Ontological approach to automating the processes of question generation and knowledge control in intelligent learning systems

Wenzu Li, Natalia Grakova, Longwei Qian
Belarussian State University Informatics and Radioelectronics
Minsk, Belarus
lwzzggml@gmail.com, grakova@bsuir.by, qianlw1226@gmail.com

***Abstract*—This article proposes an approach for designing a general subsystem of automatic generation of questions and automatic verification of answers in intelligent learning systems. The proposed approach is based on the existing approaches and ontology theory, using this approach it is possible to generate various types of questions from the knowledge bases automatically and automatically verify the correctness and completeness of user answers. Compared with the existing approaches, the approach proposed in this article has certain advantages, and the subsystem designed using this approach can be used in systems built using OSTIS Technology.**

***Keywords*—question generation, answer verification, knowledge base, ontology, learning systems**

I. Introduction

With the rapid development of computer technologies and Internet, the way people acquire and store information has changed greatly. Various multimedia software and examination systems have been used by many educational institutions, which has greatly changed the way users learn. However, with the rise of artificial intelligence technologies in recent years, learning methods have moved from the traditional multimedia training mode to the era of intelligent education [5].

The use of advanced learning methods provided by artificial intelligence technologies in the learning process can form a new learning mode, stimulate the user's learning interest and improve the user's learning efficiency. Artificial intelligence methods are used in the field of education, although they can not change the nature of the learning process, but can produce new learning methods and greatly improve the learning efficiency.

Compared with traditional multimedia training systems (MTS), intelligent learning systems (ILS) have the following advantages:

- the learning materials of the MTS are pre-selected by experienced teachers and then stored in the database, so the test question database will appear redundant and duplicated. ILS uses semantic network and ontology approach to store learning materials, which are rich in content and easy to expand and modify;
- users of the MTS cannot ask questions to the system, and the system cannot answer questions to users. But this is a basic function in the ILS;
- when testing users, ILS can use the knowledge in the knowledge base to automatically generate various types of questions, and can also verify the user answers including correctness and integrity checks. ILS can give prompt steps of the test questions according to user requirements, and finally give system reference opinions based on user test results;
- ILS allows users to use natural language to ask questions and answer various types of test questions automatically generated by the system.

Using the knowledge base to automatically generate various types of questions and the automatically verify the user answers is one of the most important features of ILS, but the approaches of question generation proposed by most researchers only allow to generate very simple questions (choice questions; fill in the blank questions and etc.), and the correlation between the correct options and the incorrect options (disturbing options) of the generated choice questions is not high. Therefore, most of the approaches for automatic generation of questions do not meet the requirements of practical applications, and the proposed approaches are implemented on specific systems, so they are not universal. At present, most approaches of answer verification are based on keyword matching and probability statistics and these approaches rely heavily on the quality of the corpus and do not consider the semantic similarity between the answers. Therefore, this article proposes an approach for automatic generation of questions and automatic verification of user answers to solve the above problems. The approach proposed in this article is aimed at developing a general subsystem for automatic generation of questions and automatic verification of answers for systems built on OSTIS technology [1]. The results of the work of the subsystem are implemented in an intelligent learning system for discrete mathematics.

II. Existing approaches and problems

A. Automatic question generation

Automatic question generation method (AQGM) studies how to automatically generate test questions from electronic documents, corpus or knowledge bases through computer technologies.

Unlike traditional approaches of using database, AQGM does not directly extract pre-stored test questions from the database, but uses various types of knowledge sources (including electronic documents, corpus or knowledge bases) to automatically generate questions that meet test requirements. Compared with the traditional approach of using database to extract questions, AQGM is a knowledge-based question generation method, so the generated questions are more flexible [5], [6], [8].

Approaches for automatic generation of questions can be divided into the following categories:

- based on electronic documents;
- based on conceptual corpus;
- based on knowledge base.

Within the approach of extracting questions using electronic documents it is easy to obtain knowledge resources, so the approach has great freedom and the scope of knowledge resources involved is also wide [11]. The knowledge in the corpus is selected by professionals, and this type of knowledge has been filtered, so the knowledge in the corpus is of high quality, which ensures that the quality of the automatically generated questions is high, too. Knowledge in the knowledge base is well structured and verified. Knowledge base is developed after the domain experts analyze the knowledge, this knowledge is filtered and has a certain structure, so the questions automatically generated using the knowledge base are more flexible and diverse, and the quality is the highest [5]. The current research on AQGM is still in its infancy, here are some research results:

- Ding Xiangmin developed an automatic generation system of multiple choice questions, firstly the system summarizes some common Chinese sentence pattern templates by using statistical method, and then uses these sentence pattern templates to extract multiple choice questions from the electronic documents in the aviation field, finally the incorrect (disturbing) options of multiple choice questions are automatically generated by the ontology base in the aviation field developed by Protégé [7], [8];
- The automatic generation system of choice questions developed by Andreas Papasalouros mainly uses the relation between parent class and subclass, class and element, element and attribute in OWL ontology to automatically generate choice questions [10];
- Based on the approaches proposed above, Li Hui used Protégé to create an ontology in the field of computer theory, using various relations between parent class and subclass, element and attribute in the ontology can automatically generate choice questions, fill in the blanks questions and judgment questions, where incorrect options (disturbing) of choice questions are also automatically generated through these relations [5], [7].

Although these approaches discussed above have many advantages, there are also many problems:

- although the approach of using electronic documents and sentence templates to automatically generate questions has abundant sources of knowledge, due to the many types of electronic documents, a large number of sentence templates are required and the types of questions generated are fixed and of low quality;
- the approach of using corpus to automatically generate questions requires the combination of electronic documents and English dictionaries to generate complete questions. The scope and quality of the generated questions depend heavily on the size and quality of the corpus, and the correlation between the incorrect and correct options of the generated choice questions is not high;
- at present, there is no unified standard for the development of most knowledge bases, so different knowledge bases have different knowledge structures and they are not compatible with each other. Because the knowledge bases are not compatible with each other, the approach of using the knowledge base to automatically generate questions can only be used in the corresponding knowledge base, and for the knowledge base developed by other approaches, only new question generation approaches can be developed;
- the approaches introduced above allow to automatically generate only the simplest objective questions, and the generated choice questions answer options are not highly correlated or independent of each other, so these approaches have not reached the conditions for practical application.

Based on the considered existing approaches and in accordance with the OSTIS Technology [1], [2], [3], [4], an approach is proposed for the automatic generation of various types of questions for ILS. This approach allows automatic generation of several types of questions from existing knowledge bases, and then save them to the ontology of questions according to the strategies for automatic generation of questions. The proposed approach will be implemented in the discrete mathematics learning system.

B. Automatic verification of answers

Answer verification is divided into subjective question answer verification and objective question answer verification. Objective questions refer to a type of question with a unique standard answer, and the user answers for this type of question can only be right or wrong, so the answer verification of the objective question only needs to compare whether the user answers and the standard answers are the same. Objective questions include: choice questions, judgment questions, etc. Most subjective questions do not have a unique standard answer, as long as the user's answer conforms to the logic and meets the requirements of the question [17]. Common subjective questions include definition explanation questions, theorem proving questions, etc. In order to verify the answers of subjective questions, it is necessary to compare the similarity between the standard answer and the user's answer [9], [19], [20]. If the similarity is higher, the user's answer is closer to the standard answer, and the score the user gets is higher. The essence of similarity comparison of subjective questions is text similarity comparison. Text similarity comparison is currently divided into two directions:

1) comparison of text similarity based on natural language;
2) comparison of text similarity based on semantic graph.

Approach for comparing text similarity based on natural language:

1) Based on keywords and keyword combinations:

- N-gram similarity
  The N-gram approach divides two texts or sentences according to N-tuples, and determines the similarity between the texts or sentences by calculating the ratio of the same number of N-tuples and the total number of N-tuples between the texts or sentences [12], [19];
- Jaccard similarity The Jaccard approach uses the idea of set theory to determine the similarity between texts or sentences based on the ratio of the number of identical words or word groups to the number of all non-repeating words or word groups between texts or sentences [12], [20].

2) Based on vector space model (VSM):

The core idea of VSM is to first convert the text into a vector in space by mathematical modeling, and then calculate the similarity value between the spatial vectors through cosine similarity, Euclidean distance, etc [12], [15]. VSM includes the following approaches:

- TF-IDF
  TF-IDF regards text after the participle as an independent feature group, and gives a certain weight according to its importance in the whole document, then TF-IDF transforms the feature group into a space vector and calculates the similarity between the space vectors to get the text similarity [13], [14].
- Word2vec
  Word2vec is a word vector training tool that uses neural networks and corpora to predict the probability of a context based on the current word or the probability of the current word based on the context. The trained word vector can be combined with other approaches (WMD, Sentence2Vec, etc.) to obtain the sentence vector and then used to calculate the text similarity [16], [19].
- Doc2Vec
  Doc2vec tool is an extension of word2vec tool, which is used to learn the distribution vector representation of any length of text [16].

3) Based on deep learning

- In recent years, many researchers have begun to use deep learning for natural language processing. This approach mainly uses DSSM, ConvNet, Tree-LSTM, Siamese LSTM and other multi-layer neural networks to model words or sentences to obtain word vectors or sentence vectors, and then calculate the text similarity [17], [18].

Similarity comparison approach based on semantic graph:

The core idea of the text similarity comparison approach based on semantic graphs is to first convert natural language text into semantic graphs through tools such as syntactic dependency trees or natural language interfaces, and then calculate the text similarity by comparing the similarities between semantic graphs [21].

- SPICE
  SPICE (Semantic Propositional Image Caption Evaluation) approach is mainly used to evaluate the quality of automatically generated image caption. The main working principle of this approach is to compare the similarity between the automatically generated image caption (candidate caption) and the image caption (reference caption) manually labeled by the staff [21]. The main feature of SPICE is the comparison of similarities through semantic content, and its working process mainly includes the following three steps:
  - first, the candidate caption and the reference caption are labeled with grammatical dependency relation. The annotation contents include: subject, predicate, preposition and other grammatical information;
  - secondly, the candidate caption and reference caption with grammatical annotation information are converted into scene graphs by relying on the parse tree, and the scene graphs are encoded according to the relations between objects, attributes and objects;
  - finally, the candidate caption scene graph and reference caption scene graph are decomposed into set tuples according to the objects, the attributes of the objects, and the relations between the objects. The system calculates the similarity by calculating the ratio between the number of identical tuples

in the candidate set tuples and the reference set tuples and the total number of tuples.

Although the approaches discussed above can compare text similarity to some extent, these approaches also have many shortcomings:

- the text similarity comparison approach based on keywords only compares the similarity between texts by words or word groups, and cannot distinguish the synonymy and polysemy of words or word groups. This approach is now mainly used for spell checking and error correction;
- TF-IDF approach assumes that each feature word in the text exists independently, and does not consider the relation between words and their positions in the sentence. When the corpus is large, this method will generate a high-dimensional sparse matrix, resulting in increased computational complexity;
- although the approach based on deep learning has greatly improved the accuracy compared with other approaches, it is also a main research direction now, but this approach relies heavily on the quality of corpus, and when the corpus changes and updates, it needs to retrain the neural network model. This approach also does not have language independence, when using corpus of different languages (such as English, Russian, etc.) to express the same semantic content, the neural network model also needs to be redesigned;
- although the SPICE approach compares text similarity from the semantic level, this approach can only describe simple semantic relations such as the attributes of objects and the connection relation between objects, so many knowledge structures in real life cannot be described.

Based on the SPICE approach, this article proposes an approach for comparing text similarity using OSTIS Technology [1] and unified knowledge coding language SC-code [4]. The approach proposed in this article is to decompose the semantic graph of various types of knowledge represented by SCg-code into various substructures according to certain rules, and calculate the ratio between the same number of substructures and the total number of substructures between different semantic graphs to get the similarity between semantic graphs. Because the approach proposed in this article uses system external language SCg as the representation language of natural language text semantic graph, so it has language independence. The system uses natural language interfaces to convert various natural language (English, Chinese, Russian, etc.) answers into the system's external language SCg representation.

III. Proposed approach

The main task of this article is to introduce the design approach of subsystem of automatic generation of questions and automatic verification of user answers for ILS. Because the subsystem needs to complete two basic functions, so it can be divided into two parts: automatic generation of questions and automatic verification of answers. Next, we will consider the specific approaches of implementing this subsystem using a discrete mathematical system as an example.

A. Proposed question generation approach

Combining the previously discussed approaches for automatic generation of questions and the structural characteristics of the OSTIS learning system knowledge base, this article proposes an approach for automatically generating various types of questions for the ILS and a design approach for the ontology of questions. By using the approaches for automatic generation of questions and the OSTIS technology [1], [2], [3], subjective and objective questions can be automatically generated from the OSTIS learning system knowledge base. The generated subjective questions include: the questions of definition interpretation; the questions of axiomatic interpretation; the questions of proof; The objective questions generated include: choice questions; fill in the blank questions; judgment questions; The generated questions are stored in the ontology of questions in the form of semantic graph using the system external language SCg. Because the system external language SCg-code does not rely on any external natural language, these generated questions can be transformed into corresponding natural language questions through different natural language interfaces.

Consider in more detail the strategies for generating questions:

1) Question generation strategy based on classes

This strategy uses various relations satisfied between the classes to automatically generate objective questions.

- Based on inclusion relation
  In the knowledge base of the OSTIS learning systems, many classes satisfy the inclusion relation, and some classes contain many subclasses, so the inclusion relation between classes can be used to automatically generate objective questions. The set theory expression form of inclusion relation between classes is as follows: $S_i \subseteq C (i \geq 1)$, ($S$-subclass, i-subclass number, C-parent class). Taking the generated judgment questions as an example, its set theory expression is: $S_i \subseteq C$ is "TRUE" or $S_i \subseteq C$ is "FALSE".
- Based on subdividing relation
  Subdividing relation is a quasi-binary oriented relation whose domain of definition is a family of all possible sets. The result of set subdivision is to get pairs of disjoint sets, and the union of these disjoint sets is the original set. There are also many classes in the knowledge base that satisfy the subdivision relation, so this relation can be used to automatically

generate various types of objective questions. The expression form of set theory of subdividing relation between classes is as follows: $S_1 \cup S_2 \cup S_3 ... \cup S_i = C$ $(i > 1, S_i \cap S_j = \phi)$. Taking the generated fill in the blank questions as an example, its set theory expression is: Set C is subdivided into S_1, S_2, ____ and S_i.

- Based on strict inclusion relation
  Strict inclusion relation is a special form of inclusion relation, it is also a very important relation in knowledge base. Using strict inclusion relation to automatically generate objective questions is similar to using inclusion relation. The expression form of set theory of strict inclusion relation between classes is as follows: $S_i \subset C(i \geq 1)$, (S-subclass, i-subclass number, C-parent class). Taking the generated choice questions as an example, its set theory expression is: Set C strictly contains_____? The correct options of choice questions are the subclasses strictly contained in set C, and the incorrect options (disturbing) are the disjoint sibling and parent classes of set C.

2) Question generation strategy based on elements

- Based on role relation
  Role relation is a type of relation often used when building a knowledge base, so role relation between elements can be used to automatically generate objective questions;
- Based on binary relation
  There are many kinds of binary relations between elements in knowledge base, so system can use these binary relations to generate objective questions automatically.

3) Question generation strategy based on identifiers

Usually some sets and relations in the knowledge base have multiple identifiers in addition to the system identifier and the main identifier, so multiple identifiers of sets and relations can be used to automatically generate objective questions.

4) Question generation strategy based on axioms

Many axioms and their mathematical expressions are stored in the discrete mathematical knowledge base, so these axioms and their mathematical expressions can be used to generate objective questions automatically.

5) Question generation strategy based on multiple relations

Many relations in the knowledge base satisfy the attributes of reflexivity, symmetry and transmission, so system can use these relations and their attributes to generate objective questions. The following is a fragment in the knowledge base that satisfies this type of relation in the SCn-code:

maximal clique*
∈ *reflexive relation*
∈ *transitive relation*
∈ *binary relation*

6) Question generation strategy based on image examples

This approach uses some concepts, relations and theorems in the knowledge base and their explanatory image examples to automatically generate some objective questions.

7) Subjective question generation strategy

First, the system stores various definitions and statements interpretation questions automatically generated by this approach into the ontology of questions according to a certain structure, and then the definitions and statements using natural language expressions are manually converted into SCL-code (a special sub-language of the SC language intended for formalizing logical formulas) expressions and stored in the logical ontology.

Using these approaches proposed above can automatically generate choice questions, fill in the blank questions, judgment questions and definition explanation questions, etc. These questions, which are automatically generated using the discrete mathematical knowledge base, are stored in the ontology of questions [1], [2]. When the user needs to be tested, the system extracts specific types of questions from the ontology of questions according to the user's requirements, and then transforms them into natural language form through natural language interface.

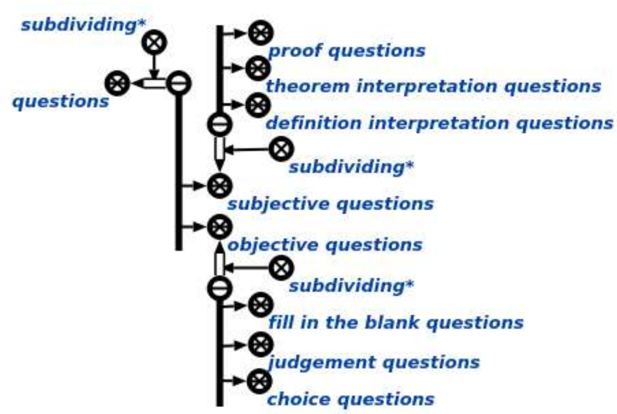

Figure 1. Hierarchy of the ontology of questions

As the carrier of various types of questions generated by storage, the ontology of questions is an important part of the automatic generation of questions and automatic verification of answers subsystem. The structure of the ontology of questions determines the efficiency of extracting the questions from the ontology of questions. Because the generated questions are divided into subjective questions and objective questions, and objective questions and subjective questions are divided into specific types, so the structure of ontology of questions can be constructed according to the type and hierarchy of the generated

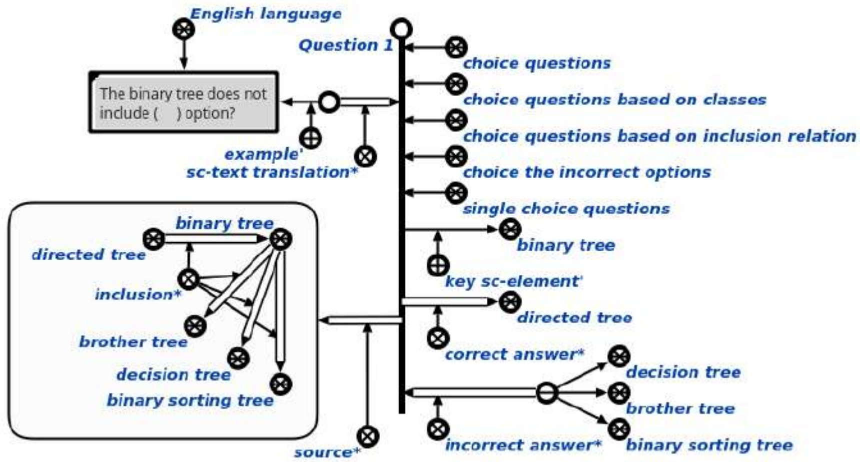

Figure 2. The semantic graph structure of choice question

questions. (Fig. 1) shows the hierarchy of the ontology of questions.

The ontology of questions can be divided into the following types according to the stored question types: choice question ontology, fill in the blank question ontology, judgment question ontology and definition interpretation question ontology, etc., and each type of ontology is used to store the corresponding types of questions. Each type of ontology is created according to the question generation strategies and the characteristics of each type of question [5]. For example, the characteristics of choice questions include:

- choice questions are single choice questions or multiple choice questions;
- for example, when using question generation strategy based on classes, if the number of subclasses that meet the conditions exceeds two, a type of question that selects options that do not meet the requirements of the question can be generated.

Lets consider the choice question ontology of the SCn-code syntax:

choice questions
⇐ *subdividing**:
{
• *choice questions based on multiple relations*
• *choice questions based on axioms*
• *choice questions based on image examples*
• *choice questions based on identifiers*
• *choice questions based on elements*
⇐ *subdividing*:*
{
• *choice questions based on role relation*
• *choice questions based on binary relation*
}
• *choice questions based on classes*
⇐ *subdividing*:*
{
• *choice questions based on subdividing relation*
• *choice questions based on inclusion relation*
• *choice questions based on strict inclusion relation*
}
}
⇐ *subdividing**:
{
• *multiple choice questions*
• *single choice questions*
}
⇐ *subdividing**:
{
• *choice the incorrect options*
• *choice the correct options*
}

Various types of automatically generated questions are stored in the ontology of questions in the form of semantic graph. (Fig. 2) shows the semantic graph structure of choice question in the SCg-code. (Fig. 3) shows the semantic graph structure of definition interpretation question in the SCg-code [1], [4].

The automatically generated various types of questions have a structure similar to the above two types of questions, and they are stored in the corresponding ontology according to this type of structure.

The approach for automatic generation of questions and the approach of using ontology to storage the generated questions, which proposed in this article have the following advantages:

- because the knowledge bases developed using OSTIS technology have the same knowledge storage structure, so only a simple modification to the approach for automatic generation of questions proposed in

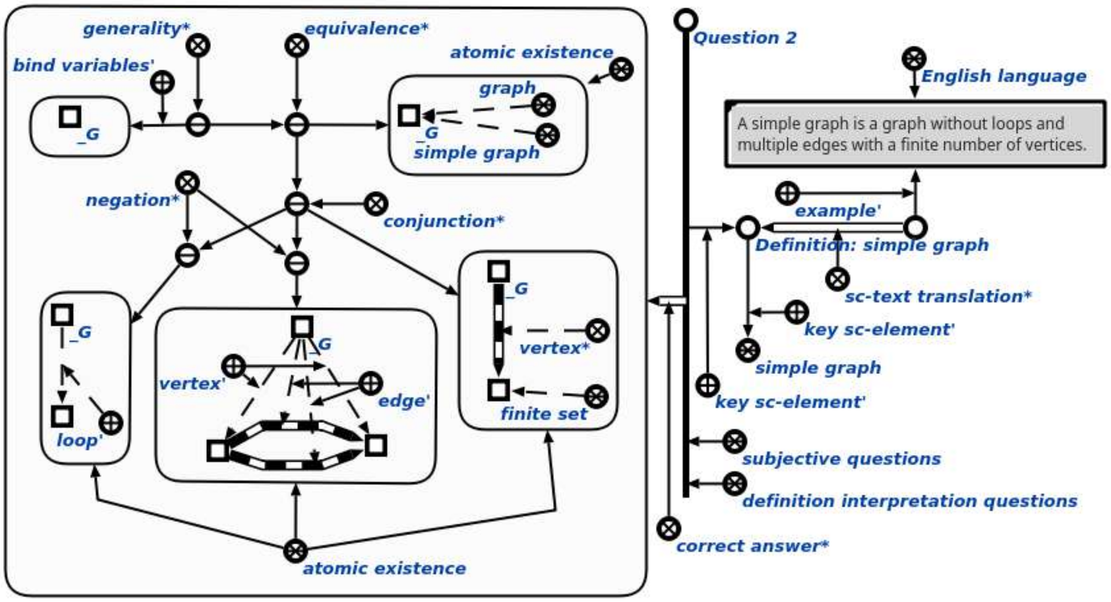

Figure 3. The semantic graph structure of definition interpretation question

this article can be used in other OSTIS systems;

- the generated questions are expressed using a unified knowledge coding language SC-code, so they do not depend on natural language, that is, the generated questions can be converted into different natural language forms only through different natural language interfaces;
- the generated questions are stored in the ontology of questions according to the types of the questions and the generation strategies, so when the questions need to be extracted from the ontology of questions, the efficiency of question extraction can be greatly improved;
- using the approach proposed in this article it is possible not only generate subjective and objective questions, but also the quality of the generated questions is very high.

The approaches for automatic generation of questions and using ontology to storage the generated questions proposed in this article solved the existing problems of the approaches for automatic generation of questions in the previous section, so the approaches proposed in this paper has certain advantages.

B. Proposed answer verification approach

In this article, the answer verification is divided into subjective question answer verification and objective question answer verification. Because objective questions have definite standard answers, so it only needs to directly compare standard answers with user answers. There are no definite answers to the subjective questions, as long as the user answers conform to the logic and meet the requirements of the questions, so it is necessary to compare the similarity between the standard answers and the user answers. According to the types of knowledge, subjective question answer verification can be divided into:

- factual knowledge answer verification;
- logical knowledge answer verification.

Factual knowledge refers to knowledge that does not contain variable types, and this type of knowledge expresses facts. Logical knowledge usually contains variables, and there are logical relationships between knowledge. Most of the answers to subjective questions are logical knowledge [1], [4], [21].

Based on the SPICE approach [21], this article proposes an approach for automatically verifying the answers to subjective questions using OSTIS technology. According to the task requirements, the approach proposed in this article needs to verify the correctness and completeness of user answers (for example, the answer is correct but incomplete, and the answer is partially correct, etc.). The answer verification approaches of factual knowledge and logical knowledge are similar, the answers described by the semantic graph are divided into sub-structures according to certain approaches and then the similarity is compared. The implementation process of the answer verification approach proposed in this article is as follows:

1) First, the knowledge base developers transform the natural language answers of subjective questions into SC-code forms and store them in the ontology of questions. We use s to represent standard answers in the form of semantic graphs.

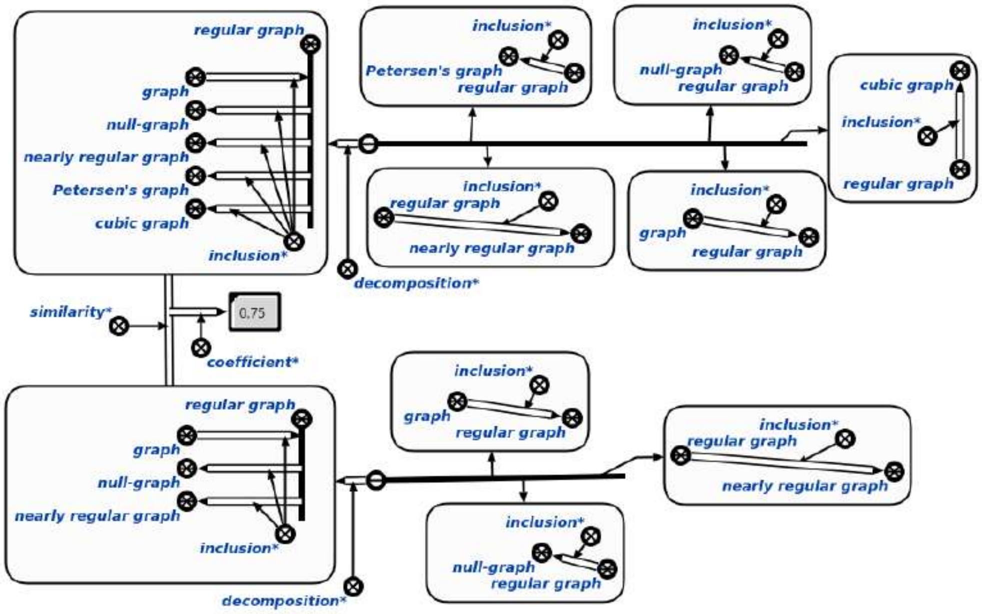

Figure 4. An example of similarity calculation of knowledge of factual types

2) Then the standard answers s and user answers u expressed using SCg-code/SCL-code [1], [2] are decomposed into sub-structures according to the rules of knowledge representation. u refers to the expression form after the natural language user's answers are transformed to SC-code through the natural language interface. The set $T_{sc}(s)$ represents all the sub-structures after the decomposition of the standard answers s, and the set $T_{sc}(u)$ represents all the sub-structures after the decomposition of the user answers u.
3) Finally, the similarity is calculated by comparing the ratio between the number of identical sub-structures and the total number of sub-structures between the standard answer and the user answer. The main calculation parameters include: precision P_{sc}, recall R_{sc}, and similarity F_{sc}. Their specific calculation process is shown in formulas (1), (2), (3).

$$P_{sc}(u,s) = \frac{|T_{sc}(u) \otimes T_{sc}(s)|}{|T_{sc}(u)|} \tag{1}$$

$$R_{sc}(u,s) = \frac{|T_{sc}(u) \otimes T_{sc}(s)|}{|T_{sc}(s)|} \tag{2}$$

$$F_{sc}(u,s) = \frac{2 \cdot P_{sc}(u,s) \cdot R_{sc}(u,s)}{P_{sc}(u,s) + R_{sc}(u,s)} \tag{3}$$

(Fig. 4) shows an example of similarity calculation of knowledge of factual types in the SC-code.

Although the verification process of answers of logical knowledge and factual knowledge is basically similar, it is necessary to consider some unique characteristics of logical semantic graphs when performing substructure decomposition, such as:

- there is a strict logical order between the nodes in the logical semantic graph, so the integrity of the logical formula cannot be broken when sub-structure decomposition is performed;
- each substructure of the decomposed logical semantic graph must have logical meaning.

The logical expression of inclusion of set is: $\forall a, a \in A \rightarrow a \in B$. (Fig. 5) shows an example of the decomposition of definition of inclusion of set into substructures in the SC-code.

Compared with existing answer verification approaches, the proposed approach in this article has many advantages:

- using semantics to compare the similarity between answers;
- because the OSTIS technology [1] has a unified way of knowledge expression, so the approach proposed in this article can calculate the similarity between texts with more complex content than the SPICE approach;
- compared with the approach of deep learning, the approach proposed in this article can judge the completeness and find the incorrect parts of the user answers through the sub-structures;
- the approach proposed in this article uses a unified knowledge coding language SC-code, so it can be used in other OSTIS systems only by modifying a few rules, and does not depend on natural language.

These advantages make up for the existing problems of answer verification in the previous section, so the answer

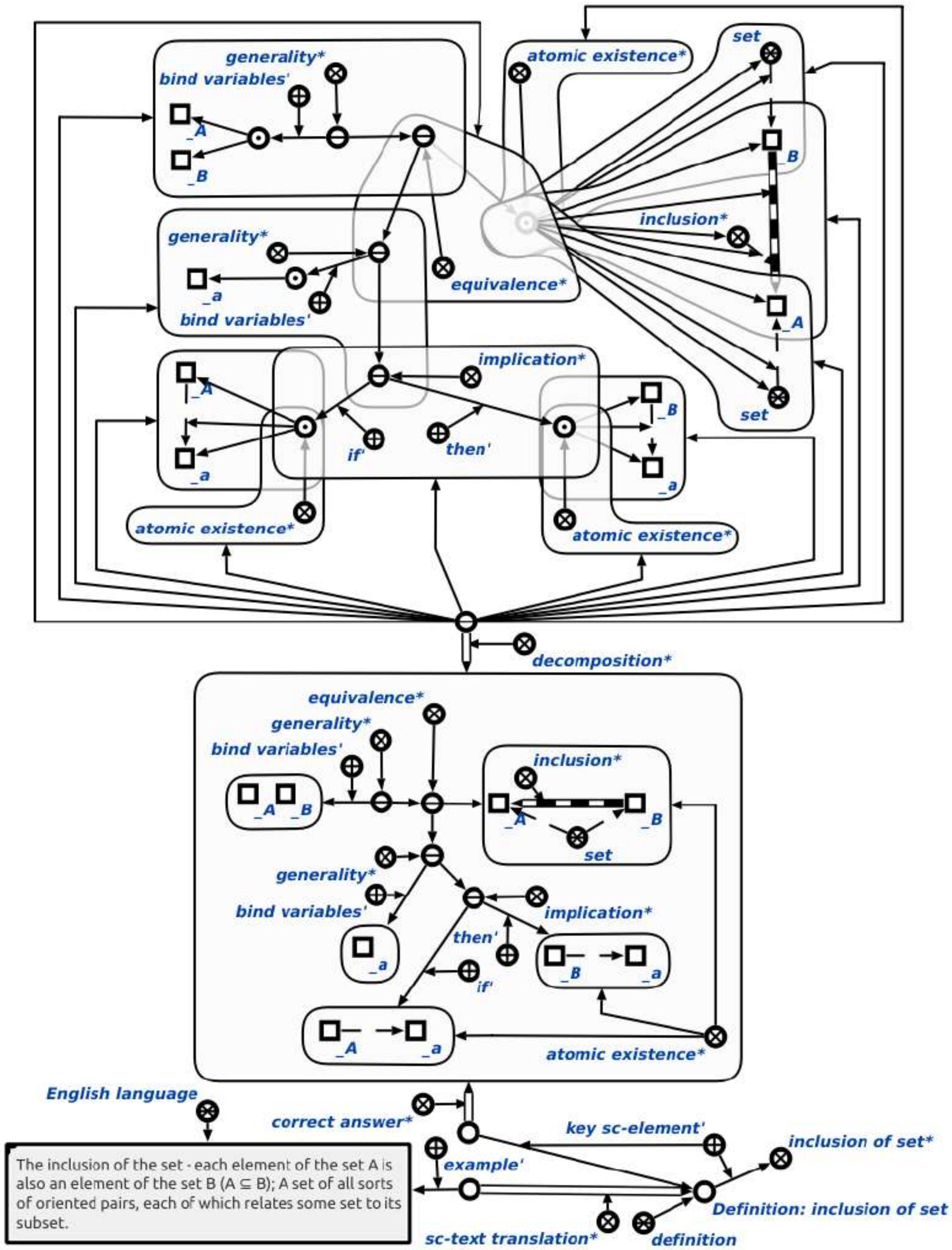

Figure 5. An example of the decomposition of definition of inclusion of set

verification approach proposed in this article has certain advantages.

The automatic question generation approach and automatic answer verification approach proposed in this article have many advantages, but there are also many problems to be solved, such as:

- how to control the quality and repetition rate of automatically generated questions in the ontology of questions, and the similarity threshold setting between standard answers and user answers;
- usually, a logical proposition may have several equivalent propositions, so the answer verification of between equivalent propositions needs to be solved;
- how to verify the answers when the concepts expressed between the standard answers and the user answers are the same but the terms are different;

IV. Conclusion and further work

This article first analyzes the development status and advantages of ILS, and then discusses the problems of the existing automatic question generation and answer verification approaches, finally, combining existing methods and OSTIS technology, this article proposes an approach for automatic generation of questions and automatic verification of answers. The proposed approach solves some existing problems well, and the subsystem of question generation and answer verification designed using this approach has good compatibility and natural language independence, so it can be used well with other OSTIS learning systems. Although the approaches proposed in this article have several advantages, there are also quite a few problems. For example, how to control the quality and rate of repetition of questions and etc. The next step is to introduce the proposed approach into the discrete mathematics learning system.

Acknowledgment

The work in this paper was done with the support of research teams of the Department of Intelligent Information Technologies of Belarusian State University of Informatics and Radioelectronics. Authors would like to thank every researcher in the Department of Intelligent Information Technologies.

References

[1] V. V. Golenkov and N. A. Guljakina, “Proekt otkrytoj semanticheskoj tehnologii komponentnogo proektirovanija intellektual’nyh sistem. chast’ 1: Principy sozdanija project of open semantic technology for component design of intelligent systems. part 1: Creation principles],” *Ontologija proektirovanija [Ontology of design]*, no. 1, pp. 42–64, 2014.

[2] V. Golenkov, N. Guliakina, I. Davydenko, and A. Eremeev, "Methods and tools for ensuring compatibility of computer systems," in *Otkrytye semanticheskie tehnologii proektirovanija intellektual’nyh sistem [Open semantic technologies for intelligent systems]*, V. Golenkov, Ed., BSUIR. Minsk , BSUIR, 2019, pp. 25–52.

[3] V. Golenkov, D. Shunkevich, I Davydenko, and N. Grakova, "Principles of organization and automation of the semantic computer systems development," in *Otkrytye semanticheskie tehnologii proektirovanija intellektual’nyh sistem [Open semantic technologies for intelligent systems]*, V. Golenkov, Ed., BSUIR. Minsk, BSUIR, 2019, pp. 53–90.

[4] (2018, Dec) Ims.ostis metasystem. [Online]. Available: https://ims.ostis.net

[5] Li H. Research on item automatic generation based on DL and domain ontology. *Journal of Changchun University of Technology (Natural Science Edition)*, 2012, Vol. 33(04), pp. 461-464.

[6] Ruslan M., Le A. H., Nikiforos K. A computer-aided environment for generating multiple—choice test items. *Natural Language Engineering*, 2006, Vol. 12(02), pp. 177-194.

[7] “Protégé [Electronic resource],” mode of access: http://protege.stanford.edu. Date of access: 01.12.2019.

[8] Din X. M., Gu H. B. Research on automatic generation of Chinese multiple choice questions based on ontology. *Computer Engineering and Design*, 2010, Vol. 31(06), pp. 1397-1400.

[9] Yuan X. L. Comparison and analysis of commonly used intelligent test paper grouping algorithms. *Computer Education*, 2016, Vol. 2016(05), pp. 121-123.

[10] Andreas P., Konstantinos K., Konstantinos K. Automatic generation of multiple-choice questions from domain ontologies. *IADIS International Conference e-Learning*, 2008.

[11] Arjun S. B., Manas K., Sujan K. S. Automatic Generation of Multiple Choice Questions Using Wikipedia. *International Conference on Pattern Recognition and Machine Intelligence*, 2013, Vol. 8251, pp. 733-738.

[12] Wan C. L., Yang Y. H., Deng F. A review of text similarity calculation methods. *Information science*, 2019, Vol. 37(33), pp. 158-168.

[13] "Application of TF-IDF and cosine similarity [Electronic resource]," mode of access: http://www.ruanyifeng.com/blog/2013/03/tf-idf.html. Date of access: 01.12.2019.

[14] "NLP text similarity (TF-IDF) [Electronic resource]," mode of access: https://www.cnblogs.com/liangjf/p/8283519.html. Date of access: 05.09.2019.

[15] "Vector Space Model (VSM) Algorithm [Electronic resource]," mode of access: https://zhuanlan.zhihu.com/p/31886934. Date of access: 02.11.2019.

[16] "Principle and practice of Doc2vec [Electronic resource]," mode of access: https://blog.csdn.net/u010417185/article/details/80654558. Date of access: 02.12.2019.

[17] Li X. J. Realization of automatic scoring algorithm for subjective questions based on artificial intelligence. *Journal of Jiangnan University (Natural Science Edition)*, 2009, Vol. 08(03), pp. 292-295.

[18] Yang M., Li P. F., Zhu Q. M. Sentence similarity calculation method based on Tree-LSTM. *Journal of Peking University (Natural Science Edition)*, 2018, Vol. 54(03), pp. 481-486.

[19] Wan H. R., Zhang Y. S. Review on Research Progress of Text Similarity Calculation. *Journal of Beijing Information Science (Technology University)*, 2019, Vol. 34(01), pp. 68-74.

[20] Qin C. X., Zhu T., Zhao P. M., Zhang Y. Research progress in semantic analysis of natural language. *Library and information work*, 2014, Vol. 58(22), pp. 130-137.

[21] Peter A., Basura F., Mark J. SPICE: Semantic Propositional Image Caption Evaluation. *Computer Vision and Pattern Recognition (cs.CV)*, 2016.

Онтологический подход к автоматизации процессов генерации вопросов и контроля знаний в интеллектуальных обучающих системах

Ли Вэньцзу, Гракова Н.В., Цянь Лунвэй

В данной работе предложен подход к проектированию общей подсистемы автоматической генерации вопросов и автоматической верификации ответов в интеллектуальных обучающих систем. Данный подход предлагается на основе существующих подходов и теории онтологий, с помощью предлагаемого подхода можно автоматически генерировать различные типы вопросов из баз знаний и автоматически проверять правильность и полноту ответов пользователей. В сравнении с существующими подходами, предложенный в данной работе подход имеет определенные технические преимущества, а подсистема, разработанная с использованием предложенного в данной работе подхода, может быть применена в различных обучающих системах, построенных по технологии OSTIS.

Received 10.01.2020