ON METHODOLOGY OF APPLICATION OF LINKED DATA TO PERSONALISE LEARNING

Svetlana Kubilinskiene[1], Eugenijus Kurilovas[2]

[1]Vilniaus kolegija – University of Applied Sciences (LITHUANIA)
[2]Vilnius Gediminas Technical University (LITHUANIA)

Abstract

The paper is aimed to present a methodology of learning personalisation based on applying Semantic Web technologies and particularly Linked Data. Semantic Web technologies and Linked Data are changing the way information is stored, described and exploited.

The "Linked Data" term refers to a set of best practices for publishing and connecting structured data on the Web. The advantages of Linked Data web are used to support semi-automatic classification of educational resources. The relations of learning objects (resources) are encoded in Resource Description Language and stored in the repository, a query language is used to retrieve data, and the knowledge of organisational systems and Linked Data is used to classify the web resources according to the domain. The Linked Data principles are applied for semantic integration and social interconnecting of educational data, resources and actors. Linked Data movement promises to significantly improve existing practices of system integration, resource sharing and personalisation to support learning. The Linked Data approach is a promising approach to establish relationships between learning resources and student's personal characteristics.

Linked Data approach and Resource Description Framework (RDF) standard model are already well-known in scientific literature, but only few authors have analysed its application to personalise learning process. Many authors agree that OWL, Linked Data, ontologies, recommender systems, and RDF-based learning personalisation trends should be further analysed.

In the paper, first of all, systematics review on application of Semantic Web and particularly Linked Data to personalise learning is presented. After that, methodology how to apply Linked Data and the other Semantic Web technologies such as RDF triples, OWL, ontologies, and recommender systems to personalise learning is presented. This personalisation should be based on applying students' personal preferences (e.g. learning styles) and intelligent technologies. Interconnections between students' learning styles and suitable learning components (i.e. learning objects and learning activities) are analysed in the paper in more detail.

According to presented methodology, after identifying particular students' learning styles and particular learning components' (learning objects' and learning activities') suitability indexes, one could create a number of analysed RDF triples, corresponding OWL-based ontologies and, finally, a recommender system to recommend learners those learning components that fit their personal preferences mostly. Probabilistic suitability indexes applied in the paper show the level of suitability of learning components to particular students and are based on probabilistic analysis of particular students' learning styles as well as on analysis of correspondence of particular learning components to learning styles.

This methodology based on applying Semantic Web technologies is aimed at improving learning motivation and thus – learning quality and effectiveness.

Keywords: Semantic Web, Linked Data, Resource Description Framework, learning styles, personalisation, learning objects, learning activities.

1 INTRODUCTION

The paper analyses and presents a methodology to personalise learning applying Semantic Web technologies and particularly Linked Data.

In Lithuania, learning personalisation topic applying intelligent technologies has become very popular during several recent years ([1] – [5]).

ISBN: 978-84-09-17939-8

According to the methodology presented in the paper, first of all we should identify individual students' learning styles and particular learning components' (learning objects' and learning activities') suitability indexes. Then we should create a number of RDF triples interconnecting students' learning styles and learning components, corresponding OWL-based ontologies and, finally, a recommender system to recommend learners those learning components that fit their personal preferences mostly.

The rest of the paper is organised as follows: systematic literature review on the analysed topic is provided in Section 2, an original learning personalisation methodology is presented in Section 3, and Section 4 concludes the paper.

2 SYSTEMATIC REVIEW

In order to specifically find research on linked data and learning personalisation, an exhaustive search conducting a systematic review was performed. This systematic review was conducted following the process proposed by [6] and [7].

According to [7], the term systematic review is used to refer to a specific methodology of research, developed in order to gather and evaluate the available evidence pertaining a focused topic. In contrast to the usual topic of literature review, unsystematically conducted whenever one starts a particular investigation, a systematic review was developed, as the term denotes, in a formal and systematic way.

The methodological steps, the strategies to retrieve the evidence, and the focus of the question are explicitly defined by [7].

According to [6], this process presents three main phases: (1) Phase 1 – Planning: In this phase, the research objectives and the review protocol are defined. The protocol constitutes a pre-determined plan that describes the research questions and how the systematic review will be conducted; (2) Phase 2 – Conduction: During this phase, the primary studies are identified, selected and evaluated according to the inclusion and exclusion criteria established previously. For each selected study, data are extracted and synthesized; and (3) Phase 3 – Reporting: In this phase, a final report is formatted and presented.

Search history in Clarivate Analytics Web of Science database is presented in Figure 1:

Set	*Results*	*Save History / Create AlertOpen Saved History*
# 2	14	(TS=(linked data AND learning personalisation)) *AND* **LANGUAGE:** (English) *AND* **DOCUMENT TYPES:** (Article) *Indexes=SCI-EXPANDED, SSCI, A&HCI, CPCI-S, CPCI-SSH, ESCI Timespan=All years*
# 1	10	(TS=(linked data AND learning personalisation)) *AND* **LANGUAGE:** (English) *AND* **DOCUMENT TYPES:** (Article) *Indexes=SCI-EXPANDED, SSCI, A&HCI, CPCI-S, CPCI-SSH, ESCI Timespan=Last 5 years*

Fig. 1. Search history in Clarivate Analytics Web of Science

During Phase 2, 14 primary articles were identified in Web of Science database in December 2019, selected and evaluated according to the research question. We can notice that the majority of suitable articles were publishet during last 5 years.

Finally, 11 articles were selected for detailed examination. The results of conduction of Phase 2 are as follows:

In [8], a new method towards automatic personalised recommendation based on the behavior of a single user in accordance with all other users in web-based information systems is introduced. The proposal applies a modified version of the well-known Apriori data mining algorithm to the log files of a web site (primarily, an e-commerce or an e-learning site) to help the users to the selection of the best user-tailored links. The paper mainly analyses the process of discovering association rules in this kind of big repositories and of transforming them into user-adapted recommendations by the two-step modified Apriori technique. A first pass of the modified Apriori algorithm verifies the existence of association rules in order to obtain a new repository of transactions that reflect the observed rules. A second pass of the proposed Apriori mechanism aims in discovering the rules that are really inter-associated. This way the behavior of a user is not determined by "what he does" but by "how he does". An efficient implementation has been performed to obtain results in real-time. As soon as a user closes his session in the web system, all data are recalculated to take the recent interaction into account for the next recommendations. Early results have shown that it is possible to run this model in web sites of medium size.

In [9], MyBehaviorCBP is a mobile phone app that uses machine learning on sensor-based and self-reported physical activity data to find routine behaviors and automatically generate physical activity recommendations that are similar to existing behaviors.In the pilot study, MyBehaviorCBP's automated approach was found to have positive effects. Specifically, the recommendations were actualised more, and perceived to be easier to follow. To the best of the authors' knowledge, this is the first time an automated approach has achieved preliminary success to promote physical activity in a chronic pain context. Further studies are needed to examine MyBehaviorCBP's efficacy on a larger cohort and over a longer period of time.

According to [10], Personalised medicine is expected to yield improved health outcomes. Data mining over massive volumes of patients' clinical data is an appealing, low-cost and noninvasive approach toward personalisation. Machine learning algorithms could be trained over clinical big data to build prediction models for personalized therapy. To reach this goal, a scalable big data architecture for the medical domain becomes essential, based on data standardisation to transform clinical data into FAIR (Findable, Accessible, Interoperable and Reusable) data. Using Ontologies and Semantic Web technologies, the authors attempt to reach mentioned goal. In [10], the authors developed an ontology to be used in the field of radiation oncology to map clinical data from relational databases. The authors combined ontology with semantic Web techniques to publish mapped data and easily query them using SPARQL. The authors combined the ontology with Semantic Web technologies showing how to efficiently and easily integrate and query data from different (relational database) sources without a priori knowledge of their structures. When clinical FAIR data sources are combined (linked data) using mentioned technologies, new relationships between entities are created and discovered, representing a dynamic body of knowledge that is continuously accessible and increasing.

The authors of [11] consider that with the emergence of cloud-based technology, personalised learning mechanism has increasingly become a fundamental requirement for most learning systems. This study aimed to identify the key factors that influence user adoption of cloud-based collaborative learning technology in the educational context. Grounded on the Unified Theory of Acceptance and Use of Technology, personalisation construct was linked to the behavioral intention, performance expectancy and effort expectancy. This research applied a new methodological approach combining both Fuzzy Analytic Hierarchy Process (FAHP) and Structural Equation Modelling (SEM) to determine the relative weight and importance of the factors as well as to test the proposed hypotheses in the research model. The findings of FAHP demonstrated that performance expectancy, social influence, and personalisation were the most important factors predicting behavioral intention to adopt cloud-based collaborative learning technology from experts' point of view. The results of the SEM showed that users' behavioural intention was significantly influenced by performance expectancy, effort expectancy, social influence and personalisation. Although, personalisation performed a direct influence on behavioural intention, its indirect influence through performance expectancy and effort expectancy was also considerable. This study and its findings can serve as a baseline by which cloud service providers, ministry of education, and educational institutions can make strategic and strong decisions about adoption of cloud-based technology in educational environments.

Results of [12] have shown that overall students were satisfied with their clinical learning experience across all placement areas. This was linked to the 6 constructs of the clinical learning environment inventory; personalisation, innovation, individualisation, task orientation, involvement, satisfaction. Significant differences in student experience were noted between age groups and student year but there was no difference noted between placement type, age and gender. It was concluded that nursing students had a positive perception of their clinical learning experience, although there remains room for improvement. Enabling a greater understanding of students' perspective on the quality of clinical education is important for nursing education and future research.

In [13}, the authors address the problem of providing an order of relevance, or ranking, among entities' properties used in RDF datasets, Linked Data and SPARQL endpoints. The authors first motivate the importance of ranking RDF properties by providing two killer applications for the problem, namely property tagging and entity visualisation. Moved by the desiderata of these applications, the authors propose to apply Machine Learning to Rank (MLR) techniques to the problem of ranking RDF properties. Their devised solution is based on a deep empirical study of all the dimensions involved: feature selection, MLR algorithm and Model training. The major advantages of this approach are the following: (a) flexibility/personalisation, as the properties' relevance can be user-specified by personalising the training set in a supervised approach, or set by a novel automatic classification approach based on SWiPE; (b) speed, since it can be applied without computing frequencies over the whole dataset, leveraging existing fast MLR algorithms; (c) effectiveness, as it can be applied even when no ontology data is available by using novel dataset-independent features; (d) precision, which is high both in terms of f-measure and Spearman's rho. Experimental results show that the proposed MLR framework outperform the two existing approaches found in literature which are related to RDF property ranking.

According to [14], an important application of web usage mining is mining web log data. The authors propose a new optimised technique for web mining, in the realm of an e-learning site to recommend the best links for a learner to visit the next. It optimises web mining, by partitioning the database, on the basis of the learner's knowledge level, to create a suffix tree(s) from the existing sequences of previous 'n' learners' path. To further reduce the overhead of re-mining the web patterns, the authors propose that a web traversal pattern should be regarded as significant, only if it qualifies the minimum threshold of length and frequency in the database. These significant patterns are added to suffixes. They are then mined, using the most efficient mining algorithm after a comparative analysis of various algorithms, to find the most frequent navigation paths for recommendation to n + 1th new learner. The authors conducted experiments on a real case study of an Indian e-learning site. This is verified by experiments with promising results on computational time. This speed up obtained, in web pattern mining, is a meaningful approach for building recommender based e-learning system.

The authors of [15] consider that learners learn differently because they are different and they grow more distinctive as they mature. Personalised learning occurs when e-learning systems make deliberate efforts to design educational experiences that fit the needs, goals, talents, and interests of their learners. Researchers had recently begun to investigate various techniques to help teachers improve e-learning systems. In this paper the authors present their design and implementation of an adaptive and intelligent web-based programming tutoring system – Protus, which applies recommendation and adaptive hypermedia techniques. This system aims at automatically guiding the learner's activities and recommend relevant links and actions to him/her during the learning process. Experiments on real data sets show the suitability of using both recommendation and hypermedia techniques in order to suggest online learning activities to learners based on their preferences, knowledge and the opinions of the users with similar characteristics.

According to [16] , nowadays, the application of Web mining techniques in e-learning and Web-based adaptive educational systems is increasing exponentially. In this paper, the authors proposed an advanced architecture for a personalisation system to facilitate Web mining. A specific Web mining tool is developed and a recommender engine is integrated into the AHAI system in order to help the instructor to carry out the whole Web mining process. The authors' objective is to be able to recommend to a student the most appropriate links/Web pages within the AHAI system to visit next. Several experiments were carried out with real data provided by Eindhoven University of Technology students in order to test both the architecture proposed and the algorithms used. Finally, the authors have also described the meaning of several recommendations, starting from the rules discovered by the Web mining algorithms.

In [17], a new method towards automatic personalised recommendation based on the behavior of a single user in accordance with all other users in web-based information systems is introduced. The proposal applies a modified version of the well-known Apriori data mining algorithm to the log files of a web site (primarily, an e-commerce or an e-learning site) to help the users to the selection of the best user-tailored links. The paper mainly analyses the process of discovering association rules in this kind of big repositories and of transforming them into user-adapted recommendations by the two-step modified Apriori technique, which may be described as follows. A first pass of the modified Apriori algorithm verifies the existence of association rules in order to obtain a new repository of transactions that reflect the observed rules. A second pass of the proposed Apriori mechanism aims in discovering the rules that are really inter-associated. This way the behavior of a user is not determined by "what he does" but by "how he does". Furthermore, an efficient implementation has been performed to obtain results in real-time. As soon as a user closes his session in the web system, all data are recalculated to take the recent interaction into account for the next recommendations. Early results have shown that it is possible to run this model in web sites of medium size.

According to [18], recently, research in individual differences and in particular, learning and cognitive style, has been used as a basis to consider learner preferences in a web-based educational context. Modelling style in a web-based learning environment demands that developers build a specific framework describing how to design a variety of options for learners with different approaches to learning. It was found that learners do have a preference regarding their interaction, but no obvious link between style and approaches offered, was detected. Derived from an examination of this experimental data, the authors suggest that while style information can be used to inform the design of learning environments that accommodate learners' individual differences, it would be wise to recommend interactions based on learners' behaviour. Learning environments should allow learners or learners' interaction behaviour to select or trigger the appropriate approach for the particular learner in the specific context. Alternative approaches towards these directions are also discussed in [18].

Systematic literature review reveals that Linked Data and other Sematic Web technologies such as RDF, OWL, and recommender systems are alreagy used in education.

The Linked Data approach is a promising approach to establish relationships between learning objects and student’s personal characteristics (unless, this point was not discussed in studies we examined). It is based on a set of well-established principles and (W3C) standards, e.g. RDF, SPARQL, aiming at facilitating Web-scale data interoperability.

Through the last years, vast amounts of educational metadata collections and university data have been provided according to Linked Data principles. In addition, the Linked Data approach allowed to provide knowledge and offers significant potential for its exploitation in educational contexts.

Some studies describe interlinking and mapping study documents representation as RDF and Linked Data. Some studies utilise open educational resources and Linked Data cloud to enrich online courses, reuse and recommend open resources for users.

RDF and Linked Data technologies are used to annotate and classify resources according to similarity and other criteria, building categories of learning resource/object [19].

Although Linked Data approach and RDF standard model are already well-known in scientific literature, only few studies have analysed its application to personalise learning process. Usually, user modelling and past experience, interests are used to provide personalisation. We did not encounter sound studies dealing with these technologies application for personalisation according to student’s learning styles according to different learning styles models, e.g. [20].

In the next section, a new methodology to personalise learning applying RDF triples, Linked Data, ontologies, and recommender system is presented.

3 LEARNING PERSONALISATION METHODOLOGY APPLYING SEMANTIC WEB TECHNOLOGIES

According to [19], the Linked Data approach is closely related to RDF, and has and will have a strong impact on the educational field and has already started to replace the fragmented landscape of educational technologies and standards with a more unified approach, which allows to integrate and interlink educational data of any kind. The strongest side of the Linked Data approach is that it does not require particular schemas to be used, but instead, accepts heterogeneity and offers solutions on the links between schemas and datasets. The learning objects, exposed as Linked Data, can be effectively enriched with metadata and interlinked.

OWL is standard ontology language which could use RDF triples to create ontologies linking students’ learning styles, learning objects and learning activities. These ontologies should be the main part of personalised recommender system that should recommend learning components and scenarios suitable to particular students according to their learning styles.

According to [19], in personalised learning, first of all, integrated learner profile (model) should be implemented based on students’ learning styles, e.g. [20].

After that, interlinking of learning components (learning objects and learning activities) with learners' profiles should be performed, and ontologies-based personalised recommender system should be created to suggest learning components that are the most suitable to particular learners according to their profiles.

After interlinking and ontologies creation stage, recommender system should be created to link students' personal data in their profiles, relevant learning objects according to corresponding metadata fields, and learning activities suitable to particular students according to their learning styles and other profiles' data.

Interlinking and ontologies creation should be based on the expert evaluation results. Experienced experts should evaluate learning components in terms of their suitability to particular learners according to their learning styles and other preferences / needs.

And, finally, recommender system should form the preference lists of the learning components according to the expert evaluation results. Probabilistic suitability indexes [21] should be identified for all learning components in terms of their suitability level to particular learners. Probabilistic suitability indexes could be easily calculated for all learning components and all students if we would multiply learning components' suitability ratings by probabilities of particular students' learning styles. These suitability indexes should be included in the recommender system, and preference lists of learning components should be created in the recommender system according to those suitability indexes. All learning components should be linked to particular students according to those suitability indexes. The higher suitability indexes the better learning components fit the needs of particular learners.

Thus, personalised learning units / scenarios (i.e. personalised methodological sequences of learning components) could be created for particular learners. An optimal learning unit / scenario (i.e. learning scenario of the highest quality) for particular student means a methodological sequence of learning components having the highest suitability indexes [21].

A number of Semantic Web and intelligent technologies should be applied to implement this approach, e.g. OWL-based ontologies, recommender systems, intelligent agents, decision support systems to evaluate quality and suitability of the learning components, personal learning environments etc.

There are three RDF triples used while creating the methodology: "student's learning style – requires – suitable learning objects", "student's learning style – requires – suitable learning activities", and "suitable learning activities – require – suitable learning objects".

Thus, in order to implement presented learning personalisation framework, first of all, RDF triples-based OWL ontologies should be created to interlink all learning components with students' learning styles: (1) Linking students' learning styles and suitable learning objects; (2) Linking students' learning styles and suitable learning activities; and (3) Linking suitable learning activities to learning objects.

4 CONCLUSION

In the paper, both systematic review results and methodology on applying Linked Data and other Semantic Web technologies such as RDF triples, ontologies and recommender systems to personalise learning are presented.

While creating learning personalisation methodology, the authors have identified three RDF triples used: "student's learning style – requires – suitable learning objects", "student's learning style – requires – suitable learning activities", and "suitable learning activities – require – suitable learning objects". In the last triple, "suitable learning activities" being the object in the 2nd triple, becomes the subject in the 3rd triple.

According to presented methodology, after identifying particular students' learning styles and particular learning components' (learning objects' and learning activities') suitability indexes, we could create a number of analysed RDF triples and corresponding OWL-based ontologies. Finally, a recommender system could be created based on these ontologies to recommend the most suitable learning objects and learning activities for particular students according to identified learning objects' and learning activities' probabilistic suitability indexes. Thus, the most suitable learning units / scenarios should be recommended to particular students.

REFERENCES

[1] J. Mamcenko, E. Kurilovas, and I. Krikun, "On Application of Case-Based Reasoning to Personalise Learning," *Informatics in Education*, vol. 18, no. 2, pp. 345–358, 2019.

[2] E. Kurilovas, "Evaluation and Optimisation of e-Learning Software Packages: Learning Object Repositories," in *Proceedings of the 4th International Conference on Software Engineering Advances (ICSEA 2009).* Porto, Portugal, September 20–25, 2009, pp. 477–483, 2009.

[3] E. Kurilovas, and I. Vinogradova, "Improved Fuzzy AHP Methodology for Evaluating Quality of Distance Learning Courses," *International Journal of Engineering Education*, vol. 32, no. 4, pp. 1618–1624, 2016.

[4] E. Kurilovas, "On Data-Driven Decision-Making for Quality Education," *Computers in Human Behavior.* DOI: 10.1016/j.chb.2018.11.003 in press.

[5] E. Kurilovas, "Advanced Machine Learning Approaches to Personalise Learning: Learning Analytics and Decision Making," *Behaviour & Information Technology*, vol. 38, no. 4, pp. 410–421, 2019.

[6] B. Kitchenham, "Procedures for performing systematic reviews," *Joint technical report Software Engineering Group*, Keele University, United Kingdom and Empirical Software Engineering, National ICT Australia Ltd, Australia, 2004.

[7] J. Biolchini, P.G. Mian, A.C.C. Natali, and G.H. Travassos, "Systematic review in software engineering," *PESC, Technical Report RT – ES 679/05*, May 2005.

[8] T. Matusch, A. Schneibel, L. Dannwolf, and A. Siegmund, "Implementing a Modern E-Learning Strategy in an Interdisciplinary Environment-Empowering UNESCO Stakeholders to Use Earth Observation," *Geosciences*, vol. 8, no. 12, 2018.

[9] M. Rabbi, M.S.H. Aung, G. Gay, M.C. Reid, and T. Choudhury, "Feasibility and Acceptability of Mobile Phone-Based Auto-Personalized Physical Activity Recommendations for Chronic Pain Self-Management: Pilot Study on Adults," *Journal of Medical Internet Research*, vol. 20, no. 10, article number: e10147, 2018.

[10] A. Traverso, J. van Soest, L. Wee, and A. Dekker, "The Radiation Oncology Ontology (ROO): Publishing Linked Data in Radiation Oncology Using Semantic Web and Ontology Techniques," *Medical Physics*, vol. 45, no. 10, pp. E854-E862, 2018.

[11] E. Yadegaridehkordi, M.H.N.B. Nasir, N.F.B.M. Noor, L. Shuib, and N. Badie, "Predicting the Adoption of Cloud-Based Technology Using Fuzzy Analytic Hierarchy Process and Structural Equation Modelling Approaches," *Applied Soft Computing*, vol. 66, pp. 77-89, 2018.

[12] E. Shivers, F. Hasson, and P. Slater, "Pre-registration Nursing Student's Quality of Practice Learning: Clinical Learning Environment Inventory (Actual) Questionnaire," *Nurse Education Today*, vol. 55, pp. 58-64, 2017.

[13] A. Dessi, and M. Atzori, "A Machine-Learning Approach to Ranking RDF Properties," *Future Generation Computer Systems-The International Journal of Escience*, vol. 54, pp. 366-377, 2016.

[14] R. Mahajan, J.S. Sodhi, and V. Mahajan, "Optimising Web Usage Mining for Building Adaptive E-Learning Site: A Case Study," *International Journal of Innovation and Learning*, vol. 18, no. 4, pp. 471-486, 2015.

[15] B. Vesin, A. Klasnja-Milicevic, M. Ivanovic, and Z. Budimac, "Applying Recommender Systems and Adaptive Hypermedia for E-Learning Personalization," *Computing and Informatics*, vol. 32, no. 3, pp. 629-659, 2013.

[16] C. Romero, S. Ventura, A. Zafra, and P. de Bra, "Applying Web Usage Mining for Personalizing Hyperlinks in Web-Based Adaptive Educational Systems," *Computers & Education*, vol. 53, no. 3, pp. 828-840, 2009.

[17] E. Lazcorreta, F. Botella, and A. Fernandez-Caballero, "Towards Personalized Recommendation by Two-Step Modified Apriori Data Mining Algorithm," *Expert Systems with Applications*, vol. 35, no. 3, pp. 1422-1429, 2008.

[18] K.A. Papanikolaou, A. Mabbott, S. Bull, and M. Grigoriadou, “Designing Learner-Controlled Educational Interactions Based on Learning/Cognitive Style and Learner Behaviour,” *Interacting with Computers*, vol. 18, no. 3, pp. 356-384, 2006.

[19] T. Jevsikova, A. Berniukevičius, and E. Kurilovas, “Application of Resource Description Framework to Personalise Learning: Systematic Review and Methodology,” *Informatics in Education*, vol. 16, no. 1, pp. 61–82, 2017.

[20] R.M. Felder, and L.K. Silverman, “Learning and Teaching Styles in Engineering Education,” *Engineering Education*, vol. 78, no. 7, pp. 674-681, 1988.

[21] E. Kurilovas, J. Kurilova, and T. Andruskevic, “On Suitability Index to Create Optimal Personalised Learning Packages,” in *G. Dregvaite and R. Damasevicius (Eds.): ICIST 2016, Communications in Computer and Information Science (CCIS)*, vol. 639, pp. 479–490, 2016.