

# ECOLE: Student Knowledge Assessment in the Education Process

Dmitry Mouromtsev  
ITMO University  
St. Petersburg, Russia  
d.muromtsev@gmail.com

Liubov Kovriguina  
ITMO University  
St. Petersburg, Russia  
lkovriguina@gmail.com

Fedor Kozlov  
ITMO University  
St. Petersburg, Russia  
kozlovfedor@gmail.com

Olga Parkhimovich  
ITMO University  
St. Petersburg, Russia  
olya.parkhimovich@gmail.com

## ABSTRACT

The paper concerns estimation of students' knowledge based on their learning results in the ECOLE system. ECOLE is the online eLearning system which functionality is based on several ontologies. This system allows to interlink terms from different courses and domains and calculates several educational rates: term knowledge rate, total knowledge rate, domain knowledge rate and term significance rate. All of these rates are used to give the student recommendations about the activities he has to undertake to pass a course successfully.

## Categories and Subject Descriptors

I.2.4 [Knowledge Representation Formalisms and Methods]: Semantic networks

## General Terms

eLearning system; knowledge rating; semantic web technologies in education

## Keywords

Semantic Web; Linked Learning; terminology extraction; education; educational ontology population

## 1. INTRODUCTION

The ECOLE System<sup>1</sup> from the very beginning has been developed as the eLearning system having Semantic Web technologies in its core. They have been chosen because of its excellence in describing educational content. Besides, they also provide interoperability between modules of the system and access to already existing outer data. Students

<sup>1</sup><http://ecole.ifmo.ru/>

Copyright is held by the International World Wide Web Conference Committee (IW3C2). IW3C2 reserves the right to provide a hyperlink to the author's site if the Material is used in electronic media.

WWW 2015 Companion, May 18-22, 2015, Florence, Italy.

ACM 978-1-4503-3473-0/15/05.

<http://dx.doi.org/10.1145/2740908.2741743>.

of ECOLE system can join courses, study subject fields by reading books and lectures with videos, slides and text content. Students can check their knowledge by a passing the tests, practices and exams. Users of ECOLE are not only supplied by various educational content (lectures, videos, tests), but they are surrounded by the competitive environment with a friendly interface. Every student has a personal account where all his activities and results are gathered. He knows his Just-in-Time (JIT) learning results, can check his knowledge of any of the terms, look up his points and rank among the results of other users, get other statistics related to his learning.

Thus, students are engaged in a kind of game, provoking them to perform actions anyhow connected to the learning process. We see it a way to provoke knowledge acquisition, especially for the students indifferent to knowledge and learning.

Students' activity in the system is collected and analysed automatically to learn the system to give students recommendations on the terms, concepts or domains they know excellent or poor, advice them what educational program to choose and what scientific activities in the university are close to his best performed courses.

System educational content comes from various sources: linked open data, tests already published on the WEB, video lectures and tests manually uploaded by the university staff. Manually added content, as a rule, has terms assigned to it. Tests collected from the web are parsed and sent to the terminology extraction module inside ECOLE system, at its output each task from the test is linked to system terms[1].

## 1.1 Related work

Semantic technologies are widely used in education in developed countries. The most notable projects concerning the usage of Linked Data for education are Linked Universities<sup>2</sup> initiative, an alliance of European universities engaged into exposing their public data as Linked Data and the Open University<sup>3</sup>, supporting distance and open learning, - a research university with over 240.000 students.

An excellent example of using semantics to make education materials reusable and flexible is the SlideWiki sys-

<sup>2</sup><http://linkeduniversities.org/>

<sup>3</sup><http://www.open.ac.uk/>

tem[2]. The SlideWiki is an educational platform dealing with presentations and assessment tests.

Another initiative of using semantic technologies in education field is the mEducator. It is a content sharing approach for medical education, based on Linked Data principles. It enables educators and learners organise, re-purpose, re-use and share medical educational resources.[3].

## 1.2 Motivation

The major task in developing and maintaining ECOLE eLearning system is choosing and interlinking relevant materials, e.g. associating terms in lectures and tests. When data from different external resources are integrated to the course, they can reduce the quality of the course, i.e. content may become unbalanced. The way to maintain and improve the quality of the course is to use students' activity in the system - it can be helpful in recommending each student the appropriate material. In ECOLE any sophisticated statistics can be gathered, e.g. statistics about students' correct/incorrect answers, allowing to filter out troublesome terms and topics, number of student's attempts to pass the course, etc. Educators can use this statistics to screen unused or rarely used resources, add explanatory materials, improve the quality of his course altogether. The student can use personal statistics to fill his knowledge gaps. ECOLE seems to be quite a new eLearning environment where educational content is combined with highly personalized recommendations concerning individual learning strategy. Developers concentrate highly on providing students with detailed information concerning their knowledge and scores. During ECOLE development the following was done:

- developed student activity ontology to keep student rates and scores for every term and the whole subject;
- simulated knowledge acquisition process;
- simulated student knowledge rate of a single term;
- obtained knowledge rate of course terms when student has passed the test and send it to the student's profile;
- increased students knowledge rate after he has passed the test;
- calculated term "importance" in the domain (temporary basing on the links between the terms in the domain ontology);
- ranked domains according to the excellence of student's knowledge of a domain.

Current paper mainly concerns the problem of students' knowledge assessment which is a significant part of the education process as a whole[4],[5],[6],[7]. The method we propose is based on the analysis of student activity in the system. The data about student's activity refer to the subject fields through the learning content. The method uses predefined metrics for evaluation of student's knowledge of a specific domain.

To develop an architecture adequate to the set problems, 1) the ontology of knowledge rate has been developed, 2) the method of setting the initial knowledge rate of the term has been developed, 3) partial knowledge rate of domain terms was calculated basing on the statistics of attempts to pass the test, 4) total knowledge rate of domain terms has

been calculated using partial rate, 5) the weights of domain terms have been obtained using analysis of relations between domain terms and 6) the domain knowledge rate has been calculated using total knowledge rate of the domain terms and their weights.

## 2. ONTOLOGY DEVELOPMENT

### 2.1 Educational Ontology

Educational ontology, which could be divided into four modules, has been developed. Diagrams are created in OntoDia<sup>4</sup>. Each module represents a subject area: education (Fig. 1), tests, terms (Fig. 2) and student activity. Division into modules allows to reduce dependencies between entities and use only a part of the ontology.

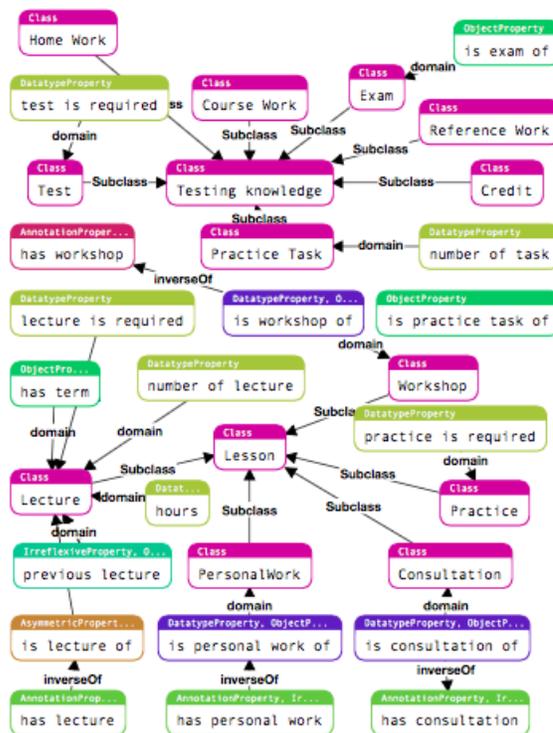


Figure 1: Ontology module of education

The ontology module of education describes relations between courses, modules, lectures and terms and helps to represent its properties and media content. It is built on top of top-level ontologies[8] such as AIISO<sup>5</sup>, BIBO<sup>6</sup>, TEACHNS<sup>7</sup> and MA-ONT<sup>8</sup>. The most outstanding feature of this ontology module is its ability to create direct and indirect interdisciplinary relations between courses[9]. E.g., physics test "Interference and Coherence" includes math terms as well ("vector", "vector product"). Thus, if a student can't pass this test, the system advices to repeat not only the

<sup>4</sup><http://www.ontodia.org>

<sup>5</sup><http://purl.org/vocab/aiiso/schema#>.

<sup>6</sup><http://purl.org/ontology/bibo/>.

<sup>7</sup><http://linkedscience.org/teach/ns/teach.rdf>

<sup>8</sup><http://www.w3.org/ns/ma-ont#>.

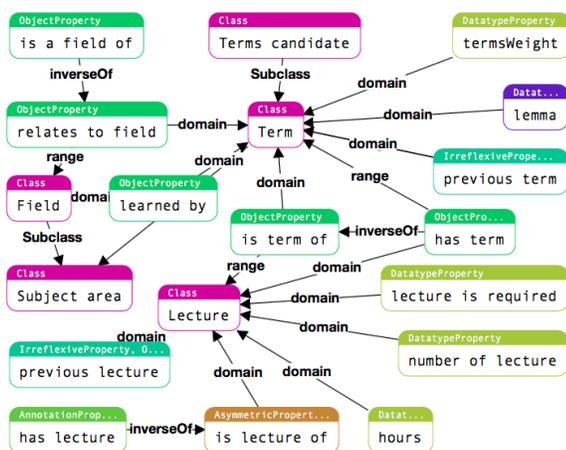


Figure 2: Ontology module of terms

lecture "Occurrence of Interference" in the "Physics" course, but also corresponding lectures from the "Vector algebra" course. This is an example of indirect links between physics and vector algebra via the domain terms "vector" and "vector product".

To describe tests' content ontology module representing the test structure has been developed. This module<sup>9</sup> has the following classes: Test, Testing Knowledge Item, Group of Tasks, Task, Answer, Question, Fill-in the Blank, Matching, Multiple Choice, Single Answer, Text Answer, True/False. The main purpose of the developed ontology module is to represent structural units of a test and provide automatic task matching by defining semantic relations between tasks and terms[10]. The ontology module has a class "Test" to store common test characteristics, e.g. its title and description, and class "Testing Knowledge Item" to describe test elements. The class "Testing Knowledge Item" has subclass "Task". The class "Group Of Tasks" [11] was added to group questions by parameters, e.g. by difficulty. The class "Task" has subclasses "Answer". The class "Question" has subclasses describing question types: "Fill-in the Blank", "Matching", "Multiple Choice", "Single Answer", "Text Answer", and "True/False". The class "Answer" has object properties "is wrong answer of" and "is right answer of". Using these two object properties except one data property "has answer" allows to use one set of answers for many questions.

The ontology module of student activity<sup>10</sup> has been developed to store information about the student's learning process and results. Two top-level ontologies have been used for its development: ontology module of tests, as described above, and FOAF ontology<sup>11</sup> that describes people and relationships between them.

The class "Learning process" stores information about every action performed by the student in the system. Students can watch the video (subclass "Video"), try to pass tests (subclass "AttemptToPassTest"), learn terms (subclass "Term") and pass a course (subclass "Course"). The ontology module also has class "Student" to store information

<sup>9</sup><http://purl.org/ailab/testontology>

<sup>10</sup><http://purl.org/ailab/learningresults>

<sup>11</sup><http://www.foaf-project.org>

about users and their activity in the system. This class is a subclass of class "Person" determined in FOAF ontology. The object properties "enrolled course", "finished course", and "subscribed course" describe relationships between the class "Student" and the class "Course". The class "Learning results" was added to store information about students educational activities and answers. Class "TestElement" contains information about "Task" (class of test ontology) and about student's "Answer" (subclass of class "LearningResults"), which can be correct or incorrect. A set of test elements constitutes attempt to pass the test. The properties "timestamp of attempt" and "percent complete of test" allow eLearning system to store information about the time in which an attempt was made and to determine the result of the test. The eLearning system uses the ontology of tests and answers given by the user to build a list of terms that the user knows.

## 2.2 Ontology module of knowledge rates

Knowledge rates ontology module (Fig. 3) is intended for keeping information about rates of term and domain knowledge rates for each student. Term's rate shows whether the student assimilated it. For example, if a student has watched or read the lecture with this term and has passed a test with this term successfully, we can consider, that student knows it. Ontology module contains class 'Rate' and 5 subclasses: 'Lecture Term Rate' computed as the number of lectures, containing this term and viewed by the student; 'Test Attempt Term Rate' keeping attempts to pass a test with this term and number of correct answers to the task with a term; 'Average Test Term Rate' based on average result of all attempts to pass one test or to pass all tests with this term; 'Total Term Rate' based on sum of rates of this term; 'Domain Rate' based on all rates of all terms from the domain student is learning. Each class contains data property 'value' to store numeric values of rates. Also ontology contains object properties which link rates to the class 'Student' from the educational ontology, 'Test' from test ontology and 'Term' from terms ontology. Ontology allows to add additional 'Rate' subclasses storing new metrics and change or add formulas to calculate.

With the described modules we retain all the data associated with the training of students. Let's begin with a general example. John Smith, our imaginary student, has started "Optics" course. This course contains lectures and practical tasks (from ontology module of education), several tests with different groups of tasks and information about wrong and right answers (from ontology module of tests), each task from tests and each lecture have terms, which were described in ontology module of terms. When the student tries to pass the test new individual of class "AttemptToPassTest" is created. When the student has solved the task his results are recorded in ontology module of knowledge rates. Based on wrong and right answers, metrics for terms that are checked in this tasks are compiled and changed, based on the formula, described below.

## 3. METHOD

### 3.1 Acquaintance with the term

It is supposed that at the start of his learning process the student does not know any domain terms. Therefore, a number of metrics is used to assess the student's knowledge.

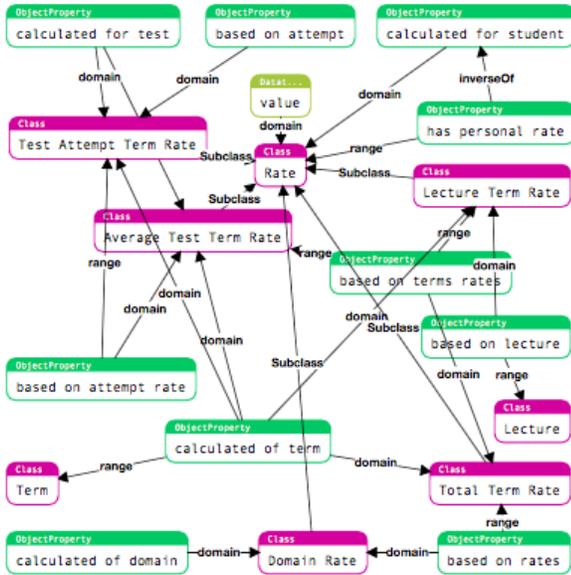


Figure 3: Ontology module of knowledge rates

To acquire some knowledge about a single term of the course our imaginary student John has to read or watch theoretical content associated to this term. The lecture he read contains subject terms *Vector Product*, *Interference*, *Light energy*. Term knowledge rate of these three terms became 0.15 after passing the lecture.

When the student reads the lecture for the first time, the system writes the fact of studying domain terms associated to this lecture to the ontology using relations between domain terms and lectures and assigns each of these terms with a 0.15 term knowledge rate. This rate can change later according to student’s results and activity in the system. Term knowledge rate of unstudied terms remains zero and student activity does not affect it. ECOLE tracks the statistics about the number of times the student watched a lecture but multiple lecture reading does not influence this initial rate. In the current implementation the system tracks only the first time when the student studied terms by reading the lecture. The ontology enables storage of the number of times the student addressed the lecture, because it provides data to understand how frequently the student faces studied terms and repeats them. In future this metrics will be used in making total knowledge rate calculation more precise.

### 3.2 Knowledge in the test

When the user has finished the test, the system sends the SPARQL Update Query[12] with his answers to the SPARQL-endpoint. When user statistics is gathered, objects having type "AttemptToPassTest" and user’s answers to the test’s tasks are created in the system. The object with type "AttemptToPassTest" is bound to hash data of user’s e-mail.

After test completion the information about amount of correct answers is displayed to the student. Also the list of domain terms for repeating is presented to the student. The system generates the list of embarrassing terms for the student using test results and relations between domain terms and tasks of the test. For each subject term of the test

Table 1: The calculation of partial term rate

| Task Number | Answer             | Impact on the rate |
|-------------|--------------------|--------------------|
| 1           | correct            | 0.515              |
| 2           | correct            | 0.128              |
| 3           | correct            | 0.057              |
| 4           | incorrect          | -0.032             |
| 5           | correct            | 0.021              |
| 6           | incorrect          | -0.014             |
|             | Total partial rate | 0.676              |

the system counts the partial term rate based on student’s answers. The list is sorted using knowledge rate of each subject term. Partial term rate can help the student to find his knowledge gaps and eliminate them.

Partial term rate is calculated using series as a model of student’s term knowledge. We assume, that a student refines his knowledge in the process of learning so scores for each correctly answered term are generated dynamically using the series

$$R_t = \sum_{n=1}^{\infty} \frac{0.515}{n^2}$$

This means, when the student gives correct answer for the term the first time ( $n=1$ ), he is given 0.515 for this term, for the second correct answer he is given 0.132 and his partial term rate is 0.643, etc.

Imagine, John Smith passed the test on optics, which includes 5 tasks with the term *Interference*. He answered correctly the tasks number 1, 2, 3, 5 and gave incorrect answers to the tasks 4 and 6. Calculation of partial term rate is given in table 1. Scores and charges correspond to the partial sum of the series. If all answers were correct, John Smith would get partial term rate 0.768.

Using series is determined by two reasons. Firstly, student’s score after passing the test can not exceed 0.85. This will encourage student to read theoretical material if he has skipped it previously. Moreover, being good in passing tests does not imply students good term understanding. Secondly, this is the remedy from dishonest students who may wish to drive up the score by passing the test endlessly. Thirdly, educational model assumes that a student can give incorrect answers and he is charged for that. This charge is subtracted from the partial term rate and it gets less and less ‘heavy’ with the number of correctly answered term because the charge is calculated as the  $N$ th term of the series.

After passing the test detailed results are displayed to the student. These scores do not interfere with term knowledge rate and show the percent of correct answers for each term in respect to the number of tasks on this term in the test. It informs the student about his results in a separate test. The user interface of the test result page is shown in Fig. 4.

### 3.3 Total knowledge rate

The statistics of student’s activity in the system and his learning results allows to formulate a list of metrics for student knowledge evaluation for every domain presented in the system. The list of metrics includes 1) answers to the tasks of the tests, 2) passing the practice, 3) learning theoretical material, 4) number of attempts to pass the test, etc. The total knowledge rate for each subject term is stored in the

| Tests / Test Of Interference And Diffraction Frenet |          |
|---|----------|
| Results of the test                                 | 65 / 100 |
| Your best result                                    | 60 / 100 |
| Subject terms for repeating                         |          |
| Wavefront   | 0        |
| Newton's Rings                                      | 0        |
| The interference pattern                            | 20       |
| Intensity   | 50       |
| Interference  | 50       |
| Diffraction   | 50       |
| Fringes   | 75       |
| Wavelength  | 75       |
| Light   | 83       |
| Spatial coherence                                   | 100      |

Figure 4: The user interface of the test result page

student profile. The system tracks activity of the students and stores it into the knowledge base. The classifier analyzes all the activity of the student and, given the activities, increases or decreases the total knowledge rate of the certain subject term.

In this paper calculation of two metrics is mentioned: familiarity with the term (Section 3.1) and the partial term rate (Section 3.2). When the student passed the test, the system collects the partial knowledge rates for each domain term linked with this test. After that the system counts the total knowledge rate of this domain terms adding 0.15 to the partial rate if the student read the lecture before tackling the test.

### 3.4 Term importance

On the first level of ECOLE data model the data about subject fields, domain terms and relations between them are stored. The semantic relation *previous term* is specified for the terms which should be learned before learning the current term. For some terms these relations coincide with their hyponym-hypernym relationship but they concern guidance through the education process more than taxonomic relations. Terms and their relations are independent from the educational material and are reused in several courses (domains). Its change does not influence other levels. Educational material such as courses, lectures or tests is stored on the second level of the data model being linked to the domain terms. This relations provide an opportunity to evaluate the student knowledge using analysis of links between the entities of the ontology.

Term is the object that describes domain concepts. The examples of such objects are "Vector", "Matrix", "Pythagorean Theorem". Domain terms are interlinked. In this paper we consider only the number of links between the terms to calculate their significance for the acquisition of the educational material. For example, the domain term "Collinearity" depends on domain term "Vector" since the knowledge about "Vector" is necessary to study "Collinearity". The domain structure is represented by the graph of domain terms.

Using the links between domain terms the system counts the significance (its quantitative correlate) of each term in learning process. The higher the significance of the domain term, the more domain terms depend from it. The knowledge of domain terms with large significance is necessary for successful completion of the course. Knowledge of domain terms with small significance is necessary to refine ones

knowledge and get the highest scores because it has little effect on the course completion.

The system recalculates terms significance automatically by the predefined schedule. Every term is characterized by 3 dependency levels: number of terms that depend on it directly (child), depend from the directly depending terms (sub-child) and depend from the depending from directly depending (terms sub-sub-child).

The term significance is count by the formula:

$$W_t = \frac{1}{e} + \sum_{k=1}^3 \frac{n_k}{e^k}$$

where  $n$  - number of terms linked to this term on each of dependency levels,  $k$  - number of dependency level,  $\frac{1}{e}$  - a constant.

$\frac{1}{e}$  was introduced to avoid zero term significance for the terms that have no children. This formula has to be normalized to compare significance of terms from different domains.

### 3.5 Domain Knowledge rate

This rate is calculated by multiplying term significance in the domain and term knowledge rate. Before its calculating, relative ratio in the specific domain is calculated for every term considering its weight. Firstly, sum of term weights in the domain is calculated

$$W_f = \sum_{n=1}^{\infty} W_t$$

where  $W_t$  - domain significance,  $n$  - number of terms in the domain.

Term relative ratio is calculated using the formula

$$D_t = \frac{W_t}{W_f}$$

where  $W_t$  - term significance,  $W_f$  - term weight.

Domain knowledge rate is the sum of term rates multiplied by its relative ratio:

$$R_f = \sum_{n=1}^{\infty} R_t D_t$$

where  $R_t$  term rate.

## 4. RESULTS AND DISCUSSIONS

The paper describes a novel approach to analyze students' knowledge in the eLearning system using Semantic Web and Linked Data. The presented method allows to obtain several knowledge rates which deliver the student just-in-time results about how his (un)successful his education is: he can see how good is his knowledge of any term or domain on the whole, what terms he has to repeat, etc.

Term significance was calculated for the optics domain (it contains 43 concepts now). Five terms with the highest term significance rate in optics are given in table 2. Obtained results show term importance in the learning process rather than its importance in the optics terminological system.

Future work plans include empirical data obtaining and analysis of students' experience with the system which is possible only after the academic term completes (about May,

**Table 2: The term significance rate in optics**

| Domain Term         | Level 1 | Level 2 | Level 3 | Rate |
|---------------------|---------|---------|---------|------|
| Interference        | 6       | 9       | 2       | 3.52 |
| Interference effect | 7       | 2       | 0       | 2.85 |
| Wave process        | 3       | 8       | 9       | 2.63 |
| Illumination        | 1       | 6       | 9       | 1.63 |
| Light               | 1       | 6       | 9       | 1.63 |

15). Testing and overall analysis of the proposed model on students will inevitably arise its changes and improvements.

Anyhow, there are already some rather significant things to be improved. Firstly, domain ontology needs refinement and unification, because it has been developed by different people: some domains are more detailed, the last increases the density of term links and influences term weight. To improve the quality of term significance calculation domain ontology has to be integrated with external resources, i.e., Mathematics Subject Classification (MSC)[13]. Secondly, broad experiments with learners are critical for quality improvement: new metrics can be added and learning strategies can be revealed bases on the analysis of users statistics.

Thirdly, a list of new metrics is ready to be added: a) Test-depending

- number of correct answers to the term;
- number of attempts to pass the task;

b) Lecture-depending:

- number of times student addressed the lecture;
- time expense to read the lecture;
- total time of lecture reading;

c) Term-depending:

- number of lectures linked to the term;
- number of tests linked to the term;
- number of domains having the term;
- number of domains having terms depending from the term;

To conclude plans for the future work, ECOLE was designed as a system capable of giving recommendation as to the student, what extra courses he needs to pass to get the best scores at the exams, as to the university scientists, informing them which students they can choose for scientific work. Future work will be connected to broaden the recommendation force of ECOLE.

Acknowledgements: *This work was partially financially supported by the Government of Russian Federation, Grant 074-U01.*

## 5. REFERENCES

- [1] Dmitry Mouromtsev, Fedor Kozlov, Liubov Kovrignina, and Olga Parkhimovich. A combined method for e-learning ontology population based on nlp and user activity analysis. *CEUR Workshop Proceedings: LILE 2014 Workshop Proceedings: Linked Learning Meets Linked Up: Learning and Education with the Web of Data*, 1254(5), 2014.
- [2] Ali Khalili, Sören Auer, Darya Tarasowa, and Ivan Ermilov. Slidewiki: elicitation and sharing of corporate knowledge using presentations. In *Knowledge Engineering and Knowledge Management*, pages 302–316. Springer, 2012.
- [3] Maurice Hendrix, Aristidis Protopsaltis, Ian Dunwell, Sara de Freitas, Panagiotis Petridis, Sylvester Arnab, Nikolas Dovrolis, Eleni Kaldoudi, Davide Taibi, S Dietze, et al. Technical evaluation of the meducator 3.0 linked data-based environment for sharing medical educational resources. In *2nd International Workshop on Learning and Education with the Web of Data, Lyon, France*, volume 4, 2012.
- [4] Eileen O’Donnell, Mary Sharp, Vincent Wade, and Liam O’Donnell. Personalised e-learning: The assessment of students prior knowledge in higher education. 2014.
- [5] Sanja Maravić Čisar, Dragica Radosav, Branko Markoski, Robert Pinter, and Petar Čisar. Computer adaptive testing of student knowledge. *Acta Polytechnica Hungarica*, 7(4):139–152, 2010.
- [6] Anna N Rafferty, Jodi Davenport, and Emma Brunskill. Estimating student knowledge from paired interaction data. *Proc. EDM*, 2013.
- [7] Krzysztof Gierlowski and Krzysztof Nowicki. A novel architecture for e-learning knowledge assessment systems. *International Journal of Distance Education Technologies (IJDET)*, 7(2):1–19, 2009.
- [8] Carsten Keßler, Mathieu d’Aquin, and Stefan Dietze. Linked data for science and education. *Semantic Web*, 4(1):1–2, 2013.
- [9] Dmitry Mouromtsev, Fedor Kozlov, Olga Parkhimovich, and Maria Zelenina. Development of an ontology-based e-learning system. *Knowledge Engineering and the Semantic Web*, pages 273–280, 2013.
- [10] Larisa Soldatova and Riichiro Mizoguchi. Ontology of test. *Proc. Computers and Advanced Technology in Education*, pages 173–180, 2003.
- [11] Réka Vas. Educational ontology and knowledge testing. *The Electronic Journal of Knowledge Management of*, 5(1):123–130, 2007.
- [12] Paul Gearon, Alexandre Passant, and Axel Polleres. Sparql 1.1 update. *Working draft WD-sparql11-update-20110512, W3C (May 2011)*, 2012.
- [13] Christoph Lange, Patrick Ion, Anastasia Dimou, Charalampos Bratsas, Joseph Corneli, Wolfram Sperber, Michael Kohlhase, and Ioannis Antoniou. Reimplementing the mathematics subject classification (msc) as a linked open dataset. In *Intelligent Computer Mathematics*, pages 458–462. Springer, 2012.