

Mining Ontologies for Analogy Questions: A Similarity-based Approach

Tahani Alsubait, Bijan Parsia, and Uli Sattler

School of Computer Science, The University of Manchester, United Kingdom
{alsubait,bparsia,sattler}@cs.man.ac.uk

Abstract. In this paper, we propose a new approach to generate analogy questions of the form "A is to B as ... is to ?" from ontologies. Analogy questions are widely used in multiple-choice tests such as SATs and GREs and are used to assess student's higher cognitive abilities. The design, implementation and evaluation of the new approach are presented in this paper. The results show that mining ontologies for such questions is fruitful.

1 Introduction

Learning may be seen as its own reward; however assessment is usually required to provide various types of reward and recognition. This notion of assessment is usually referred to as summative assessment compared to formative assessment which is mainly for providing necessary feedback to students to support the learning process.

Assessment items (i.e. questions) can be classified into two widely used formats: (i) Objective (e.g. Multiple Choice Questions (MCQs) or True/False questions) and (ii) Subjective (e.g. essays or short answers). Each family of questions has its own advantages/disadvantages w.r.t. different phases of testing (i.e. Setting, Taking and Marking). On the one hand, objective tests can be used to assess a broad range of knowledge and yet require less administration time. In addition, they are scored easily, quickly and objectively either manually or automatically and can be used to provide instant feedback to test takers. On the other hand, objective questions are hard to prepare and require considerable time per each question [26]. For example, Davis [8] & Lowman [19] pointed out that even professional test developers cannot prepare more than 3-4 items per day. In addition to the considerable preparation time, manual construction of MCQs does not necessarily imply that they are well-constructed. See for example the study carried out by Paxton [23] who analysed a large number of MCQs and reported that they are often not well-constructed.

A major challenge in preparing MCQs is the need for good distractors that should appear as plausible answers to the question for those students who have not achieved the objective being assessed. At the same time, distractors should appear as implausible answers for those students who have achieved the objective [3]. Moreover, a well-written MCQ is a question that does not confuse students,

and yields scores that can be used in determining the extent to which students have achieved educational objectives [3, 17].

Many guidelines have been proposed to ensure the effectiveness of distractors; however, many major issues are still debatable such as the appropriate number of distractors [10, 23].

Before the effectiveness of MCQs is discussed further, the difficulty of evaluating such questions should be mentioned. The difficulty of evaluation lies, among other things, in the need for administering those questions to real students in normal settings and analyzing their grades according to well-defined procedures. For example, one can follow the procedures described in Item Response Theory (IRT) [16, 21, 20] which is a theory that explains the statistical behavior of good/bad questions. According to IRT, good test questions have the following three characteristics: (i) prevent students from guessing the correct answer, (ii) function towards proper discrimination between good and poor students and (iii) different questions in the test have different difficulties.

In addition to the above mentioned characteristics of good questions, the need for having questions that assess different levels of learning objectives should be also mentioned. For example, a test developer might be interested in knowing the level of which a student has achieved a learning objective (ranging from the ability to recall information to the ability to analyse and judge the provided information) [2]. Note that questions that address a specific level of learning objectives can be of different difficulties for a specific set of students. Note also that questions that assess lower level objectives are not necessarily questions of a lower quality as long as they meet the determined learning objectives [11].

Given the considerable time and effort required to develop MCQs, we propose to automate the generation of these questions by using an ontology-based approach. Our motivation to use OWL ontologies in particular is their precise semantics, available reasoning services and considerable efforts put into their development. One of the promises of representing knowledge in such ontologies is that it can be used for different applications. In this paper, we investigate the potential benefit of ontology-based question generation.

Recently, a handful of studies [12, 13, 22, 29, 30, 1] explored the generation of MCQs over ontologies. A brief overview of these approaches is provided in section 5. A general critique of these approaches is the lack of pedagogic theory backing which we try to overcome in this paper. Moreover, most of these approaches generate questions of the type "What is X?" or "Which of the following is an example of X?" based on class-subclass and/or class-individual relationships. This type of questions can only assess lower levels of learning objectives [2]. Therefore, it is crucial to design approaches capable of generating questions of other types.

In this paper, we present a new approach for generating multiple-choice analogical questions from ontologies. Such questions aim to assess the analogical reasoning ability of students (i.e. the ability to determine the underlying relation between a pair of concepts and identifying a similar pair that has the same underlying relation). We also describe the notion of relational similarity and how

to use it to control the difficulty of the generated questions. In addition, we report on some experiments carried out to evaluate the new approach using a large corpus.

2 Preliminaries

To understand the procedure required to generate MCQs, we first present a simple, yet general, definition for MCQs in what follows.

Definition 1. *A multiple choice question MCQ is a tool that can be used to evaluate whether (or not) our students have achieved a certain learning objective. It consists of the following parts:*

- A statement S that introduces a problem to the student (i.e. stem).
- A number of functional options $A = \{A_i | 2 \leq i \leq \max\}$ that can be further divided into two sets:
 1. A number of correct options $K = \{K_m | 1 \leq m \leq i\}$ (i.e. key)
 2. A number of incorrect options $D = \{D_n | n := i - m\}$ (i.e. distractors).

To generate good MCQs we need a psychologically plausible theory that guides us in the generation. In this paper, we propose to use the notion of similarity to control the characteristics of the generated questions. For example, consider a question that has a stem that is similar to the key and different from the distractors. We would expect that the students will find this question to be an easy one (assuming that they notice the clues provided with the correct answer). Similarly, we would expect the question to be difficult if the stem is very similar to one (or all) of the distractors and different from the key.

There are at least two major types of similarity. In addition, similarity is different than the general notion of relatedness. For example, we say that cars and fuel are closely related compared to cars and bicycles that are closely similar. This notion of similarity is usually referred to as semantic similarity. A number of measures have been proposed to measure semantic similarity between concepts. See for example [24, 25, 18, 15] for general similarity measures and [7] for a semantic similarity measure that was designed for DL ontologies. Another important type of similarity is relational similarity [28, 27] which corresponds to similarities in the underlying relations. For example, food is to body as fuel is to car. When two pairs of concepts have a strong relational similarity, we say that they are analogous. In analogical reasoning, we compare two different types of objects and identify points of resemblance.

Different types of similarity can play different roles in question generation. For example, semantic similarity can be used to generate plausible distractors for simple recall questions of the form "What is X?". Also, controlling the degree of similarity between the stem, key and distractors allows us to generate questions of different difficulties. Along similar lines, relational similarity plays a major role in generating questions that assess higher cognitive abilities. As an example of such questions that can be generated using our proposed similarity-based

approach, we consider analogy questions that have the form "X is to Y as:". The alternative answers to such questions take the form " X_i is to Y_i " where the key is the pair (X_i, Y_i) that has the same underlying relation as the pair (X, Y) in the stem. See Table 1 for a sample multiple-choice analogy question taken from the GRE exam. For our purposes, we define analogy questions as follows (detailed explanation for Relatedness and Analogy functions will be provided later):

Definition 2. Let Q be a multiple-choice analogy question with stem $S = (X, Y)$, key $K = (V, W)$ and a set of distractors $D = \{D_i = (A_i, B_i) | 1 < i \leq \max\}$. We assume that Q satisfies the following conditions:

1. The stem S , the key K , the distractor D_i are all good (i.e. $\text{Relatedness}(X, Y) \geq \Delta_R$, $\text{Relatedness}(V, W) \geq \Delta_R$, $\text{Relatedness}(A_i, B_i) \geq \Delta_R$).
2. The key K is significantly more analogous to S compared to the distractors (i.e. $\text{Analogy}(S, K) \geq \text{Analogy}(S, D_i) + \Delta_1$).
3. The key K is sufficiently analogous to S (i.e. $\text{Analogy}(S, K) \geq \Delta_2$).
4. The distractors should be analogous to S to an extent (i.e. $\text{Analogy}(S, D_i) \geq \Delta_3$).
5. Each distractor D_i is unique (i.e. $\text{Analogy}(S, D_i) \neq \text{Analogy}(S, D_j)$).

Table 1. A sample multiple-choice analogy question [9]

Stem:	Cat: Mouse
Options:	(A) Lion: Tiger (B) Food: Warmth (C) Bird: Worm (D) Dog: Tail
Key:	(C) Bird: Worm

We would like to be able to control the difficulty of the generated questions. According to Definition 2 and Propositions 1.a, 1.b, 1.c we can control the difficulty of Q by increasing or decreasing Δ_1 , Δ_2 and Δ_3 .

- Proposition 1.**
- a. Increasing Δ_1 decreases the difficulty of Q .
 - b. Increasing Δ_2 decreases the difficulty of Q .
 - c. Decreasing Δ_3 decreases the difficulty of Q .

To generate analogy questions, we need to define two functions, *Relatedness* and *Analogy*. A very basic example for the *Relatedness* function is to consider concepts that are both referenced in one (or more) of the ontological axioms in the source ontology as sufficiently related concepts (e.g. $X \sqsubseteq \exists r.Y \rightarrow \text{Relatedness}(X, Y) > 0$). However, such a syntax-based notion of relatedness is sensitive to tautologies and therefore cannot be adopted without further considerations. For simplicity, we currently adopt a simple relatedness notion that considers a pair of named classes to be sufficiently related if they have one of the structures in Figure 1. These structures have at most one change in direction in the path connecting the two nodes and at most two steps in each direction. Other

structures were discarded to avoid too difficult (and probably confusing) questions. While in the most general case, one should consider pairs with arbitrary related classes (e.g. by considering user-defined relations), for current purposes we only consider class-subclass relations. This simplifies the problem considerably in several dimensions while still generates reasonable number of candidate pairs (as we will see later). In addition, we need to define the function *Analogy*

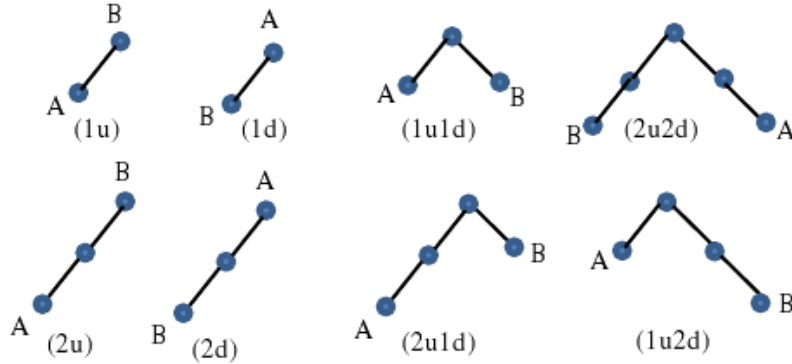


Fig. 1. Closely related structures of class-subclass relations (labels represent no. of steps and direction (up or down) in the path that connects the two concepts)

which is the core function for generating multiple-choice analogy questions. This function is defined as follows:

Definition 3. Let $Analogy(x, y)$ be the function that takes two pairs of concepts and returns a numerical score for their analogy value according to the equation:

$$Analogy(x, y) = \frac{SharedSteps(x, y)}{TotalSteps(x, y)} \times \frac{SharedDirections(x, y)}{TotalDirections(x, y)} \quad (1)$$

3 Extracting Analogy Questions from Ontologies

One of the questions that arise here is how many MCQs can be generated from a given ontology? To answer this question we need to first determine what parts of the ontology (i.e. classes, individuals, properties, and annotations) will be considered in the generation process. Secondly, we need to determine whether (or not) a filtering mechanism is used to differentiate between good and bad questions and to generate questions that are supposed to be good questions only. As an example, the following equation (2) can be used to count the number of possible multiple-choice questions of the form "What is [class name]?" with one key and three distractors (all are class names), assuming that no filtering mechanism is used (n is the number of classes in the given ontology and T_i is

the number of correct answers (i.e. super-classes) for class i):

$$\sum_{i=1}^n \binom{T_i}{1} \times \binom{n-1-T_i}{3} \quad (2)$$

Needless to say, the number of questions increases rapidly as n grows (see Figure 2 for some examples). Also, it reaches its maximum value when T_i equals $\frac{n-1}{4}$ (i.e. the ratio of correct answers to wrong answers is 1:3). The number of possible questions that can be generated from a given ontology can further be increased if we consider other parts of the ontology (e.g. individuals, properties). Having said this, we should also mention that generating a large number of questions is not desirable unless the generated questions are expected to be all good. A similar analysis of the number of possible analogy questions is part of future work. In what follows, we provide an algorithm (see Algorithm 1) that can be

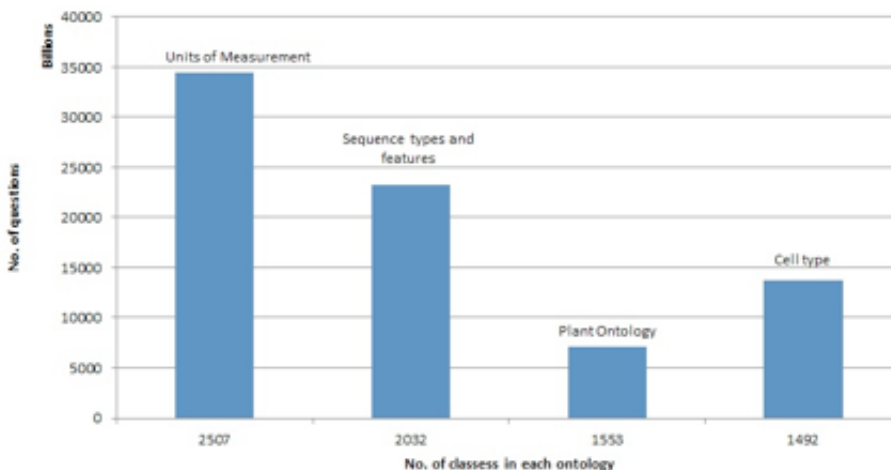


Fig. 2. Number of possible questions from some BioPortal ontologies

used to generate multiple choice analogy questions from a given ontology \mathcal{O} . The algorithm is founded on the premise that varying the relational similarity (i.e. the analogy degree) between the stem, the key and distractors allows us to control the difficulty of the generated questions. This can be achieved by setting the parameters Δ_1 , Δ_2 and Δ_3 to different values. The proposed approach consists of two phases: (i) extraction of interesting pairs of concepts which can be determined using the proposed *Relatedness* function, those pairs can be used as stems, keys or distractors and (ii) generation of multiple-choice questions based on the similarity between pairs which can be derived from the proposed *Analogy* function. Note that this approach can be generalized to generate other types of questions such as finding the antonyms/odds.

Algorithm 1 Generate_Analogy_Question($\mathcal{O}, \Delta_R, \Delta_1, \Delta_2, \Delta_3, \max$)	
1	AQ = {};
2	For each pair of classes (X,Y) in \mathcal{O} s.t. $\text{Relatedness}(X,Y) \geq \Delta_R$
3	For each pair of classes (V,W) in \mathcal{O} s.t. $(A,B) \neq (V,W)$ and $\text{Relatedness}(V,W) \geq \Delta_R$
4	and $\text{Analogy}((A,B),(V,W)) \geq \Delta_2$
5	Q.S = (X,Y);
6	Q.K = (V,W);
7	For each i ($0 \leq i < \max$)
8	Get a pair of classes (A_i, B_i) in \mathcal{O} s.t. $(X,Y) \neq (A_i, B_i)$ and $\text{Relatedness}(A_i, B_i) \geq \Delta_R$
9	and $\text{Analogy}((X,Y),(A_i, B_i)) \geq \Delta_3$
10	If $\text{Analogy}((X,Y),(V,W)) \geq \text{Analogy}((X,Y),(A_i, B_i)) + \Delta_1$ and $(A_i, B_i) \notin Q.D$
11	Q.D = Q.D + (A_i, B_i);
12	End If
13	Next i
14	AQ = AQ + Q;
15	Next (V,W)
16	Next (X,Y)
17	Return AQ;

4 Empirical Evaluation

To evaluate the proposed approach, we implemented a question generation engine that utilizes algorithm 1 and used the implemented engine to generate analogy questions from three ontologies (one specialized ontology and two tutorial-based ontologies). The three ontologies are presented in Table 2 below with some basic ontology statistics. The first ontology is the Gene Ontology which is a structured vocabulary for the annotation of gene products. It has three main parts: (i) molecular function, (ii) cellular component and (iii) biological role. The two other ontologies are the People & Pets Ontology and the Pizza Ontology which are very simple ontologies that are usually used in ontology development tutorials. The table shows the number of classes in each ontology and the number of sample questions generated by the engine (this is only a representative sample of all the possible questions). The table also shows the percentage of questions that our proposed solver agent can solve correctly. The details of the approach used to simulate question solving are explained in what follows. Other ontologies can be used as input for our implemented question generation engine; however we tried to avoid ontologies that use difficult-to-read labels (e.g. labels that have no spaces between words).

Table 2. Ontologies used to generate analogy-questions along with basic statistics

	No. of Classes	No. of questions	%Correct
Gene Ontology	36146	25	8%
People & Pets	58	15	67%
Pizza Ontology	97	16	88%

In order to evaluate the proposed similarity-based approach defined in Algorithm 1, we need at least to simulate students while solving the generated questions and check whether (or not) the proposed approach can be used to successfully control the difficulty of questions. To do this, we follow the method

explained by Turney & Littman [28, 27] for evaluating analogies using a large corpus.

In their study, Turney & Littman reported that their method can solve about 47% of multiple-choice analogy questions (compared to an average of 57% correct answers solved by high school students). The solver takes a pair of words representing the stem of the question and 5 other pairs representing the answers presented to students. Their proposed method is inspired by the Vector Space Model (VSM) of informational retrieval. For each provided answer, the solver creates two vectors representing the stem (R_1) and the given answer (R_2). The solver returns a numerical value for the degree of analogy between the stem and the given answer. Then, the answers are ranked according to their analogy value and the answer with the highest rank is considered the correct answer. To create the vectors, they proposed a table of 64 joining terms that can be used to join the two words in each pair (stem or answer). The two words are joined by these joining terms in two different ways (e.g. "X is Y" and "Y is X") to create a vector of 128 features. The actual values stored in each vector are calculated by counting the frequencies of those constructed terms in a large corpus (e.g. web resources indexed by a search engine). To improve the accuracy of their proposed method, they suggested to use the logarithm of the frequency instead of the frequency itself.

In this paper, we follow a similar procedure to evaluate the difficulty of our generated MCQs. First, we constructed a table of joining terms relevant to the relations considered in our approach (e.g. "is a", "type", "and", "or"). Based on these joining terms, we create vectors of 10 features for the stem, the key and each distractor. The constructed terms are sent as a query to a search engine (Yahoo!) and the logarithm of the hit count is stored in the corresponding element in the vector. The hit count is always incremented by one to avoid getting undefined values. Following this procedure, our proposed solver agent solved 8% of the questions generated from the Gene Ontology, 67% of the questions generated from the People and Pets Ontology and 88% of the questions generated from the Pizza Ontology. We argue that this is caused by the specific terminology used in the Gene Ontology and lack of web resources that have information regarding it compared to the other ontology.

Examples of the questions that were generated using our proposed approach are presented in Tables 3 & 4. Those questions were generated from the People & Pets ontology and Pizza ontology respectively. Moreover, we varied the difficulty-control parameters to generate different sets of questions (i.e. questions of different difficulties) from the two tutorial ontologies. The results (See Table 5) show that the proposed parameters sufficiently controlled the difficulty of the generated questions.

5 Related Work

Chung, Niemi, and Bewley (2003) [4] developed the Assessment Design and Delivery System (ADDS). The purpose of ADDS is to assist non-expert physics

Table 3. Sample question generated from the People & Pets Ontology

Stem:	Haulage Truck Driver : Driver
Options:	(A) Quality Broadsheet : Newspaper (B) Giraffe : Sheep (C) Bus : Vehicle (D) Giraffe : Cat Liker
Key:	(C) Bus : Vehicle

Table 4. A sample question generated from the Pizza Ontology

Stem:	Sloppy Giuseppe : Pizza De Carne
Options:	(A) Cogumelo : Pizza Vegetariana (B) Pizza : Food (C) Cogumelo : Napoletana (D) Cogumelo : Sorvete
Key:	(B) Pizza : Food

teachers in designing appropriate assessments by constraining the design process by structure-based and cognitive-based rules derived from an ontology that was specifically designed for the system. In addition, ADDS's domain ontology has links to a set of reusable assessment tasks or components of tasks (i.e. text, graphic, multimedia) along with information to guide teachers practice.

Holohan et al. (2005) [12] described the OntAWare system which is an ontology-based authoring environment for learning content. It employs an ontology graph traversal algorithm that generates MCQs of the form "Which of the following items is (or is not) an example of the concept, X?". The alternative answers will be generated randomly and the question as a whole can be exported to external systems that conform to the IMS/QTI [14] standard. One of the central problems in OntAwar, other than the highly constrained forms of questions, is that the ontology graph transformations employed in the system are hardcoded (in Java) to incorporate implicit instructional strategies and therefore their approach is not ready to be generalized and adopted in other systems. They extended their work in Holohan et al. (2006) [13] by focusing on the generation of SQL exercise problems for database students using domain-dependent algorithms.

Stankov and Zitko (2008) [29] proposed templates and algorithms for automatic generation of objective questions (i.e. MCQs, T/F) over ontologies. The

Table 5. Difference in the percentage of correctly answered questions (generated from People & Pets Ontology and Pizza Ontology)

	Types of questions	Δ_1	Δ_2	Δ_3	No. of questions	%Correct
People & Pets	Decreased Difficulty	0.5	1	0	15	67%
	Increased Difficulty	0.75	0.5	0.1	11	27%
Pizza Ontology	Decreased Difficulty	0.5	1	0	16	88%
	Increased Difficulty	0.75	0.5	0.1	16	50%

focus in their work was to extend the functionality of a previously implemented tutoring system (Tex-Sys) by concentrating on the assessment component. The proposed methodology generates a set of random alternative answers for each MCQ without an attempt to filter them according to their pedagogical appropriateness.

Papasalouros et al. (2008) [22] presented various ontology-based strategies for automatic generation of MCQs. These strategies are used for selecting the correct and wrong (distracting) answers of the questions. The answers are later transformed into English sentences using simple natural language generation techniques. The evaluation of the produced questions by domain experts shows that the questions are satisfactory for assessment but not all of them are syntactically correct. The major problem related to this approach is the use of highly constrained rules with no theory backing that would motivate the selection of these rules. For example, the distractors in each MCQ are mainly picked from the set of siblings of the correct answer while there might be other plausible distractors.

Cubric and Tasic (2009) [5] reported their experience in implementing a Portege plugin for question generation based on the strategies proposed by Papasalouros et al. (2008) [22]. More recently, Cubric and Tasic (2010) [6] extended their previous work by considering new ontology elements (i.e. annotations). In addition, they suggested employing question templates to avoid syntactical problems in the automatically generated questions. This also enables the generation of questions in different levels of Blooms taxonomy [2].

6 Conclusion and Future Work

A handful of studies have already proposed some approaches to generate MCQs over ontologies, however little have been done on the theoretical and evaluation aspects. In this paper, we propose a new approach to generate multiple-choice analogy questions from ontologies. The paper describes the foundations of the proposed approach from a psychological point of view. In addition, the paper reports on some evaluations carried out to evaluate the proposed approach. The results show that mining ontologies for analogy questions in particular and for assessment questions in general is fruitful. Moreover, the results show that the proposed approach can be used to control the difficulty of the generated questions.

For future work, we aim to generalize the proposed approach for generating analogies and consider arbitrary relations found in existing ontologies (i.e. user-defined relations instead of only class-superclass relations). To evaluate such analogies, we suggest to use Latent Relational Similarity (LRS) [27] which has the ability to learn relations instead of using predefined joining terms.

References

1. M.Al-Yahya. Ontoque: A question generation engine for educational assessment based on domain ontologies. In 11th IEEE International Conference on Advanced

- Learning Technologies, 2011.
2. B.S. Bloom and D.R. Krathwohl. Taxonomy of educational objectives: The classification of educational goals by a committee of college and university examiners. Handbook 1. Cognitive domain. New York: Addison-Wesley, 1956.
 3. S.Burton, R.Sudweeks, P.Merrill, and B.Wood. How to prepare better multiple-choice test items: Guidelines for university faculty. Brigham young university testing services and the department of instructional science. Retrieved November 22, 2011, from <http://testing.byu.edu/info/handbooks/betteritems.pdf>, 1991.
 4. G.Chung, D.Niemi, and W.L. Bewley. Assessment applications of ontologies. In Paper presented at the Annual Meeting of the American Educational Research Association, 2003.
 5. M.Cubric and M.Tosic. SEMcq: Protege plugin for automatic ontology-driven multiple choice question tests generation. In 11th Intl. Protege Conference, Poster and Demo Session, 2009.
 6. M.Cubric and M.Tosic. Towards automatic generation of e-assessment using semantic web technologies. In Proceedings of the 2010 International Computer Assisted Assessment Conference, University of Southampton, July 2010.
 7. C.d'Amato, S.Staab, and N.Fanizzi. On the influence of description logics ontologies on conceptual similarity. In EKAW 08 Proceedings of the 16th international conference on Knowledge Engineering: Practice and Patterns, 2008.
 8. B.B. Davis. Tools for Teaching. San Francisco, CA: Jossey-Bass, 2001.
 9. GRESampleQuestions. Best sample questions. Retrieved March 10, 2012, from <http://www.bestsamplequestions.com/gre-questions/analogies/>.
 10. T.M. Haladyna and S.M. Downing. How many options is enough for a multiple choice test item? Educational & Psychological Measurement, 53(4):9991010, 1993.
 11. M.Hufler, M.AL-Smadi, and C.G. Investigating content quality of automatically and manually generated questions to support self-directed learning. In In White-lock, D., Warburton, W., Wills, G., and Gilbert, L. (Eds.), CAA 2011 International Computer Assisted Assessment Conference, University of Southampton, 2011.
 12. E.etal. Holohan. Adaptive e-learning content generation based on semantic web technology. In Proceedings of Workshop on Applications of Semantic Web Technologies for e-Learning, pages 2936, Amsterdam, The Netherlands, 2005.
 13. E.etal. Holohan. The generation of e-learning exercise problems from subject ontologies. In Proceedings of the Sixth IEEE International Conference on Advanced Learning Technologies, pages 967969, 2006.
 14. IMS. IMS question & test interoperability. ASI best practice & implementation guide. final specification version 1.2. IMS global learning consortium Inc., June 2002.
 15. J.Jiang and D.Conrath. Semantic similarity based on corpus statistics and lexical taxonomy. In In: Proc. of the 10th International Conference on Research on Computational Linguistics, Taiwan, 1997.
 16. J.Kehoe. Basic item analysis for multiple-choice tests. Practical Assessment, Research & Evaluation, 4(10), 1995.
 17. K.King, D.Gardner, S.Zucker, and M.Jorgensen. The distractor rationale taxonomy: Enhancing multiple-choice items in reading and mathematics. Assessment Report. Pearson, July 2004.
 18. D.Lin. An information-theoretic definition of similarity. In In: Proc. of the 15th International Conference on Machine Learning, page 296-304, San Francisco, CA, 1998. Morgan Kaufmann.
 19. J.Lowman. Mastering the Techniques of Teaching (2nd ed.). San Francisco: Jossey-Bass, 1995.

20. M.Miller, R.Linn, and N.Gronlund. *Measurement and Assessment in Teaching*, Tenth Edition. Pearson, 2008.
21. R.Mitkov, L.AnHa, and N.Karamani. A computer-aided environment for generating multiple-choice test items.cambridge university press. *Natural Language Engineering*, 12(2):177194, 2006.
22. A.Papasalouros, K.Kotis, and K.Kanaris. Automatic generation of multiple-choice questions from domain ontologies. In *IADIS e-Learning 2008 conference*, Amsterdam, 2008.
23. M.Paxton. A linguistic perspective on multiple choice questioning. *Assessment & Evaluation in Higher Education*, 25(2):109119, 2001.
24. R.Rada, H.Mili, E.Bicknell, and M.Blettner. Development and application of a metric on semantic nets. In *In: IEEE Transaction on Systems, Man, and Cybernetics*, volume19, page 17-30, 1989.
25. P.Resnik. Using information content to evaluate semantic similarity in a taxonomy. In *In Proceedings of the 14th international joint conference on Artificial intelligence (IJCAI'95)*, volume1, pages 448453, 1995.
26. J.T. Sidick, G.V. Barrett, and D.Doverspike. Three-alternative multiple-choice tests: An attractive option. *Personnel Psychology*, 47:829835, 1994.
27. P.Turney. Measuring semantic similarity by latent relational analysis. In *IJCAI is the International Joint Conference on Artificial Intelligence*, 2005.
28. P.Turney and M.Littman. Corpus-based learning of analogies and semantic relations. *Machine Learning*, 60(1-3):251278, 2005.
29. B.Zitko, S.Stankov, M.Rosi, and A.Grubi. Dynamic test generation over ontology-based knowledge representation in authoring shell. *Expert Systems with Applications: An International Journal*, 36(4):81858196, 2008.
30. K.Zoumpatianos, A.Papasalouros, and K.Kotis. Automated transformation of SWRL rules into multiple-choice questions. In *FLAIRS Conference11*, 2011.