System Design for Automatic Generation of Multiple-Choice Questions Adapted to Students' Understanding
Shin'ichi Tsumori and Kenji Kaijiri
Interdisciplinary Graduate School of Science & Technology, Shinshu University, Nagano, JAPAN
E-mail: tsumori@shotoku.ac.jp

ABSTRACT
This paper reports the system that has a characteristic of generating multiple-choice questions with adapted level of difficulty automatically. The system has two major features. One is to generate a lot of questions using two databases, technical-term database and question-template database. The other is to control the difficulty of question dynamically by selecting the adequate alternatives when questions are generated. Alternative terms are selected from the technical-term database that is hierarchically configured by technical-term information. It enables the system to gather similar or not similar alternatives in order to control the difficulty of questions.

INTRODUCTION
In recent years, many universities and colleges in Japan have been making some kinds of special plan, for example, giving preference to examinees with licenses on entrance examinations, and opening new courses for students who prepare for certifying examinations. As the number of students who want to become qualifiers increases, the demand for such a course increases. As the test method for many qualifying/certificate examinations, multiple-choice questions is adopted. Multiple-choice questions are hardly used in order to confirm in-depth knowledge with an analogy and so on because it restricts examinees to answer with a few alternatives. However, multiple-choice questions are suitable for confirming knowledge like the relation of the name of a word and its definition. Moreover, it is often used for the large-scale examinations because of its high objectivity and its easiness of mechanical grading. When students study for qualifying/certificate examinations, they often execute exercises using the questions in the past examinations. It is thought that the acquisition of domain knowledge by solving past examinations is effective to the examination grid because actual examinations often use a lot of same/similar questions in the past examinations. On the other hand in order to study effectively within a limited time, it is necessary for students to find their weak points and to execute exercise repeatedly until they understand. In order to support this type of learning, a large amount of questions need to be prepared by teachers beforehand, and they should be given to students according to their understanding levels. However, it is difficult for teachers to prepare such a huge question pool beforehand and to grasp the each student’s understanding level dynamically. Some researches have tried to solve such problems. Sherman X.Huang [1], Anh Nguyen Viet, et al [2], Edmond Holohan, et al [3], Hongchi Shi, et al [4] and Bernhard Thalheim [5] proposed methods that select teaching materials for students individually. These are approaches that select and offer teaching materials and questions corresponding to the understanding levels dynamically by measuring students’ understanding levels and difficulty of questions. However, these researches have problems that teachers need to make a lot of teaching materials and questions beforehand. Jose Antonio Gonzalez, et al [6], Edmond Holohan, et al [7], Enrique Lazcorreta, et al [8], Hans Christian Liebig, et al [9], and Gerald F. Braun, et al [10] proposed methods for generating questions automatically. These approaches aim to generate a lot of questions automatically by using ontology or knowledge base. However, they have problems at the point of learning effect because their adaptation is not based on students’ understanding levels, and as the result generated questions may be too easy or too difficult for students. We have been examining the method of generating multiple-choice questions that are adapted to student’s understanding level automatically. Multiple-choice questions can be based on the fixed forms by question sentences, so it is easy to generate questions mechanically. Moreover, difficulty of multiple-choice questions can be controlled by selecting confusing alternatives. In addition, it is possible to make questions about the relationship between a child class and its parent class by defining terms hierarchically. This method is expected to solve the trade-off problem between the adaptation load of questions according to students’ understanding and the teachers’ load of generating questions.

In this paper, automatic generation of multiple-choice questions according to students’ understanding level is discussed. First, the generation method for multiple-choice questions and the conceptual distance are explained. Next, the computational method for students’ understanding level and the difficulty of
question are discussed. The conclusion and the future problems are described in the last section.

HOW TO GENERATE QUESTIONS

The multiple-choice question is the type of question that is selected one answer (sometimes two or more answers) among two or more alternatives that are the candidates of correct answers. It is difficult to be used for the question that purpose is to confirm the student’s in-depth knowledge, because the way of inputting is limited in a few alternatives only. However, it has been widely used in qualifying/certificate examinations and educational systems such as CAI (Computer Assisted Instruction). One of the methods of presenting the multiple-choice question in CAI is to make questions, alternatives, and knowledge of results (KR) as the frame that is the unit of the display. However, because the multiple-choice question is expressed in the fixed form composed by a question sentence and two or more alternatives, questions can be generated automatically by using term information stored in the database. In addition, when new terms are added to the database, new questions will be generated in the same way. That is why this method enables to generate a lot of questions easier than a method of preparing a lot of frames by teachers beforehand.

In this paper, we propose a method to generate multiple-choice questions automatically by using term information shown in Figure 1, which is stored in the technical-term database.

![Figure 1: term information](rom.png)

### Figure 1: term information

- **symbol**: ROM
- **alias**: Read-Only Memory
- **definition**: a memory that can only be read
- **feature**: data isn't deleted even if the power supply is turned off
- **figure file**: rom.png
- **is-a**: memory
- **part-of**: mother board

A question is generated by applying term information to a template stored in question-template database. Figure 2 is an example of the template to ask correspondence between the “definition” of the term and the “symbol” of the term.

![Figure 2: question template](rom.png)

### Figure 2: question template

Q. Select the term that is defined by the following sentence:

" 

a. 

b. 

c. 

d. 

The question shown in Figure 3 (a) can be generated by filling the symbol “ROM” and the definition “a memory that can only be read”, and other symbols, "RAM", "DRAM", and "Flash Memory", chosen from technical-term database into the template shown in Figure 2. It is also possible to generate different types of questions by using other templates. Figure 3 shows different types of sample questions generated using various templates in the question-template database and term information in the technical-term database. Questions from (b) to (f) are to ask the correspondences between two attributes. For example, question (b) is a type of question that asks correct definition of the term. Question (g) and (h) are the samples using “is-a” and “part-of” attributes. “Is-a” and “part-of” attributes are used to show the relations between concepts. “Is-a” is used to compose the conceptual space described in the next chapter. “Part-of” composes another conceptual space.

CONCEPTUAL SPACE

In general, we can adapt the difficulty level of multiple-choice questions based on the selection of alternatives. For example, difficulty levels of the question in Figure 3 (a) and the other question in Figure 4 (both correct alternatives are “ROM”) are different because their alternatives are not the same each other, even if the question sentences are the same. The student might feel that the question in Figure 3 (a) is more difficult than one in Figure 4, because the question in Figure 3 (a) includes many alternatives whose concepts are near to “ROM” that is the correct alternative. The multiple-choice question containing a lot of alternatives whose concepts are similar to the correct alternative is more difficult than the question whose concepts are far from the correct alternative. For the purpose of considering the similarity between
concepts, we allocate terms in the learning domain on the two dimensional space, that is, we allocate terms within a tree by using "is-a" attributes. Figure 5 is an example that expresses a hierarchy of concepts in the field of "memory". In this paper, we call the tree expressed by Figure 5 "conceptual space".

<table>
<thead>
<tr>
<th>(a) definition → symbol</th>
</tr>
</thead>
<tbody>
<tr>
<td>Select the term that is defined by the following sentence: &quot;a memory that can only be read&quot;.</td>
</tr>
<tr>
<td>a. RAM, b. ROM, c. DRAM, d. Flash Memory</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>(b) symbol → definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Select the description that defines &quot;ROM&quot;.</td>
</tr>
<tr>
<td>a. a memory to be able to read and write</td>
</tr>
<tr>
<td>b. a memory that can only be read</td>
</tr>
<tr>
<td>c. a memory that records data by whether there is charge in the capacitor or not</td>
</tr>
<tr>
<td>d. a memory that can be deleted content electrically</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>(c) feature → symbol</th>
</tr>
</thead>
<tbody>
<tr>
<td>Which term has the feature &quot;data isn’t deleted even if the power supply is turned off&quot;?</td>
</tr>
<tr>
<td>a. RAM, b. ROM, c. DRAM, d. Flash Memory</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>(d) symbol → feature</th>
</tr>
</thead>
<tbody>
<tr>
<td>Which description is adequate for the feature of &quot;ROM&quot;?</td>
</tr>
<tr>
<td>a. data is deleted when the power is turned off</td>
</tr>
<tr>
<td>b. data isn’t deleted even if the power supply is turned off</td>
</tr>
<tr>
<td>c. refreshing operation is necessary</td>
</tr>
<tr>
<td>d. the deletion and writing for the unit of the block are possible</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>(e) symbol → figure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Which figure represents &quot;ROM&quot;?</td>
</tr>
<tr>
<td>fig. a, fig. b, fig. c, fig. d</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>(f) figure → symbol</th>
</tr>
</thead>
<tbody>
<tr>
<td>Which term the following figure shows?</td>
</tr>
<tr>
<td>figure</td>
</tr>
<tr>
<td>a. RAM, b. ROM, c. DRAM, d. Flash Memory</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>(g) is-a relation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Which term belongs to &quot;auxiliary memory&quot;?</td>
</tr>
<tr>
<td>a. Floppy Disk, b. SRAM, c. EPROM, d. PROM</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>(h) part-of relation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Which term is a part of &quot;mother board&quot;?</td>
</tr>
<tr>
<td>a. Floppy Disk, b. ROM, c. CD, d. DVD-RAM</td>
</tr>
</tbody>
</table>

We can measure the similarity between two arbitrary concepts in the conceptual space. The similarity is defined by using a number of links between two concept nodes. Less number of the links between two concept nodes means that similarity between two concepts is higher. In this paper, we call the number of the links that connect two arbitrary nodes "conceptual distance". For example, the conceptual distance between "EPROM" and "ROM" is 1 because "EPROM" and "ROM" are connected by one link, and one between "EPROM" and "PROM" is 2 because "EPROM" and "PROM" are connected by two links via "ROM".

Q. Select the term that is defined by the following sentence: "a memory that can only be read".

a. CPU  b. ROM  c. HDD  d. CD-ROM

Figure 4: question sample (2)

Figure 5: conceptual space

STUDENT MODEL

In general, the multiple-choice question is used for the purpose of asking whether each attribute of dictionary knowledge, for example its symbol, definition, feature and so on of the term, can be appropriately connected to each other. Therefore, it is suitable to define student's understanding level at the unit of concept. In this paper, we assume that student's objective is to acquire all concepts in the conceptual space, so we use an overlay model for student model. We assign the understanding level of each concept existing in the conceptual space as shown in Figure 6. An understanding level is a numerical value from 0 to 1 when it is decided, or "*" when it is not decided.
Understanding level 0 means the student has not understood the concept at all (or system cannot explain that the student has not understood the concept). In contrast, understanding level 1 means the student has already understood the concept (or system cannot explain that the student has not understood the concept at all). When all understanding levels in the conceptual space exceed thresholds, students are assumed to be the end state of study. The understanding level in this paper does not have special quantitative meaning absolutely. For example, the understanding level 0.5 does not mean that the student understands 50% of the concept of the term. The understanding level has a meaning to be compared with other levels only. For example, the term whose understanding level is 0.5 is understood by the student better than the term whose understanding level is 0.3. The method of setting understanding levels will be described later.

**DIFFICULTY OF QUESTIONS**

![Figure 7: conceptual distance and understanding level](image)

The difficulty of questions varies depending on the student’s understanding levels even if the question and alternatives are quite the same. For example, the student who is studying the term (the student does not understand it completely) feels the question very difficult even if the question contain only alternatives having small conceptual distance. On the other hand, the student who has understood the term completely does not feel the question difficult regardless of the contents of incorrect alternatives. From the above consideration, the difficulty of questions is related with two items as follows:

a) conceptual distance between alternatives of each incorrect one and a correct one

b) understanding level of each term for each user

To discuss this point further, we assume the multiple-choice question with a correct term and two incorrect alternatives. Figure 7 shows the conceptual distance between correct alternative \( p_0 \) and incorrect alternatives. The conceptual distance of each incorrect alternatives and correct alternative \( p_0 \) is expressed by the radius. Figure 7 shows that the conceptual distance between \( p_0 \) and \( p_1 \) is 3. The value 0.1 attached to \( p_1 \) means that the understanding level of \( p_1 \) is 0.1. Based on this assumption, we compare two questions as follows,

a) question that has correct alternative \( p_0 \) and incorrect alternatives \( \{p_1, p_2\} \)

b) question that has correct alternative \( p_0 \) and incorrect alternatives \( \{p_1', p_2'\} \)

The conceptual distances from \( p_0 \) to \( p_1 \) and \( p_2 \) are 3, 4 respectively, and both conceptual distances from \( p_0 \) to \( p_1' \) and \( p_2' \) are 2. If we don’t consider the understanding levels, the difficulty of question a) is lower than one of b), because question b) includes the alternatives whose concepts are similar to correct alternative \( p_0 \). However, the understanding levels of \( p_1 \) and \( p_2 \) are low (0.1 and 0.5 respectively). We can explain this reason that the student cannot distinguish \( p_1 \) and \( p_2 \) from \( p_0 \) because she/he understands neither \( p_1 \) nor \( p_2 \) well.

From the above consideration, two factors affect the difficulty level of questions as follows,

1. If the understanding level of each alternative becomes higher, the difficulty of question becomes lower.
2. If the conceptual distance between each incorrect alternative and correct alternative becomes longer, the difficulty of question becomes lower.

Then, we assume that the difficulty of question \( \text{diff} \) is calculated using a function as shown in formula (1), where \( u_c \) is the understanding level of \( p_c \) \((c \geq 1)\), \( d_c \) is the conceptual distance from \( p_c \) \((c \geq 1)\) to \( p_0 \) (correct alternative) and \( n \) is the number of alternatives.

\[
\text{diff} = f_{\text{diff}}(u_c, d_c, n)
\]

Formula (2) is a current implementation of the formula (1). According to the formula (2), the difficulty of question a) 0.21 and that of question b) is 0.08. It shows that question a) is more difficult than question b).

\[
\text{diff} = \frac{\sum_{c=1}^{n}((1-u_c)/d_c)}{n}
\]

The examination of the validity of formula (2) is a future problem.

**UNDERSTANDING LEVELS**

The understanding levels of terms are updated after each question is answered. As for the multiple-choice question, students only select one item from several alternatives, so the possibility of the correct answer by so-called “fluke” becomes greatly higher than other test methods. Moreover, student's thought process cannot be guessed from her/him answers. Therefore, it
is impossible to make accurate student model using
only one question. However, there is no problem to
judge that the student does not understand a term
concerned in the question. In the multiple-choice
question, the randomly selected answer may be
correct, but the probability that such an event occurs
continuously is very low. Highly trusted understanding levels are acquired after exercise by
many questions.
Now, we discuss how to set/modify the understanding
levels. It is assumed that the question with alternatives
\{p_0, p_1, \ldots, p_n\} were given to the student.
As for choosing the alternatives that the system sets /
modifies, the system adjusts the understanding level
based on the answer.
(a) When the student’s answer is correct, only the
understanding level of \( p_0 \) will be updated.
(b) When the student’s answer is wrong, both the
understanding levels of the incorrect alternative
and correct alternative will be updated, for
example, the question includes correct the
alternative \( p_0 \) and the incorrect alternatives \( \{p_1, p_2\} \) and the student selects \( p_1 \), then both the
understanding levels of \( p_1 \) and \( p_0 \) will be updated.
We assume that the student who understands certain
term can answer the questions continuously that
include that term. In addition, we assume that the
student who answers difficult question correctly has a
high understanding level about the terms included in
this question. Then, the system updates the
understanding level \( u_{i,m} \) for the alternative \( p_i \) on
times \( m \) by formula (3), using \( u_{i,m-1} \) and the
difficulty of question \( \text{diff} \).
\[
\begin{align*}
  u_{i,m} &= f_{ul}(u_{i,m-1}, \text{diff}) \\
  &= \frac{c \cdot u_{i,m-1} + \beta \cdot \text{diff}}{\alpha + \beta}
\end{align*}
\]
Formula (4) and (5) are the candidates of formula (3).
Formula (4) is for when a student answers the
question correctly, and formula (5) is for when a
student answers it incorrectly.
\[
\begin{align*}
  u_{i,m} &= g \cdot u_{i,m-1} \\
  &= \gamma \cdot u_{i,m-1}
\end{align*}
\]
Both \( \alpha \) and \( \beta \) are positive constants for calculating
weighted average, and they should be decided properly from a point of view of attaching importance
to the past understanding level. \( \gamma \) is a positive
constant ranging from 0 to 1 that decide a degree of
dropping the understanding level of the term when
students mistake. Though all default values for
understandings level \( u_{0,0} \) should be "\*", they all are
set to 0.5 because understanding levels of all terms
are unknown in initial state.
The examination of the validity of formula (4) and (5)
is also a future problem.

**PROCESS OF EXERCISES**

Students execute exercises following the procedures
below.
(1) The student selects the domain to execute
exercises using questions. Term information and
the student model of the conceptual space are
loaded after the selection.
(2) The system generates a question and gives the
student it. The method for generating questions
will be explained in the next chapter.
(3) The system evaluates student answer and updates
the understanding level according to the answer.
(4) The student finishes studying the domain when
all understanding levels of terms in the
conceptual space exceed the given threshold (the
selection of appropriate threshold is a future
problem).

**STRATEGY OF GIVING QUESTIONS**

The system generates the adapted questions by using
the following procedures.
(1) The system selects one alternative from the
conceptual space for correct one. It should be
selected among the terms that the student has not
understood yet or has not studied yet.
(2) The system decides the range of the difficulty of
question for generating question. The range is
assumed that the question is neither too easy nor
too difficult for the student. For example, the
understanding level of alternative selected in
procedure (1) is assumed to be 0.6, the range
should be decided higher than 0.6. Understanding
level and difficulty level is different concept, but
we use them for the purpose of adjusting the
students’ understanding level to the difficulty
level of questions, so we treat them as
comparable concepts.
(3) The system selects a type of question (for
example, asking the correspondence between the
symbol and the definition). There are some types
for generating question, for example, selecting
one randomly, selecting one that has not given in
the past, and selecting one that has not answered
correctly. The best method of selecting type of
question according to student’s understanding
level is a future problem.
(4) The system selects some terms from the
conceptual space that are incorrect alternatives. In
this case, the conceptual distance and the
understanding level are selected according to the
difficulty level determined in step (2).
(5) The system generates a sentence which will be
given to the student as a question.
We have not decided appropriate default values for
initial understanding levels and how to examine the
validity of our method.
CONCLUSION

The method of generating multiple-choice questions that are adapted to student's understanding level automatically was discussed. Multiple-choice questions are a type of question that has been used widely because it is easy for inputting answers and for grading mechanically. In addition, it is possible to adapt to the various learning domains. On the other hand, it is difficult to generate and give adapted questions to the students, and researches for solving this problem have been not enough.

The characteristic of our research is to select alternatives dynamically based on the student’s understanding levels of each term in the conceptual space. As a result, it is possible to give adapted questions according to students’ understanding levels. In this paper, we discuss the system for the domain of information processing. However, it is also possible to apply this method to other domains by appropriately constructing the (technical-) term databases.

We need to solve the following problems in order to apply this method to real situation.

1. definition of formulas for the understanding level and the difficulty of questions
   We have to decide the specific formulas for the understanding level and the difficulty of questions shown as $f_{diff}$ and $f_{ul}$.

2. examination of setting range of understanding level
   According to this paper, students may have to solve a lot of questions to finish studying because the system needs students to set high values to all terms. In our method, students need to answer many questions in order to get understanding levels which are higher than given threshold, so some method, for example the use of hierarchy for updating the understanding level, is needed to decrease the number of questions.

3. examination of method for making term information and usage
   If certain term’s explanation involves other term information, it causes other new problem. For example, if “DRAM” is defined using “RAM”, students may solve the question includes alternative “DRAM” because the question teaches her/him that DRAM is a kind of RAM. If the question has two alternatives that have parent-child relationship each other, it may cause that two or more alternatives become correct. It is necessary to clarify the making rule of term information and the method of selecting alternatives in order to avoid these problems.

4. consideration to the conceptual space that is expressed using semantic network
   In this paper, the conceptual space is assumed to be a hierarchy model because of easy discussion. However, there are many domains that should be expressed using the semantic networks. We have to consider the method of calculating the understanding level and the difficulty of question about the semantic network.

We plan that we will develop this system, and make some simulations intended for various level students to verify the validity of technique in this paper.

REFERENCE