

Learner's Performance Evaluation based on Knowledge Extracting and Ontology

Sami A. M. Al-Radaei

Department of Computer Engineering, IIT-BHU
Varanasi-221005- India

R. B. Mishra

Department of Computer Engineering, IIT-BHU
Varanasi-221005- India

ABSTRACT

Test is one of the tools that are used to evaluate learner's achievement. Most of test scoring in e-learning systems are for true-false and fill in blank questions. Description questions are human efforts and time consuming. A courseware with its questions bank had been built based on ontology. Extracting the semantic keywords from the learner's answer would be used to score the answer. In this paper we introduce a method to score the learner's answer based on semantic keywords in the question's ontology. Position priority and frequency of occurrence the semantic keywords have been taken in our calculation. This scoring is used to evaluate the learner's performance to answer description questions.

Keywords: E-learning, Semantic keywords, Ontology

1. INTRODUCTION

Evaluating the learner's performance is one of the most frequent activity in e-learning systems [1] it is considered as evidence of greater knowledge transmission[2]. Test is one of tools that are used to evaluate learner's achievement. The traditional test involves a lot of manual operations from both lecturer and learner. These operations have been reduced in most of e-learning systems but there is a difficulty to evaluate the learning achievement of learner by reading the learner's answer [3]. However, most of the test scoring in e-learning system is for true-false, multi-choice and fill in blank questions which allow the learner to select signal answer [4-6] and they neglected the description questions which is human efforts and time consuming. The automatic scoring was based on identifying the learner's answer with standard answer in the question bank, in programming languages question the automatic scoring can perform through compiled the source file from the question bank with set of input then compare the result with the result of the learner's answer[7, 8]. Many researchers have engaged the use of semantic web in their work for different purposes in E-learning systems such as control the learners' acquired knowledge [9], generating learning path sequencing [10, 11] and storing and retrieving the learning materials [12]. An auto scoring mechanism to evaluate the learning performance using problem solve approach[9]. The learner selects the best answer for the question form the system then he locates relevant information. Other researchers evaluate the learner by browsing the relevant record in learner's profile which considered a time/effort consuming [3, 13]. The evaluation of learner's performance using intelligent scoring mechanism has been proposed for different type of questions (True/False, multiple choice, multiple response and multiple fill-in-blank) [14].

Semantic web offers a promising approach to e-learning environment which ensures machine-process ability and interchange ability. Semantic web aims at introducing metadata to describe meaning of web resources in order to

create environments that are capable of advance automatic processing of the web content by both human and software agents [15]. Ontology, defined as a representation of a shared conceptualization of a particular domain, is the backbone of semantic web. Ontology typically consists of description of concepts relevant for a particular domain, their relations and axioms about these concepts and relationships.

This paper presents an automatic test scoring for the description questions by extracting the semantic keywords in the answer for learner's performance in e-learning system. The rest of the paper is organized as follows: in section 1, we presented the calculation of keyword's semantic value in the course-ware. In section 2 we described the structure of the ontology as backbone of semantic web technology for the courseware. Then the extraction of knowledge is presented in section 4. In section 5 we show an experimental result. Finally, the paper is concluded.

2. SEMANTIC VALUE

In order to assign semantic values to the questions in the courseware, first we extract the important keywords in the course-ware then we assign the semantic values to these keywords. The computation for the important keywords in the courseware begins with eliminating the stop word and common English words. This step is to reduce the total number of keywords in the keywords-courseware list. Then the expert lecturer will select the important keywords among all the keywords in the keywords-courseware list to avoid the occurrence of unrelated keywords in the courseware. We construct the keywords-courseware list using the Vector Space Model by computing the keywords' semantic value for each courseware. The keywords and their occurrence frequencies are represented in the matrix A as follows: $A = K \times Q$ Keyword-courseware matrix is $k \times q$ matrix, where K is keywords in Q courseware. This matrix represents one-row matrices and one-column matrices and these represent the vector. The Frobenius Norm of matrix (Euclidean Norm) is defined as the square root of the absolute squares of the matrix elements, which give the length of the vector. The semantic value $SVcw_j^{k_i}$ for the i th keyword k_i in the j th courseware cw_j is calculated according to the occurrence of the keywords in the courseware using Euclidean norm as follows:

$$SVcw_j^{k_i} = \frac{k_i}{\sqrt{\sum_{i=1}^n k_i^2}} \quad (1)$$

Where k_i is the occurrence of the keyword k in the courseware j. In table 1 we have shown part of keywords-courseware list for the course-work of Java programming language, the course-work consists of eight course-ware and we have extracted total of 75 keywords for all course-ware.

The semantic values of the keywords of the questions are different from courseware to other course-ware depends on the importance of the keyword in the courseware. For

example, the keywords private has the following semantic values (0.26858, 0.1411165, 0.636119, 0.155952, and 0.556815) for the courseware cw2, cw4, cw5, cw7 and cw9 respectively. The highest values for this keyword is in courseware cw5 because the learning materials in cw5 are

contain of class and object topic. We have used the calculation of keywords semantic values to generate dynamic learning path sequencing for learner according to his course selection[16].

Table 1 keywords-courseware semantic values

Keyword	cw1	cw2	cw3	cw4	cw5	cw6	cw7	cw8
array	0.348822	0.482561	0	0.543659	0	0	0.351984	0
Boolean	0.811214	0.841768	0	0	0	0	0	0
character	0	0.369921	0	0	0	0	0	0.625712
class	0.219156	0.9114	0.387956	0.448325	0.343542	0.298392	0.123194	0.387943
data	0.29237	0.84176	0.771976	0.81259	0.339263	0	0	0.11125
implement	0.648972	0	0.231593	0.336244	0.353373	0.693742	0	0.437997
inheritance	0.486728	0.16835	0	0.141165	0.127224	0	0.211194	0
instance	0.486728	0	0	0.215757	0.161983	0	0	0.23777
interface	0	0	0.385988	0.758249	0.432568	0.146133	0	0.23777
object	0.811214	0.128668	0.787416	0.285874	0.262929	0	0	0.888585
private	0	0.26858	0	0.141165	0.636119	0	0.155952	0.556815
protected	0	0	0	0.28233	0.824792	0	0	0.375426
public	0.456693	0.112585	0.771976	0.33946	0.86717	0.18373	0.281587	0.437997
subclass	0.456693	0.857788	0.671976	0.28233	0.279892	0	0	0

3. STRUCTURE OF THE ONTOLOGY FOR COURSEWARE:

Semantic Web is the extension of the current web [15], which is built for the goal of better enabling machine and human to work in cooperation. The semantic web annotated the content with formal semantic. This make the content suitable for machine consumption, which is intended to be built for human consumption and enables automated agents to reason about web contents and produce an intelligent response. There are various Semantic web languages for knowledge representation and semantic web reasoner to infer over it are available [17]. Ontology is the backbone of Semantic web which appears as a promising technology to implement e-learning. Ontology enables to annotate the learning materials semantically.

The Ontology is an explicit specification of conceptualization to describe and represent an area of knowledge. The concepts, their relationships and properties of concepts are expressed in ontology in hierarchical structure, the ontology provides a way to encode knowledge and semantic such that the machine can understand. Ontology enables us to annotate the learning materials semantically. OWL (Web Ontology Language) is a standard for representing knowledge on the Web with a focus on both making these documents compatible with Web standards and on being useful for the modeling of knowledge using past research on ontologies and reasoning[18]. The ontology contains the following major components regardless of the difference between ontology languages: Class, individuals and relationships. The Class or concept is the

abstract description of group of different individuals. Individual is an instance of class. The relationship is defined as the way in which classes and individuals are related to each other. The simple protocol and RDF query language (SPARQL) is used for querying the RDF and ontology files[19].

Figure 1 shows the mapping ontology diagram of book chapter and their relationships to other components. The class chapter is used to describe the content fragments. The data property hasConcepts holds the information of the contents available in a particular chapter titled "Interfaces and Inheritance" has the following concepts: "Implementing an Interface", "Inheritance", "Overriding and Hiding Methods", "Polymorphism", "HidingFields", "Using the Keyword super", "Object as a Superclass", "Writing Final Classes and Methods", "Abstract Methods and Classes". Chapter represents different type of materials. It means that the materials might be a study materials (concepts) or examination materials (pre-test and final test), Chapter and PreTest and FinalTest are related through hasPreTest and hasFinalTes properties respectively. The data property DifficultyLevel identifies how hard the chapter or learning materials are to work with. We identified three levels namely Low, Medium and Hard. The data property hasSemanticValue gives the weight of the chapter based on the important keywords in the chapter as we computed in the previous section. FinalTest and PreTest are related to the subclass Question using the object property hasQuestion. Each question has semantic keywords and their values.

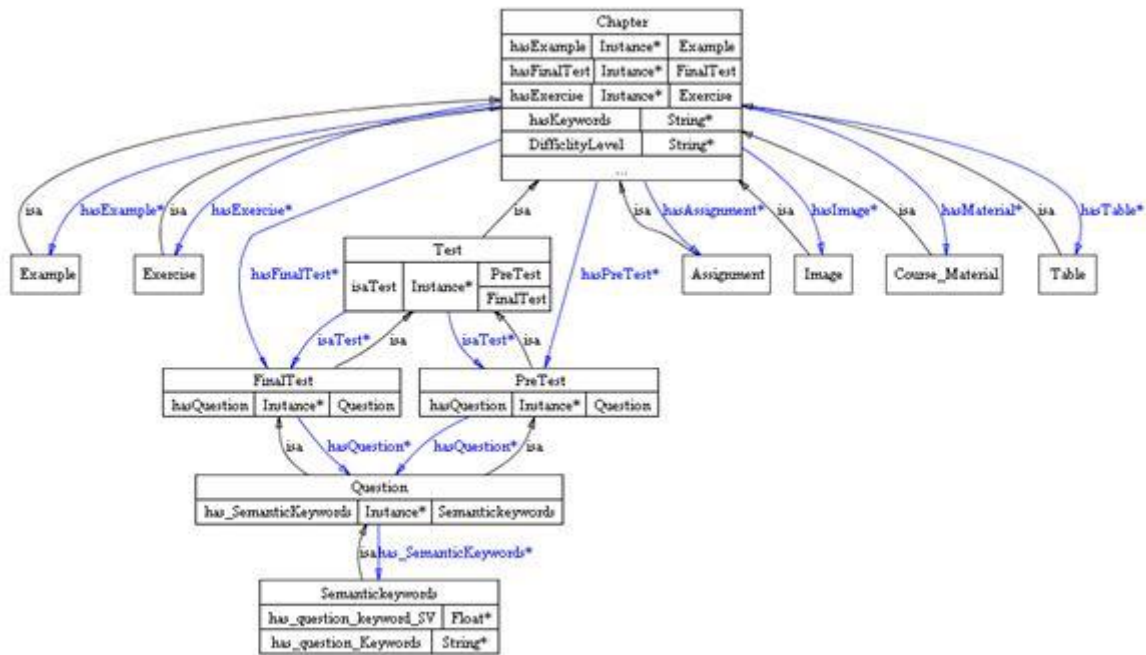


Figure 1 Class mapping of the courseware ontology

4. EXTRACTING THE KNOWLEDGE FROM STUDENT’S ANSWER:

This process is made up from several steps: 1) identifying the question, 2) collecting the answer 3) extracting the keywords 4) calculating the score and 5) updating the learner’s profile see figure 2.

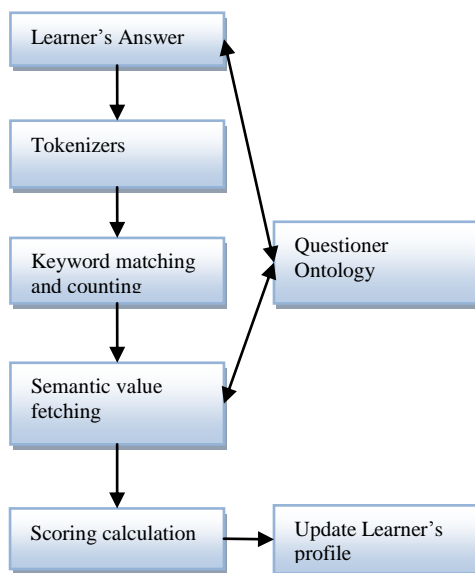


Figure 2: Test scoring flowchart

SPARQL query is used to identify the question from the ontology. Then we collect the answer from the learner which is in form of text. Extracting the keywords from the answer is

involves three stages, tokenizing, matching and fetching. A flat array of token is created for the answer using tokenizers which take the answer as input characters and segment it into tokens. In the next stage we match the tokens with the corresponding question’s semantic keywords in the ontology. In the same time we count the frequency of occurrence and location of each token. Each semantic keyword in the ontology has a position priority to appear before or after other semantic keywords in the answer. We have inserted the semantic keywords based on their priority. For calculating the position priority in the answer we have considered the first location of the semantic keywords in the answer as the position of the matched semantic keyword regardless of the appearance of the same semantic keyword in a later position. For example, in the basic java question “What is inheritance?” the learner should explain about class before sub class or base class. So sequence of keywords can be either “class, sub class, or base class” or “class, superclass, and subclass”. Thus, the size of the answer is also taken in our calculation since the short answer for description question is not acceptable for most of the lectures.

During matching the tokens with the corresponding question’s semantic keywords we fetch the semantic values for those semantic keywords, the semantic values are calculated as we explained in the previous section.

Then we calculate the score for the answer as in following equation:

$$score = \sum_{i=1}^n sv_i * \frac{1}{1 - \log_{tl} frq_i} * (1 - \log_{tl} |p_i - cp_i|)$$

Where sv_i is the semantic value of the semantic keyword i in the ontology, tl is the total length of the answer, p_i is the position of the semantic keywords in the ontology. cp_i is the current position of the semantic keyword in the answer and n is the total number of match's semantic keywords in the answer.

This equation consist of three weights are assigned to each semantic keywords in order to calculate the final score for the answer. These weights are keyword's semantic value (sv_i), keyword's frequency of occurrence weight ($\frac{1}{1 - \log_{tl} frq_i}$) and keyword's position weight ($1 - \log_{tl} (|p_i - cp_i|)$).

For the frequency weight and position weight we have taken the length of the answer as base of the logarithm in order to calculate the effect of frequency of occurrence as well as position of the semantic keyword with respect to the answer size. This score is used to evaluate the performance of the learner for the particular courseware. The result of each test is saved to the learner's profile which is in the form of ontology. After completing the test every concept learned within a range of scores will be saved in the learner's profile with their semantic keywords as *conceptlearned*. At any time we can

look into the performance of any learner by navigating his profile ontology using SPARQL query.

Extracting the semantic keywords from learner's answer lead us to know what exactly the learner has learned from the courseware. In another word the concept learned during the courseware and has been examine to the learner would be convert from (concept to learn) class to (concept learned) class in the courseware ontology. Moreover, the increasing of specific semantic keywords leads us to know the exact concept or topic.

5. EXPERIMENTAL RESULT

In this study we have developed a web-based learning system to implement the proposed method to evaluate the performance of the learners in answering description question based on the semantic web and semantic value calculation. We have built the courseware ontology for the learning material and questions and other related classes, and other ontology for the learner profile. As mentioned above, our concern in learner's profile is how to update the learner's performance. Protégé and its plug-ins are used to generate the all the ontologies, we have used SPARQL and Jena to query and update the ontologies in our web-based learning system which is built using Netbean and other tools. Figure 3 shows the question page in our web based learning system.

Table 2 Performance evaluation for 4 learners

Learner	Semantic Keyword	First position	Frequency of occurrence	Semantic Value	Frequency weight	Position weight	Score
Learner 1	class	1	5	0.9	1.473197445	1	1.325878
	superclass	2	4	0.8	1.382495573	1	1.105996
	subclasses	3	10	0.7	1.850274154	1	1.295192
	attributes	4	3	0.5	1.280829711	1	0.640415
	behaviour	5	2	0.64	1.160544085	1	0.742748
Score							9.562615
Learner 2	class	1	2	0.9	1.160544085	1	1.04449
	superclass	2	7	0.8	1.6349379	1	1.30795
	subclasses	4	3	0.7	1.280829711	0.861664811	0.772552
	attributes	5	1	0.5	1	0.861664811	0.430832
	behaviour	3	2	0.64	1.160544085	0.780743913	0.579896
Score							7.739047
Learner 3	class	1	7	0.9	1.6349379	1	1.471444
	superclass	2	3	0.8	1.280829711	1	1.024664
	subclasses	4	4	0.7	1.382495573	0.861664811	0.833873
	attributes	5	7	0.5	1.6349379	0.861664811	0.704384
	behaviour	3	6	0.64	1.556641376	0.780743913	0.777816
Score							9.004888
Learner 4	class	1	20	0.9	2.486787	1	2.238109
	superclass	0	0	0.8	1	0	0
	subclasses	0	0	0.7	1	0	0
	attributes	0	0	0.5	1	0	0
	behaviour	0	0	0.64	1	0	0
Score							4.188104



Figure 3 snapshot of the question page in our e-learning system

Here we are showing one question with its semantic keywords to demonstrate the performance evaluation. The question given to the four learners is “What is an Inheritance?” the answer for this question is “*Inheritance is a mechanism that enables one class to inherit all the behaviour and attributes of another class. Through inheritance, a class immediately picks up all the functionality of an existing class. Because of this, you only must define how the new class is different from an existing class. With inheritance, all classes those you create and those from the Java class library and other libraries are arranged in a strict hierarchy. A class that inherits from another class is called a subclass. The class that gives the inheritance is called a superclass. A class can have only one superclass, but each class can have an unlimited number of subclasses. Subclasses inherit all the attributes and behaviour of their superclasses.*” The semantic keywords and their semantic values are “class, superclass, subclass, attributes and behaviour”. In table 2 we show the performance of four different answers. We have calculated the score for each learner and normalized it with respect to score of the correct answer. From table 2 we can observe that answer from learner 1 contains the five semantic keywords and he has written them in his answer in proper sequence. Learner 2 and learner 3 have written the semantic keywords in the same sequence but the length of the answer of learner 3 is longer than the answer from learner 2. Finally the answer of learner 4 contains only one semantic keywords and his score is low.

6. CONCLUSION

Test is used to evaluate the performance of learners in e-learning environments. Evaluation of the description question is a human effort and time consuming. We have built the course-ware with their questions in ontology format. Learner's profile ontology is used to save the learner's performance, preferences, goal and learner's activities. We have introduced a method to calculate the score of the answer of description question based on semantic keywords in the answer as well as their semantic values. This score is based on predefined keywords for each question in the ontology. The position priority and frequency of occurrence of the semantic values are taken into account to calculate the score for the

answer. In future work this method would be used to evaluate coding questions.

7. REFERENCES

- [1] K. Mishra and R. B. Mishra, "Heuristic method for performance evaluation of learning in intelligent tutoring system " *INFOCOM journal of computer science*, vol. 9, pp. 1-7, 2010.
- [2] T. Beleche, D. Fairris, and M. Marks, "Do course evaluations truly reflect student learning? Evidence from an objectively graded post-test," *Economics of Education Review*, vol. 31, pp. 709-719, 2012.
- [3] C.-K. Chiou, G.-J. Hwang, and J. C. R. Tseng, "An auto-scoring mechanism for evaluating problem-solving ability in a web-based learning environment," *Computers & Education*, vol. 53, pp. 261-272, 2009.
- [4] W. Han, X. Jun, Feng, and Gang, "Implementation of A Testing System on Web-based Course of C Programming Language," *J. Computer & Digital Engineering*, vol. 31, pp. 37-41, 2003.
- [5] M. R. Del Sorbo and W. Balzano, "e-Xamina: An Experimental Multi-user Assessment Platform for Computer Adaptive Testing," 2011, pp. 237-240.
- [6] A. Bentiba, M. J. Zemerly, O. A. Hammadi, S. A. Sharif, and M. Naqbi, "Dynamic Generated Material for E-Learners (DGML)," 2011, pp. 247-250.
- [7] F. Yang and G. Liu, "Study and application of automatic scoring technology in C programming test," 2012, pp. 587-589.
- [8] D. S. Morris, "Automatic grading of student's programming assignments: an interactive process and suite of programs," 2003, pp. S3F - 1-6 vol.3-S3F - 1-6 vol.3.
- [9] A. Gladun, J. Rogushina, Garc, F. a-Sanchez, Mart, B. nez, R. jar, Fern, and J. T. ndez-Breis, "An application of intelligent techniques and semantic web technologies in e-learning environments," *Expert Syst. Appl.*, vol. 36, pp. 1922-1931, 2009.
- [10] E. Kontopoulos, D. Vrakas, F. Kokkoras, N. Bassiliades, and I. Vlahavas, "An ontology-based planning system for e-course generation," *Expert Systems with Applications*, vol. 35, pp. 398-406, 2008.
- [11] C.-M. Chen, C.-J. Peng, and J.-Y. Shiu, "Ontology-based concept map for planning personalized learning path," 2008, pp. 1337-1342.
- [12] J. Emina, "Preparation of the learning content for semantic e-learning environment," *Procedia - Social and Behavioral Sciences*, vol. 1, pp. 824-828, 2009.
- [13] C.-C. Tsai, "Taiwanese Science Students' and Teachers' Perceptions of the Laboratory Learning Environments: Exploring Epistemological Gaps," *International Journal of Science Education*, vol. 25, pp. 847-60, 2003.
- [14] H.-L. Yang and M.-H. Ying, "Scoring Effect of Online Test: Implications on KM and e-learning," *INTERNATIONAL JOURNAL OF INNOVATION AND LEARNING*, vol. 4, pp. 40--58-40--58, 2007.
- [15] T. Berners-Lee, J. Hendler, and O. Lassila. (2001, May 2001) The Semantic Web. *Scientific American*

- [16] S. A. M. Al-Radaei and R. B. Mishra, "A Heuristic Method for Learning Path Sequencing for Intelligent Tutoring System (ITS) in E-learning," *International Journal of Intelligent Information Technologies*, vol. 7, pp. 65-80, 2011.
- [17] R. B. Mishra and S. Kumar, "Semantic web reasoners and languages," *Artif. Intell. Rev.*, vol. 35, pp. 339-368, 2011.
- [18] A. Kalyanpur, B. Parsia, and J. Hendler, "A Tool for Working with Web Ontologies," *International Journal on Semantic Web and Information Systems*, vol. 1, pp. 36-49, 2005.
- [19] E. Prud and A. Seaborne. (2008). *SPARQL Query Language for RDF*. Available: <http://www.w3.org/TR/rdf-sparql-query/>