Semantic Web Game Based Learning: An I18n approach with Greek DBpedia*

Charalampos Bratsas
Web Science Program
School of Mathematics
Aristotle University of Thessaloniki
cbratsas@math.auth.gr

Despina Eleana Chrysou
Web Science Program
School of Mathematics
Aristotle University of Thessaloniki
chrysoue@math.auth.gr

Aikaterini Eftychiadou
Web Science Program
School of Mathematics
Aristotle University of Thessaloniki
eftichiadou@math.auth.gr

Dimitris Kontokostas
Web Science Program
School of Mathematics
Aristotle University of Thessaloniki
jimkont@math.auth.gr

Panagiotis Bamidis
Lab of Medical Informatics
Medical School
Aristotle University of Thessaloniki
bamidis@med.auth.gr

Ioannis Antoniou
Web Science Program
School of Mathematics
Aristotle University of Thessaloniki
iantonio@math.auth.gr

ABSTRACT

Recent web advances and progress in technology enhanced pedagogies have outlined the important role of gaming within education. The advent of Semantic Web and Linked Data, as well as, the availability of initiatives and infrastructures like that of DBpedia have facilitated the incorporation of gaming within educational knowledge bases. This paper presents a web game that uses datasets derived from Greek DBpedia. It is consisted of several different quiz types and it is used for educational purposes in native speakers. The application has a preliminary evaluation in primary education settings in north Greece. The evaluation results indicated its potential in the learning process.

Categories and Subject Descriptors

J.1 [**Administrative Data Processing**]: Education

Keywords

DBpedia, E-Learning, Semantic Game, LOD, SPARQL, I18n

1. INTRODUCTION

Playing is and should be a part of a child's daily activity and many efforts have been made to create educational activities through games. In particular, games, dating back thousands of years have been used to augment the educational practice. The last decades, the computer software industry generated a new approach to implementing educational games. Programming in particular, offers a structured way for developers to generate a game workflow, making it suitable for generating educational scenarios. Recently, the evolution of the web and the mass networking gave a boost to the education industry, by offering a huge amount of data and potential educational resources. Key indicator of the latter is the world's largest, multilingual crowd-sourced encyclopedia, that is Wikipedia. The English Wikipedia alone has almost 3.9 million articles[1] with various topics. Wikipedia's scientific articles accuracy can be compared to encyclopedia Britannica [6].

The advent of Semantic Web and Linked Data (LD) [1] can take educational software a step further. Linked Data provide a structured way for developers to access knowledge. Since the beginning of the Linking Open Data project[2], this knowledge seemed to be cultivating and the e-learning opportunities started to accumulate slowly but surely. A classic LD example is DBpedia. DBpedia extracts knowledge from Wikipedia articles. Since Wikipedia articles can be considered accurate, the extracted knowledge can be also considered of good quality. The recent DBpedia internationalization effort [8] gave rise to new language specific DBpedia editions by extracting knowledge form local Wikipedia language editions. The new internationalized DBpedias can create localized knowledge; that is of great importance, especially for non-native English speakers.

The application presented in this paper, uses the Greek DBpedia endpoint to automatically generate quiz games for native speakers. Using the recent advances in the field (cf. section 2), the application was designed with internationalization in mind. The proposed architecture of the application is discussed in section 3, and a pilot demo[3] was implemented. The educational value of this tool is examined in section 4 where we implement a preliminary evaluation in a primary school. The evaluation results indicated its potential in the

*Linked Learning 2012, 17 April 2012, Lyon, France.

[1]http://en.wikipedia.org/wiki/Wikipedia:Size_of_Wikipedia

[2]http://www.w3.org/wiki/SweoIG/TaskForces/CommunityProjects/LinkingOpenData

[3]http://el.dbpedia.org/en/blog/2012/02/03/semantic-e-learning-game

learning process.

2. STATE OF THE ART

The DBpedia project [2] - as part of the Linking Open Data project - extracts structured data from Wikipedia and publishes them as Linked Data. The knowledge extraction is achieved using the DBpedia Information Extraction Framework (DIEF). The DIEF generates an entity for every article and uses numerous extractors to retrieve specific information and assign it back to the entity. The richest source of structured information is extracted from Wikipedia infobox templates[4]. DIEF uses the infobox instances defined in every Wikipedia article to classify the article (entity) to a DBpedia ontology class. Then, the article infobox template parameters are associated with the corresponding DBpedia ontology properties (with expected datatypes). The DBpedia ontology and the mappings are created with a controlled crowd-sourced approach using an open and collaborative Mappings Wiki[5] [7].

Wikipedia's knowledge, structured in DBpedia, offers the best candidate dataset for educational games. There exist two applications that use linked data for automated generation of quiz-like questions. The *Who Knows* game, introduced by N. Ludwig et al., presents questions generated out of DBpedia RDF data [10]. Who Knows is an existing facebook application[6] that - according to [10] not only offers an opportunity for entertainment and learning, but also a method for evaluating linked data to clean up DBpedia. A similar effort was also presented by M. Foulonneau, describing "the streamline to create variables and populate simple choice item models using the IMS-QTI standard" [4]. Both applications use the English DBpedia as their sole source.

The recent DBpedia internationalization effort, starting with the Greek DBpedia [8], gave a new perspective for educational games. The internationalized DBpedias share the same DBpedia ontology, an ontology that can be resembled as "a common framework to semantically annotate and integrate infoboxes across different Wikipedia language editions". Thus, an effective way to re-use queries among different DBpedia language editions. By using a carefully designed architecture, one can easily generate queries in multiple languages, which can be proved beneficial for non-native English speakers.

3. ARCHITECTURE OF THE SEMANTIC GAME

Using the Greek DBpedia as a localized Linked Data source, we created a prototype application that automatically generates Quiz-like questions. This application has three core components (cf. Figure 1):

1. *Fact Engine*: Queries a DBpedia endpoint (i.e. `http://el.dbpedia.org/sparql`) to retrieve facts
2. *Query Engine*: Processes the output of the *Fact Engine* to generate queries

[4]`http://en.wikipedia.org/wiki/Help:Template`
[5]`http://mappings.dbpedia.org`
[6]`http://apps.facebook.com/whoknows_/`

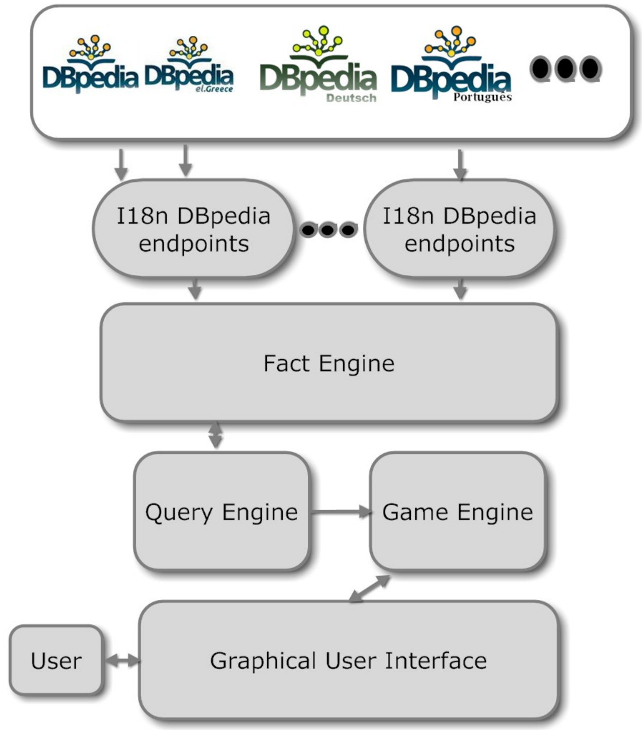

Figure 1: Semantic Web Game Based Learning Architecture.

3. *Quiz Engine*: Communicates with the *Query Engine* to randomly generate a quiz for a user.

As Figure 1 depicts, the *Fact Engine* (cf. subsection 3.1) communicates with an I18n DBpedia endpoint to generate facts. The *Query Engine* (cf. subsection 3.2) is responsible to select a fact for a question, while the *Quiz Engine* (cf. subsection 3.3) renders a question for a user and collects user feedback.

3.1 Fact Engine

The Fact Engine uses the SPARQL Query Language [9] to access data from DBpedia. A total of 30 SPARQL queries were created, that retrieve facts belonging to 8 general categories: geography, history, athletics, astronomy, general, chemistry, politics, and economy. These facts form the basis of the next processing steps of the application. A sample query is depicted in Listing 1, retrieving flag images for countries, along with their depiction and feedback links for the DBpedia and Wikipedia pages. Although this query was run in the Greek DBpedia endpoint, by carefully choosing triples adhering to the DBpedia ontology, a cross-DBpedia SPARQL query can be achieved.

Even with a relatively small number of 30 queries, we managed to produce a total of 12,000 sets of results from Greek DBpedia (cf. Table 1). For the implementation phase of our application (cf. section 4), the queries were limited to 20, generating a total of 3,204 facts.

```
PREFIX ontology: <http://dbpedia.org/ontology/>
PREFIX rdf:
<http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX rdfs:
<http://www.w3.org/2000/01/rdf-schema#>
PREFIX foaf: <http://xmlns.com/foaf/0.1/>

select distinct ?country ?flag ?dbpedia ?wiki
where {
  ?dbpedia  rdf:type ontology:Country;
  rdfs:label ?country.
  ?dbpedia ontology:thumbnail ?flag.
  OPTIONAL {?dbpedia foaf:page ?wiki.}
} ORDER BY ASC(?country)
```

Listing 1: Sparql Query retrieving image flag facts for countries.

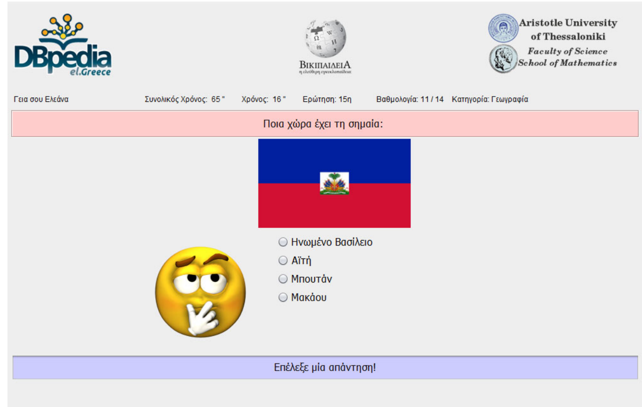

Figure 2: A multiple choice qustion.

3.2 Query Engine

The Query Engine randomly selects a fact belonging to a category - generated by the *Fact Engine* - and forwards it to the *Quiz Engine*. It also functions as a feedback generator, by providing the feedback links to Wikipedia and DBpedia for the *Quiz Engine*. The Query Engine, depending on the rendered question type, can be used to provide distractors (cf. subsection 3.3). Distractors are also facts, used to provide similar but wrong answers to a question i.e. in the case of a multiple choice question.

3.3 Quiz Engine

The *Quiz Engine* (QE) is responsible for creating a complete game for an end-user. QE controls the user game progress, keeps statistics and renders questions according to the available question types. Currently there are three question types implemented: multiple choice, matching and hangman. Depending on the question type a different strategy is implemented for the question rendering.

To produce a multiple choice question, a set of four triples is needed. Two triples contain different properties of the question fact: one used to form the question and another for the answer (i.e. "country" and "flag" from Listing 1). A DBpedia and a Wikipedia link are also used as user feedback links. Furthermore, a number of distractors is also needed; they are generated by randomly selecting other responses to the same query. Figure 2 presents a multiple choice quiz containing the question: "Which country has the following flag:", followed by a specific country's flag. The choices provided are: United Kingdom, Haiti, Bhutan, and Macau. Haiti is the correct answer, and the rest countries are used as distractors.

The set of results could produce questions in different directions, i.e. the image flag query (cf. Listing 1) could produce either "which country has the following flag" or "which flag belongs to country:". Since this can increase the questions variations, a randomly selected direction is chosen in every question rendering.

MCQ, in its simple form, consists of one correct answer and up to three distractors (cf. Figure 2). To make the game more challenging, MCQ was extended with two more variations: a) with an anagram quiz and b) with multiple correct answers. The anagrammed MCQ is a combination of a multiple choice and an anagram quiz, where the user has to discover the answer through a number of anagrammed responses. Anagrams are created on-the-fly and are removed for user feedback. In the multiple correct answers variation the number of the possible correct answers or distractors can vary from one to five. Then, a mechanism randomly produces a number of correct or incorrect answers.

In the matching question type, the query engine returns up to four facts. For example, for the question: "Which language is spoken in the country:" we retrieve a set of four countries and four languages. The countries are randomly positioned on the left column, allowing the user to select their language from a drop down menu on the right column. Each country can be matched in a unique way.

The Hangman question type is a port of the well-known hangman game (cf. Figure 3). This question type is very simple, as only one fact is retrieved and no distractors are necessary.

At the end of the game, a file with all questions and correct answers, the time spent and results is created for future

Figure 3: Hangman with feedback links.

Table 1: Datasets

#	Question	Category	Facts	Used	Facts Used
1	Countries-Flags	Geography	268	YES	249
2	Capitals-Images	Geography	196	YES	68
3	Cities-Images	Geography	38	YES	17
4	States of America-Flags	Geography	51	YES	50
5	Cultural Heritage-Images	Geography	183	YES	152
6	Archaelogical Sites-Images	History	67	YES	48
7	Monuments-Images	Geography	115	NO	0
8	Castles-Images	Geography	8	NO	0
9	Towers-Images	Geography	19	NO	0
10	Sports-Images	Athletics	39	YES	31
11	Olympic Sports-Images	Athletics	25	YES	20
12	Planets-Images	Astronomy	8	YES	6
13	Animals-Images	General	94	YES	59
14	People-Images	General	2,964	NO	0
15	Countries-Capitals	Geography	136	YES	133
16	Greece's Administrative Areas-Capitals	Geography	42	YES	40
17	States of America-Capitals	Geography	14	YES	14
18	Castles-Countries	Geography	4	NO	0
19	Islands-Population	Geography	285	NO	0
20	Chemical Elements-Symbols	Chemistry	96	YES	96
21	Space Missions-Launch Dates	Astronomy	49	NO	0
22	Wars/Battles-Dates	History	186	YES	161
23	People-Date of Death	General	746	NO	0
24	People-Place of Birth	General	3,152	NO	0
25	Cultural Heritage-Countries	Geography	179	YES	176
26	Politicians-Offices	Politics	562	YES	496
27	Countries-Languages	Geography	324	YES	320
28	Countries-Currencies	Economy	216	YES	211
29	Islands-Countries / Location	Geography	866	YES	857
30	People-Description	General	1058	NO	0
	Total		11,990		3,204

statistics reference.

4. IMPLEMENTATION

To determine the effectiveness of the game in authentic educational settings, a first pilot implementation was conducted during the academic school year 2011-2012. The game was implemented to a primary state school of *Neapoli Thessalonikis* (urban area). The participants were fifty four (54) students, 36 boys and 18 girls, attending the fourth, fifth and sixth grades of the primary school, aged between 10 and 12 years. For the purposes of the study they were divided into two distinct age groups (10-11 and 11-12). The school was chosen mainly due to easy access purposes.

The implementation of the game aimed at fostering learners' critical thinking and discovery learning while revising material taught in the conventional classroom. Moreover, emphasis was given to motivate young learners to use computers and enhance enjoyment.

4.1 The Game

To meet the needs of the particular teaching situations, specific adjustments were made to the game presented in section 3. Thus, the application was customized to generate 25 random questions, including a variety of categories, namely geography, history, athletics, astronomy, general knowledge, chemistry, politics, and economy. Taking into consideration the subjects taught in the last two grades, as well as the aims and objectives of the curriculum (DEPPS)[7] which advocates the integration of ICT, the particular subareas were selected as it was considered necessary to integrate the game into the curriculum. However, the questions were not only based on lessons previously covered but could be used as warm-ups for the teaching sessions to follow.

According to Gagne [5], the game was designed to have a colourful and pleasant interface, hence, emoticons and photos were added to capture young learners' attention. What is more, extra information regarding the correct answers was provided through links to Wikipedia and DBpedia.

4.2 Process

The game was conducted during 2 teaching sessions in the computer lab of the school. Each class consisted of 27 students. Learners were divided in pairs. The lesson followed a three phase framework, including a *warm-up phase*, a *game phase* and an *evaluation phase*.

The main aim of the *warm-up phase* is to familiarize learners with the process and activate background schemata. There-

[7] `http://www.pi-schools.gr/programs/depps/`

fore, students were asked to recognize the Wikipedia logo and provide information regarding its use. The majority of children were aware of Wikipedia as they had already used the wiki while seeking information for their homework assignments. After presenting some general information concerning Wikipedia (i.e. purpose and effective use) the game was introduced to the classroom. Detailed instructions were provided, followed by a short demonstration of the game. Furthermore, emphasis was given on the feedback links of Wikipedia and DBpedia presented after each question, to ensure that learners will use them throughout the game.

During the actual *game-phase*, learners had the opportunity to play the game in pairs, where they played in a row. The customized version included all the question types described in subsection 3.3 (multiple choice, hangman, and matching). The learners seemed to highly enjoy the whole process, helping and supporting each other.

Finally, during the third stage of the lesson, an evaluation of the effectiveness of game was attempted. A self-assessment questionnaire comprising 20 questions was administered to the students. The questionnaire aimed at detecting the general attitudes and stances of the learners regarding the use of the particular game. The questionnaire consisted of 17 close-ended questions and 3 open-ended questions to invite general comments. Regarding the format of the close-ended questions a four point Likert scale format was followed, since an even number of responses is considered to be more appropriate for young learners not allowing them to use the middle category to avoid making a real choice [3]. In order to ensure honest responses the questionnaire was completed anonymously.

The vast majority of the learners evaluated in a positive way the game. More specifically, 85% of the respondents answered that they enjoyed playing the game very much while 44% characterized the game as easy. As far as the questions are concerned, 78% of the respondents stated that the questions were interesting while 81% did not encounter any problems in understanding them. When asked to indicate the type of the quiz they liked the most the majority of the learners (63%) chose the Hangman questions, 19% expressed their preference over the multiple choice questions (simple form) and 15% for the matching quiz.

Regarding the interface of the game, almost half of the respondents, 48% stated that they loved the interface of the game and 26% expressed positively without suggesting any further changes. However, some of them suggested more colours and pictures, in order to make it more appealing.

A good majority (81%) of the kids found the link to Wikipedia and DBpedia really useful. A total of 56% of the kids followed that link more than once, to find and learn more information about the correct answer. Finally, 70% of the kids believe that the link itself increases motivation to look for further information around the correct answer.

In the question of whether the game helped them to find out and learn the correct answer (knowledge acquisition), results were very encouraging indeed: 37% answered "very positively" while another 44% just "positively". 33% answered that all questions were within their cognitive background. 22% answered that the questions were within the limits of their knowledge. Finally, 85% could hardly wait to play the game again.

Qualitative observations over children running the application, as conducted by the evaluation team, revealed reactions matching logical thinking, the joy upon answering correctly and in the appearance of the happy emoticon, as well as the opposite when they saw the sad one and their declaration to respond properly next time; all these elements form the belief that kids enjoyed the process and the game overall, while the engaged in a potentially useful learning experience.

5. CONCLUSION AND FUTURE WORK

The implementation described in section 4 was a preliminary evaluation for our pilot application and the student's comments could form a valuable feedback for user-interface and architecture future improvements. In regard to the user-interface, improvements could take place in user interaction and graphics design to make it more appealing and extend the question range with more categries. Moreover, the application could better target different platforms (i.e. mobile phones and tablets).

An important issue that we had to deal with is the incorrect data the application could retrieve from DBpedia. A completely automatic generation of questions would lead to some inconsistencies that would confuse the players. A query evaluation system should be created that would not only filter the quiz questions but, feedback errors back to Wikipedia as well. Furthermore, other Linked Data sourced (i.e. freebase[8]) could also be used for the *Fact Engine* (cf. subsection 3.1).

The application was designed with internationalization in mind. Although primary testing of the same questions with other languages indicate our approach was correct, this needs thorough investigation, and is definately a future work plan. Extending this further, multi-language quiz could be created that will foster foreign language learning.

Immediate future plans of the development team, are associated with evaluation of the games by different age groups, including adults and releasing the application as open-source. A likely extension of the game to fit tertiary and continuing medical education will also be sought.

6. ACKNOWLEDGMENTS

The administrative and financial support by the municipality of Veria is gratefully acknowledged. This work was also supported by the project mEducator (Multi-type Content Sharing and Repurposing in Medical Education), funded by the eContentplus Programme, a multiannual Community programme to make digital content in Europe more accessible, usable and exploitable (Grant ECP 2008 EDU 418006)

7. REFERENCES

[1] C. Bizer, T. Heath, and T. Berners-Lee. Linked data - the story so far. *Int. J. Semantic Web Inf. Syst*, 5(3):1–22, 2009.

[8] `http://www.freebase.com`

[2] C. Bizer, J. Lehmann, G. Kobilarov, S. Auer, C. Becker, R. Cyganiak, and S. Hellmann. Dbpedia - a crystallization point for the web of data. *Web Semantics: Science, Services and Agents on the World Wide Web*, 7(3):154 – 165, 2009.

[3] Z. Dornyei. *Questionnaires in Second Language Research: Construction, Administration, and Processing*. Lawrence Erlbaum Associates, Publishers, 2003.

[4] M. Foulonneau. Generating educational assessment items from linked open data: The case of dbpedia. In R. Garcia-Castro, D. Fensel, and G. Antoniou, editors, *ESWC Workshops*, volume 7117 of *Lecture Notes in Computer Science*, pages 16–27. Springer, 2011.

[5] R. M. Gagne, W. W. Wager, K. Golas, and J. M. Keller. *Principles of Instructional Design*. Wadsworth Publishing, 2004.

[6] J. Giles. Internet encyclopaedias go head to head. *Nature*, 438(7070):900–901, 12 2005.

[7] S. Hellmann, C. Stadler, J. Lehmann, and S. Auer. DBpedia live extraction. In *Proc. of 8th International Conference on Ontologies, DataBases, and Applications of Semantics (ODBASE)*, volume 5871 of *Lecture Notes in Computer Science*, pages 1209–1223, 2009.

[8] D. Kontokostas, C. Bratsas, S. Auer, S. Hellmann, I. Antoniou, and G. Metakides. Internationalization of linked data: The case of the greek dbpedia edition. *Web Semantics: Science, Services and Agents on the World Wide Web*, (0):–, 2012.

[9] E. Prud'hommeaux and A. Seaborne. SPARQL query language for RDF, 2008.

[10] J. Waitelonis, N. Ludwig, M. Knuth, and H. Sack. Whoknows? evaluating linked data heuristics with a quiz that cleans up dbpedia. *Interact. Techn. Smart Edu.*, 8(4):236–248, 2011.