Enhancing e-Learning content by using Semantic Web technologies

Herminio García-González, José Emilio Labra Gayo, and MPuerto Paule-Ruiz

Abstract—We describe a new educational tool that rely on Semantic Web technologies to enhance lessons content. We conducted an experiment with 32 students whose results demonstrate better performance when exposed to our tool in comparison with a plain native tool. Consequently, this prototype opens new possibilities in lessons content enhancement.

Index Terms—e-Learning, Semantic Web, didactic effectiveness, learning management system, semantic enhancement, enriching learning content

✦

1 INTRODUCTION

E-Learning has supposed a huge advance in learning environments allowing educational community to rely on new technologies to give an improved experience and empower their students with better materials [1]. In this new era of learning, new learning environments have arisen such as Learning Management Systems (LMS) which enable users to share contents, create courses, collaborate with each other through forums or wikis, create and fulfil assignments, give and receive feedback and some others. They have been integrated in many universities as part of courses and degrees and many students and teachers are, nowadays, familiar with them. Nevertheless, with these novel tools new challenges arise. Among the diverse changes that may be covered in this area, we will focus on Semantic Web and content enhancement. Teachers contents on e-Learning platforms are contributing to enhance the knowledge of the attendants. But related with this main content there is more information that can be emerged using the appropriate tools. For example, if some content is mentioning Obama, a student may be wondering who is Obama or confused if Obama is mentioned in various ways (e.g. Barack, Obama, Barack Obama, B. Obama or even Barack Hussein Obama II). This problem is derived from the lack of semantics in the uploaded content. Our proposal is to take advantage of Semantic Web in order to: provide more information about outstanding entities, reconcile entities and enrich pages with RDFa[1] (Resource Description Framework in Attributes) and microformats. The main contribution of this work is a new technology that uses a set of Semantic Web techniques to complement and expand the learning courses content. This technology allows to enhance learning content hosted at LMS, favouring the increment of courses didactic effectiveness [2] as this work states.

2 RELATED WORK

The most similar architecture to that shown in our work is presented by [3], where authors show an architecture to enhance government data and then publish these enhanced data as Linked Data. An enhancement centred on museums was reported by [4], where authors use Semantic Web to link and add contents to museum objects. In [5], the authors propose an enhancement of user-generated content using geospatial Linked Open Data to improve tagging of Social Media platforms, like Facebook. The use of ontologies to recommend new personalised contents to the students depending on their fails and progress, is described in [6]. Enhancement for media management systems including videos, images and articles is described in [7] where they used a Red Bull Content Pool for the demonstration. Using Semantic Web for interactive Relationship discovery is addressed in [8] where authors highlight its use in technology enhanced learning. In [9], the authors use Web Semantic mining techniques to provide different personalised e-Learning experiences. A use of Web Semantic to discover and share content in OpenCourseWare environments is described in [10]. Ontologies as a way for describing content, for defining learning material and for structuring learning material is presented in [11]. Annotating videos with Linked Open Data (LOD) vocabularies and therefore improve search of educational videos is described in [12].

Content enhancement has also been performed using adaptative techniques from the Adaptive Hypermedia proposed in [13] with differents approaches like the creation of adaptative languages [14] [15] [16] or using learning objects [17] [18].

3 PROPOSED PROTOTYPE

We have developed a prototype called LODLearning to enhance lessons contents within LMS tools. Enhancements

• _Herminio García-González is with the Department of Computer Science, University of Oviedo, C/Calvo Sotelo, S/N 33007 Oviedo, Spain._
  E-mail: herminiogg@gmail.com
• _José Emilio Labra Gayo is with the Department of Computer Science, University of Oviedo, C/Calvo Sotelo, S/N 33007 Oviedo, Spain._
  E-mail: labra@uniovi.es
• _MPuerto Paule-Ruiz is with the Department of Computer Science, University of Oviedo, C/Calvo Sotelo, S/N 33007 Oviedo, Spain._
  E-mail: paule@uniovi.es

1. https://rdfa.info/

Fig. 1. Example of Miguel de Cervantes' card

Fig. 2. Component diagram of LODLearning prototype.

in this context refers to the addition and linking of related latent content into lessons material. That enhancement offers the opportunity to learn new knowledge without leaving the platform, providing the students with a new way of searching for related content. LODLearning performs a NLP (Natural Language Processing) entity recognition algorithm that extracts the most relevant known entities from the given text. It also searches through the Semantic Web for new content to add to these entities. Therefore, the principal idea behind LODLearning is to take advantage of the Semantic Web to complement and expand the learning content within courses.

3.1 Prototype use case

LODLearning takes the lessons content from the LMS tool and analyses it in order to retrieve meaningful entities that are shown to the user, enriching the present content with expanded information. For the hypothesis demonstration we have integrated the LODLearning prototype with Sakai LMS, which supports all the learning management needed in a typical course environment providing different ways to integrate with and expand its functionality. In particular, it supports the IMS Learning Tools Interoperability protocol [19]). From a teacher point of view only a few steps are required for the enhancement of the content. The first time the teacher enters into the tool, lessons content will be displayed without any enhancement. If the teacher goes to the import section, an entity recognition algorithm will be executed, providing then the teacher with a list of checkable entities (i.e., if Cervantes is mentioned it will be displayed in the list). These entities are accompanied with a confidence percentage indicating the probability that they would indeed be present in the lessons content.

Once relevant entities have been selected by the teacher, they will be added to the system that will finally show the lessons content with the enhancement added. This makes the enhancement system dynamic because it can be adapted depending on the content and the teacher requirements.
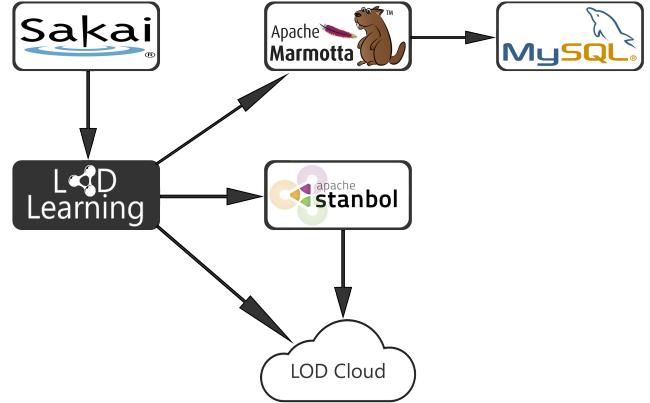
The system adds new enhanced content to lessons by using cards which show different information depending on the entity type previously selected by the teacher. New content to lessons can be added by using individual cards for every recognised entity. Cards can be designed with different information depending on entity type, for example, a photograph, a description, the birth date, the birth place, the death date, the death place and the wikipedia link for person entities. An example is shown in Figure 1.

For embedding these cards into the original content we opted for a modal based approach, showing a link when an entity is mentioned. When the link is pressed the corresponding item is displayed showing more information about the entity. With this approach new knowledge can be offered to the user without the need to leave the tool and the main content (see Figure 4).

3.2 Technological stack

The following technologies are being used:

- Sakai[2]: This is the LMS tool that is responsible for all the learning infrastructure. It offers authentication, course management, content management and an interface to expand its functionality.
- Apache Stanbol[3]: This component runs NLP and returns a list of URIs with some relevant attributes. Stanbol is used as an entity recogniser and entity disambiguator.
- Apache Marmotta[4]: This is a RDF (Resource Description Framework) triple store which offers a SPARQL[5] (SPARQL Protocol and RDF Query Language) endpoint and a set of web services for updating RDF content which is used to persist the enhanced content. Marmotta adapts a MySQL database to persist triples on it.

2. https://sakaiproject.org/
3. https://stanbol.apache.org/
4. http://marmotta.apache.org/
5. https://www.w3.org/TR/rdf-sparql-query/

Fig. 3. Sakai Lessons tool with the three topics covered in the evaluation included.

- DBpedia[6]: This project collects data from Wikipedia and transforms it into RDF. DBpedia is part of the LOD Cloud[7].

Figure 2 provides a diagram on how these technologies interact in our prototype. For the connection between Sakai and the prototype we used the LTI protocol [19], from the IMS Global Learning Consortium, in its 1.1 version. This protocol is a standard that defines how educational applications should communicate with LMSs. Between Apache Marmotta and LODLearning we used a REST API as well as between Apache Stanbol and LODLearning. DBpedia exposes a SPARQL endpoint which is queried with Apache Jena[8]. And finally, Apache Marmotta communicates with MySQL through JDBC.

The application flow is the following: once the application is invoked from Sakai, it queries the Sakai lessons API and adapts the available menus depending on the user role (i.e., admin, instructor or student). If a teacher performs an entity content importation, LODLearning sends the lesson content to Apache Stanbol which executes an entity recognition algorithm. Once Stanbol finishes, it returns a RDF graph with the entities URI, the confidence and some extra

6. http://wiki.dbpedia.org/
7. http://lod-cloud.net/
8. https://jena.apache.org

attributes. LODLearning persists this RDF and some extra attributes (queries from the DBpedia) to Apache Marmotta. Finally, once the importation is persited, whenever a user enters to content section, LODLearning will run SPARQL queries for the different persisted entities. LODLearning will also change entities appearances for links that will reveal their cards.

4 PROTOTYPE EVALUATION

This evaluation is focused on the didactic effectiveness measurement of the enhanced content performed using Semantic Web technologies. In our study, didactic effectiveness is associated with the change in students' performance while they were using the tool [20].

For the evaluation we composed a lesson into the Sakai learning system which was formed by three different topics (Spanish General Elections of 2011, Miguel de Cervantes and Federico García Lorca). The Native Sakai tool with these topics can be seen in Figure 3. In contrast, LODLearning downloads these lessons and enhances them with related content about the current lesson which is shown in form of cards. These cards will later appear whenever a student performs a click in the corresponding link (see Figure 4).

Therefore, the main difference between Sakai native tool and LODLearning lies in that more optional content that can

Fig. 4. LODLearning tool with content enhancements. The arrows show the action performed when a link is pressed, revealing its corresponding enhanced content.

be consulted by the students and in the experience that the students get from both tools.

The sample comprised 32 students pursuing the mandatory education stage in a State High School from the North of Spain and consisted of 18 women and 14 men aged from 13 to 14 years. The sample was divided into two groups in a random manner, namely control and experimental groups to perform an inter-subject study.

Control group evaluation was carried out by means of two different tasks for an intra-subject study. The first one (pretest) consisted in a questionnaire about three different topics covered in the Sakai course lesson. This first questionnaire was completed without any tool exposition in order to assess the knowledge of the sample. Then, the control group was exposed to the Sakai lessons native tool where the students read and memorised the exposed contents to perform a second questionnaire (posttest) about these topics. The experimental group evaluation was performed with the same method. However, it was exposed to our
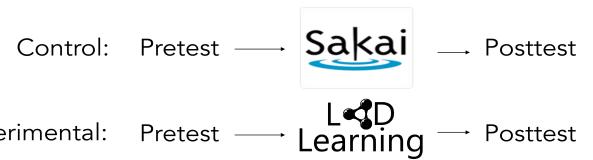
Fig. 5. Evaluation process permorfed by the students in the evuualiation of both tools.

own designed prototype. Finally, the sample was asked to complete a satisfaction questionnaire to know their impressions about the tools they were exposed. This procedure can be seen in Figure 5. Time intervals, for both groups, for the completion of every requested task were as follows: 10 minutes for the first and second questionnaires, 5 minutes for satisfaction questionnaire and 15 minutes for reading and memorising the exposed tool contents.

The two evaluation questionnaires contained 11 and 19

TABLE 1
Marks obtained by the students. Sample size(n), mean($\bar{x}$), standard deviation(s), max and min for every group. 'Before' refers to results before exposition to the tool and 'After' to results after exposition to the tool.

	Control Before	Experimental Before	Control After	Experimental After
n	16	16	16	16
$\bar{x}$	14.77273	19.88636	32.89474	42.10526
s	9.889193	14.16889	13.79175	10.52632
max	36.360	54.550	57.89	57.89
min	0	0	10.53	21.05

questions respectively, all of them assigned with the value 1 for the right answer and 0 for no response or a wrong answer. Both of them displayed queries about present content in the lessons Sakai tool, either in the native version or in the content enhanced one using the LODLearning prototype. Questions were single choice or free text where some of the questions asked about multimedia content like maps and images. The first questionnaire consisted of 6 standard questions and 5 questions about the enhanced content. The second one included the first questionnaire plus 6 questions about the enhanced content and 2 standard questions. Satisfaction questionnaires —based on a Likert scale— were composed by 6 questions about the two different tools. For both groups the questionnaires were composed of the same questions. This satisfaction questionnaire was completed by 30 students out of the 32 ones. These 2 students preferred not to complete the satisfaction questionnaire. Questionnaires were designed and completed using the Google Docs platform and then downloaded as a CSV file for transformation and calculation of final marks with our own Python script. The technological stack described in the previous section was hosted in an Ubuntu 14.04 LTS server where students had access to it through internet by using Chrome or Firefox in their latest versions.

Results were collected adding 1 point for every correct student answer and then their marks were normalised in a 100 base following the English grading system. Results are shown in Table 1.

5 RESULTS

Statistical analysis was performed using R, version 3.2.4 [21]. A Student's t-test was carried out between control and experimental groups (inter-subject study) before and after being exposed to the lessons based on Sakai native tool or to the LODLearning prototype, respectively; as well as between after and before within the same group (intra-subject study). Before exposition to the tool differences between control and experimental groups were not significant ($p = .24$). However, with the conventional level of significance ($\alpha = .05$), after exposition to the tool differences between control and experimental groups were significant ($p = .04$). Differences between the same group before and after exposition to the tool were very significant ($p = .00018$) and ($p = .00002$) for control group and experimental group respectively (see Figure 6). As part of the didactic effectiveness study for LODLearning, Cohen's d index [22] was also calculated to know the effect size
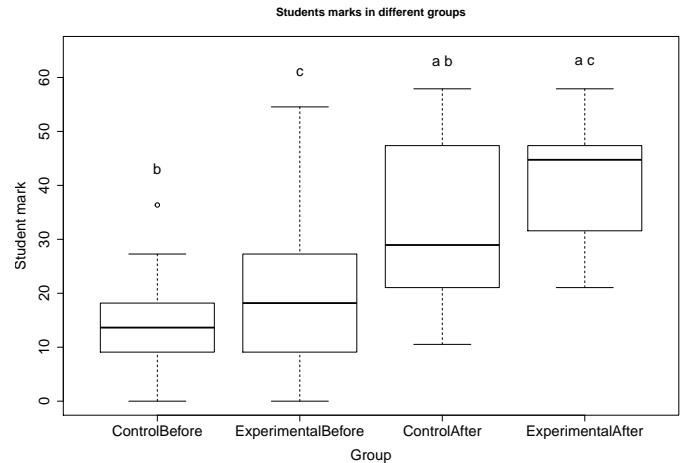

Fig. 6. Representation of the experiment results for control and experimental groups after and before exposition to the tools. b & c very significant differences ($p < .001$). a significant differences ($p < .05$) by means of Student's t-test.

between control and experimental groups after exposition to the tool ($d = .75$). A study for correct answers ratio for each question (after exposition to the tool) was performed by means of Fisher's exact test. Results, shown in Figure 7, exhibit that experimental group was significantly greater than control group for questions 9 ($p = .04484$) and 19 ($p = .004069$). Results of satisfaction test are shown in Figure 8 as well as students subject suggestions for tools inclusion, Figure 9.

6 DISCUSSION AND INTERPRETATION OF RESULTS

Pretest study indicates that students in two groups had similar performance before exposition to the tools indicating similar levels of knowledge in both groups ($p = .24$). Nevertheless, posttest results report significative differences in control and experimental groups after exposition to the tools, pointing to changes in students' performance when using our prototype in comparison to the native Sakai tool. Moreover, the effect size (Cohen's d) shows that our results are not only significant, but are relevant and close to a big effect size. This measure proves that our prototype could be worthy to be used by its positive impact on students' performance. Another facet that deserves to be highlighted is the novelty aspect which can be a motivating factor and would stimulate students' interest. Scientific literature reports it in areas such as mathematics [23] and sciences [24]. However, this novelty aspect is present in both tools as students reflected in the degree of familiarisation question (see Figure 8).

When questions are considered separately some interesting data arise. Questions 19 and 9 suggest significant differences between control and experimental groups. Both of them were part of the enhanced data included into the prototype. Questions 19, 18, 17 and 16 registered the biggest differences; these questions, about multimedia items (i.e., maps and photographs), show that students tend to perform better with multimedia learning content. The other question with a significant difference between groups, question 9,
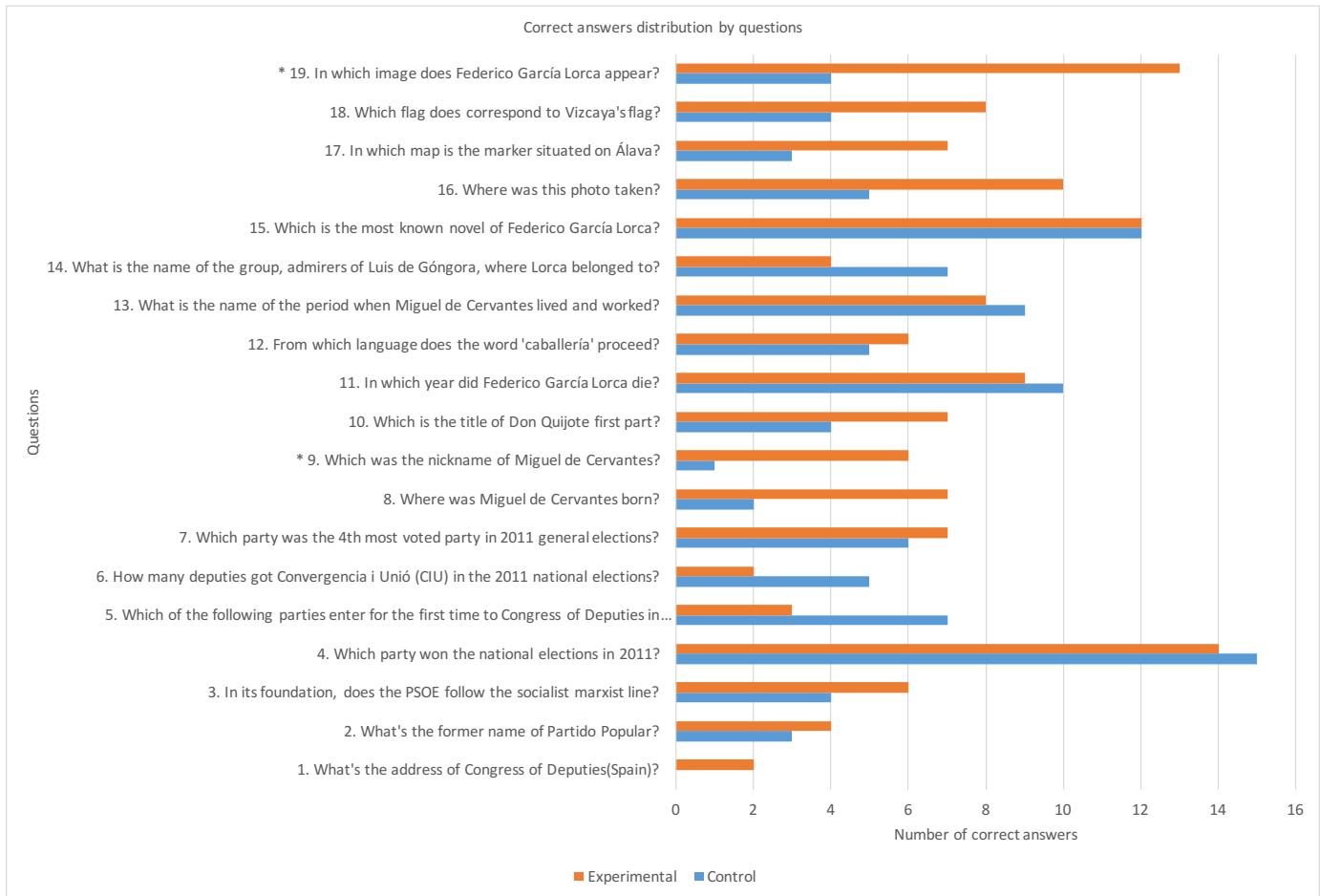
Fig. 7. Distribution of correct answers by each question for control and experimental groups after exposition to the tool. Each bar represents the number of students that gave a correct answer for the respective question. * Significant evidence for Experimental > Control ($p < .05$) by means of Fisher's exact test.

suggests that when the prototype uses a short description text (e.g., question 9 and 8) students tend to remember this text more than when using a long text (e.g., question 2). Other questions about enhanced content (i.e., questions 12, 3, 2, 1) registered some better performance in the experimental group without as big differences as the previous ones which are caused also by long description text. However, questions 14, 13, 11, 6 and 5, about standard content, registered a better performance in the control group which might be influenced by the bigger amount of contents that should be memorised by the experimental group. These results report that to obtain better content didactic effectiveness short text and multimedia content are the ones that should be prioritised which are in line with similar results reported by [2].

As Figure 8 shows, there are not significant differences among students in satisfaction levels. This supports that students are equally satisfied when using both tools. Therefore, we consider that LODLearning could be included in State High School courses without affecting notably students workflow with virtual learning environments. Moreover, these answers indicate no relevant issues in using the LODLearning prototype by the students. In contrast, LODLearning does not seem to increase satisfaction levels for the students nor their inclusion recommendation levels

even though it does increase students' performance. However, this might be influenced by their degree of familiarisation with virtual learning platforms as they exposed in the first questions of the satisfaction questionnaire. These results also report that the enhancement content can be added transparently without interfering with the student and its learning process. Moreover, they rated their experience with the tools very positively and they also recommended their inclusion in subjects they were coursing.

When asked about their recommended subjects for tool inclusion they tended to recommend subjects related to the contents of this evaluation (i.e., History & Social Sciences and Linguistics & Literature) but also subjects like Natural Sciences and Technology where enhanced content about some difficult terms might be useful. These results are in line with the control suggestions where students tended to recommend more subjects, but most rated subjects are those which the experimental group recommended. The absence of Fine Arts draws attention, as it might be an interesting subject where it would be possible to conduct a more in-depth study.

One of the main lacks of other approaches is that the teacher needs to have technical knowledge [14] [15] [16] [17] [18]. However, in our approach, the teacher only needs to choose between the recognised entities in order to enhance
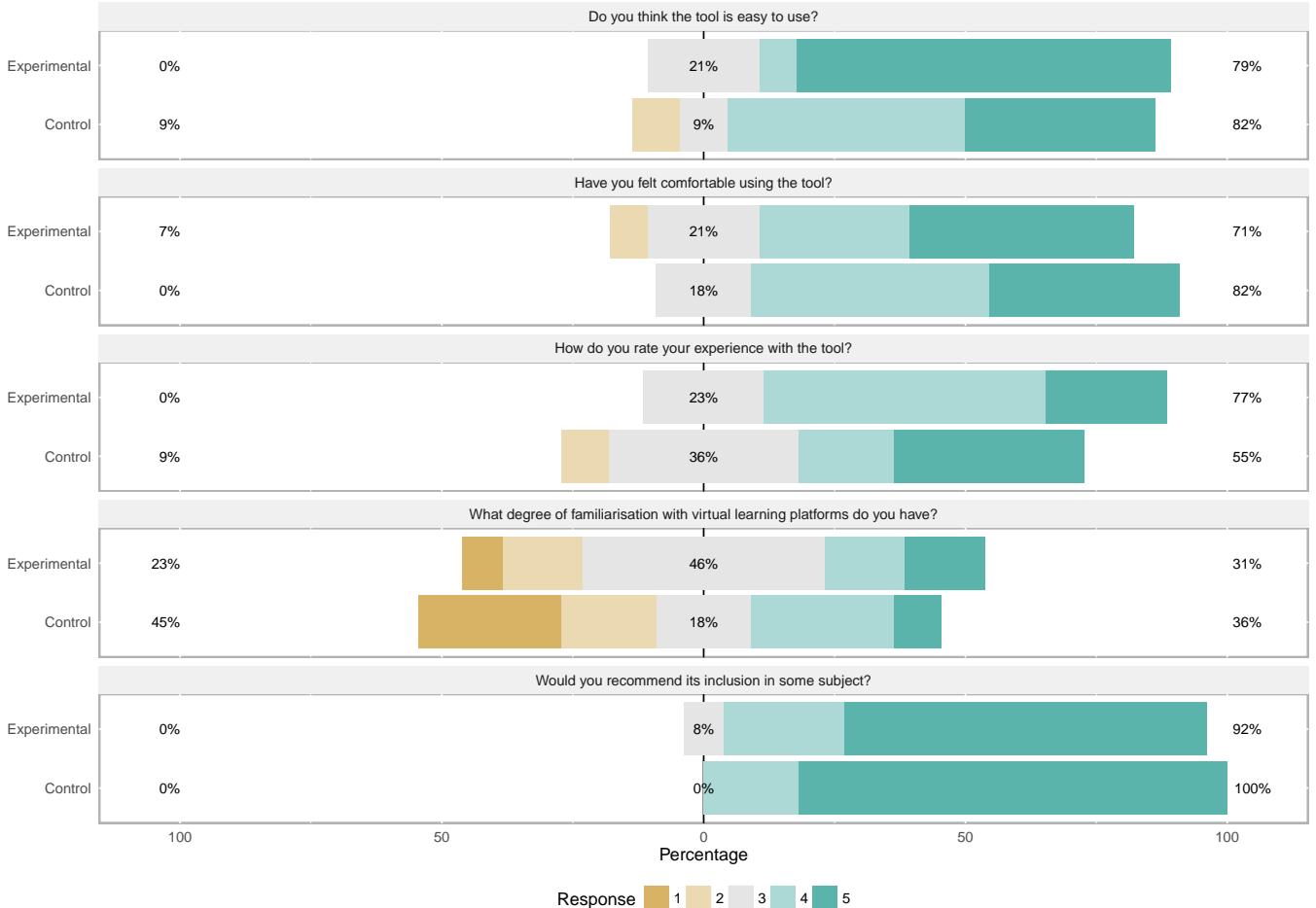
Fig. 8. Control and experimental groups satisfaction punctuations about the two different tools in a Likert scale based questionnaire. Punctuation of 1 refers to Strongly disagree/Very poor and 5 to Strongly agree/Very good
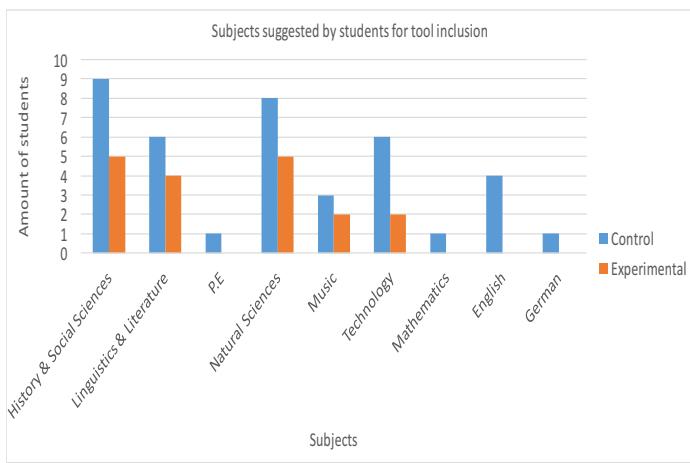

Fig. 9. Bar chart which represents the number of students that suggested the inclusion of their tested tool in different subjects they were coursing.

the content. With our tool, and the support of Semantic Web, our approach provides more flexibility due to its design. Furthermore, other approaches did not cover a numerical evaluation [3] [4] [5] [7] [9] [11] nor a didactic effectiveness evaluation [6] [8] [10] [12], whereas our work includes this type of evaluation.

7 CONCLUSIONS AND FUTURE WORK

In this work we have described the interaction of the LODLearning tool that we have developed and the way that it leverages the Linked Open Data Cloud to enhance lessons contents in the Sakai LMS. This prototype demonstrates that content enhancement can be used to improve courses didactic effectiveness. Nevertheless, support for more e-Learning platforms, inclusion of more enhancement content, an authoring system for designing new cards and more exhaustive and extended experiments should be addressed as future work in order to produce a better and more reliable platform. This work leads to a new way of use of Semantic Web Data in e-Learning platforms and highlights the combined use of e-Learning and Semantic Web in order to create more powerful learning tools.

ACKNOWLEDGMENTS

REFERENCES

[1] S. Y. Park, "An Analysis of the Technology Acceptance Model in Understanding University Students' Behavioral Intention to Use e-Learning," *Journal of Educational Technology & Society*, vol. 12, no. 3, pp. 150–162, 2009.

[2] M. Paule-Ruiz, V. Álvarez-García, J. R. Pérez-Pérez, and M. Riestra-González, "Voice Interactive Learning: A Framework and Evaluation," in *Proceedings of the 18th ACM Conference on Innovation and Technology in Computer Science Education*, ser. ITiCSE '13. New York, NY, USA: ACM, 2013, pp. 34–39.

[3] D. Glachs, V. Damjanovic, F. Strohmeier, and S. Fernández, "EAGLE–Local Government Learning Platform," *Proceedings of the Linked Learning meets LinkedUp: Learning and Education with the Web of Data, co-located with 13th International Semantic Web Conference (ISWC 2014).*

[4] M. Coccoli and I. Torre, "Interacting with annotated objects in a Semantic Web of Things application," *Journal of Visual Languages & Computing*, vol. 25, no. 6, pp. 1012–1020, dec 2014.

[5] D.-P. Deng, G.-S. Mai, C.-H. Hsu, C.-L. Chang, T.-R. Chuang, and K.-T. Shao, *Linking Open Data Resources for Semantic Enhancement of User–Generated Content*, ser. Lecture Notes in Computer Science, H. Takeda, Y. Qu, R. Mizoguchi, and Y. Kitamura, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, jan 2013, vol. 7774.

[6] A. Gladun, J. Rogushina, F. García-Sanchez, R. Martínez-Béjar, and J. T. Fernández-Breis, "An application of intelligent techniques and semantic web technologies in e-learning environments," *Expert Systems with Applications*, vol. 36, no. 2, pp. 1922–1931, mar 2009.

[7] T. Kurz, G. Güntner, V. Damjanovic, S. Schaffert, and M. Fernandez, "Semantic enhancement for media asset management systems," *Multimedia Tools and Applications*, vol. 70, no. 2, pp. 949–975, 2012.

[8] S. Lohmann, P. Heim, and P. Díaz, "Exploiting the Semantic Web for Interactive Relationship Discovery in Technology Enhanced Learning," in *2010 10th IEEE International Conference on Advanced Learning Technologies*. IEEE, jul 2010, pp. 302–306.

[9] P. Markellou, I. Mousourouli, S. Spiros, and A. Tsakalidis, "Using Semantic Web Mining Technologies for Personalized e-Learning Experiences," in *Proceedings of the web-based education*. Grindelwald, Switzerland: ACTA Press, feb 2005, pp. 461–826.

[10] P. Nelson, T. Edmundo, C.-P. Ricardo, L.-V. Jorge, and A. C. Janneth, "Consuming and producing linked open data: the case of OpenCourseWare," *Program*, vol. 48, no. 1, pp. 16–40, 2014.

[11] L. Stojanovic, S. Staab, and R. Studer, "eLearning Based on the Semantic Web," in *Proceedings of WebNet 2001, World Conference on the WWW and the Internet*, no. FEBRUARY 2002, Orlando, Florida, 2001.

[12] H. Q. Yu, C. Pedrinaci, S. Dietze, and J. Domingue, "Using Linked Data to Annotate and Search Educational Video Resources for Supporting Distance Learning," *Learning Technologies, IEEE Transactions on*, vol. 5, no. 2, pp. 130–142, 2012.

[13] P. Brusilovsky, "Methods and techniques of adaptive hypermedia," *User modeling and user-adapted interaction*, vol. 6, no. 2-3, pp. 87–129, 1996.

[14] P. De Bra, D. Smits, K. van der Sluijs, A. I. Cristea, J. Foss, C. Glahn, and C. M. Steiner, *Intelligent and Adaptive Educational-Learning Systems: Achievements and Trends*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2013, ch. GRAPPLE: Learning Management Systems Meet Adaptive Learning Environments, pp. 133–160.

[15] A. Montes García, P. De Bra, G. H. L. Fletcher, and M. Pechenizkiy, "A DSL Based on CSS for Hypertext Adaptation," in *Proceedings of the 25th ACM Conference on Hypertext and Social Media*, ser. HT '14. New York, NY, USA: ACM, 2014, pp. 313–315.

[16] P. De Bra, E. Knutov, D. Smits, N. Stash, and V. F. C. Ramos, "GALE: A Generic Open Source Extensible Adaptation Engine," *New Rev. Hypermedia Multimedia*, vol. 19, no. 2, pp. 182–212, 2013.

[17] M. del Puerto Paule Ruiz, M. J. F. Díaz, F. O. Soler, and J. R. P. Pérez, "Adaptation in current e-learning systems," *Computer Standards & Interfaces*, vol. 30, no. 1–2, pp. 62–70, 2008.

[18] I. Gutiérrez, V. Álvarez, M. Paule, J. R. Pérez-Pérez, and S. de Freitas, "Adaptation in E-Learning Content Specifications with Dynamic Sharable Objects," *Systems*, vol. 4, no. 2, p. 24, 2016.

[19] C. Severance, T. Hanss, and J. Hardin, "Ims learning tools interoperability: Enabling a mash-up approach to teaching and learning tools," *Technology, Instruction, Cognition and Learning*, vol. 7, no. 3-4, pp. 245–262, 2010.

[20] M. Loi and B. Ronsivalle, "A Particular Aspect of Cost Analysis in Distance Education," in *Distance and E-Learning in Transition*. John Wiley & Sons, Inc., 2013, pp. 161–168.

[21] R Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, 2016. [Online]. Available: https://www.R-project.org/

[22] J. Cohen, "Statistical power analysis for the behavioral sciences. Vol. 2," *Lawrence Earlbaum Associates, Hillsdale, NJ*, 1988.

[23] H. Kaufmann and D. Schmalstieg, "Mathematics and geometry education with collaborative augmented reality," *Computers & Graphics*, vol. 27, no. 3, pp. 339–345, 2003.

[24] P. Maier, M. Tönnis, and G. Klinker, "Augmented Reality for teaching spatial relations," in *Conference of the International Journal of Arts & Sciences, Toronto*, 2009.

Herminio García-González is a PhD Student at the University of Oviedo. He obtained his degree in Software Engineering from the University of Oviedo in 2014 and his MSc in Web Engineering from the same university in 2016. His current research interests are Web Semantics and e-Learning.

José Emilio Labra Gayo is an Associate Professor from the University of Oviedo, Spain. He is the main researcher of the WESO research group (Web Semantics Oviedo). The group collaborates on practical applications of semantic web and linked open data and collaborates with industrial partners and public administrations applying semantic technologies to solve practical problems.

MPuerto Paule-Ruiz is Lecturer for the Department of Computer Science at the University of Oviedo. She received her Ph.D from the University of Oviedo in 2003. She has participated actively in several regional and national projects related with adaptive educational and context-aware e-learning systems, and is at the moment involved in a national research project based on adaptive systems. She is the author of several publications in major journals.