Evaluating Multiple Choice Question Generator

Conference Paper · July 2011

DOI: 10.1007/978-3-642-32826-8_29

3 authors, including:

Ching-Chieh Kiu
Taylor's University

Dickson Lukose
GCS Agile Pty. Ltd.

Evaluating Multiple Choice Question Generator

Sieow-Yeek Tan, Ching-Chieh Kiu, Dickson Lukose

Artificial Intelligence Center, Mimos Berhad,
Technology Park Malaysia, Bukit Jalil, Kuala Lumpur, Malaysia

{tan.sy, cc.kiu, dickson.lukose}@mimos.my

Abstract. Semantic-based computer-assisted automated question generator has been increasingly popular as a tool for creating personalized assessment questions. Various question generator tools have been proposed, such as those which generate structured question from a text file, generate Multiple Choice Question (MCQ) from a text file or from ontology-based knowledge representation. A comparison framework and evaluation methodology is required to evaluate different question generator tools. This paper discusses the requirement and criteria to carry out performance comparison of different MCQ generator. A feature comparison of Question Generation (QG) tool, namely Mimos-QG[1] with the existing QG tools is presented. We have evaluated our QG tool based on several standard criteria such as the correctness of (a) distractor generation, (b) answer choice grouping strategy and (c) syntactical and pedagogical quality on three different domain ontologies. The experimental result indicated that Mimos-QG is capable of producing good quality direct type and grouping type multiple choice questions.

Keywords: Question Generator, Multiple Choice Question Generator, Ontology, Semantic Web, Education Technology

1 Introduction

In the last decade, various computer-assisted educational tools have been introduced for different teaching domains. Question-answering (Q&A) is an important method to evaluate students' understanding and performance in either electronic (online) or traditional (offline) learning process. Studies in [1], [2] and [3] show that students have substantially benefited from automated question-answering approach offered via e-learning. Studies [4] and [5] have also shown that high number of questions from the same subject drives an in-depth understanding of the topic and subject. Multiple choice question (MCQ) is one popular mean of self-assessment in learning evaluation. MCQ comprises of a short text describing the question, a number of alternative choices where one of the choices is the correct answer, and the others are wrong answers (a.k.a distractors [6]).

[1] The authors developed this system under the Semantic Technology Research & Development initiative funded by the Ministry of Science, Technology & Innovation (MOSTI), Malaysia.

The use of ontology-based knowledge representation in semantic-based question generation is becoming common nowadays. Domain ontologies can be considered as a proper formalism which serves the basis of automatic assessment. It contains the domain knowledge in the form of definitions of terms (concepts), individuals (instances) belong to these terms and the relationship between these terms and individuals. The formalism is required to conform to Semantic Web Technology standard in Web Ontology Language (OWL) representation [9]. The trend from the literature survey showing work on ontologies in e-learning domains has been primarily focused on ontological formalization. Studies in [7, 8] have proposed methods in constructing the learning objects, instructional processes and learning design by using ontology formalism. Hence, an ontology-driven QG is needed to be researched and developed to utilize the benefit of ontology in the domain subject in the e-learning system [11].

By leveraging on the pre-defined OWL standard, automatic generation of questionnaires based on existing ontological descriptions have been used to formulate multiple choice questions from MCQ generator. Papasalouros et. al. [6] proposed various strategies to automatically generate MCQ from domain ontologies. It takes OWL ontology as inputs and generates multiple choice questions. The MCQs are generated based on the basic meta-ontology relations and 'binary role' between class and instances. Their strategies were further implemented as a plugin for Protégé ontology editor by Tosic and Cubric [10]. Cubric et. al. [11] enhanced the abovementioned strategies by generating assessment for different knowledge domains. They proposed a proprietary MCQ ontology and used it as a pre-defined question template. Semantic interpretation is then applied to perform mapping between an arbitrary domain and the pre-defined MCQ ontology, in order to generate the MCQ. Bloom's taxonomy standard is used for supporting the MCQ generation process; by grouping the questions into several categories which required different comprehension levels to answer.

Inevitably, good quality question is the key to perform effective assessment over a learning process. Many criteria of evaluation have been used in evaluating MCQ generation strategies. For example, in [12], questions are evaluated by checking whether it is correctly written, meaningful, clear, appropriate to the context and etc. However, most of the evaluation processes are still done manually and are subjective to expert domain. For instance, the ontology used as an input dataset for a QG is normally constructed by the domain experts that are also going to evaluate the questionnaires generated by the QG [6]. Thus, a comparison framework and evaluation methodology is required to evaluate different MCQ generator.

The aim of the paper is to discuss the requirement and criteria to carry out performance comparison for each MCQ generator. We evaluate our QG tool, namely Mimos-QG based on several standard criteria such as distractor generation correctness, answer choice grouping strategy correctness, syntactical correctness and pedagogical quality on three different domain ontologies. In addition, we present a feature comparison of Mimos-QG with existing QGs tools.

The structure of the paper is outlined as follow: Section 2 discusses the requirement for evaluating MCQ QG tool. A literature review and comparison between Mimos-QG and the other MCQ generators are explicated in Section 3.

Section 4 presents the experimental evaluation results of Mimos-QG. Conclusion and further work are concluded in the last section.

2 Requirement of Evaluating MCQ Generator

MCQ question formulation includes generating multiple distractors against a correct answer in a list of choices. While generating MCQ from ontology, the distractors generation can be generally categorized into three techniques, these are: (1) Class-based instance strategy, (2) Property-based strategy and (3) Terminology-based strategy. Ontology consists of concepts or classes, properties and instances. The classes are connected with 'is-a' relationship to denote the super-class and sub-class hierarchy. In class-based instance strategy, distractors are generated based on classes and instances by choosing the proper classes different from those that appear in the correct answer. Meanwhile in property-based strategy, the question components and distractors are generated based on properties which denote relationship between classes in the ontology. Lastly, in terminology-based strategy, the question component and distractors are generated based on concept or sub-concept relationships.

In order to effectively evaluate an ontology independent MCQ generator tool, a set of evaluation requirement and criteria need to be defined. These criteria are listed as below but are not limited to the following:

(1) Input requirements: refer to the content used by QG tool to generate multiple choice questions. The quality of input will definitely influence the quality and type of the questions generated.

(2) Level of user interaction: some QG tools require user intervention in producing a set of MCQ questionnaires. For example, user intervention is required for defining tags or question categories and defining templates for specific tags [12]. Some tools require user intervention in selecting qualify questions to be part of a questionnaire set or in defining question template before generating questions.

(3) Type of output: different type of QG tools will generate different type of multiple choice questions. For example, the MCQ's question and answer pair can be True or False or providing with a list of answer choice for selection purposes. Some output will be in different representation and require further post-processing steps to display them in a proper MCQ format.

(4) Quality of question: pedagogical quality, linguistic or syntactical quality which is normally judged and ranked manually by instructors or domain experts [13].

In order to compare and evaluate ontology independent MCQ generator tool, we need to define and develop the following resources: source of ontologies as the input requirements, benchmarking results and metrics for comparing the performance of the tools.

3 Survey of the Multiple Choice Question Generators

3.1 Related Works

In year 2009, an international patent [14] had been filed for a method and tool for automatically generating questions for a programming language. The proposed method created a set of pre-defined question templates containing compliable codes. Some misinterpretation templates were used to generate options for a question which included various correct and incorrect answers. Questions are automatically generated by changing certain variables' value in the compliable code. The incorrect options are served as distractors which were generated by changing the tolerance limit value defined by the user.

Study in [15] defined an innovative approach for generating MCQs by adapting to a student understanding model. A question is generated by applying term information to a pre-defined template stored in question-template database. The questions are generated using "is-a" and "part-of" attributes, which show the relations between concepts. In addition the proposed methodology introduced a method to generate questions of varying levels of difficulty. More difficult questions are generated when other alternatives are conceptually near to the correct alternative. More importantly, the difficulty of the generated questions is adapting to different students based on their understanding level for a studied subject.

In [11], an ontology driven QG for serving the purpose of automated generation of computer-assisted assignment (CAA) from the Semantic Web was proposed. The method defines a proprietary MCQ ontology to act as the basis for development of question types. Different question type consists of its own specification to produce its template. The question template then is used for MCQ generation by mapping to subject knowledge described in the underlying domain ontology. Also, the idea of categorizing the generated questions at multiple level of comprehension on a subject was introduced. Question that is categorized at higher level of comprehension will require higher level of understanding on a subject to answer. The level of comprehension is represented from one of the most influential models of educational competencies, namely Bloom's Taxonomy [16]. These levels are noted as "Knowledge", "Comprehension", "Application" and "Analysis". The abovementioned method successfully links the generated question to different comprehension level by using annotation-based and semantic measurement strategies when generating the question-answer pairs. For example, distractors from a MCQ which are conceptually having "high-similarity" to the correct answer will require higher level of comprehension to answer it correctly.

3.2 Similarities and Differences in Question Generation Strategy

In general, the approach from [14] described at previous section holds certain similarities to our Mimos-QG. One of them is that it employed a question template technique in generating the correct answer and distractors. However our QG implementation enables dynamic question template selection which is independent from subject domain. Our approach comply with ontology standard in generating question which indirectly allow conceptual intelligent to be taken into consideration during question generation process. Unlike method proposed by [14], in which the questions are generated by just changing certain programming variable values in a set of pre-defined compliable code templates, our QG strategy that leverages on ontology standard composes question by making use of the additional relationship between the concepts described under OWL formalism.

The QG approach from the study in [15] applies question template technique in MCQ generation. As the QG approach also leverages on ontology formalism, the method allows conceptual space consideration in generating the MCQ options. The evaluation of the difficulty level of the generated question has done indirectly by measuring the conceptual space between the correct answer and the distractors. This strategy is similar to Mimos-QG approach. The difference falls in the techniques used for generating MCQ options. In [15], the generation of question options is based on pre-defined contents. The contents are structurally organized and stored in a terms-information database. Meanwhile, our approach generates question options based on a rule-based conceptual relationship. In other words, the generation of correct answer and the distractors is governed by a set of pre-defined rules baseline to ontology formalism.

In [11], an ontology driven QG was proposed which holds certain similarity to Mimos-QG methodology. The author defined a new MCQ ontology as a basis for the development of the MCQ question format specification, which is similar to the question template database used in the Mimos-QG. In addition, the author proposed rules to govern the semantics mapping between the domain knowledge and MCQ to enable conceptual intelligent consideration when generating questions. The differences between the QG tool when compared to Mimos-QG are: the question template defined in MCQ ontology results in a rigid format specification as compared to Mimos-QG approach which has maintained flexibility by enabling dynamic question template selection in MCQ generation. Also, they use annotation-based strategies in generating question-answer pair as well as the distractors instead of rule-based.

3.3 Features Comparison of MCQ Generator Tools

Different QG adopt different strategies and methodologies to automatically generate multiple choice questions. In this section, a comparison of features between the above mentioned QG tools and Mimos-QG are outlined in Table 1. The comparison study is defined based on the knowledge obtained from the reference citations of the tools. Fourteen features have been identified for the comparison as shown in Table 1. If the QG tool contained the specific feature, it will be noted as '*Yes*', otherwise it will be noted as '*No*'.

Table 1. Features Comparison in Multiple Choice Question Generators

Question Generator Features	**Tool-1** [14]	**Tool-2** [15]	**Tool-3** [11]	**Mimos -QG**
1. Able to identify topic level ontology for question generation	No	No	No	Yes
2. Enable conceptual intelligent consideration between correct answer and generated question	No	Yes	Yes	Yes
3. Applying question template technique in MCQ question generation	Yes	Yes	Yes	Yes
4. Enable dynamic question template selection in question generation	No	Yes	No	Yes
5. Enable dynamic question template selection independent from target concept	No	No	No	Yes
6. Generating question-answer pair based on conceptual relationship	No	Yes	Yes	Yes

7. Generating question-answer pair based on rule-based conceptual relationship	No	No	No	Yes
8. Generating distractors in question generation	Yes	Yes	Yes	Yes
9. Generating distractors by considering conceptual space relationship	No	Yes	No	Yes
10. Generating distractors govern by rule-based conceptual relationship	No	No	Yes	Yes
11. Implementing method to evaluate question difficulty level	No	Yes	No	Yes
12. Evaluate question difficulty level by considering conceptual space between correct answer and distractors	No	Yes	No	Yes
13. Evaluate question difficulty level by rule-based conceptual relationship calculation on correct answer and distractors	No	No	No	Yes
14. Categorize difficulty level based on Bloom Taxonomy (BT) level using quantified difficulty value	No	No	Yes	Yes

As summarized from the features comparison table, a qualitative feature comparison between the three QG tools to Mimos-QG is depicted in Fig. 1. The qualitative feature (QF) comparison can be categorized into several aspects, namely: Various types of question generation using template (QF1), Conceptual space consideration in generating correct answer and question (QF2), Conceptual space consideration in generating distractors (QF3), Dynamic question generation rule independent from question template (QF4) and Evaluate question difficulty (QF5).

Qualitative Features	Tool-1	Tool-2	Tool-3	Mimos-QG
QF 1	4	3	4	5
QF 2	4	2	5	5
QF 3	0	5	4	5
QF 4	2	3	3	5
QF 5	0	0	0	5

(Rated 1 to 5) where 1 is the lowest and 5 is the highest

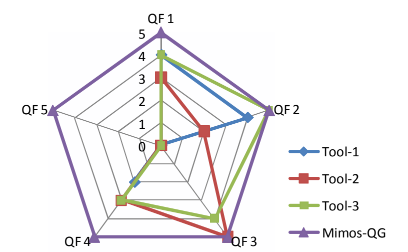

Fig. 1. Qualitative Feature Comparison

4 Mimos-QG Evaluation

Mimos-QG is an ontology independent Multiple Question Generator tool. The class-based instance strategy and terminology-based strategy are adopted to generate MCQ. The class-based strategy is categorized into three approaches known as: direct approach, negative approach and group approach. Fig. 2 shows the direct answer type of MCQ compared to group answer type of MCQ. The algorithm and framework of the Mimos-QG can be referred from [17].

4.1 Data Set

We have identified three ontologies from different domain: Accounting, Travel and Conference which were used as input dataset for evaluating the Mimos-QG. The details of ontology are presented in Table 2 and conceptualized in Fig. 3.

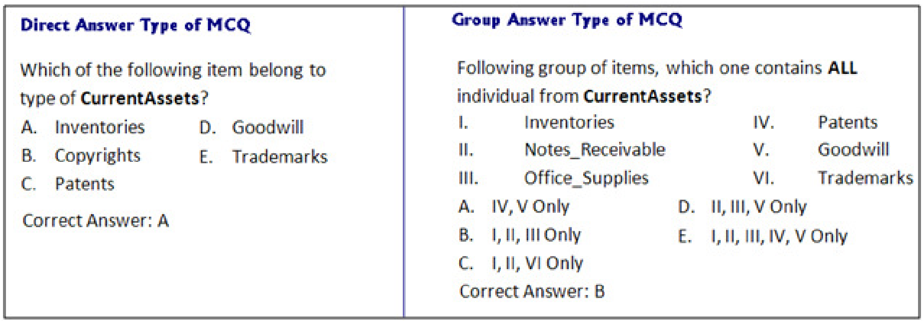

Fig. 2. Direct answer type of MCQ and group answer type of MCQ

Table 2: Dataset for Mimos-QG experimental evaluation

Ontology	Class	Object Property	Datatype Property	Instance
Accounting	21	2	8	59
Conference	98	60	32	0
Travel	30	6	4	10

4.2 Experimental Result

During experimental study, twenty questions were generated from each ontology assigned as input dataset. The questions were evaluated based on the capability of the ontology to generate each type of question from the Mimos-QG's question generation methods, quality of distractors and grouping answer choice, syntactically correctness and pedagogical quality. As shown in the Table 3 and Fig. 4, the MCQ generation from ontology is subjective to the nature of the ontology, whereas by the hierarchy of concepts and number of concepts and instances of the ontology has affected the evaluation criteria of the MCQ questions. As shown in Fig. 4(a), four types of multiple choice questions (Direct, Negative and Group approach of Class-Based Instance Strategy and Terminology-based Strategy) are generated from the Accounting ontology. As depicted by Fig. 4(b), all type of multiple choice questions except Negative approach of Class-Based Instance Strategy are generated from the Travel ontology. On the other hand, Conference ontology was only able to generate Group approach of Class-Based Instance Strategy and Terminology-based Strategy type of multiple choice questions as shown in Fig. 4(c).

The Mimos-QG performance on syntactical correctness and pedagogical skill measurement was good and acceptable as shown in Fig. 4(d). The results have proven that the Mimos-QG is capable of generating quality multiple choice questions to support assessment of different subject learning and understanding. However, the distractors and grouping answer choices generated from the Conference ontology is poor due to the lack of instances in the ontology itself.

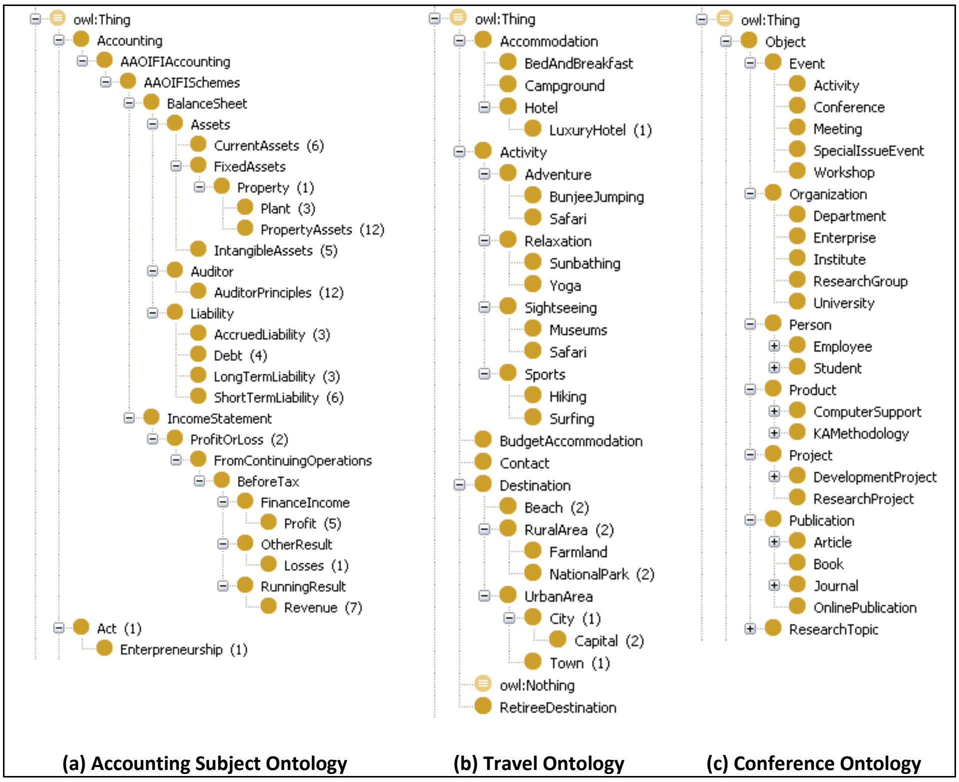

Fig. 3. Sample Ontologies from Different Domains

5 Conclusion

In this paper, we have discussed the criteria and requirements for evaluating an Independent Ontology Question Generator. We have also presented the experimental result on Mimos-QG. The experimental result have indicated that Mimos-QG is capable of generating quality MCQ questions to support assessment of different subject learning and understanding

For the future direction of Mimos-QG development, we intend to extend it in several prospective directions: (1) extend and enrich the question template to achieve higher level of syntactically correct question generation and (2) extend the MCQ generation strategies to incorporate other ontology components such as rules, axioms, restrictions, events and etc.

Table 3. Result of the Mimos-QG based on three ontologies

Question	Accounting				Travel				Conference			
	(a)	(b)	(c)	(d)	(a)	(b)	(c)	(d)	(a)	(b)	(c)	(d)
Q1	3	Y	N	N	3	Y	Y	Y	3	N	Y	Y
Q2	1	Y	Y	Y	1	Y	Y	N	3	N	Y	Y
Q3	3	N	Y	N	4	Y	Y	Y	1	Y	Y	Y
Q4	3	Y	N	Y	4	Y	Y	Y	1	Y	N	N
Q5	3	Y	N	Y	4	Y	Y	N	3	Y	Y	Y
Q6	3	Y	Y	Y	3	Y	Y	Y	1	Y	N	Y
Q7	2	Y	Y	Y	3	Y	Y	N	1	N	N	Y
Q8	1	Y	Y	Y	4	Y	N	N	3	Y	Y	Y
Q9	4	Y	N	Y	3	Y	Y	Y	3	N	Y	Y
Q10	3	Y	Y	N	3	N	Y	Y	3	N	Y	Y
Q11	3	Y	Y	Y	4	Y	Y	Y	3	N	Y	Y
Q12	3	Y	N	N	4	Y	N	N	3	N	Y	Y
Q13	3	N	Y	Y	4	Y	Y	N	3	N	Y	N
Q14	3	Y	N	N	4	Y	N	Y	3	N	Y	Y
Q15	3	Y	Y	Y	3	N	Y	Y	3	Y	Y	Y
Q16	3	Y	Y	N	4	Y	N	N	3	N	N	Y
Q17	2	Y	Y	Y	3	Y	Y	Y	1	Y	Y	Y
Q18	3	Y	Y	Y	3	N	Y	Y	3	N	Y	N
Q19	3	Y	Y	N	3	Y	Y	Y	3	N	Y	Y
Q20	1	Y	Y	Y	3	N	Y	Y	3	Y	Y	Y

Label
(a) Type of question
(b) Distractor & Grouping
(c) Syntatically
(d) Pedagogical
Y - Yes
N - No

Type of question:
Type 1: Class-based Instance Strategy (Direct Approach)
Type 2: Class-based Instance Strategy (Negative Approach)
Type 3: Class-based Instance Strategy (Group Approach)
Type 4: Terminology-based Strategy Methodology

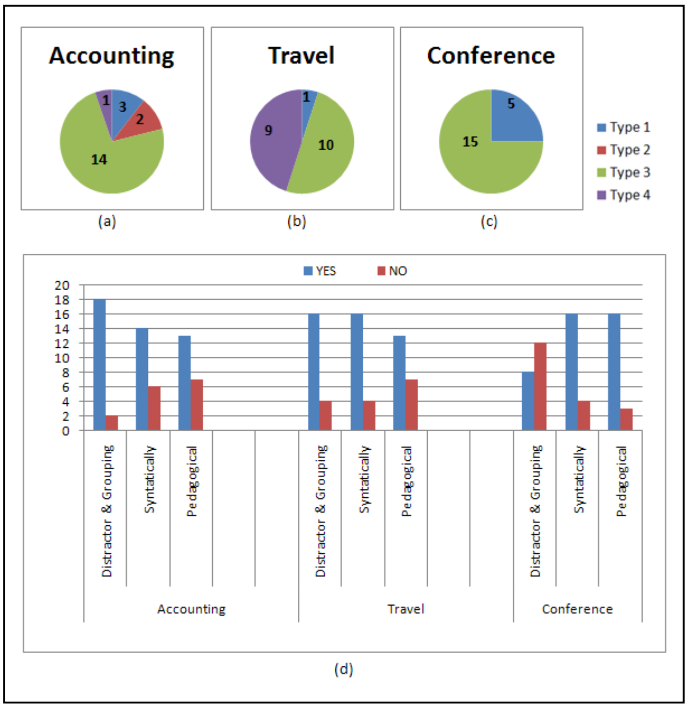

Fig. 4. Evaluation result of MCQ questions generated by Mimos-QG

References

1. Andrenucci, A., Sneiders, E.: Automated Question Answering: Review of the Main Approaches. 3rd International Conf. on Information Technology and Applications, 514–519 (2005)
2. McGough, J., Mortensen, J., Johnson, J., Fadali, S.: A Web-Based Testing System with Dynamic Question Generation. 31st ASEE/IEEE Frontiers in Education Conference, 3, 3–23 (2001)
3. Wang, W.M., Hao, T.Y., Liu, W.Y.: Automatic question generation for learning evaluation in medicine. ICWL 2007. LNCS, Springer, Heidelberg, 4823, 242–251 (2008)
4. Jonathan, C., Gwen, A., Eskenazi, E. M.: Automatic question generation for vocabulary assessment. Proceedings of the conference on Human Language Technology and Empirical Methods in Natural Language Processing, 819-826, Canada (2005)
5. Burdescu, D. D., Mihaescu, M. C., Building a decision support system for students by using concept maps. Proceedings of International Conference on Enterprise Information Systems (ICEIS'08), Spain (2008)
6. Papasalouros A., Kotis K., Kanaris K.: Automatic generation of multiple-choice questions from domain ontologies. IADIS e-Learning 2008 conference, Amsterdam (2008)
7. Sicilia M.A, Barriocanal E.G., On the convergence of formal ontologies and standardized e-learning, International Journal of Distance Education. 3(2), 13-29, (2005)
8. Knight, C., Gasevic, D., Richards, G.: An ontology-based framework for bridging learning design and learning content. Educational Technology & Society, 9 (1), 23-27 (2006)
9. OWL Web Ontology Language Overview, http://www.w3.org/TR/owl-features/
10. Tosic M., and Cubric M.: SeMCQ – Protégé Plugin for Automatic Ontology-Driven Multiple Choice Question Tests Generation, 11th Intl. Protégé Conference, Poster and Demo Session, (2009)
11. Cubric, M., Tosic. M.: Towards automatic generation of eAssessment using semantic web technologies. International Computer Assisted Assessment (CAA) Conference, Southampton, (2010)
12. Stanescu, L., Spahiu, C. S., Ion, A., Spahiu, A.: Question Generation for Learning Evaluation, International Multi-Conference on Computer Science and Information Technology, 509-513, (2008)
13. Heilman, M., Noah, S. A.: Question Generation via overgenerating transformations and ranking. The 14th Annual Conference on Artificial Intelligence in Education Workshop on Question Generation (2009)
14. Ishaq, A., Nanjappa S. B., Acharya, S.: Method and system for automatically generating questions for a programming. Infosys Technologies Limited, assignee. Patent 12/319566 (2009)
15. Tsumori, S., Kaijiri, K.: System Design for Automatic Generation of Multiple Choice Questions Adapted to Students' Understanding. Proceedings of the 8th International Conference on Information Technology Based Higher Education and Training, Japan, 541-546, Japan (2007)
16. Anderson, L.W., Sosniak, L. A.: Bloom's Taxonomy: A forty-year retrospective The National Society For The Study Of Education, (1994)
17. Tan, S.Y, Dickson, L.: System and method for ontology-based question generation. Mimos Berhad. (Patent - Pending)