JOURNAL OF E-LEARNING AND KNOWLEDGE SOCIETY
Vol. 16, No. 01 (2020), pp. 72-77

Enabling smart learning systems within smart cities using open data

Antonella Carbonaro

University of Bologna - Bologna (Italy)

(submitted: 07/01/2020; accepted: 13/04/2020; published: 30/04/2020)

Abstract

Deploying ad-hoc learning environments to use and represent data from multiple sources and networks and to dynamically respond to user demands could be very expensive and ineffective in the long run. Moreover, most of the available data is wasted without extracting potentially useful information and knowledge because of the lack of established mechanisms and standards. It is preferable to focus on data availability to choose and develop interoperability strategies suitable for smart learning systems based on open standards and allowing seamless integration of third-party data and custom applications. This paper highlights the opportunity to take advantage of emerging technologies, like the linked open data platforms and automatic reasoning to effectively handle the vast amount of information and to use data linked queries in the domain of cognitive smart learning systems.

KEYWORDS: Smart Learning Systems, Linked Open Data, Cognitive Computing in Education

DOI
https://doi.org/ 10.20368/1971-8829/1135239

CITE AS
Carbonaro A., (2020), Enabling smart learning systems within smart cities using open data.Journal of E-Learning and Knowledge Society, 16(1), 72-77.
https://doi.org/ 10.20368/1971-8829/1135239

1. Introduction

The connection between smart learning systems and smart cities (SC) is strong. From 2010 the interest in SC began to grow exponentially. In (Harrison et al., 2010) Harrison defined a SC as an instrumented, interconnected, and intelligent city. Instrumented refers to the collection and integration of data from the use of sensors, applications, personal devices, and other resources. Interconnected refers to the integration of all such data and to the provision of a set of services. Intelligent refers to the complex elements, such as analytical calculations, modelling and visualization of services for better operational decisions. Some important technology trends are favouring the dissemination of data in SC: i) data provision from sensors that are smaller, less expensive and more reliable ii) sensors and data processing capabilities that are increasingly interconnected via the Internet of (every) Thing iii) open data innovation leading to the creation of a global data space containing billions of assertions (Lim et al., 2018). Only by making data linked and open will it be possible to integrate them and establish truly interoperable SC. The same features describe smart learning systems: available educational data can be collected and integrated in order to enhance decision-making and deliver better learning resources to the users, in particular, personalisation and inference reasoning (Andronico et al., 2004). In the context of technology-enhanced learning, such a data model could be employed to analyse information from multiple data sources, like generic or domain-specific datasets, and unify them in an interlinked data-processing area, in a Linked Open Data (LOD) approach. In this way, smart learning systems can effectively provide services by the public value it creates for users (Riccucci et al., 2007). Frequently, the intrinsic potential of the available online educational data and datasets are wasted without extracting potentially useful information and knowledge because of the lack of established mechanisms and standards (Riccucci et al., 2005). Rather, they can be exploited using sophisticated data analysis techniques such as automatic reasoning to find patterns and extract information and knowledge (Carbonaro, 2010), (Coccoli et al., 2016, 2017), (Gomede et al., 2018), (Muniasamy and Anandhavalli, 2020). Moreover, education information sharing and analysis in conjunction with non-traditional data sources (e.g., social media, web content, and linked data) can provide an important component to facilitate the development of the next generation of smart technology-enhanced learning services. However, due to the high heterogeneity of data representation and serialisation formats, and a lack of commonly accepted standards, the education landscape is characterised by the ubiquitous presence of data silos, which prevents domain experts from obtaining a consistent representation of the whole knowledge. Without a shared data model for such concept integration, it is impossible to actuate automatic data analysis processes like inference reasoning, especially within inter-domain contexts.

This paper aims to illustrate the characteristics that a framework for the integration and interoperability of data and concepts in the field of cognitive smart learning systems should have. This paper highlights the opportunity to take advantage of emerging technologies, like the LOD platforms to effectively handle the vast amount of information and to use data linked queries. One of the most important features of the linked data is their ability to enrich internal data archives with external data that could come from open data, statistics, DBpedia, Massive Open Online Courses and other LOD cloud sources. In particular, we illustrate a promising approach and an example of its application in the context of an Introduction to Algorithms and C Programming Language course taken as an example addressing the personalisation of all learning process components to different learners' characteristics. The paper is organized as follows: the next Section shows recent research efforts to apply LOD principles and technologies to solve problems around learning environments, such as the interoperability of educational data and resources, enrichment of material and recommendation and personalisation content. Section 3 describes the proposed creation of a common model interconnecting a variety of heterogeneous data sources and establish a personalized ontology-based framework, considering a C programming case study. Section 4 presents built ontologies used to formally describe the entities and relationships involved in concepts and the associated background knowledge. Section 5 shows how SPARQL and SWRL can be used to create innovative personalised-learning content systems that are connected to the SW. Finally, some considerations and conclusions close the paper.

2. Open and Connected Smart Learning Environment

LOD principles have been adopted by an increasing number of data providers over the last years, about geographic locations, people, companies, books, scientific publications, proteins, online communities, etc. Topics such as ontology building, use of semantics technologies and future applications that will be supported by these technologies are becoming important research areas in their own right (Ristoski and Paulheim, 2016). Open data has significant potential to foster innovation (Carbonaro, 2012), (Carbonaro et al., 2018) (Bizer et al., 2011)). For example, innovation and significant economic and social value can be created by using datasets such as map data, public transport timetables, statistical data and data on international trade or crime. To carry out this data opening an agreement is necessary to harmonize and organize the structure or format, data catalogues and metadata, which allow the relationships between all open databases available to be searched for and established. The interconnection of the variety of publicly available data sources can significantly facilitate reuse, exploitation, and possible extension of data.

The mEducator project (http://www.meducator.net) demonstrates how the LD principles can be applied to model and expose metadata of both educational resources and services/APIs. The metadata of educational resources, retrieved from different services, are transformed from their native formats into RDF and are made accessible via URIs (Uniform Resource Identifiers). The mEducator metadata schema covers the most frequently used aspects of educational resources from basic ones such as title and descriptions to more sophisticated ones such as learning outcomes and licensing models. The results of the experimental evaluation demonstrated improved interoperability and retrievability of the resource descriptions, presented as part of an interlinked resource graph.

Vega-Gorgojo et al. (Vega-Gorgojo et al., 2015) underline the importance of use of LD for improving the visibility of course offerings, recommendation of educational material or expert matching; the study doesn't consider personalization according to the learning styles. The authors state that LOD movement promises to improve existing practices of system integration, resource sharing and personalization to support learning.

An interesting review conducted by Pereira et al. (Pereira et al., 2018) shows efforts to apply the LOD principles and technologies to solve known problems such as interoperability of educational data and resources, access to data for various analyses, enrichment of material, and recommendation and personalization content. They obtained evidence regarding how the results of the recent researches in semantic technologies in education, more specifically the movement of LOD, are being applied in practice.

ELaCv2 (Chrysafiadi et al., 2020) is a recent integrated adaptive educational environment that provides e-training in programming and the language 'C'; the system schedules dynamically the learning material for each individual learner and proposes which domain concepts are known and unknown and which domain concepts need revision.

From the above-mentioned papers, it is possible to underline some main issues: the semantic interoperability of smart learning connected concepts and their data is crucial but still poorly widespread. Often, concepts and data are represented in a disjointed context resulting into vertical application development.

3. Source Code Semantic Annotation, Interlinking and Enrichment

This paper suggests a smart learning framework addressing the personalisation of all learning process components (learning objects, relevant learning/teaching methods and preferred learning activities) to different learners' characteristics. Let's take as an example the domain of teaching and learning to program in a specific programming language, for example, C computer programming. To identify the different elements and key concepts of the C source code, a phase of analysis of the source code is fundamental. This is usually done by creating a concrete syntax tree (CST) and an abstract syntax tree (AST). The first tree represents the syntactic structure of a string according to a certain formal grammar; parsers convert a CST into a more contextualized AST for practical use. The different elements identified by AST including arrays, pointers, functions, variables and inline documentation are then annotated according to their type. Eclipse CDT and its class and concept framework were used for the parsing and AST construction. Semantic annotation process assigns metadata to source code elements to allow subsequent processing and understanding. The semantic annotation process begins after the parser has analysed the source code to identify the different constructs. The task of semantic annotation is to annotate the elements identified with the appropriate concepts defined in the used ontology. The annotation component models the Resource Description Framework graph and adds semantic descriptions using the SE ontology. The source code elements represented in the AST are used for construct RDF statements as triples <subject, predicate, object> or <subject, predicate, literal value> to semantically describe resources (in this case, resources are the elements of learning knowledge of the source code such as class, method, parameter, parameter, return like, etc.).

Considering that the re-use of ontologies and interconnection with other relevant entities promotes the interoperability of knowledge, where possible, re-used ontologies can be adopted and used in the community to produce network effects. Within the interlinking and enrichment process, different vocabularies are used to enrich the semantic description of resources annotated by the semantic annotation component. In the context of smart learning systems, two types of datasets should be considered: (1) those directly related to educational information, containing, for example, educational resources, institutional data and educational indicators; and (2) datasets that can be used in teaching and learning scenarios, while not being directly published for this purpose, for example, datasets from different domains (libraries, encyclopedias, etc.), such as the ones made available by the Europeana project (http://www.europeana.eu/), as well as the collection of British Museum, available as LD (Piedra et al., 2015). Some of the most cited datasets are DBpedia (http://wiki.dbpedia.org/), GeoNames (http://www.geonames.org/), MusicBrainz (https://musicbrainz.org/), Bio2RDF (http://bio2rdf.org/) and PubMed (http://pubmed.bio2rdf.org/ sparql). This most cited datasets, also known as data hubs in the LOD cloud, are datasets that are interlinked with many others. Interlinked datasets with hubs are essential, mainly to make data findable and reusable. The Dublin Core (http://dublincore.org/ specifications/) is one of the simplest and most widely used metadata schema. Originally developed to describe web resources, Dublin Core has been used to describe a variety of physical and digital resources. Dublin Core is comprised of 15 core metadata elements; whereas the qualified Dublin Core set includes additional metadata elements to provide for greater specificity and granularity.

One specific vocabulary is Learning Object Metadata Ontology (a mapping of IEEE LOM - Learning Object Metadata elements to RDF based on the LD principles) (http://kmr.nada.kth.se/static/ims/md-lomrdf.html). This standard focuses specifically on the syntax and semantics of digital or non-digital learning objects.

The number of academic institutions publishing LD is constantly growing, as the growing lists of partners at the LinkedUniversties.org and LinkedEducation.org websites show. An institution publishes information about the courses it offers through its website and makes them available in RDF format, such as Open University (Qing, Dietze et al., 2012).

4. Framework Personalisation

Semantic Web describes a new way to make resources content more meaningful to machines, whereas the meaning of data is provided by the use of ontologies (Reda et al., 2018), (Patel and Jain, 2019). Ontologies, as a source of formally defined terms, play an important role within knowledge-intensive domains overcoming the problem of interpreting homonyms and synonyms in different sources. Ontologies can also be reused, shared, and integrated across applications. They provide a common agreed understanding of the domain by specifying a formal representation of the entities and relationships involved in concepts and the associated background knowledge. Some recommendations are provided on the use of ontologies: describe resources with previously defined ontologies whenever possible, extend standard ontologies and, where necessary, create ontology mappings to reconcile terms. We used these recommendations and identified the main following C programming concepts: i) the preprocessor directives, which define actions to be performed before the compilation process, ii) the functions, which can be invoked in the programs, iii) expressions, which define pieces of code that can be reduced to a value, iv) modifications applicable to declarations of functions and variables, v) declarations, understood as both declarations of functions, variables and types, but also as expressions of the program control flow, vi) data types, both primitive and user-defined and vii) the variables, understood as all the memory locations accessible from the program. So, the built ontology includes the classes and properties characteristic of a C language source code. For example, we added the Conditional Directive class, Define Directive class, and Include Directive class as subclasses of the Directive class. We added the Dereferencing Expression class and Referencing Expression class as subclasses of the Expression class. We added the Enum Type Declaration class, Struct Type Declaration class, Union Type Declaration class and Type Alias Declaration class as subclasses of the Type Declaration class. Moreover, regarding properties, we added hasFunctionDeclaration property, hasTypeDeclaration property and hasVariableDeclaration property as subproperties of the hasDeclaration property. We added the isAliasedBy property as a subproperty of the isReferencedBy property. We added pointsTo as a subproperty of the owl:topObjectProperty.

Our framework personalises the learning environment and provides learners with suitable learning objects, learning activities, and teaching methods based on their different preferences and needs. All components are explicitly and precisely represented using ontologies. We built the Learner Model Ontology to describe all characteristics related to the learner, which influence how she/he interacts with a learning system. For example, the Learner Model Ontology comprises the PersonalInformation class, Learning-Style class, BackgroundKnowledge class, and Preferences class. There are different models to represent learning styles; including the Myers-Briggs Learning Style (Extroverts, Introverts, Intuitions, Sensing, Thinking, Feeling), VAK learning style model (visual [verbal], visual [non-verbal], auditory and kinaesthetic), and the Felder Silverman Learning Style Model (Sensory/intuitive, Visual/verbal, Active/reflective, Sequential/global). The developed Learning Activities Ontology represents the variety of learning activities that help learners learn better. For example, the learning activities ontology comprises the Experiments, GroupWork, LabWork and Lecture classes. Finally, the built Learning Object Ontology consists of different Modules that introduce the learning module and its suitability for various learners at various levels. For example, the learning object ontology comprises the DifficultyLevel class, LearningObjectType class, and Structure class.

5. Inference Mechanisms in Proposed Framework

The procedure of semantic annotation is applied to C programs on the basis of the built ontologies. The source codes are extracted from the archive of the programs that the author of this paper provides to first-year computer science students within the Introduction to Algorithms and C Programming Language course offered during the first semester at the University of Bologna, Italy. Input data represented in semi-structured formats are transformed into an RDF graph and the datasets are thus semantically annotated according to reference ontologies and stored within a triple-store server. Users can interact with the system using either web-based access or the SPARQL endpoint. SPARQL endpoint promotes single standardized access and creates an inter-linked source graph. When educational institution offers SPARQL endpoint, we can reach to its data using standard schemas and link its own data with other open education linked data based on "one data, multiple applications" idea. For example, we could use a SPARQL query to navigate the linked datasets and ask for each student, the courses that deal with the topics of interest, divided by concept. Listing 1 illustrates the code snippet of the described SPARQL query.

```
  SELECT ?Student ?Course ?ConceptName
(count(distinct ?Source) as ?Sources)
  WHERE {
     ?Student :isInterestedIn ?Concept.
     ?Source :hasConcept ?Concept.
     ?Concept rdfs:label ?ConceptName.
     ?Source :hasLearningStyle ?Learning.
     ?Course :hasLearningObject ?Source.
     ?Student :hasLearnerProfile ?LearningM.
     ?LearningM lm:isIntuitive ?Learning.
     ?Student a ?StudentLevel.
     ?Course :isFor ?StudentLevel.}
```

```
GROUP BY ?Student ?LearningM ?Course
?ConceptName
ORDER BY ?Student ?ConceptName ?Sources
```

Listing 1 - An example SPARQL query that retrieves the names of students and courses that contain concepts in which the students are interested.

We decided to investigate the SWRL potential in our application as an essential step to formalise the logic layer for ontologies. The existence of a standardised rule language for the SW and using it with ontologies will create innovative personalised-learning content systems that are connected to the SW. For example, through the SWRL rule below, it is possible to make the reasoner deduce in a fully automatic way which are the most suitable teaching materials for the specific student, based on his learning goal and learning style. Listing 2 illustrates the code snippet of the described SWRL rule.

```
Student (?p) ^ hasLearnerProfile (?p , ?lp ) ^
hasLearningGoal (?lp , ?g )
^ hasCategory (?lp , ?s ) ^ ComposedBySubject (?g ,
?subj )
^ ComposedByConcept (?subj , ?concept )
^ ComposedByLearningObject (?concept , ?o )
^ recommendedForLearner (?o, ?s )
--> canUse (?p , ?o )
```

Listing 2 - An example SWRL rule that retrieves the most suitable teaching material for the specific student, based on his learning goal and learning style.

6. Conclusions

Technology is playing a key role in SC providing them with robust solutions that are of benefit to citizens. SC aim to incorporate smart solutions in their industrial, infrastructural, social and educational activities. Managing SC with intelligent technologies can allow to improve the quality of the services offered to citizens and make all processes more efficient. Linked data offers common representation based on ontologies and vocabularies to allow understanding and interoperable access of information resources in SC. One of the most important features of the linked data is their ability to enrich internal data archives with external data that could come from open data, statistics, DBpedia, Massive Open Online Courses and other LOD cloud sources. In this paper, we have presented the use of a LD framework using ontologies in the educational processes and learning activities in an Introduction to Algorithms and C Programming Language course. The approach performs a semantic annotation process assigning metadata to source code elements and key concepts to allow processing and understanding. The procedure of semantic annotation is applied on the basis of the developed ontologies; SPARQL queries and SWRL rules can be used to navigate the linked datasets and perform inference reasoning. The proposed framework makes use of open data, published ontologies and domain knowledge to construct a domain ontology consisting of common constructs, concepts, and instances. SWRL, commonly used for building inference mechanisms in ontology-based knowledge systems, allows to represent additional attributes that cannot naturally be inferred using traditional ontological models. Reasoning on enriched information can lead to interesting considerations and conclusions.

References

Andronico, A., Carbonaro, A., Colazzo, L., Molinari, A., Ronchetti, M., & Trifonova, A. (2004) Designing models and services for learning systems in mobile settings, Mobile and Ubiquitous Information Access Workshop 2003, LNCS 2954, Springer-Verlag: Berlin Heidelberg, 90-106

Bizer C., Heath, T., Berners-Lee (2011). Linked data: The story so far. In Semantic services, interoperability and web applications: emerging concepts. IGI Global, 205–227.

Carbonaro A, F. Piccinini, R. Reda (2018) Integrating heterogeneous data of healthcare devices to enable domain data management. Journal of E-Learning and Knowledge Society 14(1).

Carbonaro A. (2010) Improving web search and navigation using summarization process, Communications in Computer and Information Science Volume 111 CCIS, Issue PART 1, 131-138.

Carbonaro A. (2010) WordNet-based summarization to enhance learning interaction tutoring. Journal of e-Learning and Knowledge Society 6(2), 67–74.

Carbonaro A. (2012) Interlinking e-learning resources and the web of data for improving student experience, Journal of E-Learning and Knowledge Society, 8(2), 33-44.

Carbonaro A., M. Ravaioli, (2017) Peer assessment to promote Deep Learning and to reduce a Gender Gap in the Traditional Introductory Programming Course, Journal of e-Learning and Knowledge Society, 3(13).

Chrysafiadi K., Virvou M., Sakkopoulos E. (2020) Optimizing Programming Language Learning Through Student Modeling in an Adaptive Web-

Based Educational Environment. In: Virvou M., Alepis E., Tsihrintzis G., Jain L. (eds) Machine Learning Paradigms. Intelligent Systems Reference Library, vol 158. Springer, Cham.

Coccoli, M., Maresca, P., and Stanganelli, L. (2016). Cognitive computing in education. Journal of E-Learning and Knowledge Society, 12(2), 55-69.

Coccoli, M., Maresca, P., and Stanganelli, L. (2017). The role of big data and cognitive computing in the learning process. Journal of Visual Languages & Computing, 38, 97-103.

Gomede, E., Gaffo, F. H., Briganó, G. U., de Barros, R. M., & Mendes, L. D. S. (2018). Application of Computational Intelligence to Improve Educationin Smart Cities.Sensors,18(1).

Harrison C., B. Eckman, R. Hamilton (2010) Foundations for smarter cities, IBM Journal of Research and Development, vol. 54, no. 4, 1–16.

Lim C., Kim, K.-J., P.P. Maglio (2018) Smart cities with big data: reference models, challenges, and considerations. Cities 82, 86–99.

Muniasamy, Anandhavalli; Alasiry, Areej. (2020) Deep Learning: The Impact on Future eLearning. International Journal of Emerging Technologies in Learning, ISSN 1863-0383, 15(1), 188-199.

Patel A. and S. Jain (2019) Present and future of semantic web technologies: A research statements, International journal of computers and applications, Talyor and Francis.

Pereira, K. C., Wolfgand Matsui Siqueira, S., Pereira B. Nunes, S. Dietze (2018) Linked data in Education: a survey and a synthesis of actual research and future challenges, IEEE Transactions on Learning Technologies, 11(3), 400-412.

Piedra N., Chicaiza, J., Lopez, J., E. Tovar (2015). Seeking open educational resources to compose massive open online courses in engineering education an approach based on Linked Open Data. Journal of Universal Computer Science, 21(5), 679–711.

Qing, H., Dietze, S., Giordano, D., Taibi, D., Kaldoudi, E., and N. Dovrolis (2012) The Open University's repository of research publications Linked education: interlinking educational resources and the web of data. in The 27th ACM Symposium On Applied Computing, Special Track on Semantic Web and Applications.

Reda R., F. Piccinini, A. Carbonaro (2018) Towards consistent data representation in the IoT healthcare landscape. In ACM DH'18: International Digital Health Conference, France.

Riccucci S., A. Carbonaro, G. Casadei (2007) Knowledge acquisition in intelligent tutoring system: A data mining approach. In Mexican International Conference on Artificial Intelligence. Springer, 1195–1205.

Riccucci, S., Carbonaro, A., Casadei, G. (2005) An architecture for knowledge management in intelligent tutoring systems. In: Proceedings of IADIS International Congress on Cognition and Exploratory Learning in Digital Age, 2005, 473-476.

Ristoski P., H. Paulheim (2016) Semantic Web in data mining and knowledge discovery: A comprehensive survey, Web Semantics: Science, Services and Agents on the World Wide Web. vol. 36.

Vega-Gorgojo, G., Asensio-Perez, J.I., E. Gomez-Sanchez (2015) A review of Linked Data proposals in the learning domain. Journal of Universal Computer Science, 21(2), 326–364.