we propose a three-layer goal-oriented water conservancy disciplines knowledge graph structure. Regarding the relationship extraction technique in the process of knowledge graph construction, the dependency between the Name Entity Recognition and Relationship Classification tasks performed successively by the traditional pipeline form of relationship extraction method will bring the error accumulation problem [5]. Inspired by UIE [6], this paper designed WDERJE, a water conservancy discipline entity-relationship joint extraction model based on the ERNIE (Enhanced Representation through kNowledge IntEgration) [7]. The joint extraction is used in WDERJE to extract both entities and inter-entity relations present in the input statements to avoid error accumulation. This knowledge graph covers all levels of knowledge in the water conservancy discipline in an orderly and complete way with a structured hierarchy of objectives–courses–knowledge units and establishes course links by deeply mining the inter-entity relationships between courses, thus assisting water conservancy students in developing personalized course plans according to their objectives. The organized units of knowledge improve the readability and utilization of knowledge and facilitate students to efficiently complete the learning tasks of pre-course, in-course, and post-course review. Combined this knowledge map with the classroom lectures to form an online and offline hybrid teaching model [8]. This is of great significance to enriching the wisdom teaching of water conservancy in colleges and universities, training high-quality water conservancy talents, and accelerating the new stage of high-quality water conservancy development [9].

The main contributions of this paper are as follows:

- We effectively improve the readability and utilization of water conservancy education resources. Inspired by Outcome Based Education, we propose a novel knowledge graph structure, a goal-oriented three-level water conservancy disciplines knowledge graph structure, and a construction scheme for this graph.
- We are the first to propose WDERJE, a water conservancy discipline entity-relationship joint extraction model based on the ERNIE, to simultaneously extract the entities and relationships in water conservancy education resources through WDERJE.

- We built the first water conservancy disciplines knowledge graph, and the upper applications, such as course link retrieval, knowledge unit retrieval, and terminology query, were built based on this knowledge graph.

2. Related Work

Cognitive intelligence is a new stage in the development of computer intelligence, and the knowledge graph is defined by many scholars as the cornerstone of computer perceptual intelligence leading to cognitive intelligence, which can completely establish the deep connection between entities in the objective world [10]. The development of the knowledge graph concept has gone through the semantic network–ontology–Web–semantic web–linked data [11–14] until 2012, when Google commercialized implementations based on some of the ideas in the Semantic Web, and its proposed knowledge graph concept is still in use today [15]. Knowledge graphs are large-scale knowledge engineering maps that represent entities and inter-entity relationships in the objective world in terms of nodes and edges. They incorporate natural language processing techniques such as named entity recognition, relationship extraction, and entity fusion and a complex knowledge system such as knowledge storage, complementation, inference, and application. The innovation of various deep learning models advancing the high speed of the above technologies and the emergence of various large-scale graph databases effectively dealing with complex semantic association problems have resulted in knowledge graphs attracting the attention of more and more scholars [16].

With the in-depth application of knowledge graph technology, academic research on it has gradually shifted from open fields to vertical fields, driving the development of intelligence within several specific fields [17]. In the field of water conservancy, at present, most of these studies are metrology analyses of the subject words and keywords of existing literature and sci-tech texts based on knowledge graphs to investigate the research trends in specific directions of water conservancy, and no scholars have carried out any research related to the application of the knowledge graph to water conservancy discipline. The effective integration of the knowledge of the whole water conservancy discipline through the knowledge graph has not yet been realized. In other studies on subject knowledge graphs, Zhengyin et al. use knowledge mapping to discover Subject Knowledge in Life and Medical Sciences [18]. Su et al. investigated a BERT-BiLSTM-CRF-based method for the automatic construction of a subject knowledge graph [3]. Chen et al. proposed a neural sequence labeling algorithm to extract instructional concepts from instructional data [19]. Yang et al. built an intelligent question-and-answer system based on the knowledge map of the high school curriculum [20]. These works on subject knowledge graphs use existing deep learning models to construct coarse-grained subject knowledge maps and do not detail the fine-grained subject knowledge graph by knowledge level. In this paper, we have made a fine-grained hierarchical division of the knowledge graph of water conservancy disciplines and propose a goal-oriented three-level water conservancy disciplines knowledge graph structure. Through the knowledge graph, the highly dispersed knowledge units and professional terms are organized and integrated, and intelligent retrieval and visual display are realized, which is of great significance in assisting in improving the quality of water conservancy personnel training.

3. Materials and Methods

The knowledge source of the water conservancy discipline knowledge graph is water conservancy educational big data, which is dominated by semi-structured and unstructured text corpus, supplemented by a small amount of structured data. This graph belongs to the vertical domain knowledge graph and is constructed using a top-down approach. The graph ontology layer is first constructed based on the clearly defined entity and relationship types.

After the ontology layer is built, start the water conservancy discipline entity-relationship extraction. WDERJE is used to extract both entities and relationships from the input state-

ments. The original corpus is annotated by the doccano annotation tool according to the pre-built ontology layers, and the WDERJE model is trained using a small number but correctly labeled data, and then the best model weights are applied to the large-scale source data for target knowledge extraction. In terms of knowledge fusion, different entities describing the same object are fused by techniques such as entity alignment for unstructured text, and relational databases are merged for structured data. Finally, the storage of the data layer of the knowledge graph of the water conservancy discipline is realized based on the graph database Neo4j.

Based on the graph construction, the knowledge within the graph is expanded by combining techniques such as knowledge complementation [21] and entity linking. At the application level, this graph is used as the underlying knowledge background, combined with Vue + Flask to realize basic knowledge visualization, and knowledge inference and link recommendation algorithms are used to realize the construction of upper-layer applications such as course relationship path retrieval and personalized course link recommendation. The construction scheme for the goal-oriented water conservancy discipline knowledge graph is shown in Figure 1.

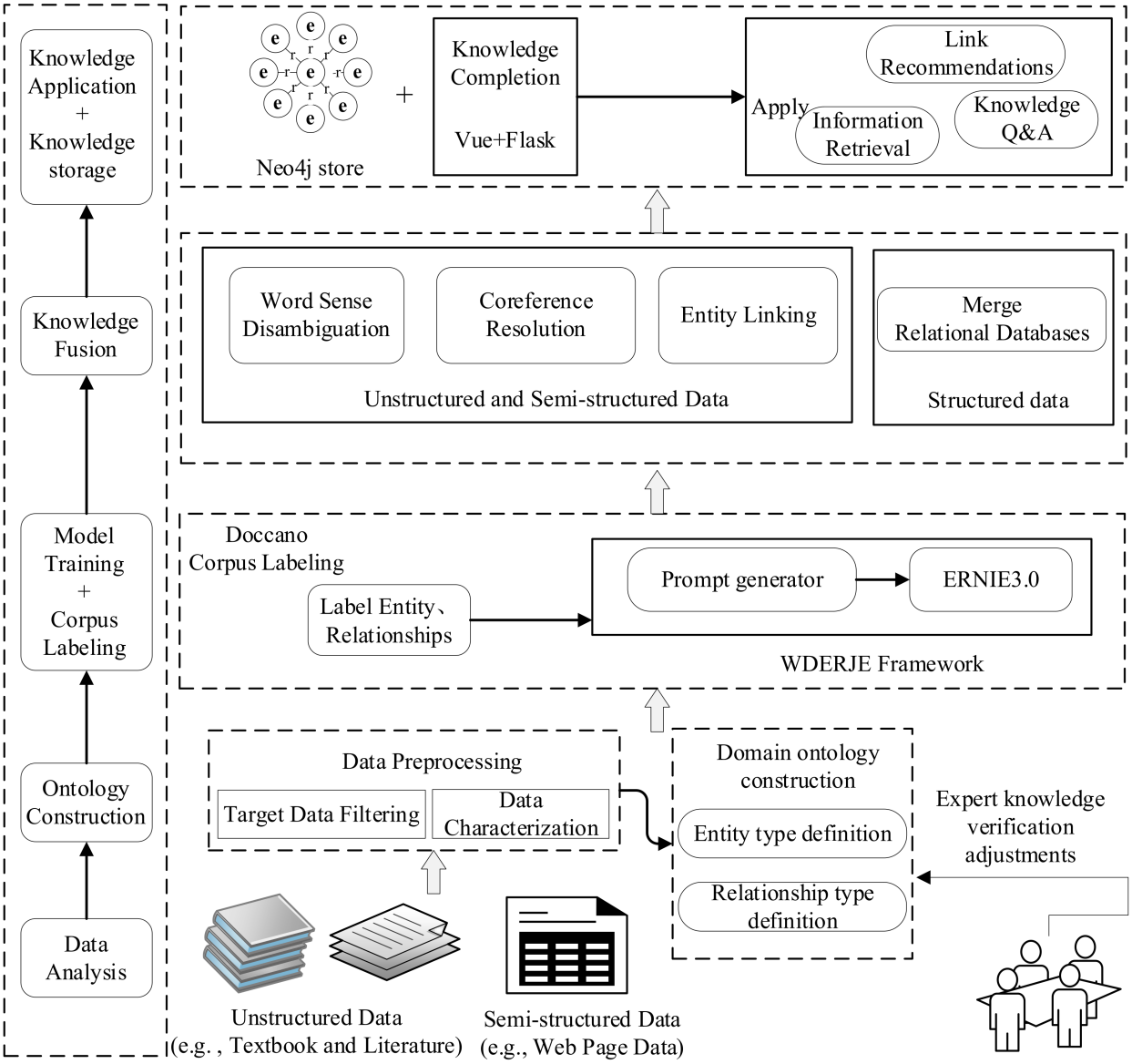


Figure 1. Overall Construction Scheme.

The next four subsections of this paper describe the research process in accordance with this graph construction, including knowledge analysis of water conservancy disciplines, ontology layer construction, knowledge extraction, and knowledge fusion.

3.1. Knowledge Analysis of Water Conservancy Disciplines

3.1.1. Level of Knowledge Analysis

Knowledge is the induction, deduction, comparison, excavation, and condensation of information formed after processing and editing raw data. The professional knowledge of water conservancy disciplines contains the general principles and laws summarized in the process of scientific research and engineering practice in the field of water conservancy over history, as well as a large number of condensed professional terms, which belongs to cognitive knowledge (cognition means knowing why and how to do). This knowledge is clustered in different structures in the education data related to water conservancy. The knowledge source of the water conservancy knowledge graph is mainly water conservancy education data, which is based on various unstructured texts used in training water conservancy professionals, including the requirements of water conservancy institutions for the introduction of talents, training programs and curriculum plans of water conservancy-related majors developed by universities, syllabus developed for each course, and the recommended textbooks, teaching materials, and extended reading materials at the time of the lecture, and a wide range of literature and journals in the field of water conservancy.

We consulted five water conservancy-related students, and the knowledge characteristics of the above knowledge sources were analyzed and summarized. For uncertain types of knowledge, we consulted two professional tutors for analysis and adjustment. The final knowledge layer is shown in Table 1: that is, the different subfields under the field of water conservancy and the main disciplines under these directions, as well as the outputs and graduation internships that need to be taken under the objectives. The textbooks and their associated author and publisher information that need to be known before the course is taken are grouped at the course level. The specific learning process of course knowledge points, course experiments, terminology, etc., are classified at the knowledge unit level. The types of knowledge at each level of the water conservancy profession are shown in Table 1.

Table 1. Definition of the knowledge type.

Knowledge Layer	Knowledge Type	Knowledge Description
Directions layer	Study Direction	Directions of study in the field of water conservancy
	Main disciplines	The main disciplines under each direction of study
	Engineering practice	Theoretical studies are followed by awareness, production, and graduation internships
Course layer	Course name	Courses to be taken under the direction of study
	Textbooks	Recommended textbooks for lectures
	Further reading	Further reading to consolidate course learning
Knowledge unit layer	Course knowledge points	Theoretical knowledge points when learning the course
	Course experiments	Operational experiments designed to consolidate knowledge points
	Terminology	Terminology in the field of water conservancy

3.1.2. Knowledge Graph Structure

Based on the knowledge levels summarized in Table 1 and combined with the OBE concept, a three-layer goal-oriented water conservancy disciplines knowledge graph structure is proposed, as shown in Figure 2. This graph digs deep into potential connections between courses and fully integrates knowledge units within the course. It helps students

to analyze the course chain and grasp the specific learning content, and to grasp how the theoretical knowledge can be applied to solve practical engineering problems through course practice and internship. We also extracted the water conservancy terminology from the water conservancy literature and journals to enrich the content of the knowledge graph. The direction of water conservancy disciplines, courses required for the discipline, course material information, course content, course experiments, and terminology are organized according to the principles of science, purpose, and order. The logic of "direction–course planning–specific knowledge unit" is followed by the water conservancy students to develop their learning and investigation, to better develop their self-directed learning ability and ability to solve practical engineering problems, and achieve the educational purposes of cultivating students to become the people they need in a given society.

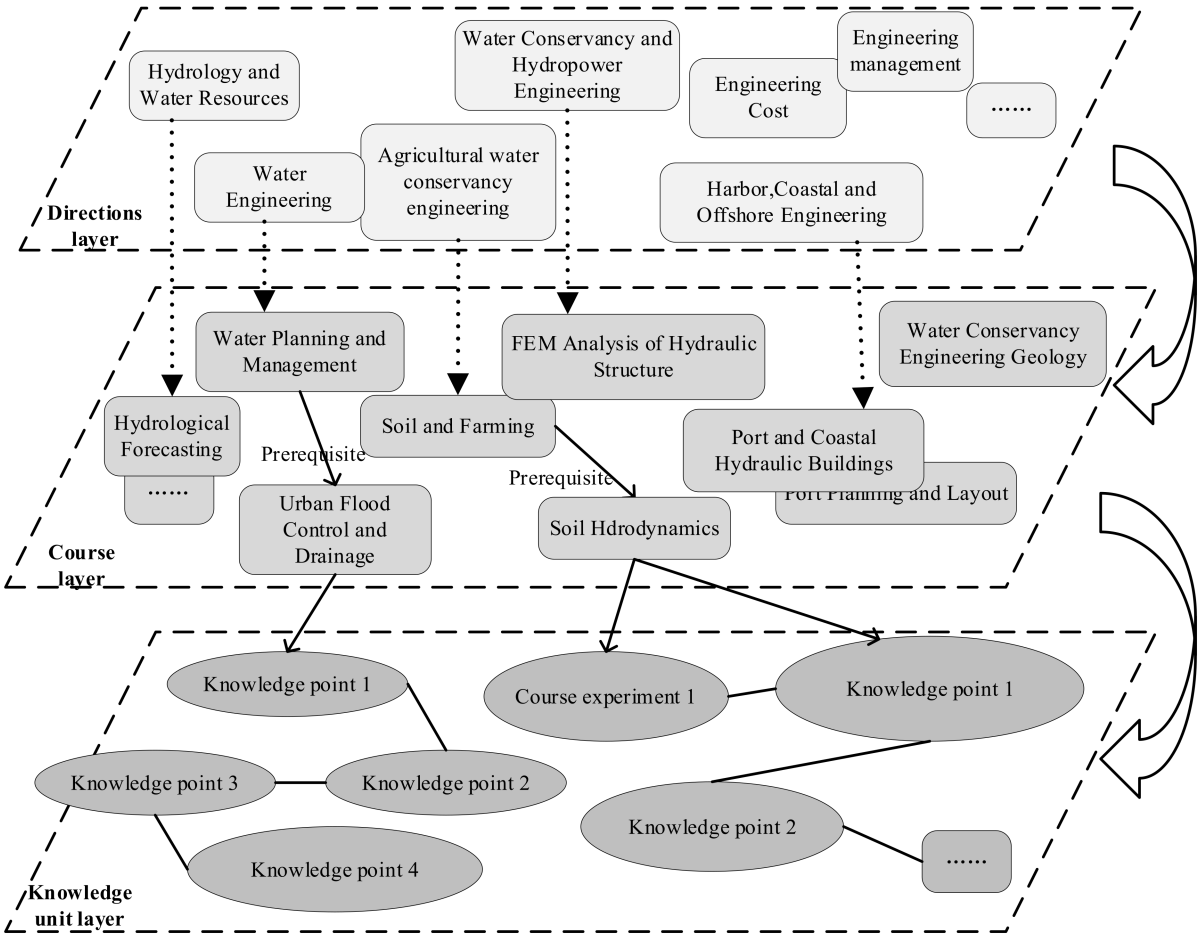


Figure 2. Three-layer Goal-oriented Water Conservancy Disciplines Knowledge Graph Structure.

3.2. Ontology Layer Construction

3.2.1. Triples Representation

According to the graph structure properties of the nodes and edges, the triple is identified as the basic data model for the knowledge representation of water conservancy disciplines [22]. Defining this graph as G, then the knowledge graph can be represented as:

$$G = <\text{subject, predicate, object}>, \quad (1)$$

where subject, predicate, and object denote the head entity, relation, and tail entity, or head entity, attribute, and attribute value of the triple, respectively. For example, the text "Water conservancy engineering construction, as a core course of major of Water Conservancy and Hydropower Engineering, is the theoretical basis for learning Water Conservancy and Hydropower Engineering" contains the entities "Water conservancy engineering construction"

and "Water Conservancy and Hydropower Engineering" and the attribute "core course", which is extracted as a triple in the form of <Water Conservancy and Hydropower Engineering, professional course, Water conservancy engineering construction> and <Water conservancy engineering construction, course properties, core course>. The head and tail entities and attribute values in the triple are presented as nodes in the water discipline knowledge graph, and the relationships or attributes are presented as edges.

3.2.2. Ontology Construction

The domain knowledge graph adopts a top-down construction model, then the definition of the ontology is the basis for the construction of the domain knowledge graph. In this work, we have defined 12 entity types based on the knowledge types of water conservancy disciplines in Table 1, namely "study direction", "main disciplines", "engineering practice", "course name", "bibliography", "author", "publisher", "knowledge points", "experiments name", "terminology name", "literature journal", and "terminology content". We have defined 14 types of relationships between entities and, from this, construct the ontology layer as shown in Figure 3.

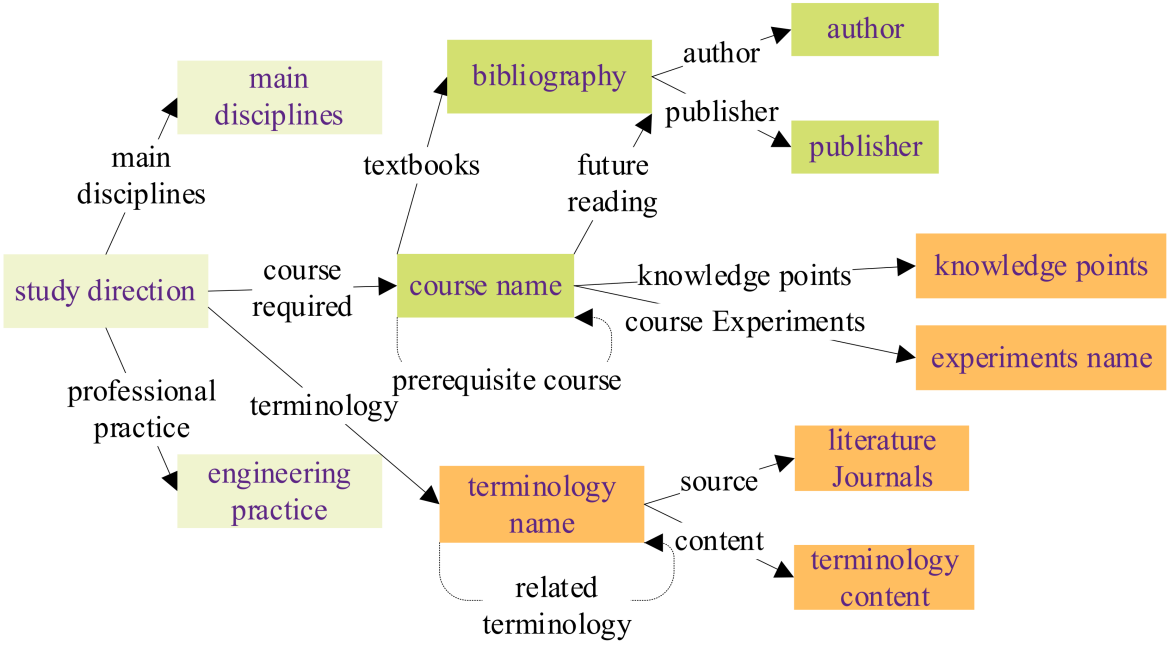


Figure 3. Ontology of water conservancy disciplines knowledge graph.

3.3. Knowledge Extraction

Knowledge graphs focus on expressing independent abstract entity-entity relationships, and Relationship Extraction (RE) is one of the keys to building knowledge graphs. The goal of relationship extraction is to extract all the triples it contains from an input text [23]. Traditional relationship extraction is performed in pipeline form, i.e., Named Entity Recognition (NER) is completed to pair entities first, and then Relationship Classification (RC) is implemented based on a predefined relationship library. The input of the RC task of this pipeline form of relationship extraction method depends on the output of the NER task. Then, the correct rate of NER will directly affect the accuracy of subsequent RC tasks [5], and the dependence between successive tasks will cause error accumulation. To address the error accumulation problem brought by the traditional pipeline form of relationship extraction methods, inspired by UIE [6], the water conservancy disciplines entity-relationship extraction task is considered as a sequence-to-sequence generation task, and a joint extraction framework for water conservancy disciplines entity-relationship, the WDERJE framework, is designed. As shown in Figure 4, the overall WDERJE consists of two parts: a prompt generator and an entity-relationship joint extractor. In the first

stage, the prompt generator generates a prompt based on the original input text, and in the second stage, the complete input containing the prompt is sent to the entity-relationship joint extractor, which finally outputs a structured extraction language in the form of a triple. The specific work of the two components is:

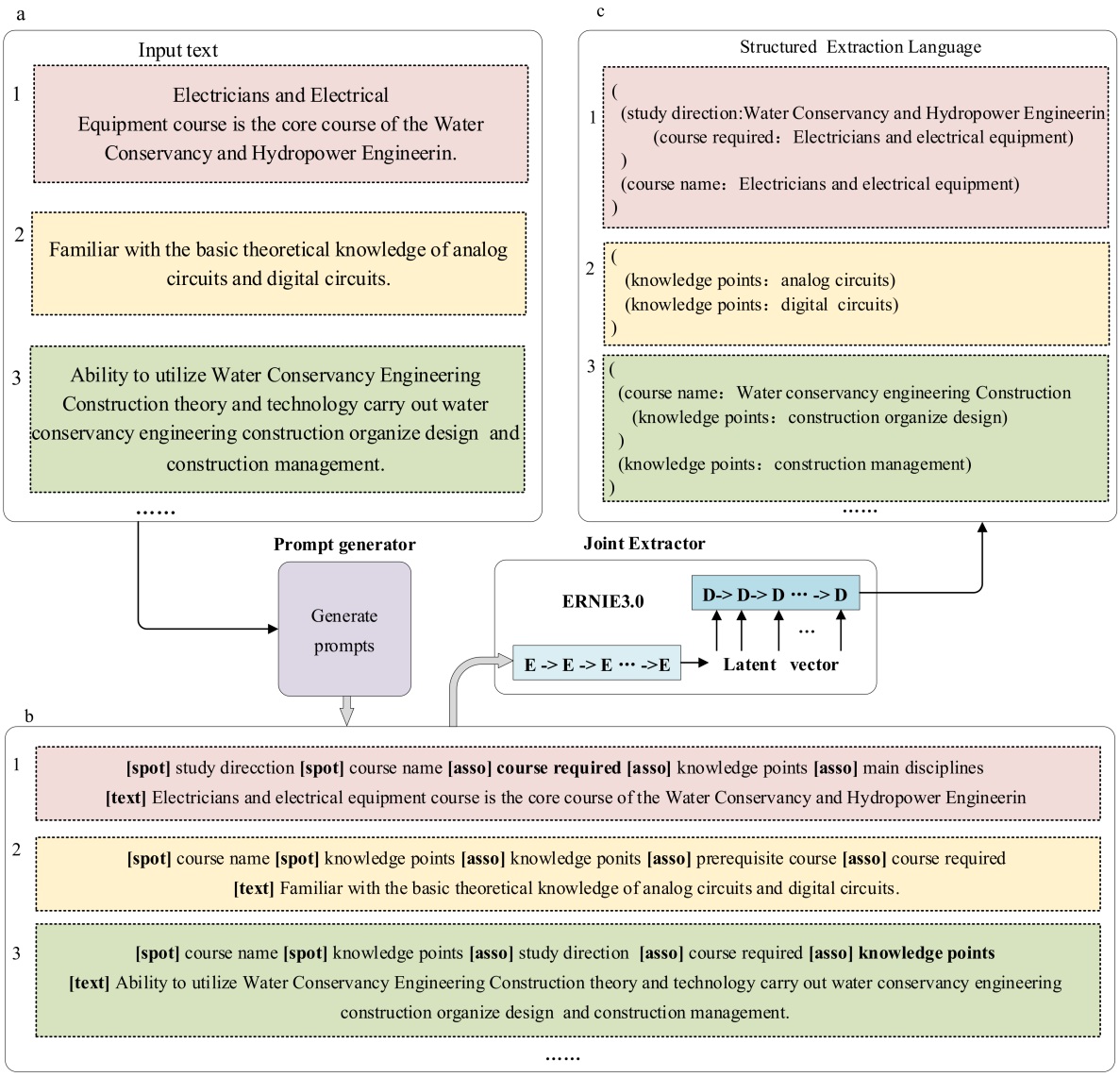


Figure 4. Overall Framework of WDERJE. Part (**a**) is the input text; Part (**b**) is the sequence of prompts; Part (**c**) is the sequence of SEL.

3.3.1. Prompt Generator

The graph ontology layer (Figure 3) has clearly defined the entity and relationship types at each level of the water conservancy disciplines domain. To accurately extract all the existing entity-relationship triples of the water conservancy disciplines from the original input text, the entity types, relationship types, and input statements in the original input statements are marked using three types of cue marks [spot], [asso] and [text] based on the definitions in the ontology layer. For example, in "[spot]study direction [spot] course name [asso] courses required [asso] knowledge points [asso] main disciplines [text] Electricians and electrical equipment course is the core course of the Water Conservancy and Hydropower Engineering", entity categories are "study direction" and "course name", relationship categories are "courses required", "knowledge points" and "main disciplines", and an input text "Electricians and electrical equipment course is the core course of the

Water Conservancy and Hydropower Engineering". The prompt generation process can be expressed as follows:

$$\begin{aligned} PG\left([x_1,x_2,\ldots\ldots x_{|n|}]\right) = [&[\text{spot}]\text{ entity type},\ldots.\ [\text{asso}]\text{ relation type},\ldots \\ &[\text{text}],\ x_1,x_2,\ldots\ldots x_{|n|}] \\ = [&P_1,P_2,\ldots\ldots P_{|p|}] \end{aligned} \quad (2)$$

where $X = [x_1,x_2,\ldots x_{|n|}]$ denotes the original input text, PG denotes the prompt generator, and $[P_1, P_2,\ldots P_{|p|}]$ denotes the complete input after generating the prompt, the exact form if which can be seen in Figure 4b.

3.3.2. Entity-Relationship Joint Extractor

WDERJE uses the large-scale knowledge augmentation pre-training model ERNIE3.0 as an entity-relationship joint extractor and uses a pre-training + fine-tuning migration learning approach to achieve joint extractor training in low-resource situations. A structured extraction language (SEL) [6] is used to encode the structures of entity extraction and relation extraction into a unified representation, which allows for the unified modeling of heterogeneous water conservancy discipline entity-relationship extraction tasks within a sequence-to-sequence generation framework. The first sequence refers to the complete input containing the prompt, and the second sequence refers to the generated structured extraction language. The structured extraction language consists of two statement patterns. One is that "Spot Name: Info Span" denotes various entity types and objects of a specific entity type, and the other one is that "Asso Name: Info Span" denotes various relationship entity types and the associated object of a specific relationship type. The joint extractor accepts $[P_1, P_2,\ldots\ldots P_{|p|}]$ as input to this phase of the task and generates a structure extraction language as output, eliminating the need for entity matching and reducing error accumulation. The generation process can be expressed as follows:

$$JE\left[[P_1,P_2,\ldots P_{|p|}]\right] = [y_1,y_2,\ldots,y_{|m|}] \quad (3)$$

where JE denotes the ERNIE3.0, $[y_1,y_2,\ldots,y_{|m|}]$ represents the returned structured extraction language, and the specific form of the structured extraction language can be seen in Figure 4c.

3.4. Knowledge Fusion

Knowledge fusion is a high-level organization of knowledge to improve the quality of knowledge base construction [24]. The main components of knowledge fusion are entity disambiguation, entity fusion, etc. Knowledge fusion is needed in water discipline professional knowledge extraction to solve the problem of multiple words with one meaning (entity duplication) or one word with multiple meanings (entity description inconsistency problem) caused by the extracted corpus of multiple heterogeneous sources. Entity disambiguation is to determine the specific point of an entity through contextual information, which mainly solves the problem of multiple meanings of a word. For example, "Hydraulic" can be described as water management engineers, but it also is short for Water Conservancy and Hydropower Engineering. In the face of such ambiguity, we first give a list of target entities, set rules to filter out the target entities that are not possible to point to determine the candidate entities, and then use the entity linking method in the candidate set to determine the real target of the allegation entity [25].

Entity fusion mainly solves the problem of multiple words with one meaning, including structure alignment, entity alignment, and attribute fusion, and structure alignment refers to the conversion of heterogeneous data into homogeneous data, and the key part is entity alignment [26]. For example, the entity of "Agricultural water engineering" extracted from the training program of the university and the entity of "Agricultural water" extracted from the text of the major, both entities are semantically different descriptions of the object of "Agriculture Water Conservancy engineering", and need to fuse these two entities. To

address this problem, we first mine the key attributes related to the entities, use the equal matching algorithm for numeric attributes and the edit distance for string attributes to calculate the similarity of entity properties, set the weights of different attributes manually, and compare the key attribute sets through the weighting algorithm similarity calculation to determine whether the two entities can be fused.

4. Results

4.1. WDERJE Training

Based on the definition of entity and relationship types in Figure 3, the corpus is annotated using the doccano annotation tool, and it is partitioned into a training set and a test set in the ratio of 8:2 according to the commonly used partitioning methods, the training set, and the validation set are for fitting the parameters of the WDERJE model, and the tuning of the model hyperparameters, and finally the performance of the model is evaluated on the test set. In the process of water conservancy disciplines knowledge extraction, it is desired that the quality of the triple extracted by the model is as high as possible, i.e., it is desired that the knowledge is extracted as accurately as possible to ensure the quality of the knowledge graph. Therefore, the weight of the Precision rate is adjusted upward in the evaluation index, and the $F_{0.5}$-score is the metric used to evaluate the performance of the model. The $F_{0.5}$-score expression is:

$$F_{0.5} = (1.25*P*R)/(0.12*P + R), \quad (4)$$

where P denotes the Precision and R denotes the Recall. The output of the evaluation on the test data shows that the 12 types of entities extracted $F_{0.5}$-score are all above 0.76, where author and publisher type entities $F_{0.5}$-score can reach 0.98, and the WDERJE framework performs better on the water conservancy disciplines entity relationship extraction task with the knowledge enhancement pre-training model ERNIE3.0 as the extractor.

4.2. Results of Knowledge Extraction

Unstructured texts such as training programs, course syllabi, textbooks, reference books, and literature journals related to the field of water conservancy within universities are used as the extracted corpus, while semi-structured data is converted into a triplet structure. Based on the 12 types of entities and 14 types of relationships defined in Figure 3, 14 types of <entity, relationship, relationship> triples were constructed, such as <study direction, course required, course name>, <course name, prerequisite course, course name>, <course name, textbooks, bibliography>, etc. The best model weights are applied to the large-scale text corpus for entity-relationship triple extraction. The extraction results are shown in Table 2. The number of terminologies is not consistent with the number of terminology content due to the absence of some terminology content.

4.3. Illustration of the Knowledge Graph

The knowledge in the knowledge graph is utilized on the basis that this knowledge is efficiently represented and stored. Figure 3 clearly defines this graph's ontology layer, followed by storing the extracted knowledge in a database based on the ontology layer. Considering that the subsequent building of semantic retrieval, intelligent question, and answer applications need to face complex semantic queries, the Neo4j graph database is used to store the knowledge of the water conservancy disciplines knowledge graph. Neo4j has complete data storage and management functions, and the graph structure model of nodes and edges can fully reflect the association between entities and can efficiently handle complex relational problems.

As Figure 5 shows the abstract storage structure of some entities and relationships, such as the terminology and the main disciplines and professional courses under the direction of Water Conservancy and Hydropower Engineering. Under the specialization, there are basic courses Advanced Mathematics and University Physics, which are at the same time prerequisites for the core courses Electricians and Electrical Equipment and

Engineering Hydrology, and other courses similarly, thus establishing course-to-course connections. Course entities are connected to knowledge units such as course experiments and recommended textbooks and authors and publishers, thus allowing the course to be linked to the complete information within it. In terms of terminology, terms that appear together in the journal, such as locating piles and bucket wheels, are connected under the same journal, thus establishing a link between related terms that are potentially related, with similar forms of data storage in other directions. Note that each circle in Figure 5 represents an entity, and in the actual store, there are connecting lines between each entity to represent the relationship (see Figure 6). Here, for clarity of illustration, only the abstract structure is represented. That is, entities with the same type of relationship are represented by a single line.

Table 2. Entity-relationship Extraction Results.

Entity Type 1	Relationship	Entity Type 2	Number of Triples
study direction	course required	course name	198
course name	prerequisite course	course name	396
course name	textbooks	bibliography	207
course name	future reading	bibliography	312
bibliography	author	author	368
bibliography	publisher	publisher	312
course name	knowledge points	knowledge point	6352
study direction	Terminology	Terminology name	57,149
study direction	Professional practice	Engineering practice	68
study direction	Main disciplines	Main disciplines	17
Terminology name	source	Literature Journals	57,463
Terminology name	content	Terminology content	57,032

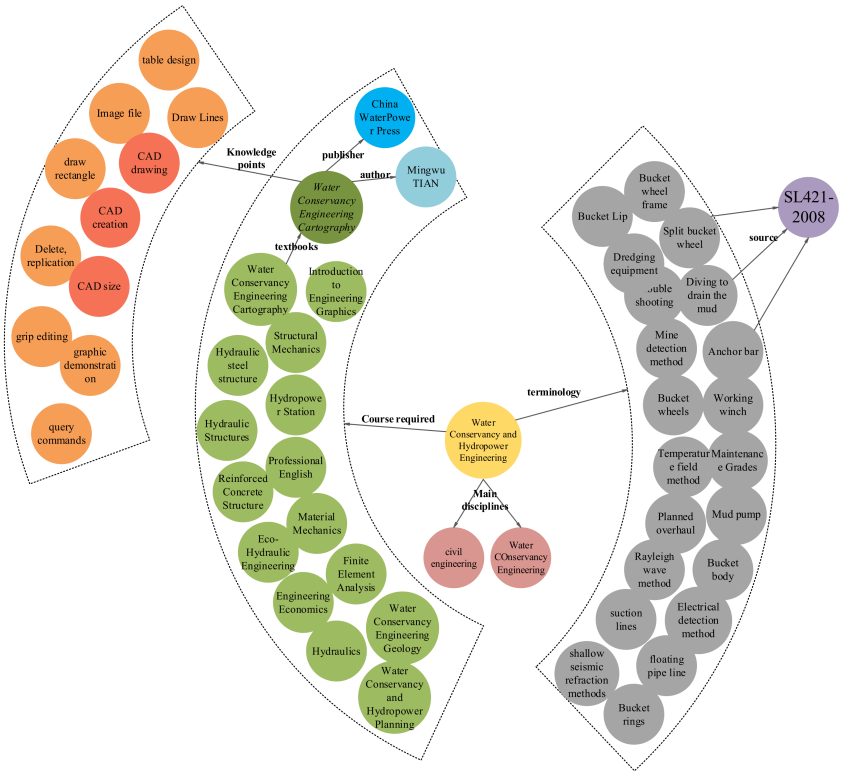


Figure 5. Abstract Storage Structure of Some Entities and Relationships.

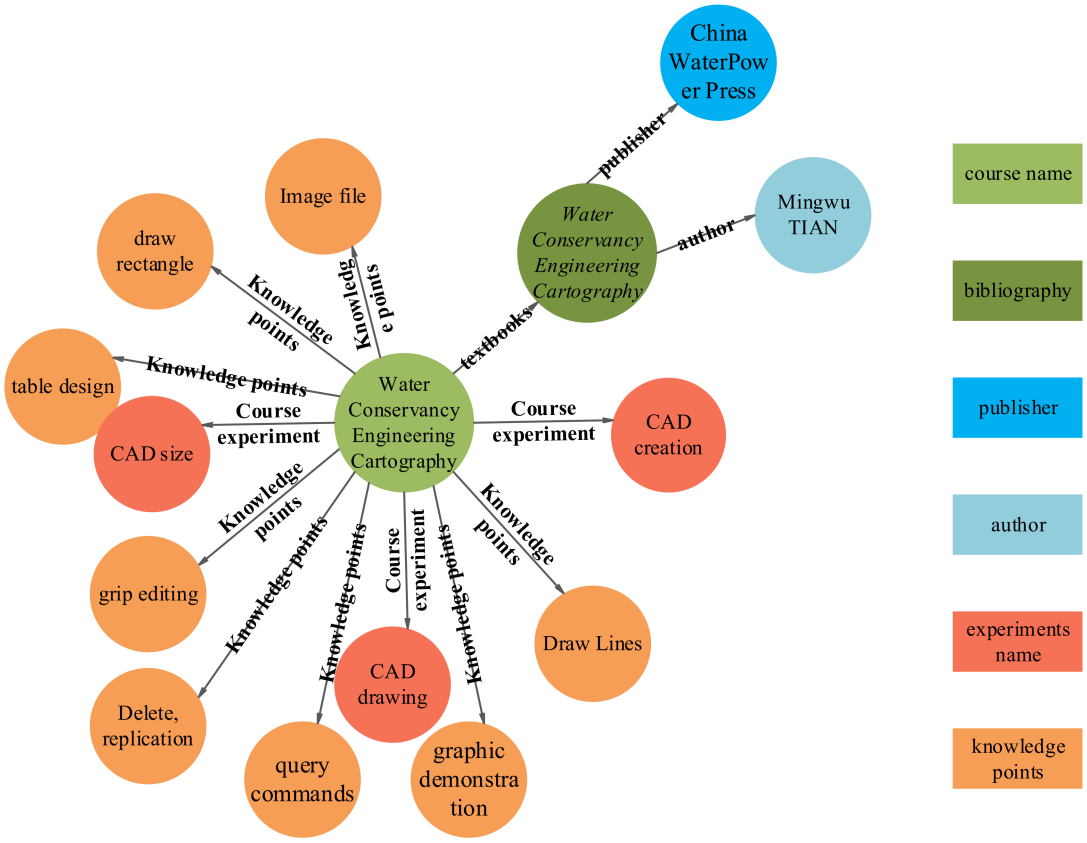


Figure 6. Illustration of Course Information Query Results.

In contrast to the redundant and unstructured water conservancy education big data, in the knowledge graph constructed from this, the lessons and knowledge units are built with a clear hierarchical focus on the objectives to assist students in grasping the "Why learn" as a whole.

5. Application

5.1. Inter-Course Relationship Pathway Search

A quick search of inter-course relationship paths can be performed based on the relationship between directions and courses and between courses and courses established in the knowledge graph. Enter course entity1 and course entity2 to visualize the results of the queried inter-entity relationship paths. For example, a search for the relationship between the Hydroelectric Station course and University Physics courses results in a visual display of "University Physics- > Theoretical Mechanics- > Hydraulics- > Water Conservancy and Hydropower Planning- > Hydropower Station". The visualization of the diagram pattern assists students in grasping the orderliness of the course learning, thus solving the problem of "How to learn".

Based on the collection and inclusion of inter-course relationships and attribute data such as credits and hours, the personalized course chain can be returned by the link recommendation algorithm, and the total credits of the course chain can be returned by knowledge calculation. Students can use the returned results to develop a personalized course plan by modifying and improving based on the course clues.

5.2. Full Course Information Inquiry

During the specific learning process of the course, a visual preview of the complete information of the course can help students to prepare for the course and form a preliminary knowledge of the recommended textbook, the main knowledge units, and the course experiments in their minds in advance.

For example, as Figure 6 shows (note that the display drawings only show some of the knowledge points of the course), if the system searches for the course entity "Water Conservancy Engineering Cartography", the search result shows the textbook used in the course, the author and publisher of the textbook, as well as the course knowledge points and course experiments associated with it. The different types of entities are distinguished using different colors, and the type of entity represented by each color is shown on the right side in Figure 6. The arrows between the two entities in Figure 6 represent the relationship between them. In this way, a visual preview of the complete information of the course can be realized. After learning, the complete knowledge units are more conducive to students' self-testing of the learning effect, and the overall knowledge can be reviewed to deepen their understanding, to solve the problem of "What to learn".

5.3. Related Terminology Query

The water-related professions have accumulated a large number of professional terms in long-term scientific research and engineering practice, and many terms are related to each other to some extent, so memorizing related terms at the same time can help students understand their original meaning more thoroughly. When a term entity is inputted for a query, the association between related terms is established, and the query results are displayed visually with the literary journal in which the current query term appears. For example, there are various river embankment hazard detection methods, including Time domain electromagnetic methods, Electric Profiling methods, etc. When querying the specific content of the electric profiling method, the results return other related detection methods and their contents retrieved from the literature of the SL436-2008 river embankment hazard detection protocols where the electric profiling method is located. In this way, the contents of the terminology and related terminology are queried.

6. Discussion

In this paper, we first propose a novel idea for constructing a goal-oriented three-level water conservancy disciplines knowledge graph and the overall construction scheme. Next, we summarize the knowledge concept by analyzing the knowledge characteristics of water conservancy disciplines and constructing the graph ontology layer. We proposed the WDERJE framework and made a good performance in the entity-relationship joint extraction in water conservancy disciplines through small sample training. Finally, we implement the upper-level application based on this graph, which can both preview the whole and retrieve the details.

The next step is to focus on the knowledge inference techniques to derive new knowledge from the concepts and principles in the discipline based on the data included in this graph and to dig out the deeper meanings of the entities in the associated entities and their attribute data. We will also build specific applications such as knowledge point annotation and knowledge assessment based on a complete water conservancy disciplines knowledge graph to assist water conservancy students in conducting self-directed learning and self-assessment with clear objectives, clear logic, and complete systems more efficiently. It will also help them to become higher-quality professionals to accelerate the new stage of high-quality water conservancy development.

Author Contributions: Conceptualization, Y.Y.; methodology, Y.Y.; software, Y.Z.; validation, Y.Y. and Y.Z.; formal analysis, Y.Z.; investigation, Y.Z.; resources, Y.Y.; data curation, Y.Z.; writing—original draft preparation, Y.Y.; writing—review and editing, Y.Y.; visualization, Y.Z.; supervision, Y.Y. and P.J.; project administration, Y.Y. and P.J.; funding acquisition, P.J. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Youth Project of the National Natural Science Foundation of China, grant number 62107014, and the 2021 Henan Province key research and development and promotion projects, grant number 212102210147.

Data Availability Statement: Our data can be found here: https://github.com/WDKG-data/data, accessed on 11 October 2022.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

References

1. GRG (Government Affairs Reporting Group). The Ministry of Water Resources issued the "14th Five-Year Plan for Water Conservancy Sci-Tech Innovation". *China Water Conservancy News*, 11 January 2022. [CrossRef]
2. Su, Y.; Zhang, Y. Automatic construction of subject knowledge graph based on educational big data. In Proceedings of the 2020 the 3rd International Conference on Big Data and Education, London, UK, 1–3 April 2020; pp. 30–36. Available online: https://dl.acm.org/doi/abs/10.1145/3396452.3396458 (accessed on 11 October 2022).
3. Rizun, M. Knowledge graph application in education: A literature review. *Acta Univ. Lodziensis. Folia Oeconomica* **2019**, *3*, 7–19. [CrossRef]
4. Killen, R. Outcomes-Based Education: Principles and Possibilities. *Unpubl. Manuscr. Univ. Newctle. Fac. Educ.* **2000**, 1–24. Available online: https://academic.payap.ac.th/pyu/uploads/userfiles/file/KM/2558/1_-2-Killen_paper_good-kena-baca1.pdf (accessed on 11 October 2022).
5. Wei, Z.; Su, J.; Wang, Y.; Tian, Y.; Chang, Y. A novel cascade binary tagging framework for relational triple extraction. *arXiv* **2019**, arXiv:1909.03227.
6. Lu, Y.; Liu, Q.; Dai, D.; Xiao, X.; Lin, H.; Han, X.; Sun, L.; Wu, H. Unified Structure Generation for Universal Information Extraction. *arXiv* **2022**, arXiv:2203.12277.
7. Sun, Y.; Wang, S.; Li, Y.; Feng, S.; Chen, X.; Zhang, H.; Tian, X.; Zhu, D.; Tian, H.; Wu, H. ERNIE: Enhanced Representation through Knowledge Integration. *arXiv* **2019**, arXiv:1904.09223.
8. Seoane, R.C.; Jiménez, J.E. Effectiveness of online and blended delivery methods on preservice teachers' knowledge and beliefs for writing instruction. *J. Educ. Teach.* **2022**, *48*, 178–196. [CrossRef]
9. Ministry of Water Resources of the People's Republic of China. Guiding opinions and implementation plans on promoting smart water conservancy construction were printed and distributed by MWR (Ministry of Water Resources). *Water Resour. Dev. Manag.* **2022**, 5. [CrossRef]
10. Ji, S.; Pan, S.; Cambria, E.; Marttinen, P.; Philip, S.Y. A survey on knowledge graphs: Representation, acquisition, and applications. *IEEE Trans. Neural Netw. Learn. Syst.* **2021**, *33*, 494–514. [CrossRef] [PubMed]
11. Lehmann, F. Semantic networks. *Comput. Math. Appl.* **1992**, *23*, 1–50. [CrossRef]
12. Berners-Lee, T.J. *Information Management: A Proposal*; European Organization for Nuclear Research (CERN): Geneva, Switzerland, 1989.
13. Berners-Lee, T. Semantic Web Road Map. 1998. Available online: https://www.emse.fr/~{}beaune/websem/SWRoadmapLee.pdf (accessed on 11 October 2022).
14. Berners-Lee, T.; Chen, Y.; Chilton, L.; Connolly, D.; Dhanaraj, R.; Hollenbach, J.; Sheets, D. Tabulator: Exploring and analyzing linked data on the semantic web. In Proceedings of the 3rd International Semantic Web User Interaction Workshop, Athens, GA, USA, 5–9 November 2006; p. 159.
15. Steiner, T.; Mirea, S. SEKI@ home or crowdsourcing an open knowledge graph. In Proceedings of the First International Workshop on Knowledge Extraction and Consolidation from Social Media (KECSM2012), Boston, MA, USA, 12 November 2012.
16. Fensel, D.; Şimşek, U.; Angele, K.; Huaman, E.; Kärle, E.; Panasiuk, O.; Wahler, A. Introduction: What is a knowledge graph? In *Knowledge Graphs*; Springer: Cham, Switzerland, 2020; pp. 1–10.
17. Zou, X. A survey on application of knowledge graph. *J. Phys. Conf. Ser. IOP Publ.* **2020**, *1487*, 012016. [CrossRef]
18. Hu, Z.; Liu, L.; Bing, D.; Qin, X. Discovering Subject Knowledge in Life and Medical Sciences with Knowledge Graph. *Data Anal. Knowl. Discov.* **2020**, *4*, 1–14.
19. Chen, P.; Lu, Y.; Zheng, V.W.; Chen, X.; Yang, B. KnowEdu: A System to Construct Knowledge Graph for Education. *IEEE Access* **2018**, *6*, 31553–31563. [CrossRef]
20. Yang, Z.; Wang, Y.; Gan, J.; Li, H.; Lei, N. Design and Research of Intelligent Question-Answering (Q&A) System Based on High School Course Knowledge Graph. *Mob. Netw. Appl.* **2021**, *26*, 1884–1890.
21. Shi, B.; Weninger, T. Open-world knowledge graph completion. In Proceedings of the AAAI Conference on Artificial Intelligence 2018, New Orleans, LA, USA, 2–7 February 2018; Volume 32.
22. Xu, B.; Wang, Q.; Lyu, Y.; Shi, Y.; Zhu, Y.; Gao, J.; Mao, Z. EmRel: Joint Representation of Entities and Embedded Relations for Multi-triple Extraction. In Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Seattle, WA, USA, 10–15 July 2022; pp. 659–665.
23. Li, A.; Wang, X.; Wang, W.; Zhang, A.; Li, B. A survey of relation extraction of knowledge graphs. In *Proceedings of the Asia-Pacific Web (APWeb) and Web-Age Information Management (WAIM) Joint International Conference on Web and Big Data, Chengdu, China, 1 August 2019*; Springer: Cham, Switzerland, 2019; pp. 52–66.

24. Banerjee, P.; Baral, C. Knowledge fusion and semantic knowledge ranking for open domain question answering. *arXiv* **2020**, arXiv:2004.03101.
25. Eshel, Y.; Cohen, N.; Radinsky, K.; Markovitch, S.; Yamada, I.; Levy, O.; Levy, R.; Specia, L. Named entity disambiguation for noisy text. *arXiv* **2017**, arXiv:1706.09147.
26. Sun, Z.; Zhang, Q.; Hu, W.; Wang, C.; Chen, M.; Akrami, F.; Li, C. A benchmarking study of embedding-based entity alignment for knowledge graphs. *arXiv* **2020**, arXiv:2003.07743. [CrossRef]