Automatic Answering System to Learners' Questions in e-Learning Contents

Rong Ma[1], Kenji Yoshida[2], Tomoko Nishishita[3] and Hirotaka Nakayama[3]
[1]Institute of Intelligent Information and Communications Technology (IICT), Konan University
8-9-1 Okamoto Higashinadaku, Kobe 658-8501 Japan
marong001jp@yahoo.co.jp
[2]Guraduate School of Natural Science, Konan University
yoshida@konan.ed.jp
[3]Faculty of Science and Engineering, Konan University
8-9-1 Okamoto Higashinadaku, Kobe 658-8501 Japan
forestgreen194@hotmail.com
nakayama@konan-u.ac.jp

ABSTRACT

The purpose of this study is to propose an automatic answering system to give proper answers to learners' questions in using e-Learning contents. The proposed method utilizes TF-IDF, a peculiar synonym dictionary and the data set of template questions growing adaptively. It has been observed that the proposed system yields good results in our e-Learning contents on graphs of quadratic functions of high-school mathematics.

Keywords: term weighting, automatic answering system, term frequency (TF), inverse document Frequency (IDF), vector space model

1. INTRODUCTION

In Information Retrieval (IR) systems based on Vector Space Model (VSM), the TF-IDF scheme is widely used due to its simplicity and robustness as well as its tractability to enhancement. This paper proposes a method using TF-IDF to give proper answers automatically to learners' questions in using e-Learning contents. The system was built so that it may give proper answers automatically to learners as quickly and correctly as possible in an e-Learning content on graphs of quadratic functions of high-school mathematics.

2.THE E-LEARNING CONTENT

Our e-Learning content is about quadratic functions of high-school mathematics. This system was developed by means of MySQL and PostgreSQL. And, we use Eclipse 3.1 vision as Java Platform. The computation module was created by Java. The composition of this e-Learning system is shown in Figure 1. Figure 2 shows an example of this e-Learning content.

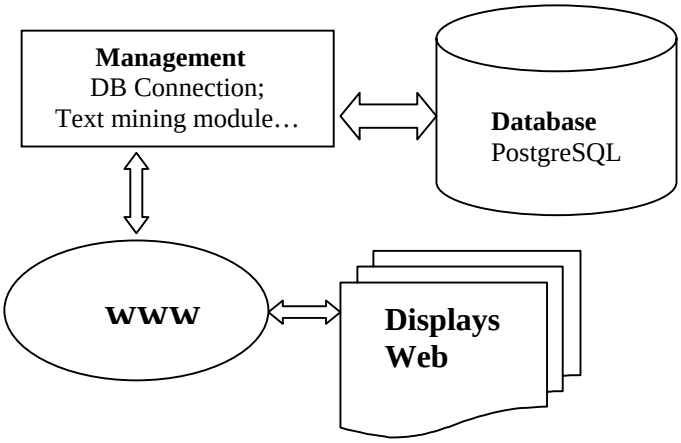

Fig.1. Composition of the e-Learning system

3. RELATED WORKS AND PROPOSED METHOD

As we know, the syntactic or semantic analysis of the natural language processing is often used in problems like automatic answering. However, in our research here, without using these methods, we tried a new method and gained a better result by changing weights of indexing terms of the querying vectors according to a text mining method based on the VSM of IR, and let the template question database changes itself when the e-learning content is used in practice.

We set 10 template questions as 10 categories. The automatic answering system is composed of the following processes:

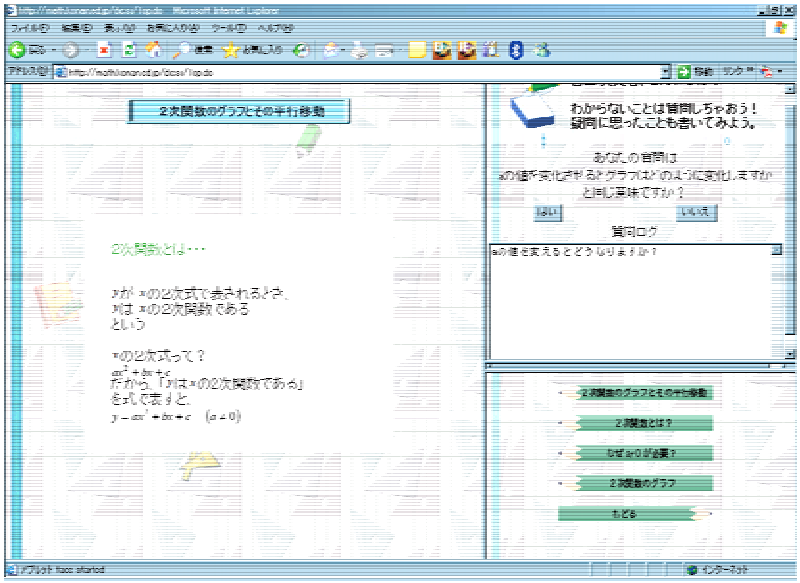

Fig.2. A main window of the e-Learning content

Step 1: morphological analysis
Step 2: extract of the indexing terms
Step 3: synonym change
Step 4: term weighting
Step 5: calculation of similarity degree
Step 6: distinction, etc.

3. 1. Morphological analysis

Morpheme analysis is one of the basic technologies of the natural language processing. It divides sentences, into morphemes (the minimum units of sentences) and shows parts of speech.

In this research, we analyze morphemes of questions from students using a Japanese morpheme analysis system "Chasen". Table 1 shows an example of parts of speech given by “Chasen”.

3. 2. Extraction of indexing terms

Words can be mainly classified into two kinds of words: the content words and the function words in the natural language processing. The content word is a word that means itself like noun and verb and so on. The function word is a word without a meaning, which shows relationships among words,etc.

Japanese includes lots of function words, such as auxiliary and preposition. For example, it has “ga, ne, ni, yo…”, which are similar to “of, in, at…” in English.

Table 1. Parts of speech by Chasen

品詞	Parts of speech
未知語	An unknown word
助詞-連体化	Auxiliary word - attributively
助詞-格助詞-一般	Auxiliary word -Standard auxiliary-general
助詞-格助詞-引用	Auxiliary word - Standard auxiliary-quotation
助動詞	Auxiliary verb
助詞-副助詞／並立助詞／終助詞	Auxiliary word, Vice- auxiliary, suffix-particle, suffix,
動詞-自立	Verb - independent
動詞-接尾	Verb - a suffix
連体詞	Attributive form
名詞-非自立-助動詞語幹	Noun- non independent - auxiliary stem
名詞-サ変接続	Noun - A "sa" connection
名詞-一般	Noun – general

In this research, we delete such words, which appear at the high frequency, because these unnecessary words cannot be as indexing terms.

Table 2 is an example of parts of speech extracted from one student's question. Here, the question is decomposed into morphemes. An example of extracted indexing terms of the student question is shown in Table 3. Table 4 shows all the 28 indexing terms used in our e-Learning content.

Table 2. An example of part of speech of one student question

Part of Speech	Morpheme	Meaning in English
unknown word	a	a
a particle / auxiliary	の	
a noun, general	値	number
a case particle / Standard auxiliary	を	
a noun, "sa" connection	変化	change
a verb, independent	する	do
a verb, suffix	せる	
a particle / auxiliary	と	and
a noun, general	グラフ	graph
a case particle / Standard auxiliary	に	
a attributive form	どの	what
a noun, non-independent	よう	
a case particle / Standard auxiliary	だ	
a noun, "sa" connection	影響	influence
a case particle / Standard auxiliary	が	
a verb, independent	ある	be
a case particle/ Standard auxiliary	ます	
a particle, suffix / auxiliary	か	?

Table 3. An example of indexing terms extracted from one student question

Part of Speech	Indexing terms	Meaning in English
an unknown word	a	a
a noun, general	値	number
a noun, "sa" connection	変化	change
a verb, independent	する	do
a noun, general	グラフ	graph
a noun, "sa" connection	影響	influence
a verb, independent	ある	be

Table 4. All indexing terms

Indexing terms	Part of speech in Japanese
a	unknown word
値(value)	a noun, general
変化(change)	a noun, general
する(do)	a verb, independent
グラフ(graph)	a noun, general
b	unknown word
c	unknown word
切片(slice)	a noun, general
a,b,c	unknown word
関係(relation)	a noun, general
ある(be)	a verb, independent
頂点(vertex)	a noun, general
なる	a verb, independent
座標(coordinate)	a noun, general
わかる(know)	a verb, independent
2	unknown word
次(dimension)	a noun, general
関数(function)	a noun, general
左右(symmetry)	a noun, general
対称(symmetry)	a noun, general
横(horizontal)	a noun, general
移動(transfer)	a verb, independent
(a,b,c)	unknown word
縦(vertical)	a noun, general
放物線(parabola)	a noun, general
横向き (transverse)	a verb, independent
描く(draw)	a verb, independent
方法(way)	a noun, general

3. 3. Weighting indexing terms

In our research, we use TF-IDF method for weighting indexing terms. We introduce the following notations:
m: number of indexing term;
n: number of the student question;
f_{ij}: frequency of the indexing term

(i) In local weighting, we use logarithmic Term Frequency (TF) here.

$$\text{TF}: \quad l_{ij} = \log(1 + f_{ij}) \qquad \begin{matrix}(i = 0, \cdots, m) \\ (j = 0, \cdots, n)\end{matrix}$$

So we can get the term-weight l_{ij}, which gives the local weight of an indexing term in document

(ii) In global weighting, we use probabilistic Inverse Document Frequency (IDF).

$$\text{IDF}: \quad g_i = \log \frac{n - n_i}{n_i} \qquad (i = 0, \cdots, m)$$

(iii) At last, we get the TF-IDF weight of indexing terms.

$$\text{TF-IDF} \ : \ d_i = l_{ij} g_i$$

3.4. Computation of the degree of similarity

To get a degree of similarity, now apply vector space model using the weight strings of the gotten indexing terms weights. The flow of the computation is: computes unique norm, normalizes each vector, and computes the cosine measure. We introduce the following notations:
s: Term vector of student's question
q_j: Term vector of template question

$$\cos \theta = \frac{s \cdot q_j}{\|s\| \cdot \|q_j\|}$$

3.5. Proposal of a peculiar synonym dictionary

In dealing with the question from the students who study from the content about the graph of quadratic function of mathematics in high school, we create a peculiar synonym dictionary.

The synonym dictionary is necessary because there are some words that are synonymous in mathematics and different in their usual meaning. For example, the synonymous sentences of "to move parallel to the direction of the horizontal axis" and "to move parallel to the direction of the vertical axis"; "to move parallel to the x axis direction " and " to move parallel to the y axis direction" respectively.

When extracting indexing terms from "to move parallel to the direction of the horizontal axis", "horizontal, axis, direction, parallel, move" are extracted and from the sentence of "to move parallel to the x axis direction", we have "x, axis, direction, parallel, move" as indexing terms.

In the context of mathematics, it is natural to consider that "x" is a synonym for "horizontal". However, it is difficult to regard these two words as synonymous generally. Table 5 shows some examples of our synonymies.

Table 5. Examples of the synonym dictionary

Synonyms	
Before	**After**
影響(influence)	関係(relation)
線(line)	左右(symmetry)
X	横(horizontal)
Y	縦(vertical)
90	横向き(transverse)
動かす(move)	移動(transfer)
変わる(change)	変化(shift)
わかる(know)	理解(understand)
°	度(degree)
逆さ(inversion)	逆(reverse)
目盛り(scale)	座標(coordinate)

3.6. Proposal of machine-learning

The definition of learning is: "things learn when they change their behavior in a way that makes them perform better in the future." In machine learning, it is believed that learning to performance rather than to knowledge. In our research, we propose a method to let the template

database grow adaptively. Our purpose is, the more using, the better the automatic answering system will be.

As mentioned before, we use the largest Cosine value as a criterion of the degree of similarity in computation of that. Here, if the value of the Cosine is bigger than 0.6, we have a good matching with the template database. If it is less than 0.6 and more than 0.2, all possible candidates are shown with giving a message box including "yes" and "no" buttons. When clicking "yes", this question's proper answer will be shown and it is added to the template database properly. It is expected that the template database will perform better in the future.

4. EXPERIMENTS

4.1. Experiments

To verify the effectiveness of this method, experiments were performed. The main flow-chart of text mining is shown in Fig.3. We compared results of experiments with/without the synonym dictionary. Throughout these experiments, it is expected that using the synonym dictionary can raise the percentage of correct answers.

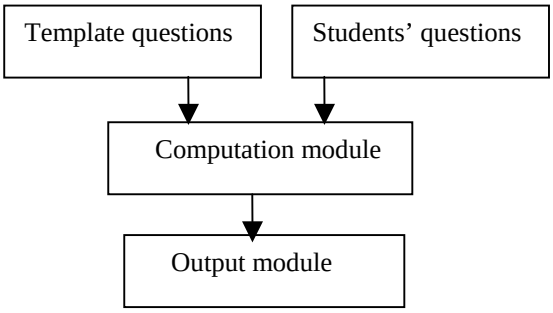

Fig.3 Main flow-chart of text mining

We make 10 template questions as 10 categories. Table 6 shows the result of experiments without a synonym dictionary. The right answer percentage is 74.07%.

Table 6. Result of experiment without a synonym dictionary

Numbers of template questions	10
Numbers of students' questions	27
Right answer percentage	74.07%

Table 7 shows the result of experiments using a synonym dictionary. It shows a better result. Fig.4 shows flow-chart of the experiment of using a synonym dictionary. The right answer percentage of this experiments is 92.59%.

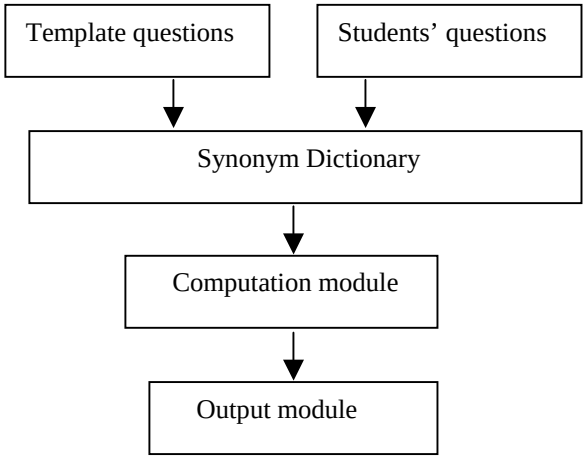

Fig.4. Flowchart of the experiment of using a synonym dictionary

Table 7. Result of experiment using a synonym dictionary

Numbers of template questions	10
Numbers of students' questions	27
Right answer percentage	92.59%

4.2. Using a machine learning method

As real usage, we believe that our template database should not be static; it needs to be dynamic, namely, to grow adaptively. As mentioned before, in computation of the degree of similarity, the largest Cosine value plays an important role as the criterion of similarity degree.

In our research, if the value of the Cosine is less than 0.6 and more than 0.2, we will take indexing terms of this candidate question into our template question database properly. The image of the machine learning method is shown in Fig.5.

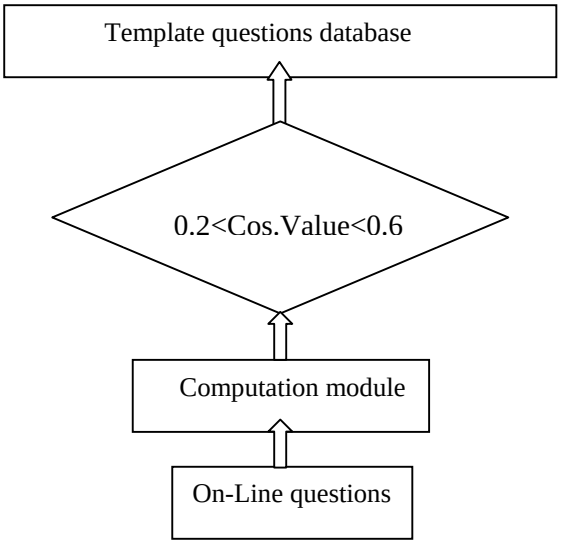

Fig.5. Image of the machine learning method

Furthermore, we conducted experiments using our method of the machine learning of template question database, with the synonym dictionary. Figure 6 shows the flowchart of the experiments using the machine learning method.

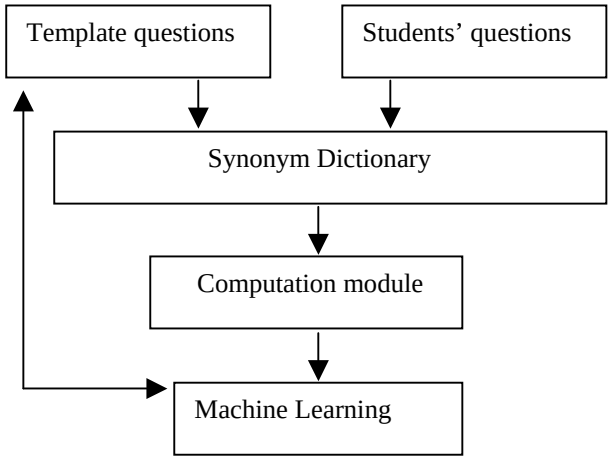

Fig.6. Flowchart of the experiment of using the machine-learning method

Also we conducted experiments of using the machine-learning model with the synonym dictionary. The result was good and right answer percentage had risen to 100% as Table 8 shows.

Table 8. Result of using machine-learning method

Numbers of template questions	10
Numbers of students' questions	27
Right answer percentage	100%

With machine learning method, for the increasing of numbers of indexing terms in this dynamic template database, and no enough students questions on line yet, we are going to gain a more reliable right answer percentage using more students questions.

5. CONCLUSIONS

As mentioned before, we made experiments of using a synonym dictionary or not. Due to these experimental results and calculations, it is expected that correct answer percentage could be raised largest by using the synonym dictionary. According to the experiment results, it is considered that the method of using probabilistic IDF way with the synonym dictionary is better, and we find the database of template questions becomes stronger to generate a more accurate vector by adding new indexing terms of the student question to the end of a category.

In order to improve the results, we want to do more research to know whether it is necessary to reduce some redundancy words in Japanese. Moreover, it is already observed that reducing some redundancy words gives us a better result.

ACKNOWLEDGEMENTS

This work was performed through ORC (Open Research Center) project (2004-2008) of MEXT (Ministry of Education, Culture, Sports, Science and Technology), Japan.

REFERENCES

[1] Information Retrieval Algorithms. Kenji Kita. 2002. Kyoritsu Publishers.
[2] Japanese morphological analysis system ChaSen, （Nara Advansed Institute of Science and Technology） http://chasen.aist-nara.ac.jp/
[3] Data Mining. Ian H.Written, Eibe Frank. Morgan Kaufmann Publishers 2000.
[4] Sen project home. https://sen.dev.java.net