

Article

Knowledge Extraction and Quality Inspection of Chinese Petrographic Description Texts with Complex Entities and Relations Using Machine Reading and Knowledge Graph: A Preliminary Research Study

Zhongliang Chen ^{1,2}, Feng Yuan ^{1,*}, Xiaohui Li ¹, Xiang Wang ², He Li ¹, Bangcai Wu ¹ and Yuheng Chen ¹¹ School of Resources and Environment Engineering, Hefei University of Technology, Hefei 230009, China² Geological Survey of Anhui Province, Hefei 230001, China

* Correspondence: yf_hfut@163.com

Abstract: (1) Background: Geological surveying is undergoing a digital transformation process towards the adoption of intelligent methods in China. Cognitive intelligence methods, such as those based on knowledge graphs and machine reading, have made progress in many domains and also provide a technical basis for quality detection in unstructured lithographic description texts. (2) Methods: First, the named entities and the relations of the domain-specific knowledge graph of petrography were defined based on the petrographic theory. Second, research was carried out based on a manually annotated corpus of petrographic description. The extraction of *N*-ary and single-entity overlapping relations and the separation of complex entities are key steps in this process. Third, a petrographic knowledge graph was formulated based on prior knowledge. Finally, the consistency between knowledge triples extracted from the corpus and the petrographic knowledge graph was calculated. The 1:50,000 sheet of Fengxiangyi located in the Dabie orogenic belt was selected for the empirical research. (3) Results: Using machine reading and the knowledge graph, petrographic knowledge can be extracted and the knowledge consistency calculation can quickly detect description errors about textures, structures and mineral components in petrographic description. (4) Conclusions: The proposed framework can be used to realise the intelligent inspection of petrographic knowledge with complex entities and relations and to improve the quality of petrographic description texts effectively.

Keywords: knowledge graph of petrography; entity sequence labelling; relation classification; complex entity separation; regional metamorphic petrology



Citation: Chen, Z.; Yuan, F.; Li, X.; Wang, X.; Li, H.; Wu, B.; Chen, Y. Knowledge Extraction and Quality Inspection of Chinese Petrographic Description Texts with Complex Entities and Relations Using Machine Reading and Knowledge Graph: A Preliminary Research Study. *Minerals* **2022**, *12*, 1080. <https://doi.org/10.3390/min12091080>

Academic Editor: Yosoon Choi

Received: 30 May 2022

Accepted: 31 July 2022

Published: 26 August 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

Due to the growing availability of massive earth observation data, the research on and application of artificial intelligence technologies, such as knowledge graphs (KGs), machine learning and deep learning, are receiving increasing attention [1–3] in the solid earth [4], remote sensing [5], geological image recognition [6–8] and metallogenic process [9] domains. In response to the rapidly increasing and varying types of filed data, scholars have proposed that studies should be guided by big data thinking and techniques commonly applied in deep information mining, such as hidden-mode processes, unknown correlations or other useful information that could be leveraged for decision making [10,11]. Meanwhile, the geological data tend to be uncertain, sparse, multiresolution and multiscale and need knowledge-rich intelligent systems for processing [12–16]. Among these approaches, machine reading based on natural language processing [17–20] and domain-specific KGs of geosciences [21–24] have also attracted increasing attention from geologists.

PaleoDeepDive, a digital library and machine reading system, is an early application of Data Mining and Knowledge Base in the geosciences [25]. To measure the relative

frequency of the occurrence of stromatolites in Macrostrat (<https://macrostrat.org>, accessed on 18 March 2022) for the North America-Caribbean region, a similar approach named GeoDeepDive was used to extract mentions of the term “stromatolite(s)” or “stromatolitic” within the published documents. A total of 10,683 papers were retrieved and 612 unique stratigraphic names were found linked to stromatolites [26]. For the Chinese geological literature, which constitutes unstructured data, researchers carried out keyword extraction based on Chinese word segmentation and word frequency statistics and showed the intrinsic information of the literature using a KG [27,28].

Google KG is the basis of Google’s semantic search and intelligent question answering service released in 2012. In general, KGs can be divided into generic KGs, such as Google KG, and domain-specific KGs. At present, domain-specific KGs have drawn the attention of some research and have been developed in some commercial applications such as intelligent question answering, intelligent decision making and intelligent detection services in health, education, geology and other fields [21]. In KGs, knowledge is a factual triple in the form of (subject, predicate, object), where each triple entity is represented as a node and edges represent the relation between nodes [29]. As large-scale probabilistic knowledge triples were extracted by information extraction tools, a probabilistic database was also proposed to associate probabilities with triples [30–32]. With the applications of KGs in different domains, geologists have also begun to study how to extract knowledge from unstructured literature sources. Knowledge extraction in the geosciences has focused on the recognition of geological entities or keywords, such as geological time. For instance, Liu et al. [33] divided the information into two types, general time entity and geological time entity, depending on the description characteristics in geological and mineral texts, and realised the structural extraction of geological time entities using a BiLSTM-CRF model. Named entity recognition (NER) using deep learning has also been applied in the extraction of information to construct a domain-specific KG of geological hazards [34]. In Western Australia, KGs were generated from the mineral exploration reports for iron ore deposits in the Chichester Range Project and gold deposits in the Coolgardie Gold Project [17]. The automated KG formulation framework showed the prospect of machine reading in knowledge extraction from unstructured geological texts.

During the process of knowledge extraction in the geosciences, entity recognition is important content, and relationship extraction is also crucial [35,36]. A traditional relation extraction task is to predict whether there is a relation between two entities in a single sentence and classify this relationship; this task is also called binary relation extraction. However, in practical applications there are also complex relation extraction and entity recognition tasks. Figure 1 shows some types of relations encountered in actual scenarios, including a binary relation, *N*-ary relation, overlapping relation (subdivided into single-entity overlapping relation and entity pair overlapping relation) and cross relation [37]. In general, professionally trained geologists usually follow certain rules to form complex entities in petrographic descriptions. For instance, dual-structure and dual-colour entities often appear in structure and colour descriptions, whereas metamorphic rocks with an equigranular blastic texture are often described as multistructure entities in Chinese petrographic descriptions.

Previous research on KG formulation in geosciences mainly focused on simple NER and relation extraction. The extraction of complex knowledge characterised by complex named entities or complex relations has not been studied. The applications of KGs in the geosciences have thus far prioritised basic queries and visualisation [36]. Smart applications, such as the automatic quality inspection of petrographic descriptions, have not been developed. A module of the intelligent mineral geological survey cloud platform, which was named as the “information release and knowledge question”, was only just designed. Research on prospective prediction based on KGs was proposed, but has not yet been carried out [38].

At present, a digital geological survey has been published for China, and a cognitive geological survey is also under development. In this paper, the massive Chinese rock

descriptions obtained through field observations are taken as the research object to carry out the geological record quality inspection using artificial intelligence. An automatic knowledge extraction and quality inspection framework based on KGs and machine reading is studied. The framework proposed in this paper will eventually provide a quality inspection service on rock description texts in the form of a web service interface.

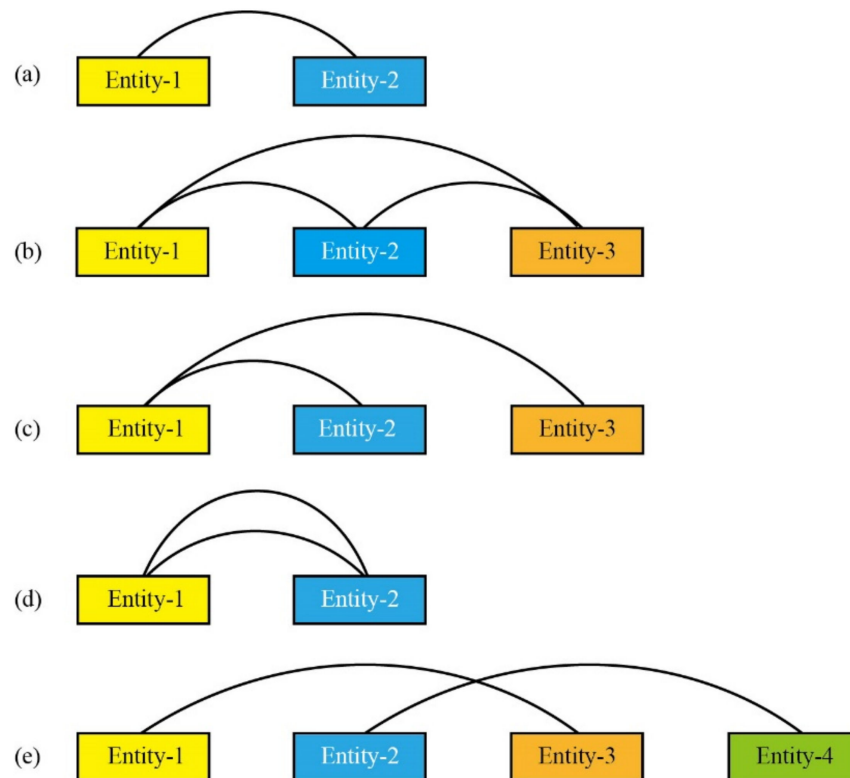


Figure 1. Diagram of the types of relations: (a) binary relation; (b) N -ary relation; (c) single-entity overlapping relation; (d) entity pair overlapping relation; (e) cross relation.

The rest of this paper is organised as follows: the framework for knowledge extraction and quality inspection is introduced in Section 2. The framework components include the definition of the named entities and relations of the petrographic descriptions based on prior knowledge, sequence labelling of rock named entities based on word embedding, N -ary relation extraction of petrographic descriptions based on an enriched pre-trained Chinese language model and complex entity separation based on prior rules. In Section 3, a case study based on the 1:50,000 sheet of Fengxiangyi located in the Dabie orogenic belt is presented. Error propagation in the pipeline mode, integration of variant data and specifications, knowledge recommendation and knowledge reasoning are discussed in Section 4. The paper is concluded in Section 5.

2. Knowledge Extraction and Quality Inspection Framework

The proposed automatic knowledge extraction and quality inspection framework for rock descriptions involves several processes, including rock named entity and relation definitions, NER, relation extraction and knowledge consistency calculation. Figure 2 shows the process of the proposed framework. First, the types of the named entities and relations of rock descriptions are defined according to the prior petrographic knowledge, and the semi-automatic formulation of the petrographic KG is completed. Second, according to the defined entity and relation types, manual annotation of petrographic descriptions is carried out to formulate the labelled corpus. The corpus is divided into a training dataset, validation dataset and testing dataset, according to the general practice of supervised learning methods. In this paper, a pipeline mode is adopted for petrographic knowledge

extraction, which consists of two closely linked components, namely, NER and relation extraction. Training and fine-tuning of rock NER and relation extraction models are carried out using the labelled corpus. After inputting a rock description, entities and relations are extracted using the trained models and entity separation is carried out in cases where the entities extracted from the description are complex. Then, the knowledge triples are created from the extracted entities and relations. Finally, using the formulated KG and the extracted knowledge triples, a consistency calculation is carried out on the petrographic knowledge obtained. Geologists verify the validity of the extracted knowledge and consistency calculation through random sampling. Some sampled descriptions which are not extracted correctly are used as an incremental annotation corpus.

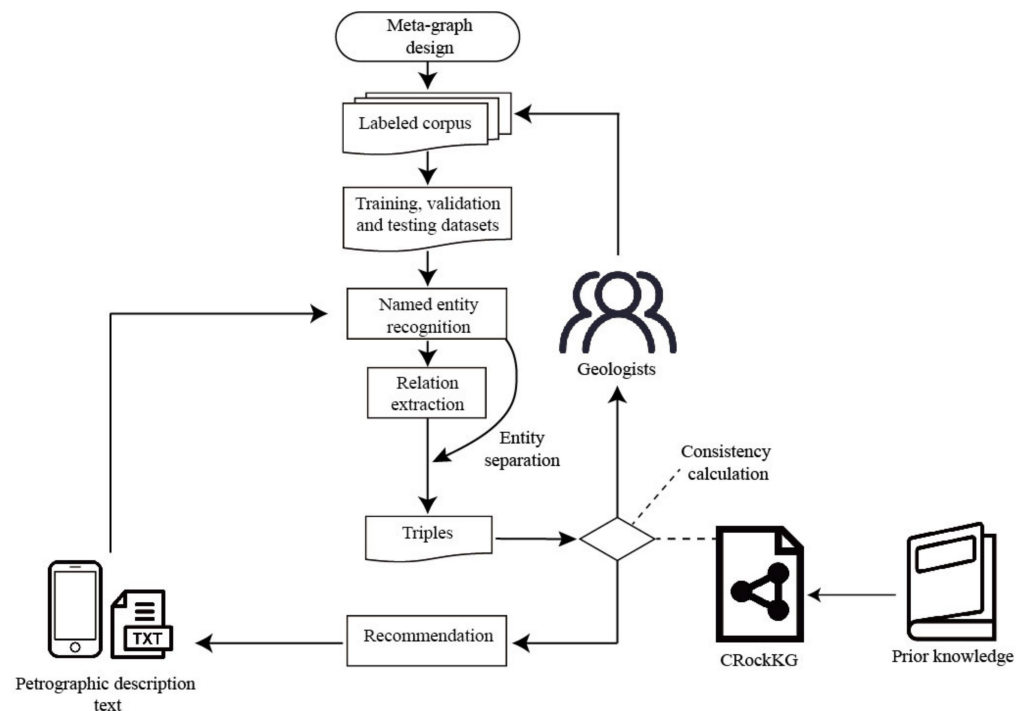


Figure 2. The framework for knowledge extraction and quality inspection of petrographic descriptions.

2.1. Predefinition of Named Entities and Relations Based on Prior Petrographic Knowledge

Scholars have different understandings of the predefinition of geological entities and relations. Wang et al. [39] opined that entity relation extraction in the geological field needs to conform to the diversity of entity and relation types in the domain's corpus. This problem makes accurate predefinition of geological entities and relations difficult. Their proposed solution was to directly extract entities and the relations from the geological texts without predefinition.

Some researchers only carried out the predefinition and annotation of geological named entities, and then the relation types were extracted from the corpus. For example, six entity types (STRAT, ORE_DEPOSIT, MINERAL, ROCK, TIMESCALE, LOCATION) were predefined and labelled in a corpus collected from WAMEX reports, and then 14 relation types (contains, in, associated_with, current_name_of, overlain_by, underlain_by, dominated_by, interbedded_with, aged, bounded_by, occur_in, located_in, intruded_by, hosted_in) were identified after filtering [9].

Geological texts usually contain basic concepts, spatial distribution, attribute information and relations [40]. Chu et al. [41] clearly defined geological named entities with four categories: entity objects (GEO), geological age (TIME), geological processes (PROCESS) and other geological indicators (OTHERS). Xie et al. [42] further subdivided geological named entities into six categories, namely, geological age, geological structure, strata, rock, mineral and location.

In the geological domain, petrographic description texts are different from the above corpus in the geological domain. In petrographic studies, the contents of rock observations and descriptions generally include colour, texture, structure and mineral composition. The type of rocks are classified based on their description and specific classification principles. Therefore, the named entity types in rock description texts can be predefined as rock, colour, texture, structure and mineral.

In rock descriptions, colour is the most striking feature; it is also an important identification characteristic and genesis marker. When observing rocks, fresh and weathered colour should be distinguished. For crystalline rocks, petrographic descriptions need to distinguish the major, minor and accessory minerals. For rocks with a porphyritic or porphyroblastic structure, the description also should contain the comparison of phenocrystic and groundmass minerals. Interstitial materials or cements are also important descriptors for rocks with clastic or granular structures. In summary, the relation types in rock descriptions can be predefined as follows: fresh colour (FRESH_COLOR), weathered colour (WEATH_COLOR), preserved texture (PRESERVE_T), preserved structure (PRESERVE_S), major mineral (MA_MINERAL), minor mineral (MI_MINERAL), accessory mineral (ACC_MINERAL), phenocrystic mineral (PHE_MINERAL) and groundmass mineral (GRO_MINERAL). There are various relation types among named entities and most relations point to the same rock entity. Hence, the relations in the rock description can be considered as N -ary relations or single-entity overlapping relations. There is also the subordinate relation type (CATEGORY_OF) between rock entities, which may also exist between mineral entities. To simplify the named entities and relations of rocks, in this study organic matter, fossils, quaternary sediments and related relations were not considered. Figure 3 shows the meta-graph for the named entities and relations of the domain-specific KG of petrography.

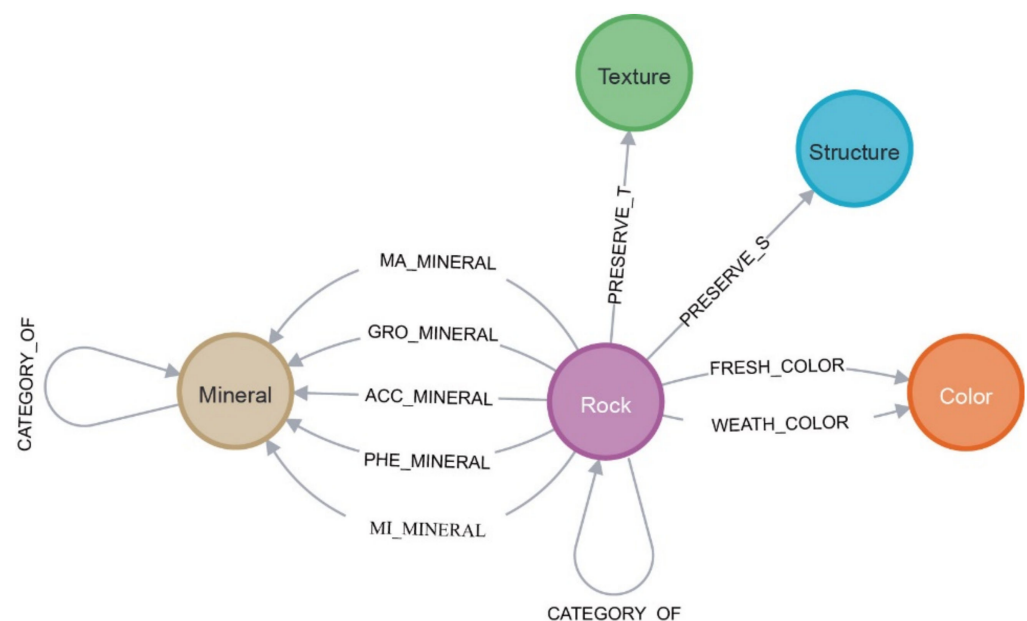


Figure 3. Meta-graph for named entities and relations of the domain-specific KG of petrography.

2.2. Petrographic Named Entity Recognition Based on Sequence Labelling Model

Existing NER approaches are based on rules and the dictionary or on deep learning. An unsupervised geological knowledge extraction method based on the geological domain vocabulary and association rules was proposed for unstructured Chinese documents [27]. In recent years, NER based on deep learning has become the mainstream method [2]. Deep learning methods transform geological NER into sequence labelling. Models, such as DBN [40], BiLSTM-CRF [33] and BiGRU-CRF [34], were used in corresponding experiments. The GRU is a variant of LSTM and its advantages are fewer parameters and faster training.

However, LSTM models are more able to strongly express large amounts of data [43]. The optimal choice between a LSTM or GRU model depends on the specific tasks at hand. With the widespread use of large-scale pre-training Chinese models, some approaches, such as ELMO-CNN-BiLSTM-CRF [41] and BERT-BiLSTM-CRF [17,44], are beginning to be adopted to identify the geological named entities in the geoscience domain. In addition, the emergence of the ELECTRA and XLNet models [45] offer more choices for downstream Chinese natural language processing tasks.

In this paper, the sequence labelling method is also adopted. Based on the labelled corpus of petrographic descriptions, the comparative experiments between bidirectional RNN models (BiLSTM-CRF and BiGRU-CRF) and pre-training Chinese models (BERT, ELECTRA, XLNet) were carried out to determine which model is suitable for NER of petrographic descriptions. The comparison processes for the rock named entity sequence labelling models is shown in Figure 4. The petrographic description texts are first labelled and saved as ANN format files and then tokenised at the Chinese character level. ANN is the file format of the BRAT (Brat Rapid Annotation Tool) [46]. The character-level representation of the input sequence is completed via an embedding layer, and the feature extraction is realised through an encoding layer. The token classification layer is finally used to determine the probability of each entity type. Models based on RNNs adopt a randomly initialised embedding layer and a CRF classification layer. However, the models based on pre-trained models only require fine-tuning of the dense layer.

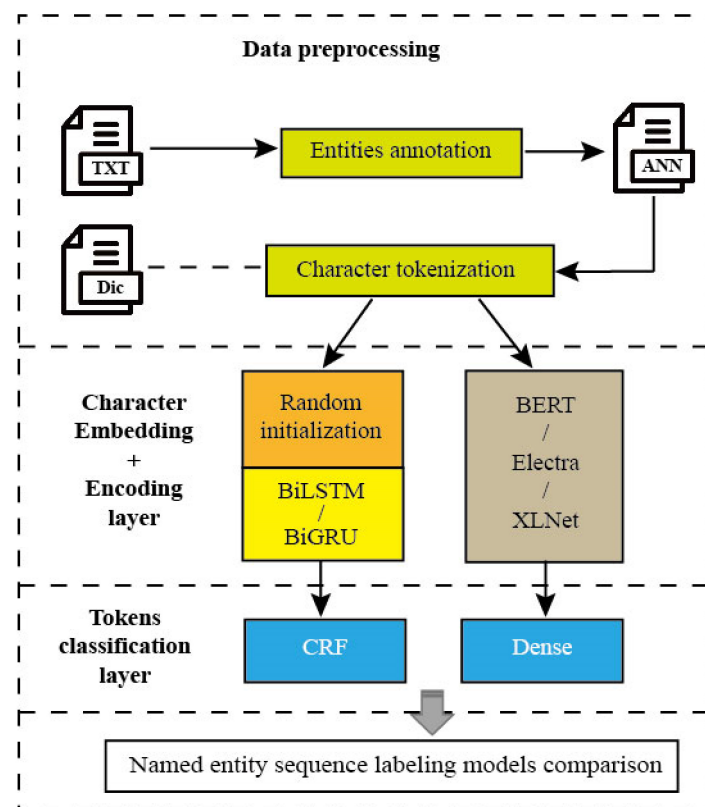


Figure 4. Framework for the named entity sequence labelling model comparison.

2.3. Petrographic Relation Extraction Based on Enriching R-Transformer Model

As mentioned above, relation extraction comprises binary relation extraction, N -ary relation extraction and entity overlapping relation extraction. Binary relation extraction was proposed earlier as a means to identify the relation between two entities in a single input sentence [47]. N -ary relation extraction pertains to the recognition of relations among n entities through multiple sentences [48]. As shown in Figure 1b, the relations among the three entities also need to be classified. The possible relation categories between entities

are predefined. In addition, “NA” is included in the predefined relation set to indicate that there is no association between entities. Overlapping relation extraction means that different relation triples in one or more sentences may have various degrees of overlap [37]. In general, there are two forms of overlap: single-entity overlap (SEO) and entity pair overlap (EPO). SEO refers to cases where triples share an overlapping entity, but they do not share overlapping entity pairs (Figure 1c). In contrast, EPO refers to the triples sharing overlapping entity pairs (Figure 1d). The extraction of cross relations (Figure 1e) is a challenging problem in geoscience, though some advanced network models were proposed for biomedical cross-sentence relation extraction [49].

Existing methods used to extract relations in the geosciences are mostly based on templates, i.e., a template library is used to match the context of two given entities in the input text. If the context fragment is successfully matched with a template in the library, the corresponding relation in the template is regarded as the relation between the two entities. Template-based methods contain two specific template implementations, one based on trigger words and one based on syntactic structure. Trigger words usually include verbs and prepositions. A word-level relation extraction approach using such trigger words was proposed to identify relations in mineral exploration reports [17]. Methods based on syntactic structure usually take verbs as the starting point to formulate rules that place entities on nodes and the dependency relations on edges. For instance, an open Chinese syntactic structure extraction model was established in the geological field, in which relations were extracted based on the syntactic structure [39]. The model uses the open Chinese language technology platform developed by the Harbin Institute of Technology to analyse the dependency syntax and obtain the syntactic structure. Based on a small number of annotated geological corpora, the syntactic structure-based patterns are automatically learned to obtain the high-frequency relation extraction templates. Finally, the learned templates are used to match the structure of the dependency relation and then identify entities and relations. However, the relation extraction templates in the model only cover the high-frequency syntactic structure, as it is difficult to achieve comprehensive templates. It can be seen that methods based on a template in the geosciences have realised unsupervised relation extraction. However, the overlapping relations, which appear often in geological knowledge descriptions, cannot be determined using syntactic-based relation extraction models and word-level relation extraction methods.

To achieve the overlapping relation extraction from the petrographic descriptions, an approach based on an enriching R-Transformer model is proposed in this paper. The method transforms the relation extraction task into a relation classification task. For single relations between rocks and mineral entities or between rocks and structure entities, relation classification mainly involves determining whether there is a relation between the two entities and the problem is considered as a binary classification problem. If there are multiple possible relations between rocks and colour entities or between rocks and mineral entities in a single rock description sentence, the relation classification is called a multiclassification problem. In this paper, the absence of a relation is considered a special relation type (marked as NA). Sequence semantic feature extraction in the R-Transformer model is based on pre-trained language models such as BERT, XLNet and ELECTRA. The framework of the proposed R-Transformer relation classification model is shown in Figure 5. First, the position of the entity pair in the sequence is marked in the input of the model; thus, the extracted vector representation of the sentence contains the position information of the entity pair. Second, the model extracts semantic information from the sentence vector and the two-entity vectors. Each entity vector is aggregated via an averaging method, and dimension reduction is realised using a fully connected dense layer with the *Tanh* activation function. Third, the two-entity vectors with reduced dimensions are concatenated with the sentence vector, and the annotation classification prediction of each sequence character is realised through the fully connected dense layer, which adopts the softmax multiclassification activation function. Considering the scale of the corpus and the total number of the entity pairs, a dropout layer is added after the combi-

nation layer to deal with the possible over-fitting problem and improve the model performance. The relationship extraction method used in this paper will be presented in detail in another article.

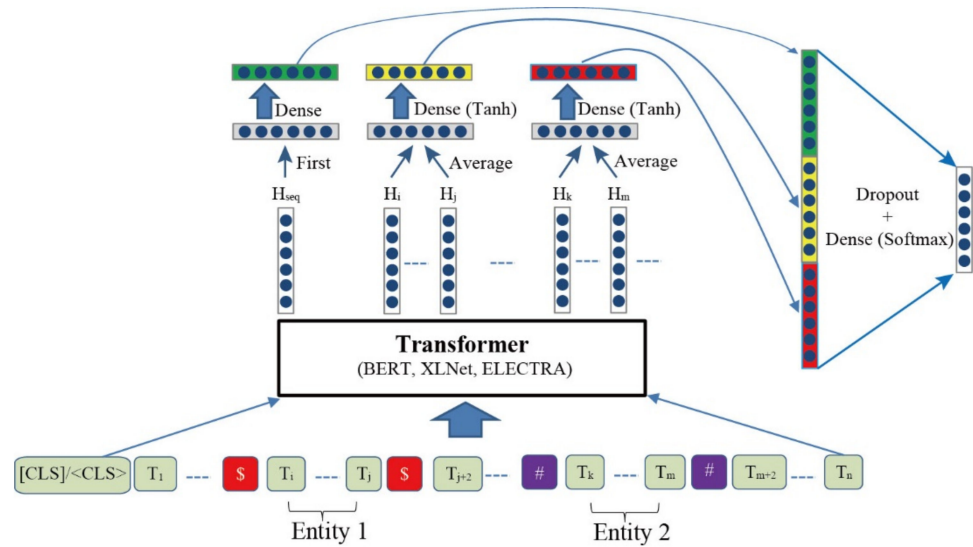


Figure 5. R-Transformer models for relation classification.

2.4. Rule-Based Complex Entity Separation

In general, geological investigators write the Chinese petrographic descriptions according to certain rules. For example, in structure and colour descriptions, often dual terms appear, such as “massive-gneissic structure” (块状-片麻状构造), “grey-light flesh red” (灰红-肉红色), etc., “块状-片麻状构造” is a term in Chinese, and “massive-gneissic structure” is the corresponding translation in English. The same below. The rule of “grain size + minor mineral morphology + major mineral morphology” is often used to describe rocks with a granoblastic structure in Chinese geological texts. Thus, the extraction of such entities with complex descriptions is an important problem to be solved in this process.

Sequence labelling models based on deep learning require manual entity annotation to realise semantic information extraction of the labelled entities. However, dual-construct entities are usually labelled as single entities, thus models trained on corpora annotated in this manner usually recognise the dual-construct entity as a single entity. To realise the extraction of dual-structure entities, it is necessary to separate entities based on rules. In this paper, dual-construct entities were split and reformed according to the concatenation character using the complex entity separation algorithm shown in Algorithm 1. For example, after splitting and reformation, “massive-gneissic structure” (块状-片麻状构造) was extracted as two entities, namely, “massive structure” (块状构造) and “gneissic structure” (片麻状构造).

Algorithm 1. Complex entity separation algorithm.

Input: a complex entity

Output: entities separated

1: input complex entity containing the entity type

2: if entity type is Texture

3: if entity is blastic texture and len (entity) > 7

4: execute extraction of grain size, minor and major mineral textures

5: else if entity type is Structure

6: if concatenation characters are present in entity

7: execute entity separation based on the concatenation character

8: return entities

3. Experimental Results

In this paper, the 1:50,000 sheet of Fengxiangyi located in the Dabie-Sulu orogenic belt in central and eastern China (Figure 6a) was selected for the empirical research. From 2014 to 2016, the Institute of Geological Survey of Anhui province carried out a digital geological and mineral survey in this area [50], thus creating a large number of electronic rock description texts. Middle-deep metamorphic strata and Neoproterozoic intermediate-acid metamorphic intrusive rocks, which are part of the core of the Dabie-Sulu orogenic belt [51], are widely distributed in this area. Figure 6b shows the major distribution of the metamorphic plutonic rocks and metamorphic supracrustal rocks in the studied sheet, including paragneiss, granitic gneiss, monzogranitic gneiss, granodioritic gneiss, eclogite, amphibolite, marble, quartz-muscovite schist and quartzite. Quaternary sediments are not studied in this paper. The rock descriptions were typical *N*-ary and single-entity overlapping relation texts. For example, the description text of the quartz-muscovite schist covers the single-entity overlapping relation between rock and structure (or texture). It also contains some *N*-ary relations between rock and mineral, including the major, minor and accessory minerals between rock and mineral. The structure and texture in the metamorphic rock description texts are typical complex entities. For example, the structural description of monzogranitic gneiss is a “massive-gneissic structure” (块状-片麻状构造), which is a double-structural entity. Its structure is also a typical granular crystal structure, which is usually described using multistructure description modes, such as the “lepidoblastic granoblastic texture” (鳞片花岗变晶结构).

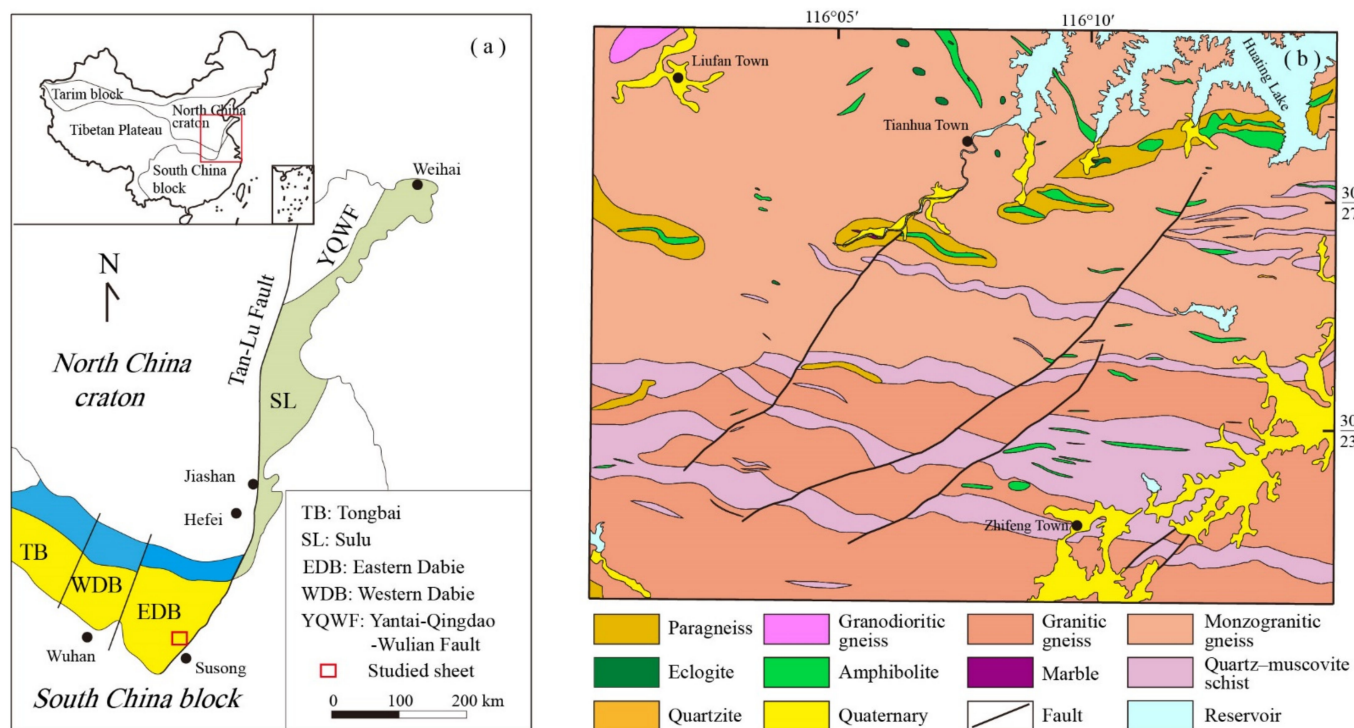


Figure 6. (a) Schematic tectonic map for the Dabie-Sulu orogenic belt in central and eastern China, showing the location of the 1:50000 Fengxiangyi sheet. The inset shows the major tectonic divisions of China (modified after Qiu et al., 2021 [52]); (b) simplified geological map of the Fengxiangyi sheet, showing the distribution of rocks.

3.1. Construction of the Prior Petrographic KG

Once a medium-scale regional geological survey has taken place in an area, e.g., at a scale between 1:200,000 and 1:250,000, the rock types in the region are generally known. According to the petrographic knowledge, previous survey reports and expertise, the textures, structures and material composition of the different rock types also are known.

Therefore, in this paper, a KG is constructed based on prior knowledge for the inspection of rock description texts. Due to the high credibility of prior knowledge, a probabilistic database approach is not adopted in this paper.

The rock types and characteristics in the experimental sheet were comprehensively summarised in the survey report, which formed the prior knowledge for the formulation of the domain-specific petrographic KG. Taking the Neoproterozoic intermediate-acid metamorphic plutonic rocks as an example, the rock types mainly contain monzogranitic gneiss, granitic gneiss and granodioritic gneiss. These plutons are ancient intrusions disintegrated from the original Susong Group, and have undergone multistage metamorphism and deformation [50]. Table 1 summarises the prior knowledge on the metamorphic deformation intrusions in the Fengxiangyi sheet, including rock type, texture, structure and mineral composition. The characteristics of rock composition are described by means of major, minor and accessory minerals.

Table 1. Prior knowledge of the metamorphic plutonic rocks in the Fengxiangyi sheet.

Rock Type	Texture	Structure	Major Minerals	Minor Minerals	Accessory Minerals
Monzogranitic gneiss (二长花岗质片麻岩)	Lepidoblastic texture (鳞片变晶结构), granoblastic texture (花岗变晶结构), porphyroclastic texture (碎斑结构), blastogranitic texture (变余花岗结构), coarse-grained blastic texture (粗粒变晶结构), medium-grained blastic texture (中粒变晶结构)	Gneissic structure (片麻状构造), massive structure (块状构造), weak gneissic structure (弱片麻状构造), ophthalmic structure (眼球状构造), streaky structure (条纹状构造), banded structure (条带状构造)	Quartz (石英), plagioclase (斜长石), k-feldspar (钾长石)	Biotite (黑云母), muscovite (白云母), epidote (绿帘石)	Zircon (锆石), sphene (榍石), magnetite (磁铁矿), garnet (石榴子石)
Granitic gneiss (花岗质片麻岩)	Lepidoblastic texture (鳞片变晶结构), granoblastic texture (花岗变晶结构), blastogranitic texture (变余花岗结构), fine-grained blastic texture (细粒变晶结构)	Gneissic structure (片麻状构造)	Quartz (石英), k-feldspar (钾长石), plagioclase (斜长石)	Biotite (黑云母), muscovite (白云母)	Zircon (锆石), apatite (磷灰石), rutile (金红石), ilmenite (钛铁矿), magnetite (磁铁矿), garnet (石榴子石)
Granodioritic gneiss (花岗闪长质片麻岩)	Blastogranitic texture (鳞片变晶结构), porphyroclastic texture (碎斑结构)	Gneissic structure (片麻状构造), massive structure (块状构造)	Quartz (石英), plagioclase (斜长石), k-feldspar (钾长石)	Biotite (黑云母), hornblende (角闪石)	Magnetite (磁铁矿), sphene (榍石), zircon (锆石)

Note: monzogranitic gneiss (二长花岗质片麻岩): “二长花岗质片麻岩” is a term in Chinese. “Monzogranitic gneiss” is the corresponding translation in English.

3.2. Knowledge Extraction and Quality Inspection

The quality inspection task requires that the computer system can accurately extract named entities and relations from the input texts. In the proposed quality inspection framework, the sequence labelling model and the enhanced R-Transformer relation classification model recognise the named entities and extract the relations from the input rock description texts in a pipeline mode. The extracted entities and relations are eventually composed into knowledge triples. The rock types in the selected sheet are mainly metamorphic rocks, which are mostly classified on the basis of texture (grain size, shape, orientation), structure and mineral composition. Based on the extraction of knowledge triples and the petrographic KG constructed based on prior knowledge, a quality inspection of the rock description texts can be realised based on the consistency between rock knowledge and knowledge triples.

Figure 7 shows the calculation process applied for knowledge alignment. The first step is the consistency calculation of the extracted texture, structure and mineral composition information. Based on the extracted rock type, Cypher, a graphic query language, is used to match the extracted textures, structures, major minerals and minor minerals with their corresponding information of this rock type in the KG one by one. To evaluate the matching results of step 1, if there are mismatched extraction knowledge triples, step 2 is executed. If all triples match, the algorithm proceeds to step 3. The unmatched triples of step 2 may be an error description or new knowledge, and the program returns the mismatch information. At the same time, the program automatically saves the rock description text to the corpus, which should be manually verified. Step 3 involves matching the rock entity extracted from the rock description text with the rock entities in the petrographic KG, which conform to the characteristics of the extracted texture, structure, major minerals and minor minerals. The output of this step is the number of rock entities that match the description. The process is terminated if only one match is found; if two or more entities are returned in step 3, then step 4 is executed. In step 4, the program indicates that there are some rock entities with the same descriptive characteristics. The knowledge identified between the rock entities is returned. This step plays the role of knowledge recommendation while conducting the consistency calculation.

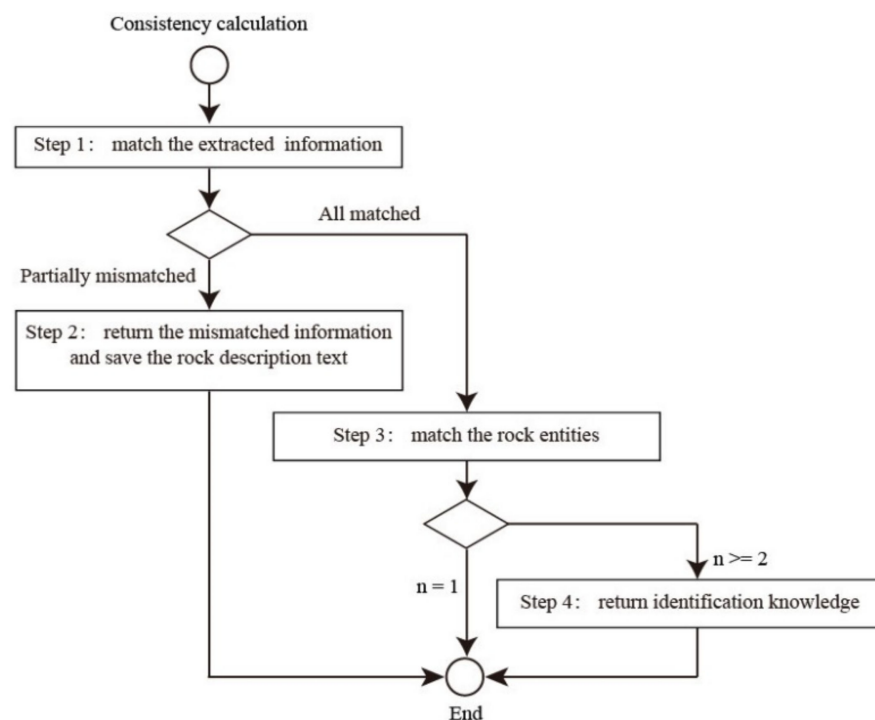


Figure 7. Knowledge consistency calculation.

As an example, in Table 2 the rock description text for a granitic gneiss outcrop is presented. The extracted rock entity name, structure entities, texture entities and mineral entities, along with the relations of the major and minor minerals, are described. Figure 8a is the subgraph for granitic gneiss. The consistency calculation for the extracted triples went through steps 1 and 3. In the petrographic KG, only granitic gneiss has the same characters extracted from the rock description text. Table 3 is the rock description text for another eclogite outcrop, and Figure 8b is the subgraph of the petrographic KG for eclogite. The consistency calculation after knowledge extraction also involves the execution and termination of steps 1 and 3.

Table 2. An example of description text extraction for granitic gneiss.

Description text in Chinese	点南为灰白色厚层状花岗质片麻岩, 细粒鳞片花岗变晶结构, 片麻状-块状构造, 主要矿物斜长石50%, 钾长石20%, 石英20%, 他形粒状, 少量黑云母.		
Description text	The south of the point is greyish-white, thick, layered granitic gneiss; fine-grained and lepidoblastic-granoblastic structure, gneissic-mass structure; major minerals: plagioclase 50%, k-feldspar 20%, quartz 20%, xenomorphic crystal, a small amount of biotite.		
Structure	Gneissic structure (片麻状构造), massive structure (块状构造)	Texture	Fine-grained blastic texture (细粒变晶结构), lepidoblastic texture (鳞片变晶结构), granoblastic texture (花岗变晶结构)
Major mineral	Plagioclase (斜长石), k-feldspar (钾长石), quartz (石英)	Minor mineral	Biotite (黑云母)
Extracted rock entity	Granitic gneiss (花岗质片麻岩)		

Note: (1) Gneissic structure (片麻状构造): “片麻状构造” is a term in Chinese, “Gneissic structure” is the corresponding translation in English. (2) The named entities are underlined in the Chinese text.

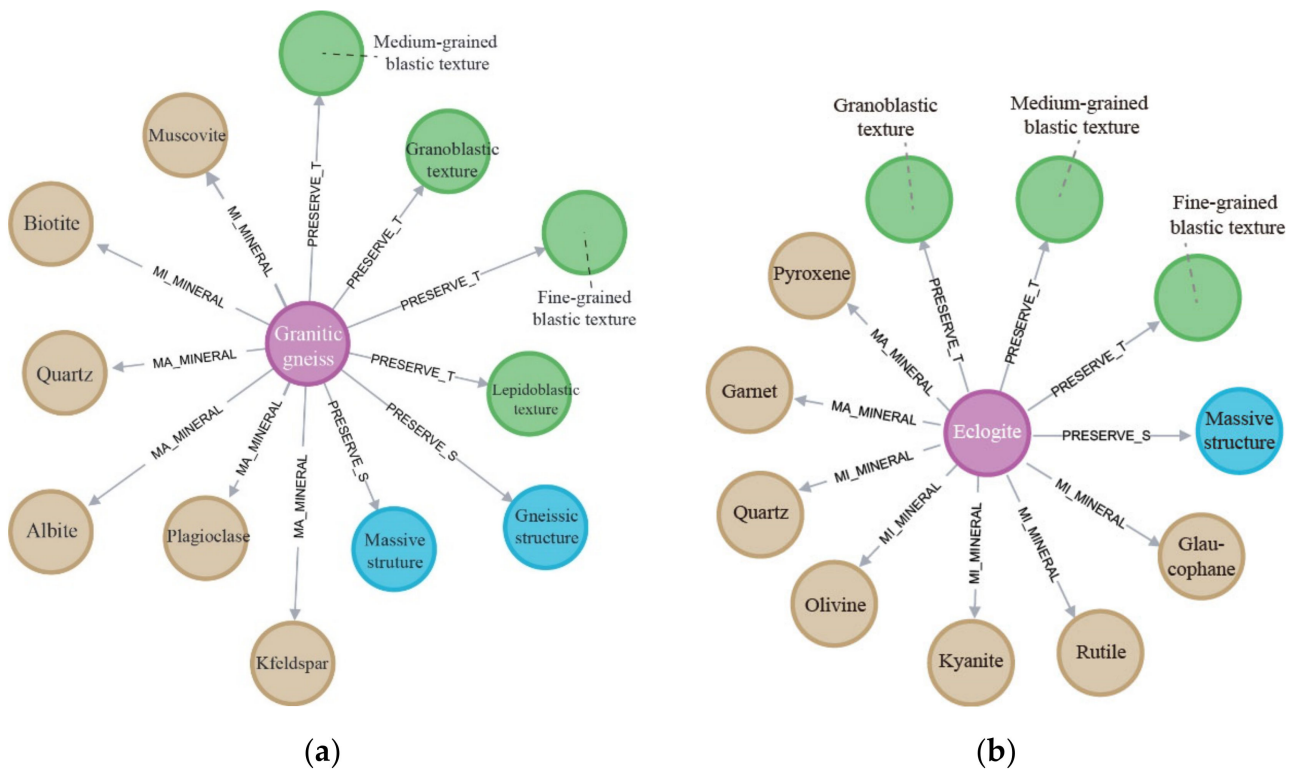


Figure 8. (a) Granitic gneiss subgraph; (b) eclogite subgraph.

Table 3. An example of description text extraction for eclogite.

Description text in Chinese	点北为榴辉岩, 灰绿色, 细粒变晶结构, 块状构造, 主要由石榴子石30%, 辉石70%组成, 矿物颗粒较细小, 多在0.5 mm左右, 石榴子石多风化呈红褐色圆粒状.		
Description text	The north of the point is eclogite. Grey-green, fine-grained blastic texture, mass structure, mainly composed of garnet 30%, pyroxene 70%. Mineral particles are small, mostly around 0.5 mm. Weathered garnet is mahogany and has a rounded grain.		
Structure	Massive structure (块状构造)	Texture	Fine-grained blastic texture (细粒变晶结构)
Major mineral	Garnet (石榴子石), pyroxene (辉石)	Minor mineral	
Extracted rock entity	Eclogite (榴辉岩)		

Note: (1) Massive structure (块状构造): “块状构造” is a term in Chinese, “Massive structure” is the corresponding translation in English. (2) The named entities are underlined in the Chinese text.

For the quartz-muscovite schist, the application of the knowledge extraction and consistency calculation process demonstrated that muscovite is described as the major mineral in most quartz-muscovite schist description texts. However, muscovite is the minor mineral in the standard description of quartz-muscovite schist. The process executed steps 1 and 2 in turn and terminated. The related rock description text was automatically stored in the corpus waiting for manual validation by users of the proposed framework. A review by geologists revealed that the reason for the mismatch was the imprecise description by investigators.

4. Discussion

4.1. Error Transformation in Pipeline Mode

In this paper, comparative experiments were carried out on the sequence labelling models used for the rock NER and the R-Transformer relation classification models used for the relation extraction. Table 4 shows the results of the comparative experiments, which shows that the sequence labelling model and the relation classification model based on BERT achieved the best performance in the naming entity and relation extraction based on the *F1* scores. In particular, the *F1* value of the BERT-based sequence labelling model reached 98.04%. This high accuracy can reduce the errors of the NER stage effectively and remedies the deficiencies of the error transmission in the pipeline mode. Meanwhile, there was relatively little difference between the performance of the various models. The possible reason is the corpus size. Further experimental studies on the comparison of different models under corpora of different scales will be carried out in the future.

Table 4. The performance of different models in a sequence labelling task and relation extraction task. Bold marks indicate the best performance in all methods.

	Indicator	BiLSTM-CRF	BiGRU-CRF	BERT	XLNet	ELECTRA
Entity	<i>p</i>	97.81	97.14	97.57	95.38	97.38
	<i>R</i>	97.81	98.33	98.51	97.63	97.81
	<i>F1</i>	97.81	97.74	98.04	96.49	97.60
Relation	<i>p</i>	-	-	91.77	91.15	90.84
	<i>R</i>	-	-	94.71	93.32	90.56
	<i>F1</i>	-	-	93.22	92.22	90.70

Note: “-”: non-execution.

4.2. Integration of Variant Data and Specifications

In digital mapping systems, apart from the unstructured data, there are also important structured data, which are more important, such as the location, landform and mapping unit of the geological observation point. At present, the objects of information extraction in

the geosciences are mostly unstructured data, including texts and documents. However, based on the experience of domain-specific KG formulation in other fields, structured data are also an important source of knowledge. The integration of structured and unstructured data to realise the rapid construction of a large-scale KG is an aspect that needs further research.

In geological texts, a common occurrence is that the use of the terms “texture” and “structure” is confused. However, in petrology specifications, such as the terminology classification and code of geology mineral resources, Part 10: Petrology (GB/T 9649.10-2009), the terms “texture” and “structure” have unambiguous definitions. In this paper, standardised terms are stored in the KG of petrography as a form of prior knowledge. If a nonstandard entity term is extracted or separated, a triple consisting of the term will have difficulty passing the consistency calculations.

4.3. Knowledge Recommendation and Knowledge Reasoning

This study takes the extraction and inspection of petrographic knowledge with complex entities and relations as the research object. As described in Section 3.2, when matching the rock description characteristics with the rock entities in the petrographic KG, there may be more than one match with the same description characteristics. In particular, some metamorphic rocks have the same fabric characteristics and mineral composition. Owing to their different geological environments (geological occurrences), the basic names of rocks may vary greatly, resulting in the phenomenon of synonymy of the same rock. For example, massive rocks mainly composed of muscovite and quartz are named muscovite quartzite formed by the regional metamorphism, but those formed through gas-liquid metamorphism of granitic rocks are also named muscovites. In such cases, the process needs not only to return the possible rock entity matches, but also to prompt the geological investigator to pay more attention to the field observation of the geological occurrence. Therefore, apart from the quality inspection of petrographic descriptions, rock identification knowledge recommendation is another possible application of petrographic KGs.

Another potential application of KGs is knowledge reasoning. Generally, metamorphic facies can be determined according to the minerals and mineral assemblages of metamorphic rocks. Since the proposed framework can be used to obtain mineral information of metamorphic rocks in the studied sheet through machine reading, the metamorphic facies of the metamorphic strata can be inferred based on the mineral information of the rocks which belong to the metamorphic strata and the computable and stored decision rules in the KG.

5. Conclusions

In this study, the methods for automatic knowledge extraction and quality inspection of petrographic description texts with complex entities and relations were investigated. A framework which contains rock named entity and relation definitions was proposed based on prior petrographic knowledge, rock NER based on a sequence labelling model, petrographic relation extraction based on an enriching R-Transformer relation classification model and rule-based complex entity separation. Considering the high accuracy of NER, the framework allows for rock named entity sequence labelling and relation classification in a pipeline mode. The petrographic descriptions of regional metamorphic rock types in the sheet of Fengxiangyi located in the Dabie orogen were selected as the experimental dataset. The experimental results showed that: (1) Large-scale pre-trained language models are suitable for complex entity recognition and complex relation extraction on small-scale petrographic description texts. (2) The framework proposed in this paper can automatically extract knowledge from petrographic descriptions of regional metamorphic rocks in the Dabie orogen. (3) Adoption of the proposed method for KG-based quality inspection can lead to improvements in rock description quality and avoid obviously inconsistent descriptions.

Author Contributions: Conceptualisation, F.Y. and Z.C.; methodology, Z.C.; software, Z.C. and B.W.; validation, X.L. and H.L.; formal analysis, X.L.; resources, Z.C. and X.W.; data curation, F.Y.; writing—original draft preparation, Z.C.; writing—review and editing, F.Y.; visualisation, Z.C. and Y.C.; supervision, F.Y.; project administration, Z.C.; funding acquisition, F.Y. and X.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research is financially supported by the National Natural Science Foundation of China (Grant No. 41820104007, 42072321) and the Natural Science Foundation of Anhui Province, China (Grant No. 2208085MD96).

Acknowledgments: We thank Xu Jinlong and Wu Heng from the Geological Survey of Anhui Province for their great help in the collection and annotation of the petrographic description corpus. The authors are also grateful to all anonymous reviewers for their comments and suggestions.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Karpatne, A.; Ebert-Uphoff, I.; Ravela, S.; Babaie, H.A.; Kumar, V. Machine Learning for the Geosciences: Challenges and Opportunities. *IEEE Trans. Knowl. Data Eng.* **2019**, *31*, 1544–1554. [[CrossRef](#)]
2. Zhou, Y.Z.; Zuo, R.G.; Liu, G.; Yuan, F.; Mao, X.C.; Guo, Y.J.; Xiao, F.; Liao, J.; Liu, Y.P. The great-leap-forward development of mathematical geoscience during 2010–2019: Big Data and artificial intelligence algorithm are changing mathematical geoscience. *Bull. Mineral. Petrol. Geochem.* **2021**, *40*, 556–573.
3. Sun, Z.; Sandoval, L.; Crystal-Ornelas, R.; Mousavi, S.M.; Wang, J.; Lin, C.; Cristea, N.; Tong, D.; Carande, W.H.; Ma, X.; et al. A Review of Earth Artificial Intelligence. *Comput. Geosci.* **2022**, *159*, 105034. [[CrossRef](#)]
4. Bergen, K.J.; Johnson, P.A.; De Hoop, M.V.; Beroza, G.C. Machine Learning for Data-Driven Discovery in Solid Earth Geoscience. *Science* **2019**, *363*, eaau0323. [[CrossRef](#)]
5. Lary, D.J.; Alavi, A.H.; Gandomi, A.H.; Walker, A.L. Machine Learning in Geosciences and Remote Sensing. *Geosci. Front.* **2016**, *7*, 3–10. [[CrossRef](#)]
6. Jia, L.; Yang, M.; Meng, F.; He, M.; Liu, H. Mineral Photos Recognition Based on Feature Fusion and Online Hard Sample Mining. *Minerals* **2021**, *11*, 1354. [[CrossRef](#)]
7. Sun, G.; Huang, D.; Cheng, L.; Jia, J.; Xiong, C.; Zhang, Y. Efficient and Lightweight Framework for Real-Time Ore Image Segmentation Based on Deep Learning. *Minerals* **2022**, *12*, 526. [[CrossRef](#)]
8. Chow, B.H.Y.; Reyes-Aldasoro, C.C. Automatic Gemstone Classification Using Computer Vision. *Minerals* **2022**, *12*, 60. [[CrossRef](#)]
9. McCoy, J.T.; Auret, L. Machine Learning Applications in Minerals Processing: A Review. In *Minerals Engineering*; Elsevier Ltd.: Amsterdam, The Netherlands, 2019; pp. 95–109.
10. Zhou, G.Y.; Zhang, M.M.; Shen, L.; Zhang, S.H.; Yuan, F.; Li, X.H.; Ji, B.; Zhou, Y.Z. Data mining of deep geological spatial information of the Yaojialing Zinc-gold polymetallic deposit. *Geotecton. Metallogenia* **2020**, *44*, 242–250.
11. Zhou, C.H.; Wang, H.; Wang, C.S.; Hou, Z.Q.; Zheng, Z.M.; Shen, S.Z.; Cheng, Q.M.; Feng, Z.Q.; Wang, X.B.; Lv, H.R.; et al. Prospects for the Research on Geoscience Knowledge Graph in the Big Data Era. *Sci. China Earth Sci.* **2021**, *64*, 1105–1114. [[CrossRef](#)]
12. Gil, Y.; Pierce, S.A.; Babaie, H.; Banerjee, A.; Borne, K.; Bust, G.; Cheatham, M.; Ebert-Uphoff, I.; Gomes, C.; Hill, M.; et al. Intelligent Systems for Geosciences: An Essential Research Agenda. *Commun. ACM* **2019**, *62*, 76–84. [[CrossRef](#)]
13. Jiang, S.J.; Zheng, Y.; Solomatine, D. Improving AI System Awareness of Geoscience Knowledge: Symbiotic Integration of Physical Approaches and Deep Learning. *Geophys. Res. Lett.* **2020**, *47*, e2020GL088229. [[CrossRef](#)]
14. Wagener, T.; Dadson, S.J.; Hannah, D.M.; Coxon, G.; Beven, K.; Bloomfield, J.P.; Buytaert, W.; Cloke, H.; Bates, P.; Holden, J.; et al. Knowledge Gaps in Our Perceptual Model of Great Britain’s Hydrology. *Hydrol. Process.* **2021**, *35*, e14288. [[CrossRef](#)]
15. Sherlock, M.J.; Hasan, M.; Samavati, F.F. Interactive Data Styling and Multifocal Visualization for a Multigrid Web-Based Digital Earth. *Int. J. Digit. Earth* **2021**, *14*, 288–310. [[CrossRef](#)]
16. Kase, S.E.; Hung, C.P.; Krayzman, T.; Hare, J.Z.; Rinderspacher, B.C.; Su, S.M.M. The Future of Collaborative Human-Artificial Intelligence Decision-Making for Mission Planning. *Front. Psychol.* **2022**, *13*, 1246.
17. Enkhsaikhan, M.; Holden, E.-J.; Durning, P.; Liu, W. Understanding Ore-Forming Conditions Using Machine Reading of Text. *Ore Geol. Rev.* **2021**, *135*, 104200. [[CrossRef](#)]
18. Berardi, M.; Amato, L.S.; Cigna, F.; Tapete, D.; de Cumis, M.S. Text Mining from Free Unstructured Text: An Experiment of Time Series Retrieval for Volcano Monitoring. *Appl. Sci.* **2022**, *12*, 3503. [[CrossRef](#)]
19. Grishman, R. Twenty-Five Years of Information Extraction. *Nat. Lang. Eng.* **2019**, *25*, 677–692.
20. Kopperud, B.T.; Lidgard, S.; Liow, L.H. Text-Mined Fossil Biodiversity Dynamics Using Machine Learning. *Proc. R. Soc. B Biol. Sci.* **2019**, *286*, 20190022. [[CrossRef](#)]
21. Abu-Salih, B. Domain-Specific Knowledge Graphs: A Survey. *J. Netw. Comput. Appl.* **2021**, *185*, 103076. [[CrossRef](#)]
22. Liu, C.; Chen, J.; Li, S.; Qin, T. Construction of Conceptual Prospecting Model Based on Geological Big Data: A Case Study in Songtao-Huayuan Area, Hunan Province. *Minerals* **2022**, *12*, 669. [[CrossRef](#)]

23. Ma, X.G. Knowledge Graph Construction and Application in Geosciences: A Review. *Comput. Geosci.* **2022**, *161*, 105082. [CrossRef]
24. Wang, B.; Ma, K.; Wu, L.; Qiu, Q.J.; Xie, Z.; Tao, L.F. Visual Analytics and Information Extraction of Geological Content for Text-Based Mineral Exploration Reports. *ORE Geol. Rev.* **2022**, *144*, 104818. [CrossRef]
25. Peters, S.E.; Zhang, C.; Livny, M.; Ré, C. A Machine Reading System for Assembling Synthetic Paleontological Databases. *PLoS ONE* **2014**, *9*, e113523. [CrossRef]
26. Peters, S.E.; Husson, J.M.; Wilcots, J. The Rise and Fall of Stromatolites in Shallow Marine Environments. *Geology*. **2017**, *45*, 487–490. [CrossRef]
27. Zhu, Y.; Zhou, W.; Xu, Y.; Liu, J.; Tan, Y. Intelligent Learning for Knowledge Graph towards Geological Data. *Sci. Program.* **2017**, *2017*, 5072427. [CrossRef]
28. Wang, C.; Ma, X.; Chen, J.; Chen, J. Information Extraction and Knowledge Graph Construction from Geoscience Literature. *Comput. Geosci.* **2018**, *112*, 112–120. [CrossRef]
29. Ji, S.; Pan, S.; Cambria, E.; Marttinen, P.; Yu, P.S. A Survey on Knowledge Graphs: Representation, Acquisition, and Applications. *IEEE Trans. Neural Netw. Learn. Syst.* **2021**, *33*, 494–514. [CrossRef]
30. Grohe, M.; Lindner, P. Infinite Probabilistic Databases. *Log. Methods Comput. Sci.* **2022**, *18*, 34. [CrossRef]
31. Ceylan, I.I.; Darwiche, A.; Van den Broeck, G. Open-World Probabilistic Databases: Semantics, Algorithms, Complexity. *Artif. Intell.* **2021**, *295*, 103474. [CrossRef]
32. Amarilli, A.; Ceylan, I.I. The Dichotomy of Evaluating Homomorphism-Closed Queries on Probabilistic Graphs. *Log. Methods Comput. Sci.* **2022**, *18*, 2. [CrossRef]
33. Liu, W.C.; Zhang, C.J.; Wang, C.; Zhang, X.Y.; Zhu, Y.Q.; Jiao, S.T.; Lu, Y.X. Geological time information extraction from Chinese text based on BiLSTM-CRF. *Adv. Earth Sci.* **2021**, *36*, 211–220.
34. Fan, R.; Wang, L.; Yan, J.; Song, W.; Zhu, Y.; Chen, X. Deep Learning-Based Named Entity Recognition and Knowledge Graph Construction for Geological Hazards. *ISPRS Int. J. Geo-Inf.* **2019**, *9*, 15. [CrossRef]
35. Qi, H.; Dong, S.C.; Zhang, L.L.; Hu, H.; Fan, J.X. Construction of earth science knowledge graph and its future perspectives. *Geol. J. China Univ.* **2020**, *26*, 2–10.
36. Zhou, Y.Z.; Zhang, Q.L.; Huang, Y.J.; Yang, W.; Xiao, F.; Ji, J.J.; Han, F.; Tang, L.; Ouyang, C.; Shen, W.J. Constructing knowledge graph for the porphyry copper deposit in the Qingzhou Hangzhou area: Insight into knowledge graph based mineral resource prediction and evaluation. *Earth Sci. Front.* **2021**, *28*, 67–75.
37. Zeng, X.; Zeng, D.; He, S.; Liu, K.; Zhao, J. Extracting Relational Facts by an End-to-End Neural Model with Copy Mechanism. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), Melbourne, Australia, 15–20 July 2018*; Association for Computational Linguistics: Melbourne, VIC, Australia, 2018; pp. 506–514.
38. Yang, M.L.; Xue, L.F.; Ran, X.J.; Sang, X.J.; Yan, Q.; Dai, J.H. Intelligent mineral geological survey method: Daqiao-Yawan area in Gansu Province as an example. *Acta Petrol. Sinica* **2021**, *37*, 3880–3892.
39. Wang, Z.G.; Wen, H.Y.; Lu, Q.; Shen, H.K. Joint extraction of open entity relation in geological field. *Comput. Eng. Design* **2021**, *42*, 996–1005.
40. Zhang, X.Y.; Ye, P.; Wang, S.; Du, M. Geological entity recognition method based on deep belief networks. *Acta Petrol. Sinica* **2018**, *34*, 343–351.
41. Chu, D.P.; Wan, B.; Li, H.; Fang, F.; Wang, R. Geological entity recognition based on ELMO-CNN-BiLSTM-CRF model. *Earth Sci.* **2021**, *46*, 3039–3048.
42. Xie, X.J.; Xie, Z.; Ma, K.; Chen, J.G.; Qiu, Q.J.; Li, H.; Pan, S.Y.; Tao, L.F. Geological entity recognition based on BERT and BiGRU-Attention-CRF model. *Geological Bulletin of China*. 2021. Available online: <https://kns.cnki.net/kcms/detail/11.4648.p.20210913.1040.002.html> (accessed on 12 March 2022).
43. Greff, K.; Srivastava, R.K.; Koutník, J.; Steunebrink, B.R.; Schmidhuber, J. LSTM: A Search Space Odyssey. *IEEE Trans. Neural Networks Learn. Syst.* **2017**, *28*, 2222–2232. [CrossRef]
44. Chen, Z.L.; Yuan, F.; Li, X.H.; Zhang, M.M. Based on BERT-BiLSTM-CRF model the named entity and relation joint extraction of Chinese lithological description corpus. *Geol. Rev.* **2022**, *68*, 742–750.
45. Cui, Y.; Che, W.; Liu, T.; Qin, B.; Wang, S.; Hu, G. Revisiting Pre-Trained Models for Chinese Natural Language Processing. In *Findings of the Association for Computational Linguistics: EMNLP 2020*; Association for Computational Linguistics: Stroudsburg, PA, USA, 2020; pp. 657–668.
46. Stenetorp, P.; Pyysalo, S.; Topic, G.; Ohta, T.; Ananiadou, S.; Tsujii, J. BRAT: A Web-Based Tool for NLP-Assisted Text Annotation. In *Proceedings of the Demonstrations at the 13th Conference of the European Chapter of the Association for Computational Linguistics, EACL 2012, Avignon, France, 23–27 April 2012*; Association for Computational Linguistics (ACL): Avignon, France, 2012; pp. 102–107.
47. Riedel, S.; Yao, L.; McCallum, A. Modeling Relations and Their Mentions without Labeled Text BT. In *Machine Learning and Knowledge Discovery in Databases*; Balcázar, J.L., Bonchi, F., Gionis, A., Sebag, M., Eds.; Springer: Berlin/Heidelberg, Germany, 2010; pp. 148–163.
48. Zhao, L.; Xu, W.; Gao, S.; Guo, J. Cross-Sentence N-Ary Relation Classification Using LSTMs on Graph and Sequence Structures. *Knowl.-Based Syst.* **2020**, *207*, 106266. [CrossRef]
49. Zhao, D.; Wang, J.; Lin, H.; Wang, X.; Yang, Z.; Zhang, Y. Biomedical Cross-Sentence Relation Extraction via Multihead Attention and Graph Convolutional Networks. *Appl. Soft Comput.* **2021**, *104*, 107230. [CrossRef]

50. Wang, X.; Guo, J.; Tao, W.; Jiang, L.; Deng, J.; Ma, C. Paleoproterozoic Tectonic Evolution of the Yangtze Craton: Evidence from Magmatism and Sedimentation in the Susong Area, South China. *Precambrian Res.* **2021**, *365*, 106390. [[CrossRef](#)]
51. Yang, Y.; Liu, Y.-C.; Li, Y.; Groppo, C.; Rolfo, F. Zircon U-Pb Dating and Petrogenesis of Multiple Episodes of Anatexis in the North Dabie Complex Zone, Central China. *Minerals* **2020**, *10*, 618. [[CrossRef](#)]
52. Qiu, X.-F.; Tong, X.-R.; Jiang, T.; Khattak, N.U. Reworking of Hadean Continental Crust in the Dabie Orogen: Evidence from the Muzidian Granitic Gneisses. *Gondwana Res.* **2021**, *89*, 119–130. [[CrossRef](#)]