



Using Case Based Reasoning to Answer Exam Questions

Tawin Tanawong

Department of Computer Science and Information Technology, Faculty of Science, Naresuan University, Pisanulok, 65000

* Corresponding author. E-mail address: tawint@nu.ac.th

Abstract

There are various uses to the knowledge source for the Tutorial System that is interesting and important to the expert system fields in an Artificial Intelligent development in the previous research studies. However we are interested in the questioning answering system which is applied to use the answering question test system in the formats of the previous exam question solving to the new exam question solving so that it can be given from the background knowledge in various ways such as the heuristics, experiences, textbooks and expertise, etc. The main point of our research is how we could give the knowledge from all of the sources to help for solving a new problem.

This paper shows how to use the case based reasoning to answer the questions. There are two parts in this paper. The first one is describing the relationship between the old case and new problem which have similar cases and the second one is applying the previous cases solving for new exam answering for the new solution. The main approaches for case retrieved with the process are the retrieve matching cases to the exam questions. The exam question forms are the key structural classification by the ontology and the other one is in the forms of the different knowledge structures but both are in the same question forms. We use the previous solving procedures to adapt the new solution problem solving and present respectively two cases, a problem description and solving procedure.

Keywords: Case-based reasoning, Ontology, Knowledge representation, Questioning-Answering system

Introduction

Tutorial system is one kind of current expert system research involved in the education system and is more interesting in the world of research. This will contribute question answering fields to the main areas question and answer group classification, comparing the question answering to the questioning similarity. These are divided into 13 groups (Lehnert, 1981) as according to answer questions by the answer group classification (Li, & Roth, 2002, pp. 1-7) and the linguists classify the question problems into 6 groups (Bloom, 1956). Moreover, there are researches that related questions to answers in the matching approach link to keywords (Lavenus, Grivolla, Gillard, & Gillard, 2004, pp. 623-637) and using ontology or bayesian networks

techniques for measuring similarity for question answering (Ramprasath, & Hariharan, 2012, pp. 218-223) (Larranaga, & Bielza, 2013). There are various researches using case based reasoning (Kolodner, 1993) from other sources, for instance, working experiences, experts, textbooks, etc that can be applied or adapted to new problem solving case based reasoning. One important way is to bring the knowledge from the case based reasoning to answer exam questions. Case retrieving, case reuse and case adaptation (Goel, 1996, pp. 67-73) (Wike, & Bergmann, 1998, pp. 497-505) (Spalazzi, 2001, pp. 3-36) (Schreiber, Wieling, & Breuker, 1993) (Peitao, Ke, Yatao, & Wei, 2011) may be the answers of these questions. Furthermore, case representation may have in forms of graphs, sets of object-oriented with attribute-values

And object-oriented in frame-based (Ralph, & Armin, 1998, pp. 25–36). The similar retrieving between the case problem in case based reasoning conceptual can be classified into three groups; the similarity assessment, improving and evaluating retrieval performance and alternatives to similarity-based retrieval (Mantaras, et al., 2006, pp. 215–240) or the nearest neighbor concepts. Likewise, question problem structural formats will be used in the ontology form. One important principle of conduct is to set question problem forms with the same Topic at the same problems but different in the Focus and emphasis on the similar measurement level, the key structure

similarity level, similar and different meanings respectively. Surface feature and structure similarity will be formed to the Concept domain, Topic and Focus of the ontological structure arrangement (Mann, 2002, pp. 1–7) (Angele, & Moench, 2003, pp. 913–928). For quick searching, Topic and Focus keywords are set to find the similarity value level to the attribute-value which will be calculated for local similarity and global similarity.

At last, the question problem adaptation of old question problems to new question problems is brought to apply for the solution procedure algorithm processing of old solution problem solving to the new question problems.

Methodology

1. The related factors and components of the system architecture

There are 3 main sections in the system architecture of the case based reasoning to answer exam questions shown in Figure1.

Section 1 discusses the exam question problems. Section 2 discusses the case retrieving and section 3 is the old problem solution procedure to new problem.

Section 1 exam question problems

1. The old exam question problems that case stored into the case based system.
2. The new exam question problems are input of the system.

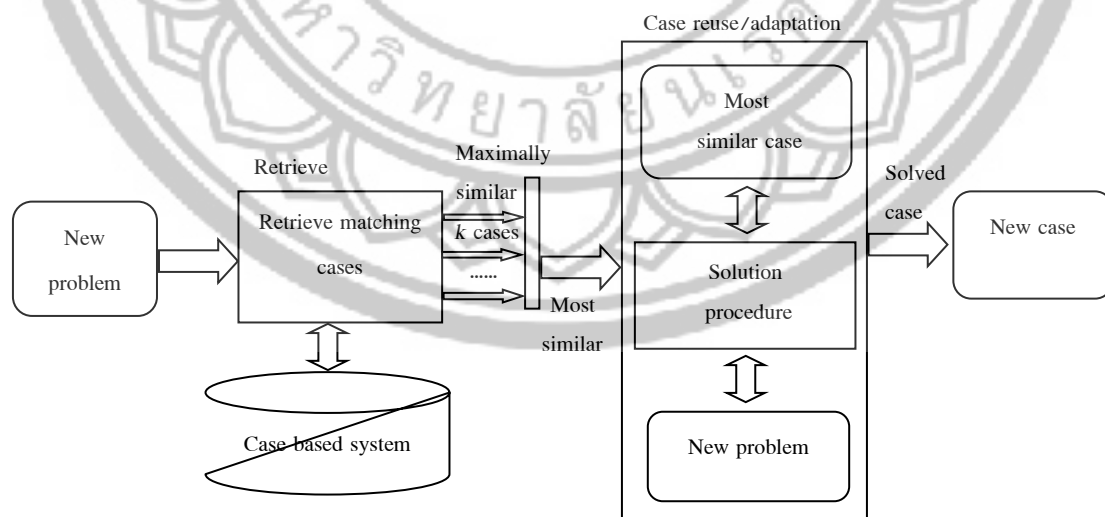


Figure 1 The system architecture to answer exam questions



Section 2 Case retrieving

1. The section of the retrieve matching cases process by finding the maximally similar k cases.

2. The selection of the most similar case has been given from the between old and new of the exam question problem.

Section 3 Old problem solution procedure

1. Applying used the old exam question problem solution procedure to new exam question problem.

2. Using an algorithms for the solution procedure.

3. Case represents to problem solving used from the old case similarity

2. Concepts to be considered about exam questions for the research

This is to find out similar cases from the questions and will use 5 forms of questions suitably through objectives in the studying as followings:

2.1 The questions which are very similar by putting the same meaning in words in other sentences.

2.2 The questions which are similar in some parts by diverting the issue a little.

2.3 The questions which are similar in some parts by diverting the issue a lot.

2.4 The questions which are similar in some parts but are opposite sides in meanings.

2.5 The questions which are very different in every parts.

The goals of using these questions in this research are database system subject and software engineering subjects which we can be applied to other subjects. For the question forms, the researcher brings examples of fundamental examination questions in order to look at the principles to find the answers from the similar questions and from one question to another. This research shows the interest

to the close relationship between the collected questions and the new ones to correct for finding answers. For answering new questions, it is necessary to indicate that there are old questions closed to the new ones. Thus for the close relations, can be case retrieving to find old cases in case based comparing to new cases to solve problems for finding the answers. The important principle of case retrieving is case matching to match old cases and new ones, after that we choose the most similar characteristics called "most similar case". This will limit under related questions. Moreover, matching will bring to interesting key structures, for example, explaining good points, and comparing or contrasting. In these example questions it presents some examples of test questions to retrieve by using interesting key structures as followings:

p1. What are the advantages of OIDs?

p2. What are the disadvantages of OIDs?

p3. What are the benefits of OIDs?

p4. What are the advantages and disadvantages of OIDs?

p5. In what situations would OIDs be a good way of representing database entities?

p6. How do OIDs make it easier to implement a database system?

p7. How do OIDs make it harder to implement a database system?

p8. How might OIDs be implemented?

p9. What extra information is needed in a database system to implement OIDs?

p10. Compare OIDs with? as ways of implementing a database system.

3. The concepts of the processing to the retrieve matching cases

3.1 The Topic and Focus features are defined as the primary key of the case representation

3.2 For the key structure of the question problem, if it contains the same or similar case, the relationship of old and new problem in term of similarity and key structure can be transformed into the ontological model.

3.3 The Topic, Focus, and Concept domain are represented as an ontological structures for case matching, so that R can be definition to $CD_1 \times CD_2 \times CD_3 \times \dots \times CD_n \times TC_1 \times TC_2 \times \dots \times TC_n \times F_1 \times F_2 \times \dots \times F_n$ where; CD_i : Concept domain, TC : Topic, F : Focus respectively.

3.4 Defining the matching properties to relationship between the question problem that have exactly matching forms , some maybe very matching, some fairly matching, some related but different matching, and bad matching.

3.5 Defining the feature types of the question problems and answer inner of cases to the similarity between the question problem calculations.

Figure 2 demonstrates the concepts of question which are based on the ontological structure for the distance similarity of the key structural question measurement. This consideration leads to the general constraint in defining the inter-class similarity [11] and the concept nodes of a class hierarchy:

$Sim(TC_1, CD_s) \leq Sim(TC_1, CD_x) \text{ IF } \langle TC_1, CD_s \rangle > \langle TC_1, CD_x \rangle$ It can be state that TC_1 and CD_x are more similar than TC_1 and CD_s . However, Every inner concepts nodes C_i is with a similar value $Sim_i \in [0...1]$ that the Following condition holds: if $C_1 > C_2$ then $Sim_1 \leq$

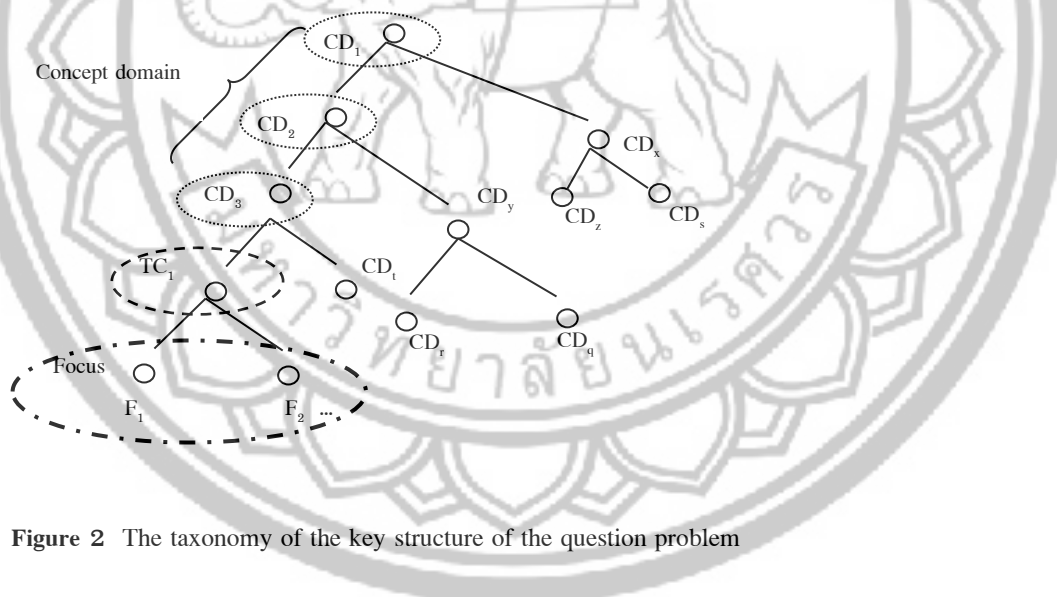


Figure 2 The taxonomy of the key structure of the question problem

Sim_2 . For example, question problem can be represented to the question relationship to ontology and the key structure of the topic as OIDs at the same point but its different to the focus of question

problem. This is that the question problem can be differentiated and calculated to find the most similar cases. We will describe the key structure of the Topic as OIDs that is shown in Figure3.

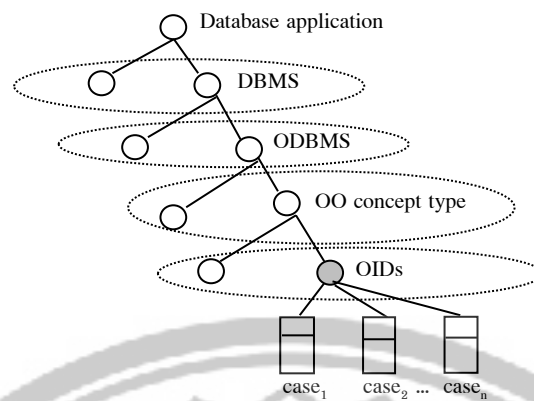


Figure 3 Taxonomy of key structure from questions

4. Case representation

This case can be divided into two main parts: problem description, solution and outputs shown in Figure4. The first question problem will describe the

case feature components within of the feature values of the question problem which has a problem description. The second is about the question problem procedural relationship

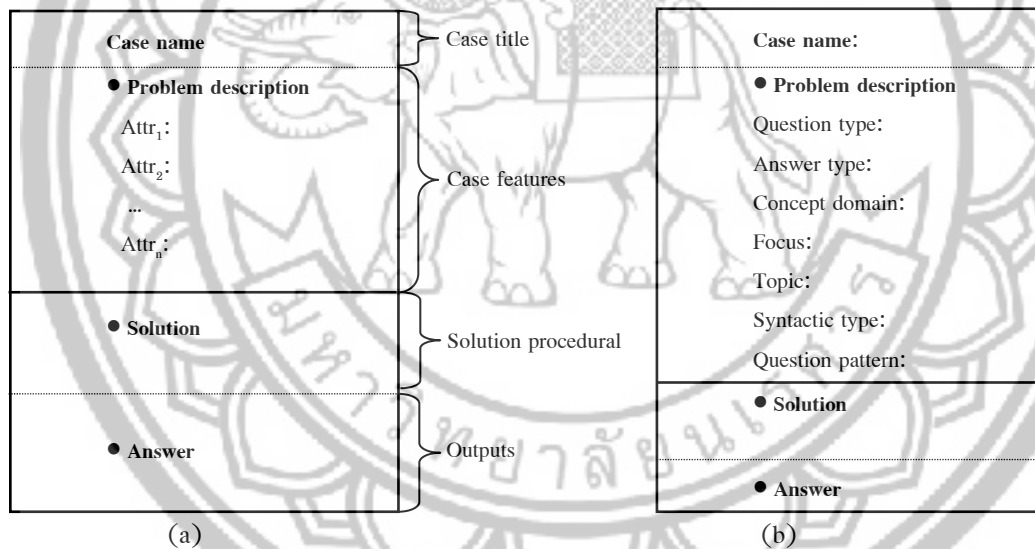


Figure 4 a) Structure of case b) Components of case formation representation

that is known as the solution procedural. However, in our research does not present the outputs section, so that we can show the question problem forms as followings:

From the examples of the question problem

p1. what are the advantages of OIDs?

p2. what are the disadvantages of OIDs?

p5. In which situations would OIDs be a good way of representing database entities?

We can present forms of the problems related to components which are the parts of “Problem description” shown in Figure5.

Question problem :1 Question type :Internalizing values Answer type :Database application Concept domain: Database application , DBMS,ODBMS, OO concept type Focus : Advantage Topic : OIDs Syntactic type : What Question pattern : P1	Question problem: 2 Question type : Internalizing values Answer type : Database application Concept domain: Database application , DBMS,ODBMS, OO concept type Focus : Disadvantage Topic : OIDs Syntactic type : What Question pattern : P1	Question problem:5 Question type : Internalizing values Answer type : Database application Concept domain: Database application ,DBMS,ODBMS, OO concept type Focus : Good way Topic : OIDs Syntactic type : What Question pattern : P5
--	---	---

Figure 5 Question problem1, question problem2 and question problem 5

5. The processing of the matching similar cases

5.1 Weighting features

It is one part of used case matching processing in the problem description to be defined as the weighting features of the case features. However, these values

are based on the main purposes of what the components of the question problem are, therefore in this research defines the level classes' values of the weighing features and explain as the followings:

Feature description	Weighting feature
Question type	4
Answer type	4
Concept domain	2
Focus	10
Topic	9
Syntactic type	1
Question phrase	1

From the relationship between the question problems in five groups, the related levels of the question problems can be presented through the Focus. From the examples of structural components of the question problems, they can be defined as the Focus values and the Topic as 10 and 9 respectively and to be the primary keys for the retrieve matching cases. Hence, this is the relationship level of the Focus feature shown in Table1.

Table1 shows the relationship of the Focus values that is at the same point of the Topic of the question problem. This research shows the examples of

presentation of the Focus based on the key properties that have exactly similar meaning, the most similar meaning, badness in the similar meaning, some partial similar meaning, and some fewer partial similar meaning respectively. The similar measure of the advantage features to the others of the Focus values that are the most similar benefit features in Good, Easier, Harder, and Disadvantages respectively which we may use all features between the similar question problem measurement by using the feature values.

**Table1** Similar case features level of the focus section

Focus	Advantage	Disadvantage	Easier	Harder	Good way	Benefit
Advantage	1.00	0.00	0.69	0.19	0.70	1.00
Disadvantage	0.00	1.00	0.31	0.82	0.30	0.00
Easier	0.69	0.31	1.00	0.00	0.77	0.69
Harder	0.19	0.82	0.00	1.00	0.24	0.19
Good way	0.70	0.30	0.77	0.24	1.00	0.70
Benefit	1.00	0.00	0.69	0.19	0.70	1.00

5.2 Similar measurements between question problems

How to use equation through similar measurements between the old question problem and new question problem is as the followings:

$$\text{Sim}(A_{\text{new question problem}}, B_{\text{old question problem}}) = \frac{\sum_{i=1}^n W_i * d(A_i, B_i)}{\sum_{i=1}^n W_i}$$

where:

d is a similar function of attribute i in cases A and B

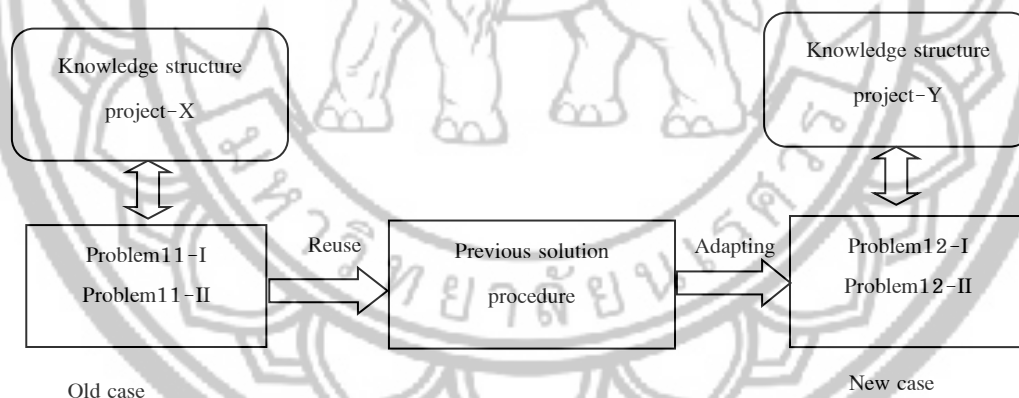
A is the target case

B is the source case

W is the importance weighting feature i

n is number of attribute in each question problem

i is an individual attribute from 1 to n

**Figure 6** The case reuse processing of the old question problem11 and new question problem12

5.3 Case reuse process to find the answers of the question problem solving

The case reuse processing to the answer of the question problem solving is based on the solution procedural processing. The study found that the question problem is the most similar between the old question problem and new question problem, and we can use this to answer new question problems in this

research. From the case representation, it is shown that there are two sections: problem description and solution. This solution section can be used to apply the solution procedural processing of the old question problem that is the most similar to the answers of the new question problem. This is one example of the solution procedural processing of the old question problem solving to new question problem shown in



Figure6. In this section, the knowledge structure of the question problem is presented in form of a data table. In question problem 11 shown in Table2, there is three attributes that are activity, duration, and dependency respectively.

Problem11. This question is about project management planning and scheduling.

1. The table below sets out a number of activities, duration and dependencies for a short software development project.

1.1 Draw an activity chart for this project.

1.2 Which sequence of tasks forms the critical path? What is the shortest time this project can be completed in?

Table 2 Characteristics of question problems

Activity	Duration (days)	Dependencies
T1	15	-
T2	10	-
T3	10	T1,T2
T4	8	T1,T2
T5	20	T3,T4
T6	10	T3,T4
T7	15	T6
T8	10	T7

Question problem:	Problem11-I	Question problem:	Problem11-II
Question type:	Analysis	Question type:	Valuing
Answer type:	Software engineering	Answer type:	Software engineering
Concept domain:	Software engineering Project management, Project plan, Project schedule, Planning	Concept domain:	Software engineering Project management, Project plan, Project schedule, Planning, Activity
Focus1:	Draw chart	Focus1:	Sequence of activities forms
Focus2:	-	Focus2:	Critical path
Topic:	Activity	Topic:	Activity
Syntactic type:	How	Syntactic type:	Which
Question pattern:	P11	Question pattern:	P12

We can classify into two subquestion problems that shows the continual relationship in question 11. From this example, the question problem components are presented shown in Figure7. The example of question problem11 solves its processing. The design of the question problem solves the answer of new question problems. It could be assumed that the defining of the question problem11 is the same as an old question problem stored in case based. The parts of the solution in the case answer the question “can we design the problem solving approaches and able to use as an example explanation”.

Figure 7 Problem description11-I and problem description11-II



It needs to note that the Focus is the specific properties of the Topic of question problems. The definition of the Focus is a drawing chart that a property of an activity Topic uses Focus to answer the steps of processing. Draw charts and activities are defined to the Focus and Topic respectively. All tasks are in a block diagram and all link together. For all of task blocks link, it is the beginning of task value inner attribute of non dependency value, and remains the link of all tasks. The Topic of the question problem11-II is the critical path values that has the shortest and sequence of task forms of Focus for question problems. The question problem processing is the beginning to define an activity of task and uses the process to the chart creating from question problem11-I. As a result it is able to compute the path of tasks just selects the longest path and a minimum of path values and compare other paths. Then, it produces the minimum time of criticizing the path values and this question problem processing and examples can be used to explain all of question problems as followings:

Solution Procedural-Problem11-I

Draw activity chart

Set activities tasks

List all tasks $T_{i;i=1,...,n}$

Create task chart

For T_i

First task where T_i non dependency

Link all tasks T_i, T_j

where T_i dependency on T_j

Solution Procedural-Problem11-II

Critical Path

Draw activity chart

Find longest path and minimum number of days

Where list all paths $P_{k=1,...,m}$

Where sum all days in tasks path P_k

Results

1. Comparing results of the question problem examples

To find the similar values between the old question problems and new question problems, the equation computation to the case features of the problem description section is employed. It is used by the weight values of each feature values. From the question problem 1 to 10 according to the example, the relationships between the question problems can be computed similar shown in Table3. From the Figure8 below shows the question problem comparison of case1 to case2, case3 to case10. The results between case1 and case3 are 1 that are the most similar case which have the similarity to case5, case4, case9, case6, case2, case8, case7 and case10 are 0.759, 0.756, 0.741, 0.706, 0.677, 0.585, 0.545 and 0.53 respectively.

Table 3 Similar question problem

	Case1	Case2	Case3	Case4	Case5	Case6	Case7	Case8	Case9	Case10
Case1	1	0.677	1	0.756	0.759	0.706	0.545	0.585	0.741	0.53
Case2	0.677	1	0.677	0.756	0.63	0.583	0.748	0.646	0.612	0.505
Case3	1	0.677	1	0.756	0.759	0.706	0.545	0.585	0.741	0.53
Case4	0.756	0.756	0.756	1	0.647	0.609	0.612	0.578	0.634	0.512
Case5	0.759	0.63	0.759	0.647	1	0.732	0.561	0.621	0.838	0.650
Case6	0.706	0.583	0.706	0.609	0.732	1	0.677	0.652	0.732	0.545
Case7	0.545	0.748	0.545	0.612	0.561	0.677	1	0.652	0.561	0.456
Case8	0.585	0.646	0.585	0.578	0.621	0.652	0.652	1	0.636	0.503
Case9	0.741	0.612	0.741	0.634	0.838	0.732	0.561	0.636	1	0.641
Case10	0.53	0.505	0.53	0.512	0.605	0.545	0.456	0.503	0.614	1

2. The solution procedure algorithm processing of old solution problem solving to the new question problems

The old question problem solving process can be used to answer new question problem, so that this

research defines an example of question problem12 as the new question problem. This case will be shown in Figure 8. For the question problem12 shown in Table 4, there is four attributes

Question problem:	Problem12-I	Question problem:	Problem12-II
Question type:	Analysis	Question type:	Analysis
Answer type:	Software engineering	Answer type:	Software engineering
Concept domain:	Software engineering Project management, Project plan, Project schedule, Planning	Concept domain:	Software engineering Project management, Project plan, Project schedule, Planning, Activity
Focus1:	Draw chart	Focus1:	Sequence of activities forms
Focus2:	Bar chart	Focus2:	Minimum number of weeks
Topic:	Activity	Topic:	Critical path
Syntactic type:	How	Syntactic type:	Which
Question pattern:	P13	Question pattern:	P14

Figure 8 Problem description12-I and problem description12-II

**Table 4** Characteristics of question problems

Activity	Description	Duration (weeks)	Dependencies
T1	Feasibility	4	-
T2	Requirements specification	2	-
T3	User interface prototype	1	-
T4	Outline design	2	T1,T2,T3
T5	Detailed design	4	T4
T6	Interface to stock database	4	T5
T7	Implementation of sales DB	2	T5
T8	Implementation of UI	4	T7
T9	Testing of system	2	T6,T8
T10	Development to customers	4	T9

Problem 12. This question is about project management planning and scheduling.

1. The table below sets out a number of activities, duration and dependencies for a short software development project.

1.1 Draw an activity chart and a bar chart for the project.

1.2 Which sequence of activities forms the critical path for the project, and what is the minimum number of weeks that the project can be completed in?

In this section the assumption is that question problem12 is the most similar to question problem11 and it is defined as an old question problem, too. Consequently, the old question problem solving process is for the new question problem12-I. The Topic and Focus of question problem are in some parts between the old and new question problems and the researcher can use the same solving processing. For this answer to question problem12-I, For this approach that use the draw chart before and problem solving addition for the bar chart which the first processing is defined as an activity of the task and the bar chart creating of each task. The task is created in two parts that are the task non dependency from the other tasks which will be created only bar chart of this task and none

linked to values for another task. The second one is the creating task linkage by dependency to other task. The critical path, sequence of activities forms and minimum number of weeks are the Topic and Focus of the question problem12-II respectively. The first is that an activity is used as a task of the question problem, and after that an idea of activity chart creating as before. The longest path can be defined as the critical path and try to find the minimally path values.

Solution Procedural-Problem 12-I

Draw activity chart

Set activity are tasks

List all tasks $T_{i:i=1,...,n}$

Create task chart

For T_i

First task where T_i non dependency

Link all tasks T_i, T_j

where T_i dependency on T_j

**Draw bar chart**

Set activity are tasks

List all tasks $T_i; i=1, \dots, n$

Create task bar chart

For T_i

Non link bar chart with duration

where T_i non dependency on T_j

Link bar chart with durations

where T_i dependency on T_j

Solution Procedural-Problem 12-II**Critical Path**

Draw activity chart

Find longest path and minimum number of weeks

Where list all paths $P_{k=1, \dots, m}$

Where sum all weeks in tasks path P_k

Conclusion

This research is about how to find the answers to the questions by using the case based reasoning, the similarity between old question problems and new question problems and algorithms of the old problem solution procedure which will be applied to the most new question problem similarity. Question examples of the topic can be defined in various differentiations of the focus. The obstacle is how to define the key structure valued level of Focus and Focus concept will be classified in the term of ontology that has different values in difficult natural language forms. Furthermore, the different values in case feature in some parts of problem description are difficult to define. We must use the heuristic experience and old knowledge to solve the problems. In conclusion, this research presents the ways to use the old solution procedural to new question problem. It is necessary to improve for better old question problem retrieving. It has been found that searching old question problems are similar in the high level and

has fewer obstacles. This brings to the difficulties to take the old question problem solution procedural to solve the new question problems.

References

- Angele, J., Moench, E., Oppermann, H., Staab, S., & Wenke, D. (2003). Ontology-based Query and Answering in Chemistry: OntoNova @ Project HaloIn. In *Proceedings of the 2nd International Semantic Web Conference (ISWC2003)* (pp. 913-928). Sanibel Island, FL, USA.: Springer.
- Bloom, B.S. (1956). Taxonomy of Educational objectives. *Handbook I: The Cognitive Domain*. New York: David McKay.
- Goel, A. (1996). Adaptive Modeling. In *International Workshop on Qualitative Reasoning* (pp. 67-73). Monterrey: AAAI
- Kolodner, J. (1993). *Case Based Reasoning*. San Mateo, CA: Morgan Kaufmann.
- Larranaga, P., & Bielza, C. (2013). Bayesian networks to answer challenging neuroscience questions, *The 26th International Symposium on Computer-Based Medical System (CBMS)*. N.P.: IEEE Conference publication.
- Lavenus, K., Grivolla, J., Gillard, L., & Gillard, P. Bellot. (2004). Question-answer matching: two complementary methods. In *Proceedings of RIAO 2004* (pp. 623-637). France: University of Avignon (Vaucluse).
- Lehnert, W. G. (1981). A computational theory of human question answering. *Elements of Discourse Understanding*, N.P.: Cambridge University.



- Li, X., & Roth, D. (2002). Learning question classifiers. In *COLING 2002, The 19th International Conference on Computational Linguistics* (pp. 1-7). USA: Association for Computational Linguistics Stroudsburg.
- Mann, G. S. (2002). Fine-Grained Proper Noun Ontologies for Question Answering. In *Proceedings of SemaNet'02: Building and Using Semantic Networks* (pp. 1-7). Taiwan: n.p.
- Mantaras, R. L. D., McSherry, D., Bridge, D., Leake, D., Smith, B., Craw, S., et al. (2006). Retrieval reuse revision and retention in case-based reasoning. *Knowledge Engineering Review*, 20(3), 215-240.
- Peitao, C., Ke, Z., Yatao, L., & Wei X. (2011). Application of Case Based Reasoning in Plane Geometry Intelligent Tutoring System, In *International Conference on Electrical and Control Engineering (ICECE)* (pp. 4369-4373). N.P.: IEEE Conference publication.
- Ralph, B., & Armin, S. (1998). Similarity Measures for Object-Oriented Case Representations. In *Proceedings of the 4th European Workshop on Case-Based Reasoning, EWCBR'98*. (pp.25-36). Dublin, Ireland.: n.p.
- Ramprasath, M., & Hariharan, S. (2012). Using Ontology for Measuring Semantic Similarity for Question Answering System. In *International Conference on Advanced Communication Control and Computing Technologies (ICACCCT)*. (pp. 218-223). N.P.: IEEE Conference publication.
- Spalazzi, L. (2001). A survey on case based planning. *Artificial Intelligence Review*, 16(1), 3-36.
- Schreiber, G., Wielinga, B., & Breuker, J. (1993). *Approach to Knowledge-Based System Development. Volume 11*. N.P.: Academic Press Inc.
- Wilke, W., & Bergmann, R. (1998). Techniques and Knowledge Used for Adaptation During Case Based Problem Solving. In *Lecture Notes in Artificial Intelligence, Vol. 1416*. (pp. 497-505). New York: Springer-Verlag, Berlin Heidelberg.