Generation of Multiple Choice Questions Including Panoramic Information using Linked Data

Fumika Okuhara, Yuichi Sei, Yasuyuki Tahara and Akihiko Ohsuga
Graduate School of Informatics and Engineering, The University of Electro-Communications, Tokyo, Japan

Keywords: Panoramic Information, Linked Data, Multiple Choice Question, Semantic Web.

Abstract: In recent years, just about all subjects require students to learn panoramic information. Because the need exists for cross-curriculum learning aimed at relating subject areas, it is useful for multiple-choice questions to include panoramic information for learners. A question including panoramic information refers to content that includes transverse related information and makes respondents grasp the whole knowledge. However, it is costly to manually generate and collect appropriate multiple-choice questions for questioners and learners. Therefore, in this research, we propose a method for the automatic generation of multiple-choice questions including panoramic information using Linked Data. Linked Data is graphical data that can link structured data, and it is used as a technology for data integration and utilization. Some attempts have been made to use Linked Data as a resource for creating teaching material, and the possibility of using Semantic Web technology in education has been verified. In this paper, we aim to realize a system for automatically generating two types of multiple-choice questions by implementing an approach to generating questions and choices. An evaluation method for the generation of questions and choices involves setting indicators for each evaluation item, such as validity and the degree of the inclusion of panoramic information.

1 INTRODUCTION

Each school has a curriculum and students should learn based on it. Regarding the importance of curriculum management, the Central Council for Education mentioned the need for "improving educational activities based on a cross-curriculum perspective" (Ministry of Education, Culture, Sports, Science and Technology, 2015). In other words, panoramic learning is necessary for learning all subjects.

The multiple-choice-question format is widely used for qualification exams, certification examinations, and the like. This format is useful because it enables the quick, easy, and objective scoring of large-scale exams. In addition, because completing these exams simply involves "choosing the correct answer from the choices," responding to each question is a highly efficient process (Ikegami, 2015). Therefore, the format may be suitable for testing a large number of people, helping learners to demonstrate their knowledge of a wide range of fields easily, and enabling questioners to evaluate a wide range of questions from units.

Based on the above, multiple-choice questions including panoramic information have been deemed useful for both learners and questioners. However, manually generating and collecting appropriate multiple-choice questions is costly. In this paper, we propose a method for automatically generating questions including panoramic information based on the given curriculum, and evaluating them.

With our proposed method, Linked Data is used as a knowledge base, and questions and incorrect choices for a correct answer set are generated based on the curriculum. Figure 1 displays an output image of the proposed system. In this system, the output is generated by setting any curriculum and selection of unit as the input, extracting the keywords included in the curriculum, selecting the keywords to set in Answers, and configuring questions and incorrect choices with Wikipedia data. The output is the components of the multiple choice question. Using the Wikipedia data which the format expressed the relationship between these data, the question can include relevant data and the relationships between the answer and can be one including panoramic information. In the current research study for evaluation of this system, requirements were set for the generation of questions and choices, and evaluation experiments were conducted based on the evaluation items. These experiments included in-

Okuhara, F., Sei, Y., Tahara, Y. and Ohsuga, A.
Generation of Multiple Choice Questions Including Panoramic Information using Linked Data.
DOI: 10.5220/0007259301100120
In *Proceedings of the 11th International Conference on Agents and Artificial Intelligence (ICAART 2019)*, pages 110-120
ISBN: 978-989-758-350-6

dices of the degree of inclusion of the panoramic information set, and in the result of the experiment, the questions made by proposed method tended to have more panoramic information than one made by comparison method.

The article is organized as follows: Section 2 the related research studies; Sections 3 and 4 highlight the purpose of this research and the proposed method used; Sections 5 and 6 provide the implementation method, some evaluation method and the results of the evaluation experiments; Section 7 present the discussion; and finally, Section 8 states the conclusion and future work.

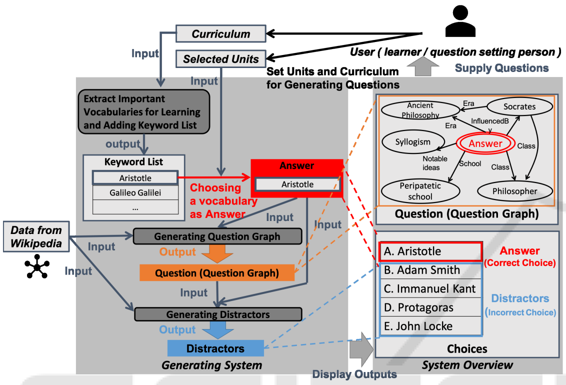

Figure 1: An Example of output image of multiple choice questions generated by the proposed system.

2 RELATED RESEARCH

Linked Data, as an existing technology, has been used for the generation of questions. Linked Data is structured graphical data that Tim Berners-Lee proposed; data sets are linked with one another using the Web mechanism. Linked open data (hereinafter referred to as LOD) is Linked Data that is published on the Web. LOD represents a data format that anyone can freely use, with various kinds of open data being linked with each other through municipalities or institutions. The LOD cloud[1], representing links between available LOD, contains more than 1,000 data sets as of May 2018. Activities related to LOD are being carried out in various fields. Among them is DBpedia[2] which turns the well-known Wikipedia data into a Linked Data format. DBpedia Japanese[3] which is a Japanese version, also exists, turning the information in the InfoBox of Japanese Wikipedia into a Linked Data format.

[1]http://lod-cloud.net
[2]http://dbpedia.org
[3]http://ja.dbpedia.org

Research (Iijima et al., 2016) has also proposed a method of presenting an unexpected connection between multiple data sets by using Linked Data technology, which can be applied to a recommendation system. (Maillot et al., 2014) presented a method for extracting the targeted subpart of resource description framework (RDF) bases, driven by a list of selected resources called the seed. In addition, a research study in the Semantic Web field, by Demarchi F. et al. (Demarchi et al., 2018) proposed an implementation that would allow agents to access ontologies that are available on the Web so as to update their beliefs based on significant content. A case study of an educational quiz is also presented that used the information to formulate the questions and to validate the answers obtained.

Several attempts have been made to use Linked Data as a resource for generating teaching materials in the education field.

In ASSESS (Bühmann et al., 2015), attempted to generate questions in several formats in the specific field of general knowledge using LOD. Based on this attempt, it is possible to generate questions in natural language by summarizing an entity based on DBpedia and verbalizing the RDF. In addition, the choices corresponding to each format of questions are implemented by using LOD. Papasalouros et al. also presented a method of generation multiple-choice questions in natural language from Semantic Web Rule Language rules which is interpreted so that if the antecedent holds, then the consequent must also hold; in the simple form $antecedent \Rightarrow consequent$ (Papasalouros et al., 2008) (Zoumpatianos et al., 2011). Also, Rocha et al. attempted generation questions that had resources that were relevant to a specific domain or topic from a dataset (Rocha et al., 2018), and Afzal et al. presented generation questions regarding the important concepts that presented in a domain by relying on the unsupervised relation extraction approach as extracted semantic relations (Afzal and Mitkov, 2014).

Furthermore, researchers in a study generated an evaluation model of the incorrect choices created in multiple-choice questions (Pho et al., 2015). The model was generated to enable the automatic evaluation of the quality of the incorrect choices that the author manually created. The model focuses on the syntactic and semantic similarity between the choices, treating them as elements related to the quality of these choices. Meanwhile, (Patra and Saha, 2018) considered closeness between the key and the possible distractors by using web information in their proposed system for automatic named entity distractor generation. A research study involved the generation of a historical ontology that used LOD to generate history questions

(Jouault et al., 2016). In this research, a question-setting system based on a learning scenario was included. Specifically, Grasser's classification method, which classified multiple knowledge bases and question formats, was used. Statistical data, such as the degree of difficulty of each test item, is used when constructing an examination test from a large number and a wide range of questions. In other words, when evaluating learning achievement through an examination test, it is necessary to set items at difficulty levels that are considered the preset passing marks. As a study on the difficulty level of an examination test, (Ikeda et al., 2013) proposed a difficulty level estimation method focusing on the similarity between the question pattern and the choices of the multiple-choice question, and evaluating this based on the difficulty parameter of the item response theory (IRT).

In these related works, the possibility of using Semantic Web technology has been verified against the theme of generating test questions. However, these proposed systems can generate only uniform questions and choices for keywords with these methods. As a result, the multiple-choice questions feature simple content, such as "What is {person's name} birth place?" and "Which work is made by {person's name}?", that contains only one or two matters concerning the answer in examination sentences. Also, there is a possibility that the questions are in a narrow field since their resource of the questions is composed of highly relevant contents. Therefore, in this research, we aimed to generate questions including more panoramic information.

3 PURPOSE

The purpose of this research is to propose a method for automatic generation of multiple choice questions including panoramic information. Panoramic information means comprehensive information that gives us macro-perspective; through which us look down at the whole learning subjects. A question including panoramic information refers to content that includes transverse related information and makes respondents grasp the whole knowledge.

In the use scenario of the question generation system, a person who sets examination questions and learners can be seen as users of the system. The exam preparer may be able to reduce costs such as time and effort for creating test questions by using a system into which a curriculum, including evaluation items, is input. In addition, learners' use scenarios include self-study and exercise test questions.

Figure 1 shows an example of an output of multiple choice questions generated by this method. In the system, output is generated by setting any curriculum and selection of unit as input, extracting the keywords included in the curriculum, selecting the keywords to set in Answers, and configuring questions and incorrect choices with Wikipedia data. The output is the components of the multiple choice question. There are "Question" which is a sentence of the test question, "Answer" which is a correct answer choice and "Distractors" which are incorrect choices. In this research, a graph is taken as a Question and referred to as the"Question Graph". From the linked structure of the graph and the sentence in question, it can be understood that the answer has some relationships that include a philosopher studied in the class, a person who had connections to peripatetic school and one of whose notable idea was syllogism, is a person in the era of ancient philosophy, and influenced Socrates. When selecting vocabulary that corresponds to the answer that satisfies all these relationships from five choices, "Aristotle" becomes the correct answer, and the remaining choices become incorrect.

4 PROPOSED METHOD

This section explains the generation approaches of Question Graph and distractors for arbitrary answer. Incidentally, it is possible to use DBpedia, DBpedia Japanese, etc. as a knowledge base.

4.1 Approaches to Generating the Question Graph

The Question Graph is generated by searching triple structures for an answer that regards an RDF graph visualizing the relationship between the acquired information as a question sentence. In other words, a Question Graph is a test question format that hides an element corresponding to the answer in the graph and guesses it from some words around it and their properties. Then, the graph itself is defined as a question. Since the matter to be questioned in the Question Graph is related to the data around Answer, there is no need to document it, and as shown in Figure 1, it is presented as "Question graph," which is a graph as a Question. There is existing research [Fionda 17] that proposes an algorithm to search all subgraph structures among multiple data sets (Fionda and Pirrò, 2017).

In our approach, a searching method considering the following requirements is devised.

4.1.1 Requirements of Generating the Question Graph

The requirements to be satisfied as the Question Graph are set as follows. In particular, the items stating (Mandatory) are essential requirements.

Requirements of Question Graph

(1) (Mandatory) Each node consists of keywords.

(2) (Mandatory) Ensure the connectivity of information around Answer.

(3) Include panoramic information as much as possible on the entire graph.

(4) Scale of the graph can help grasp all its content.

(5) The number of vocabulary words corresponding to the answer is extremely small.

4.1.2 Method of Generating a Question Graph

We devised a method to generate a Question Graph. The requirements below are based on an RDF graph using DBpedia.

Question Generation Algorithm for Single Selection Form

The following describes the method of generating question graphs for single choice questions. In this format, the Answer node has one or more answer in the graph, and respondents answer select one from some choices.

The graph for an answer is generated by extending the link structures of data around Answer by searching for neighbor nodes to the answer node, further searching from neighbor nodes to each of their neighbor nodes repeatedly. The link structure between the nodes is formed by directed links of IN and OUT. In the process of searching the link structures, the search range and the number of times are restricted preliminarily in order to make the graph, considering the scale based on requirement (4). Regarding the number of searches, when searching for neighbor nodes by 1 hop with respect to the answer node, the number of hops is determined as the search depth h. The search range defines the number of each of the directed link structures of IN and OUT at the same depth h as the search width w. This method generates a Question Graph with a scale satisfying the restrictions of h and w. In particular, considering the degree of inclusion of panoramic information based on requirement (3), we propose a search method in which all neighbor nodes after the answer node are absolutely distant from the origin.

We devised the following Algorithms 1 and 2 as the basic algorithms for generating Question Graphs for answers.

Algorithm 1: Main.

Input: KG G, $Answer$, depth h, width w
Output: KG
1: $N_S = get\_far\_nodes(Answer, \{\}, G, h, w)$
2: $M_S = get\_all\_links(N_S)$
3: **return** (N_S, M_S)

Algorithm 2: get_far_nodes.

Input: Node $target$, Set of ancestor nodes $Ancestors$, KG G, depth h, width w
Output: Set of nodes
1: $N = \{target\}$
2: **if** $|Ancestors| == h$ **then**
3:     **return** N
4: **end if**
5: $count = 0$
6: **for** $direction \in \{IN, OUT\}$ **do**
7:     $B = neighbors(target, direction)$
8:     **while** $count < w$ AND $0 < |B|$ **do**
9:         $flg = True$
10:         $n = \arg\max_{n' \in B} dist(target, n')$
11:         $B = B \backslash \{n\}$
12:         **for** $n_j \in Ancestors$ **do**
13:             **if** $dist(n_j, n) < dist(n_j, target)$ **then**
14:                 $flg = False$
15:             **end if**
16:         **end for**
17:         **if** flg **then**
18:             $count = count + 1$
19:             $N = N \cup get\_far\_nodes(n, Ancestors \cup \{target\}, G, h, w)$
20:         **end if**
21:     **end while**
22:     $count = 0$
23: **end for**
24: **return** N

The Main algorithm acquires all node sets relating to information around *Answer*, then acquires all the node sets' link structures and returns this information as the subgraphs. As input, the algorithm is given the knowledge graph KG (Knowledge Graph) G, the correct *Answer*, the search depth h, the search depth w, on the knowledge base as the output. The subgraph consists of a set of all nodes N_s and a set of link structures between nodes M_s.

The get_far_nodes algorithm returns all the nodes in the subgraphs relating to the target node. A graph on *Answer* can be finally obtained by specifying Answer as the default of *target*. From lines 6 to 23, the graph extends searching neighbor nodes for each directed link of *target*. In order to obtain the neighbor node of the *target*, the function of *neighbors* that return the neighbor node set in line 7 is defined and used. In line.10, the neighbor node whose distance to the *target* is the maximum is returned by the function *dist* that finds the distance between specified nodes. In line.12, the distance comparison between the neighbor node and all its ancestors and the distance between the *target* and all ancestors is recursively repeated up to the search depth of h. This ensures that the distance between all nodes, from the *Answer* node to the neighbor node, will always be farther away. Finally, all the node sets are returned, and a subgraph consisting of all node sets and their link structures as a Question Graph is obtained.

Based on the algorithm, we aim to generate a graph with a compact scale and a high degree of panoramic information inclusion.

Question Generation Algorithm for Multiple Selection Form

The following describes the method for generating question graphs in multiple-choice question form. Because this question format requires answers with multiple choices, it should be guaranteed that there are that two or more words corresponding to Answer nodes on the graph. Therefore, in requirement (5), the number of words corresponding to the answer is adjusted.

In this approach, the vocabulary corresponding to the Answer node is necessarily greater than the specified number nA when depth $h = 1$. First, if there are nA words in the knowledge base corresponding to nodes that have all neighbor nodes acquired by the *neighbors* function, adopt them as the neighbor nodes on the graph. The algorithm keeps a panoramic degree as much as possible by checking this from the vocabulary with the largest distance in order. After that, getting a combination of all nodes adjacent to *Answer*, and acquires multiple words that can be applied to the *Answer* node. For $h = 2$ and later, the graph is completed by connecting neighbor nodes with the *get_all_nodes* procedure.

In addition to the above basic algorithm, we also consider the major degree of vocabulary this time. For weighting measures, we used the Balanced Corpus of Contemporary Written Japanese; (BCCWJ) by "Corpus Development Center, NINJAL"[4]. The vocabulary of the textbook sub-corpus (OT) was set to 1.0, the history and social science classifications among libraries and the publication sub-corpus (LB and PB) were set to 0.5, and the other was set to 0.0. This was applied by adding it to the *dist* function so that the vocabulary existing in the corpus could be more easily adopted. In addition, each node of depth $h = 1$, is preferentially adopted when the vocabulary of corpus OT exists in the inquiry result from the knowledge base.

From the above basic procedures, a Question Graph on the answer is generated.

[4]http://pj.ninjal.ac.jp/corpus_center/bccwj/

4.2 Approach to Generating Distractors

Since distractors are nodes that do not correspond to an answer in the Question Graph, they can be generated by searching for nodes that do not satisfy all the link structures with the answer (even if one of them is satisfied). Therefore, we explore the method of generating distractors using the Question Graph generated above.

4.2.1 Requirements of Generating Distractors

The requirements to be satisfied by distractors we set as follows.

Requirements of Distractors

(1) (Mandatory) Each node consists of keywords.

(2) (Mandatory) Incorrect answer to the Question.

(3) Avoid words clearly recognized as incorrect answers.

4.2.2 Method of Generating Distractors

From requirement (3), in order to generate distractors that are not clearly recognized as incorrect answers, a vocabulary set similar to the Answer should be selected. Vocabularies similar to the answer can be thought that the link structures on the knowledge base are similar to one of the Answer. From this, distractor candidates are generated from the adjacent link structures of answers by using the Question Graph generated by the above method.

First, a set of adjacent link structures for an answer is extracted from the Question Graph. We regard the words corresponding to a node that satisfies a set of other link structures only when deleting one or more links from the extracted set as distractor candidates.

Furthermore, referring to the acquisition method of (Pho et al., 2015), narrowing down the words belonging to the same class as answer on DBpedia among the candidates. If an answer belongs to more than one class, it is immediately rated class C, which is the lowest class; that is, the direct class C of the answer is an instance, but not an instance of that subclass. Finally, candidates corresponding to the objects obtained as a property of class C are adopted.

As a method of adopting from a candidate to an option, we adopt them from the one with the smallest difference between the distances to the neighbor nodes and largest degree of popularity. Figure 2 shows the generating steps.

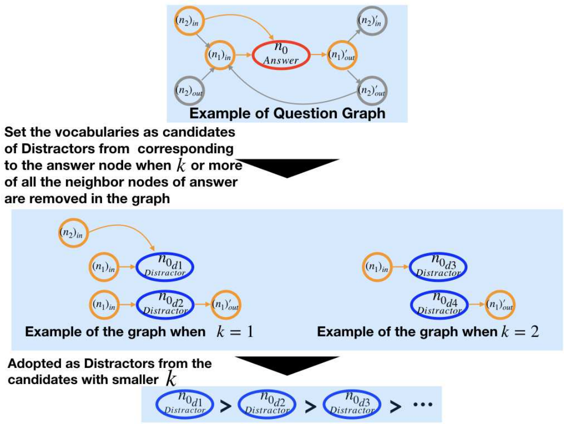

Figure 2: Generating step of Distractors based on the method.

5 IMPLEMENTATION

The implementation of generation of Question Graphs and distractors by the above approach is described below. We use DBpedia Japanese as a knowledge base and set the SPARQL endpoint to "http://ja.dbpedia.org/sparql/".

In this research, we set social studies subjects (geography, history, citizens etc.) especially as the domain of questions. In addition, we selected words defined as article titles or categories on Wikipedia in advance and used them as keywords for answers. For implementation, no curriculum is set, and the vocabulary on DBpedia are set as the keywords.

5.1 Generating the Question Graph

As described in the previous section, a Question Graph is generated by collecting information around the answer via SPARQL query and then visualizing the RDF graph.

In defining the *dist* function in Algorithm 2 above, we use value of similarity based on the pre-learned model of word2vec[5] (Suzuki et al., 2016) as a comparable index between words syntactically and semantically.

In the visualization of graphs, the words around the answer and the relationships between them are drawn using Graphviz.

[5]http://www.cl.ecei.tohoku.ac.jp/~m-suzuki/jawiki_vector/

For example, the following Figure 3 shows the Question Graph generated based on the RDF graph for Answer = "Socrates." Here, the Answer node is red, the vocabulary of the corpus OT is blue, the corpus' LB and PB are green, and other items are grey.

5.2 Generating Distractors

As in the above approach, distractors are generated by obtaining some instances belonging to the same class C as an answer by using the link structures with the answer node.

Table 1 summarizes distractor candidates generated by Answer ="Socrates".

Table 1: Distractors on Answer="Socrates".

links	Answer/Distractors candidates{total}
(0,1,2,3,4,5)	Socrates {1}
(3,4)	Heraclitus {1}
(0)	Diogenes (cynic school) {1}
(1)	Immanuel Kant {1}
(2)	John Stuart Mil {1}
(3)	Anakusagorasu {1}

6 EVALUATION

For the generated Question Graph and distractors, the evaluation method corresponding to each of the above requirements is described below. In this evaluation, the search depth of the Question Graph was set to $h = 2$ and the width to $w = 2$. We selected the appropriate Answer, but this time, for the evaluation of distractors, set words with Answers to which nobody were supposed not to know the correct answer.

6.1 Evaluation Method for Question Graph Generation

Evaluation items for the Question Graph requirements are listed below.

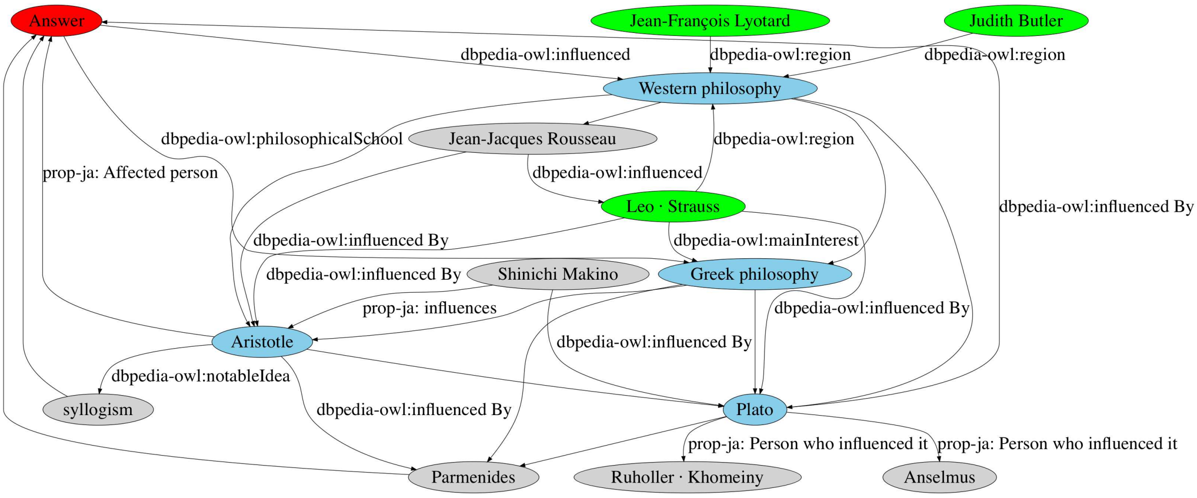

Figure 3: The Question Graph on Answer="Socrates".

Question Graph Evaluation items

① Consistency...[Requirement (2)]
- Answer corresponding to correct answer.

② The degree of inclusion of panoramic information...[Requirement (3)]
- The degree of crossing classes of each node
- The degree of crossing units on the curriculum
- The degree of crossing time

③ Specificity...[Requirement (5)]
- Smallness of words corresponding to the Answer

④ Readability...[Requirement (4)]
- Compactness of the graph scale

The following two evaluation experiments were conducted on the second item regarding the degree of inclusion of panoramic information, and on the third item regarding specificity.

6.1.1 Evaluation Method for Degree of Inclusion of Panoramic Information

Since a specific curriculum is not set, the first and third evaluation items are implemented. The first item is an index of how far across the classes in the knowledge base used to generate the Question Graph are. In this case, we examined the classes to which vocabulary belongs from Class "Thing" and below in all classes on DBpedia. In the evaluation, only nodes that can be acquired from DBpedia are subject to calculation for the classes.

6.1.2 Specificity Evaluation Method

For the generated graph, the number of words corresponding to the answer node is also evaluated. It expresses the smallness of the number of correct alternative answer from the question graph when the choices are ignored. The number of corresponding words becomes clear by searching nodes that have all of the same link structure as the Answer.

6.2 Evaluating Method of Generating Distractors

Evaluation items for the requirements of distractors are listed below.

Distractor evaluation items

① Incorrect answer to the Question Graph...[Requirement (2)]

② The validity of those who do not know the correct answer is less than one out of the number of choices...[Requirement (3)]

③ The magnitude of similarity to the answer...[Requirement (3)]

In particular, for the second and third evaluation items, the following evaluation experiments were conducted.

6.2.1 Comparison of Validities by Experimental Subject

By comparing the validity to the actual generated questions in subjective experiments, we verified whether the validity were less than one out of number of the choices.

6.2.2 Comparison of Similarity between the Answer and Choices

We compared similarities among the obtained distractors. From the syntactic and semantic aspects, the following indexes of similarity comparison are listed.

- Syntactic similarity

  For both answer and distractor candidates, compare both the parts of speech and the composition patterns of the words. In this case, we used CaboCha[6] as a parsing tool and compared both dependency and part of speech, verifying whether they matched.

- Semantic similarity

  There are methods of comparison based on indexes of evaluation models in (Pho et al., 2015) and comparison by item analysis in (Mitkov et al., 2009); by type of vocabulary {PersonLocationOrganization}; by "DBpedia entity", which is a semantic index using entities given to vocabulary on DBpedia; and by calculating and comparing "wup similarity," which is a semantic index using the distance in WordNet vocabulary hierarchical structure. For the time being, similarity was calculated by a pre-learned word2vec model.

6.3 Results

For the execution environment, we used MacBook Pro for PC and macOS High Sierra for OS. In this system, SPARQL ran with the library SPARQLWrapper[7] as the language to be used in Python. We also measured the execution time for question generation for 10 set Answers. The Question Graph required 76.66 seconds to read the pre-learned model of word2vec only once upon execution, and the time of Question Graph generation for each subsequent question was 13.09 sec. The time needed to generate distractors was on average 3.575 sec for each of the questions.

In the experiment, a Question Graphs, three Distractors and an Answer were set as a question set. In addition, we made 10 questions in a single answer format (hereinafter referred to as "single-answer form") and 5 questions for multiple answer formats (hereinafter referred to as "multiple-answer form").

[6]https://taku910.github.io/cabocha/

[7]https://rdflib.github.io/sparqlwrapper/

6.3.1 Results of Question Graph: Evaluation Method for Degree of Inclusion of Panoramic Information

For the evaluation of the degree of inclusion of panoramic information, we summarized the number of classes in one graph as the degree of crossing classes in Tables 2.

Table 2: The degree of crossing classes (single-answer form).

Q.	number of nodes		number of classes	
	proposal	random	proposal	random
1	21	21	7	5
2	21	17	5	3
3	21	21	6	5
4	21	21	6	7
5	21	21	3	8
6	21	21	8	5
7	21	19	6	7
8	21	21	6	3
9	21	21	3	3
10	21	21	6	3
Ave.	21	20.4	5.6	4.9

6.3.2 Results of Question Graph: Specificity Evaluation Method

Regarding the 10 Question Graphs generated in single-answer form, we queried the vocabulary set corresponding to the Answers node on DBpedia, and it was all empty except for Answer. In multiple-answer form, when two answers were specified, all the graphs were such that there were only two words corresponding to Answer nodes. Therefore, it can be said that the evaluation items of specificity were satisfied by this experiment.

6.3.3 Results of Distractors: Comparison of Validities by Experimental Subject

Results by subjective experiment are shown. In Table 3 and Figure 4, results obtained from 37 subjects are summarized, including the validity for single-answer form. Similarly, for the multiple-answer form, the results of 23 respondents are summarized.

6.3.4 Comparison of Similarity between the Answer and Choices

For each distractor using the *dist* function (proposal) and random generation (random); random selection of the neighbor nodes of each target node from all nodes that have links of the target node, similarity comparison by syntactic pattern (pattern) and the results of similarity comparison by pre-learned word2vec model (word2vec) are shown in Table 4. In the table, the

Table 3: The validity in experiment (single-answer form).

Q.	validity[%]	selectivity of Distractors (descending)[%]		
		D1	D2	D3
1	64.9	24.3	8.09	2.67
2	21.6	59.4	5.36	3.56
3	21.6	48.6	18.9	10.8
4	13.5	70.3	13.5	2.67
5	40.5	32.5	18.9	8.09
6	13.6	54.0	21.6	10.8
7	91.9	8.1	0.00	0.00
8	56.8	35.1	5.41	2.69
9	21.6	40.5	29.7	8.09
10	18.9	35.1	29.7	16.2

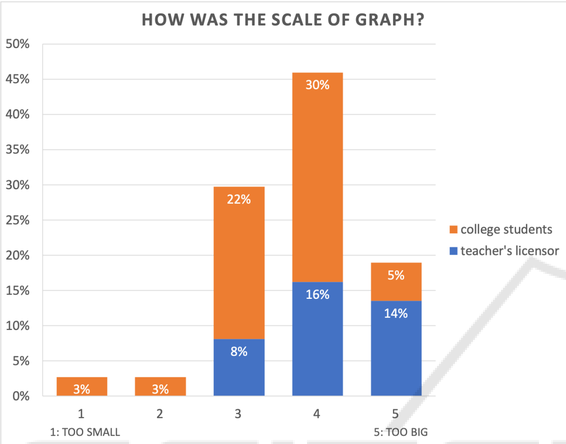

Figure 4: Questionnaire:"Scale of Question Graphs".

"pattern" value is set as 1 if the pattern matches the answer, and "word2vec" indicates the similarity with Answer. Each value is the average of the values per question; that is, it is the value of similarity per distractor.

Table 4: Comparison of similarity between Answer and Distractors generated by proposed method and random (single-answer form).

Q.	pattern		word2vec	
	proposal	random	proposal	random
1	2	2	.613	.330
2	2	2	.305	.248
3	2	1	.451	.364
4	3	1	.359	.325
5	3	2	.801	.369
6	2	0	.438	.534
7	3	0	.354	.179
8	2	1	.557	.360
9	0	2	.475	.463
10	0	2	.671	.510
Ave.	1.9	1.3	.503	.368

7 DISCUSSION

In terms of implementation, we set the vocabulary on DBpedia as the learned keyword without setting the curriculum, so it was obvious that mandatory requirements (1), "each node consists of keywords," and (2), "ensure the connectivity of information around Answer," were satisfied in the Question Graph. On the other hand, if setting any curriculum and keywords exist on Linked Data to be used, the requirements are satisfied and the method can be used. Also, if the answer is a special vocabulary, there is a possibility that any link structures may not be found and the generation will fail theoretically. However, it seems that the major vocabulary such as those that appear in the examination and its relationships are largely covered by LOD like DBpedia. Regarding requirement (3), "the degree of inclusion of panoramic information," the proposal was exceeded the random by an average of 0.7 classes (as seen in Table 2).

From this, with respect to the index on the number of classes, the degree of panoramic information was greater by proposal than by random generation in the experiment. Regarding "scale of the graph" in requirement (4), the results of the questionnaire on subject experiments in Figure 4 showed that 4 out of 5 evaluations occurred most frequently in this experiment. Therefore, the scale was large based on the subjects chosen. Finally, regarding "the number of words corresponding to Answer node" in requirement (5), the requirement was satisfied since there were no words corresponding to the answer nodes except on answers in the 15 graphs.

Regarding distractors, mandatory requirement (1), "each node consists of keywords," and (2), "incorrect answer to the question," are satisfied from the generation approach in the experiment. Regarding the second evaluation of Requirement (3), "avoid words that are clearly recognized as incorrect answers," in the case all subjects do not know the correct answer, it is desirable that the correct answer rate for each question is one out of the number of choices; 1/4 =25% or less. For that evaluation, six questions in single-answer form satisfied the index. Questions 1, 5, 7 and 8 did not. Among the four unsatisfied examples, in Q.7, only one out of the three distractors was selected, so it was a remarkable result that did not satisfy the requirement. Regarding the multiple formats, the difference in the selectivity between distractors is less than 13%, and the selectivity was not biased. Next, in Table 4, results of the random is evaluation of Distractors generated by extracting randomly from words belonging to the same class as the answer on DBpedia. In these indexes, the proposal tended to have more similar Distractors with Answer than random.

In this study, we conducted a questionnaire to compare with existing questions (The Japanese History Aptitude Testing Foundation, 2017) (The Japanese History Aptitude Testing Foundation, 2018) for 14 teachers' license holders. For the five single-answer forms, we got responses mainly on "degree of panoramic information". An average of 72.8% answered that the degree of panoramic of proposed questions were higher than the random in Questionnaire 1. In addition, Figures 5 and 6 show the evaluation results of the degree of panoramic information and content. Not only the difference in degree of panoramic information but also in content shows a big impression overall.

8 CONCLUSION

In this paper, we proposed a method of generating multiple-choice questions including panoramic information. Prospects are listed below.

In the proposed method, we considered the distance between nodes to generate a graph including panoramic information, but did not consider the meaning of links between nodes. To generate intentional test questions, not only the nodes should be considered, but also the types of links. In the evaluation of the degree of inclusion of panoramic information, we evaluated based on the three items; degree of crossing classes, the degree of crossing time, which was impossible with only these indexes. Therefore, we should review the current evaluation indexes, clarify the definition of the degree of panoramic information, and set up an evaluation index based on it to conduct experiments. Currently, it is necessary to analyze how this index influences the degree of the whole graph. In consequence, Distractors generated by proposed method tend to be more similar than the random as the evaluation indexes of pattern and word2vec.

In the generation of distractors, if the class to which an answer belongs is not unique, only one kind of candidate apparently different from the answer may be generated, so this process should be improved. Also, to deal with synonyms between choices, we will establish a verification phase using WordNet[8].

Also, as a new question form application of this proposal, a combination question is considered. The combination question is often seen in Japanese history and world history examinations of the National Center Test for University Admissions[9], which is a format that answers combinations of answers in different questions from choices. We are considering that there should be a demand for this format.

In the future, we aim to improve the method of automatically generating questions considering panoramic information by reviewing the approaches and evaluation method for Question Graphs and distractors.

[8] https://wordnet.princeton.edu

[9] https://www.dnc.ac.jp/center/

ACKNOWLEDGEMENTS

This work was supported by JSPS KAKENHI Grant Numbers JJP16K00419, JP16K12411, JP17H04705, JP18H03229, JP18H03340 and JP18K19835.

REFERENCES

Afzal, N. and Mitkov, R. (2014). Automatic generation of multiple choice questions using dependency-based semantic relations. *Soft Computing*, 18(7):1269–1281.

Bühmann, L., Usbeck, R., and Ngomo, A.-C. N. (2015). Assessautomatic self-assessment using linked data. In *International Semantic Web Conference*, pages 76–89. Springer.

Demarchi, F., Santos, E. R., and Silveira, R. A. (2018). Integration between agents and remote ontologies for the use of content on the semantic web. In *Proceedings of the 10th International Conference on Agents and Artificial Intelligence - Volume 1: ICAART,*, pages 125–132. INSTICC, SciTePress.

Fionda, V. and Pirrò, G. (2017). Meta structures in knowledge graphs. In *International Semantic Web Conference*, pages 296–312. Springer.

Iijima, T., Kawamura, T., Sei, Y., Tahara, Y., and Ohsuga, A. (2016). Sake selection support application for countryside tourism. In *Transactions on Large-Scale Data- and Knowledge-Centered Systems XXVII*, pages 19–30. Springer.

Ikeda, S., Takagi, T., Takagi, M., and Teshigawara, Y. (2013). A proposal and evaluation of a method of estimating the difficulty of items focused on item types and similarity of choices. *Journal of Information Processing*, 54(1):33–44.

Ikegami, M. (2015). Comparison of strategy use between different item format types on multiple-choice grammar test. *Studies in Language and Literature*, 35(1):55–72. (in Japanese).

Jouault, C., Seta, K., and Hayashi, Y. (2016). Content-dependent question generation using lod for history learning in open learning space. *New Generation Computing*, 34(4):367–394.

Maillot, P., Raimbault, T., Genest, D., and Loiseau, S. (2014). Targeted linked-data extractor. In *Proceedings of the 6th International Conference on Agents and Artificial Intelligence - Volume 1: ICAART,*, pages 336–341. INSTICC, SciTePress.

Ministry of Education, Culture, Sports, Science and Technology (2015). The 100th primary and secondary education subcommitteehandouts 1-1 4. necessary measures to realize the philosophy of government guidelines for teaching. http://www.mext.go.jp/b_menu/shingi/chukyo/chukyo3/siryo/attach/1364319.htm.

Mitkov, R., Ha, L. A., Varga, A., and Rello, L. (2009). Semantic similarity of distractors in multiple-choice tests: extrinsic evaluation. In *Proceedings of the Workshop on Geometrical Models of Natural Language Semantics*, pages 49–56. Association for Computational Linguistics.

Papasalouros, A., Kanaris, K., and Kotis, K. (2008). Automatic generation of multiple choice questions from domain ontologies. In *e-Learning*, pages 427–434. Citeseer.

Patra, R. and Saha, S. K. (2018). A hybrid approach for automatic generation of named entity distractors for multiple choice questions. *Education and Information Technologies*, pages 1–21.

Pho, V.-M., Ligozat, A.-L., and Grau, B. (2015). Distractor quality evaluation in multiple choice questions. In *International Conference on Artificial Intelligence in Education*, pages 377–386. Springer.

Rocha, O. R., Zucker, C. F., and Giboin, A. (2018). Extraction of relevant resources and questions from dbpedia to automatically generate quizzes on specific domains. In *International Conference on Intelligent Tutoring Systems*, pages 380–385. Springer.

Suzuki, M., Matsuda, K., Sekine, S., Okazaki, N., and Inui, K. (2016). Neural joint learning for classifying wikipedia articles into fine-grained named entity types. In *Proceedings of the 30th Pacific Asia Conference on Language, Information and Computation: Posters*, pages 535–544.

The Japanese History Aptitude Testing Foundation (2017). *The Japanese History Aptitude Testing Foundation in 2016, the 35th time, question collection of all the grade, pp.29-pp.67*.

The Japanese History Aptitude Testing Foundation (2018). *The Japanese History Aptitude Testing in 2017, the 36th time, question collection of all the grade, pp.33-pp.75*.

Zoumpatianos, K., Papasalouros, A., and Kotis, K. (2011). Automated transformation of swrl rules into multiple-choice questions. In *FLAIRS conference*, volume 11, pages 570–575.

A Questionnaire Results in Evaluation Experiment

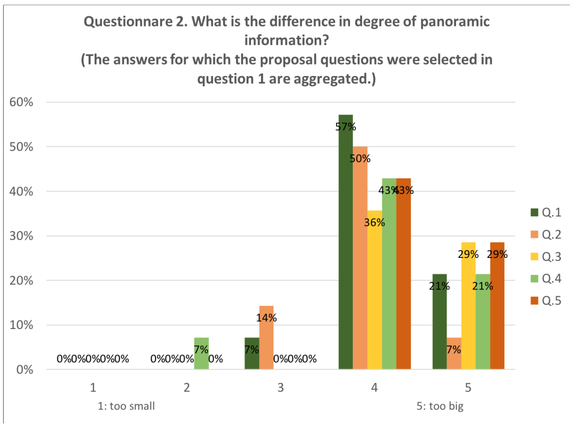

Figure 5: Questionnairem 2: “Comparison of differences in degree of panoramic information”.

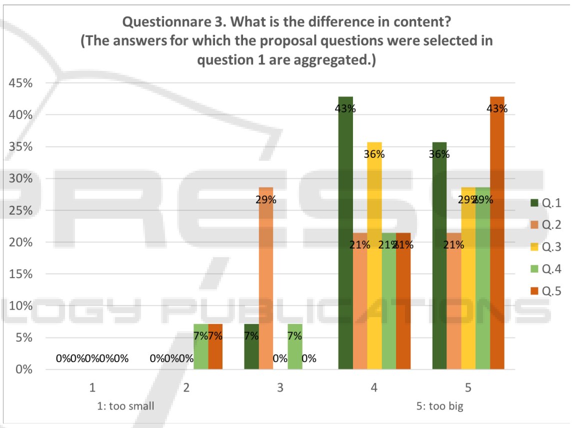

Figure 6: Questionnaire 3: “Comparison of differences in degree of information contents”.