



الجامعة الوطنية  
UNIVERSITI  
TEKNOLOGI  
MARA

# COMPASS

Compilation of Pahang Scholars' Synergy

---

Year 2011

ISSN 1985-9937

---



# **TQUEST: Determining Cognitive Level of Question Using Information Extraction (IE)**

*Muhd Eizan Shafiq Abd Aziz  
Mohamad Affendi Abdul Malek*

## **ABSTRACT**

*Each question in examination question set consists of keywords from Bloom's taxonomy. This keyword determines the cognitive level of the question. Students must be evaluated according to six cognitive levels of Bloom's taxonomy. Preparing examination questions and determines the cognitive levels take huge amount of time. A computer application is required to ease this task as it becomes critical since Outcome-Based Education (OBE) enters the education world where it requires students to be evaluated in many aspects according to related fields. Raw questions are the main input of this computer application. Using an appropriate technique, this application will extract keywords from question and determine the cognitive level by itself. Previously, these tasks are done manually by responsible committee and it takes time to complete all the questions. This computer application helps the committee until the final part of preparing question set which is preparing camera ready of question sets where it can be printed out as well.*

**Keywords:** *Bloom, cognitive, information extraction, OBE, application*

## **Introduction**

Creating examination questions is not a simple task. Several standard procedures have to be followed starting from collecting raw questions until printing out the draft of question sets. The introduction of Outcome Based Education (OBE) makes the processes of setting question sets have to be carried out carefully. In order to deliver good and understandable examination question, Bloom's taxonomy is used to design the examination question. Even though Bloom's taxonomy has already been used in Universiti Teknologi MARA (UiTM) in designing examination question, several research questions arose regarding level of difficulties of each question in the question sets. Does the question use appropriate keywords in delivering the message to student? What is the level of difficulty of each question? How many questions available according to

six cognitive levels in Bloom's taxonomy?

The research aims to propose an information system that is able to assist in creating the examination question sets. Several standard procedures are remained as usual, however, only certain procedures are going to be changed since the introduction of the information system. The purpose of the information system is to identify cognitive level of questions which previously done manually by people. It is believe to reduce amount of time used in identifying and determining the cognitive level and time required in creating the question sets.

### **Cognitive Domains in Bloom's Taxonomy**

Bloom (1956) created this taxonomy and since that it contributes a lot especially in academic field. Three main domains which are Affective domain, Cognitive domain and Psychomotor domain are identified and each domain has its own focus. This research focus on Cognitive domain which primary focus of this domain is mental skills required students' ability to recall specific facts, patterns and concepts. Bloom (1956) identified six cognitive levels in Cognitive domain as depicted in Table 1.

Table 1: Six levels in Cognitive domain

<b>Level</b>	<b>Skills Demonstrated</b>
Knowledge	Able to recall data or information of previous learnt
Comprehension	Able to describe, tell, elaborate using their own words based on scenario or material given.
Application	Able to apply what they have learnt in class in any situations.
Analysis	Able to distinguish or differentiate between facts and inferences.
Synthesis	Able to form or create a new meaning or structure from different elements.
Evaluation	Able to judge or argue about ideas.

Bloom (1956) put Knowledge level as the most basic level and this is followed by Comprehension, Application, Analysis, Synthesis and finally Evaluation. Each level demonstrates different skills and it has keywords that must be used in the examination question. This research is carried out for diploma level students which it focuses on giving basic skills for semi-professional work with firm foundation and it must be useful if students

decide to further study to the higher level (M Mustafa M Ghazali, 2008). Bloom (1956) provides keywords for each cognitive level in Cognitive domain. Appropriate keywords must be used during question construction and keywords represent level of difficulty of the question. This is to ensure students are able to understand the requirements of the question and students are tested with different kind of difficulty levels. This is also to ensure there are many difficulty levels in a set of examination questions. Table 2 depicts keywords available in those six levels.

Table 2: Verbs or keywords in six levels of Cognitive domain

<b>Cognitive Level</b>	<b>Verbs/Keywords</b>
Knowledge	Who, what, why, when, omit, where, which, choose, find, how, define, label, show, spell, list, match, name, select, recall, tell, name, relate
Compre- hension	Compare, contrast, demonstrate, interpret, explain, extend, illustrate, infer, outline, relate, rephrase, translate, summarize, show, classify
Application	Apply, build, choose, construct, develop, interview, make use of, organize, experiment with, plan, select, solve, utilize, model, identify
Analysis	Analyze, categorize, classify, compare, contrast, discover, dissect, divide, examine, inspect, simplify, survey, take part in, test for, distinguish, list, distinction, theme, relationships, function, motive, inference, assumption, conclusion
Synthesis	Build, choose, combine, compile, compose, construct, create, design, develop, estimate, formulate, imagine, invent, make up, originate, plan, predict, propose, solve, solution, suppose, discuss, modify, change, original, improve, adapt, minimize, maximize, delete, theorize, elaborate, test, happen
Evaluation	Award, choose, conclude, criticize, decide, defend, determine, dispute, evaluate, judge, justify, measure, compare, mark, rate, recommend, rule on, select, agree, interpret, explain, appraise, prioritize, opinion, support, importance, criteria, prove, disprove, assess, influence, perceive, value, estimate, deduct

These keywords have to be in place together in the examination question. That makes each question has different cognitive level. This is to ensure in each question set, it contains variety of cognitive level questions.

Diploma students should be tested at all cognitive levels but too many questions at Knowledge level will not fulfill the requirements of OBE of the course. Table 3 shows examples of questions using cognitive levels:

Table 3: Example of examination questions containing keywords from six cognitive levels

Cognitive Level	Verbs/Keywords
Knowledge	<b>List</b> five components of information cycle
Compre- hension	<b>Explain</b> briefly all five components of information cycle
Application	<b>Identify</b> five types of information system and its uses
Analysis	<b>Compare</b> and <b>give</b> example of storage media and stor- age device
Synthesis	<b>Construct</b> an Entity Relationship Diagram (ERD) ac- cording to scenario given
Evaluation	According to scenario given, <b>recommend</b> a solution to solve the problem face by the company

Examination questions are designed by lecturers appointed by faculty. Current method used is lecturers have to design the questions in softcopy format and submit the raw questions to committee using storage media such as compact disc. Committee will review the questions and determine which cognitive level of each question is. This is what we call as raw questions coming from lecturers. This research will change this procedure where lecturers no longer need to submit the questions in a compact disc. An information system runs on intranet of UiTM will be developed and it can be accessed by lecturers of UiTM. Appointed lecturers will design the raw questions and save those questions in this information system according to duration given. System will automatically identify and determine cognitive level of each question by extracting keywords available in the question. If currently this process is carried out manually, the introduction of this information system will reduce or eliminate human interaction towards this process. Questions are stored in the DBMS as structured data and it can be retrieved anytime and anywhere by committee.

# Structured, Semi-Structured and Unstructured Data

Raw questions are categorized as structured data as it is going to be stored in the DBMS. Bergman (2005) described Wood's statement said that these three categories of data can be found nowadays as depicted in Figure 1 below. Further explanation about these three categories of data can be found in the next section.



Source: <http://www.mkbergman.com/>

Figure 2: Sources of data

## Structured Data

Dorion (2007) explained that data is organized in a structured way where usually always being displayed in a tabular format. It makes the data easily identifiable by user. DBMS is used to keep data in a computer (Wang, 2010). It is organized in semantic chunks (entities) and similar entities are grouped together as classes or relations. Similar entities placed in a same group that has similar descriptions, in other words we call them attributes. It has schema that represent descriptions of all entities in a group such as format, length, must exist, and follow the same order. To retrieve information from the source, it allows users to select according to columns and rows in a field.

## Semi-structured Data

Semi-structured data has some kind of structure but it still cannot be

categorized as structured data due to certain differences. Wood explained, semi-structured data organized in semantic entities and similar entities are grouped together. However, the entities may have different attributes, not all attributes are required and order is not necessary. Wood added that size and type of similar attributes in a group may differ. Semi-structured data can be found in flat records, XML files, RSS feeds and a few others as described by Wang (2010). Semi-structured data comes from DBMS, hence, it can be searched and organized neatly regardless to any output format since it does not has any strict formatting rule (Loshin, 2005; Wang, 2010).

## **Unstructured Data**

Wang (2010) said more than 80% of data we have seen today are unstructured data. Sources of this kind of data include email, audio, video, images and many more. It can be of any type and do not have specific format or sequence as explained by Wood.

As for this research, it is required to apply information technology to ensure its success. Lecturers come from different campuses, separated by different geographical locations. By putting this system online as an intranet application, lecturers are able to contribute in supplying the raw questions into the question bank server. This is where the raw questions become as structured data because DBMS is going to be used to keep the raw questions safely (Wang, 2010). Next section is about applying information extraction (IE) to extract cognitive keywords from the sentence.

## **Extracting Cognitive Keywords from Structured Data Using Information Extraction (IE) Technique**

There are various techniques in retrieving information from sources either documents or databases. One of the techniques is Information Extraction (IE) which is very useful for this research. According to Grisham (1997) and Eikvil (1999), IE as the identification and extraction of instances of a particular class of events or relationships in a natural language text and their transformation into a structured representation such a database. The actual process in preparing the questions begins where lecturers design the raw questions and save the questions in a softcopy format using common file format, for example in Microsoft Word file and put the questions in a compact disc. Instead of saving the questions in a compact disc, once this information system gets online, those raw questions can be saved straight to the question bank server. Once the raw questions are

stored in the database, it becomes structured data and using IE technique it helps committee in retrieving and determining cognitive level of question (Gaizauskas, 2002).

Figure 3 shows the new approach and it shows the aims of this research in re-engineering the current approach where certain processes are done manually before, now will be incorporated with information technology. A graphical user interface (GUI) is going to be developed to allow appointed lecturers enter the raw questions including the answers. The committee will be able to view all raw questions in the database, however, at this moment, the committee will not do anything yet. Once the raw questions are fully submitted to the question bank server, then the committee will start their tasks in reviewing all raw questions including making any amendment if necessary.

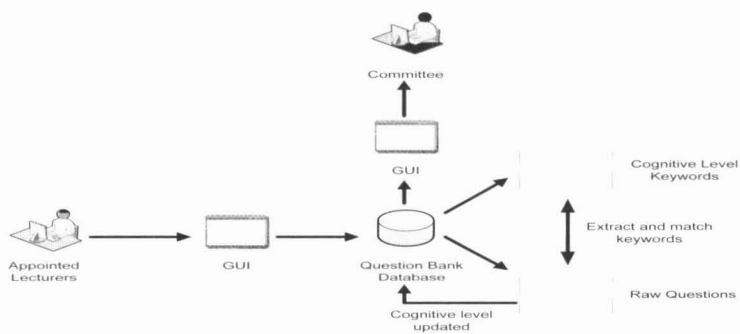


Figure 3: New approach of preparing examination questions

An algorithm is going to be designed in order to extract keyword(s) from raw questions. It then will be compared with existing cognitive keywords stored in the database. The system will update the cognitive level of each question automatically and store it in the database. However, this process can be done once the committee completely reviews all raw questions. During the reviewing process, there may be some amendment to be done to the raw questions, for example sentences re-phrasing, grammatical checking and a few others. To avoid repeating processes, it is recommended to execute the process once all raw questions have been completely reviewed through some action triggered by the committee.

Figure 4 below shows how exactly the process in determining the cognitive level will be executed. As usual, each raw question must have



at least one keyword. If there are more than one keyword exist in one raw question, system will extract those keywords in the first place then determine which keyword has higher cognitive level. If the keyword does not exist then system will assign 0 (zero) value as a cognitive level to that particular raw question. System will notify the committee of the raw questions with 0 (zero) cognitive level. Therefore the committee will be able to review the questions where there might be amendment to be done on the raw question before it can be used as examination question.

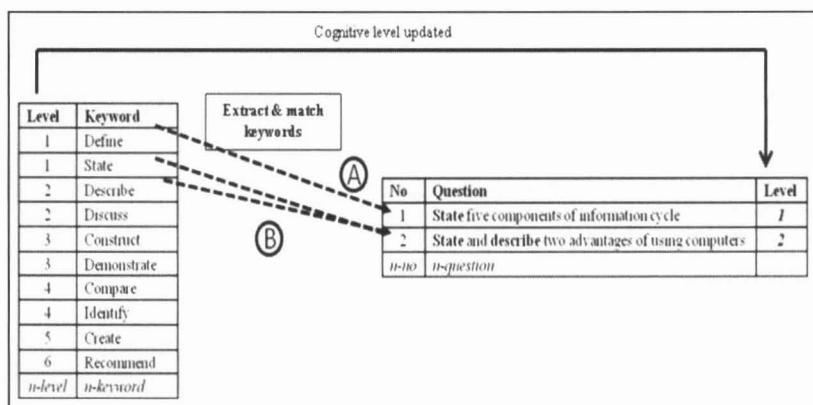


Figure 4: Determining cognitive level of raw question

Example A shows one keyword is found in the raw question. It then will be compared with existing keywords on the left table and cognitive level of the question will be updated. Example B shows another possibilities where two keywords are found in the sentence. “State” is a Level 1 keyword while “describe” is a Level 2 keyword. Since there are two levels can be assigned to the same question, system will choose highest level for raw question which is Level 2. By incorporating IT to simplify this task, it will reduce the amount of time used by the committee to determine the cognitive level of raw questions.

Each raw question is categorized according to chapters in a text book. It is important to include this feature to help the committee while preparing the question sets. Researchers realize the committee will pick any questions randomly from all chapters which normally start with choosing the chapter and displaying all questions created under the selected chapter. Then, the committee will chooses which questions to be put in the question sets. Once everything is ready, draft of final examination question sets are ready to be printed out and reviewed by faculty.

## Conclusion

Planning and designing examination questions are critical activities in evaluating student performance. It should fulfill Outcome-Based Education (OBE) standards which require students to be evaluated in all aspects. Here come the six cognitive levels of Bloom's taxonomy that will play a key role in performing student. Information Extraction (IE) technique is used to determine cognitive level of each question based on keywords exist in question. Therefore, this computer application will help responsible committee in performing their tasks especially reducing the amount of time spent to determine the cognitive level of questions. Further research will be conducted on several techniques in order to find better technique that can be applied on this computer application.

## References

- Bergman, M.K, (2005), Semi-structured Data: Happy 10<sup>th</sup> Birthday. See <http://www.mkbergman.com/>
- Bloom, B.S. (Ed.). (1956). Taxonomy of educational objectives: The classification of educational goals. Handbook 1. Cognitive domain. New York: McKay
- Bloom, B.S., Engelhart, M.D., Furst, E.D., Hill, W.H., Krathwohl, D.R. (1956). Taxonomy of Educational. New York: Longman Group Limited.
- Chowdhury, G.G., (1999), Template Mining for Information Extraction from Digital Documents.
- Doan, A., Ramakrishnan, R., Vaithyanathan, S., (2006), Managing Information Extraction. SIGMOD 2006.
- Eikvil, L., (1999), Information extraction from World Wide Web: A survey. Technical Report 945, Norwegian Computing Center
- Feldman ., (1998) Knowledge Management: A Text Mining Approach. In Proceedings of 2nd International Conference on Practical Aspects of Knowledge Management (PAKM98).
- Feldman, R., Regev, Y., Gorodetsky, M., (2008), A Modular Information

- Extraction. (Journal Intelligent Data Analysis.
- Gaizauskas, R., (2002), An Information Extraction Perspective on Text Mining: Tasks, Technologies and Prototype Applications. Euromap Text Mining Seminar. Sheffield, UK.
- Gottlob, G, Koch., (2004), Logic-based Web Information Extraction.
- Grishman, R., (1997), Information Extraction: Techniques and and Challenges, International Summer School on Information Extraction: A Multidisciplinary Approach to an Emerging Information Technology, p.10-27, January 1997
- He, Lei. (2006). A Novel Web-Based Educational Assessment System With Bloom's Taxonomy. FORMATEX 2006.
- Isbell, J., Butler, M.H. (2007). Extracting and Re-using Structured Data From Wikis. Hewlett-Packard Development Company
- Jérôme Azé, Mathieu Roche, Yves Kodratoff, Michèle Sebag., (2004), Learning Interestingness Measures in Terminology Extraction. Proceedings Of World Academy Of Science, Engineering And Technology Volume 1.
- Jones, K.O., Harland, J., Reid, J.M.V., Barlett, R. (1999). Relationship Between Examination Questions and Bloom's Taxonomy. 39th ASEE/IEEE Frontiers in Education Conference 1999.
- Li, X., Wang, Y.Y., Acero, A. (2009). Extracting Structured Information from User Queries With Semi -Supervised Conditional Random Fields. SIGIR 2009.
- M Mustafa M Ghazali., (2008), Pendidikan Tinggi. UiTM.
- M. Hearst., (1999), Untangling text data mining. In Proceedings of ACL'99: The 37th Annual Meeting of the Association for Computational Linguistics.
- Mohtashami, M., Scher, J.M., (2000). Application of Bloom's Cognitive Domain Taxonomy to Database Design. ISECON2000
- Mooney, J.M, Bunesco, R. (2005). Mining Knowledge from Text Using Information Extraction. ACM SIGKDD Explorations

- Nassif, H., Woods, R., Burnside, E., Ayvaci, M., Shavlik, J., Page, D. (2009), Information Extraction for Clinical Data Mining: A Mamography Case Study. 2009 IEEE International Conference on Data Mining Workshops
- Rohana (2009). Bengkel Penggubalan Soalan dan Skema Jawapan. iLQaM, UiTM.
- Sarawagi, S. (2008). Information Extraction. Journal Foundations and Trends in Databases.
- SH Tanalol., (2008), Mining Exam Question based on Bloom's Taxonomy. In Proceedings of KMICe 2008: Knowledge Management International Conference.
- Un Yong Nahm, Raymond J. Mooney., (2004), Text mining with information Extraction.
- Williams, D., (2002), Database Driven Discovery of Structure from Partially Structured Data.
- Williams, D., Poulouvassilis, A., (2005), Combining Data Integration and Information Techniques. NLDB 2008 Proceedings
- Witten, I.H., (2005), Text mining in Practical handbook of internet computing, edited by M.P. Singh, pp. 14-1 - 14-22. Chapman & Hall/ CRC Press, Boca Raton, Florida.
- Wood, P. School of Computer Science and Information Systems, Birbeck College, University of London. See <http://www.dcs.bbk.ac.uk/~ptw/teaching/ssd/toc.html>

---

MUHD EIZAN SHAFIQ ABD AZIZ, Fakulti Teknologi Maklumat dan Sains Kuantitatif, UiTM Pahang.eizan@pahang.uitm.edu.my