RESEARCH

Semantic Web Innovations for Higher Education

BLINDED BLINDED

Correspondence: BLINDED
BLINDED, BLINDED, BLINDED, BLINDED
Full list of author information is available at the end of the article

Abstract

Advances in artificial intelligence in the area of knowledge representation and reasoning have allowed the Web of documents to evolve into the Semantic Web. This technology enriches Web resources with formal semantic information to give them meaning and to allow software agents to exploit them in a more intelligent way. The Semantic Web provides a huge amount of open free datasets, annotated with ontologies or formal shared representations, as well as innovative applications to exploit them. In this paper we discuss some of the main advantages of using Semantic Web in different activities involved in Higher education such as course and programs planning, personalization of learning, ontology use by students and teachers, support in learning resource search and in autonomous life-long learning and in learning communities and we present the results of more than a decade of research and development in Semantic Web enhanced learning in Higher Education both at (Blinded: Institution 1) and at (Blinded: Institution 2).

Keywords: Semantic Web; personalization; elearning platform; competency; automatic learning sequence; concept prerequisite identification; knowledge graphs

1 Introduction

The Semantic Web is a field of research in which Artificial Intelligence and Web Technologies converge. On one hand, the symbolic artificial intelligence has always search for ways to formally represent knowledge and to do rich inferences from this representation. In this quest, knowledge has been represented by decision trees, by rules, by scripts, by frames and finally by semantic networks or concept graphs which represent the knowledge as concepts and as relations between concepts. These relations express various semantics such as property, association, instantiation as well as taxonomic and partonomic relations. On the other hand, Web technologies evolved from the Web 1.0 which used metadata and hypertext markup language (html) tags, intended to add syntactic and visualization information to a web resource, towards markup languages like xml, which can carry information on the semantics and structure of the domain addressed by the resource. Resources started then to be manipulated not as documents but as data interconnected through semantic graphs by means of unique identifiers, URIs (unified resource identification). The convergence of advances in these two domains gave rise to the Semantic Web.

According to its creator, Tim Berners Lee, the Semantic Web is based on shared formal knowledge representations that can evolve and on software agents that can manipulate these representations. "For the semantic web to function, computers must have access to structured collections of information and sets of inference rules that they can use to conduct automated reasoning." (Berners-Lee et al., 2001). Thus, the main purpose of the Semantic Web is to introduce explicit descriptions about the meaning of resources, to allow the machines themselves to have a level of understanding of the Web's content (Castells, 2003).

The potential of the Semantic Web depends on the representation of knowledge and on the automatic manipulation of this representations by software applications.

1.1 The representation of knowledge in the Semantic Web

The representation of knowledge in the Semantic Web has evolved from lists of concepts, taxonomies or controlled vocabularies, up to topic maps, conceptual models, and ontologies. The schema presented in Figure 1 underlines that more expressive knowledge representation enables more sophisticated reasoning capabilities and more intelligent behaviors.

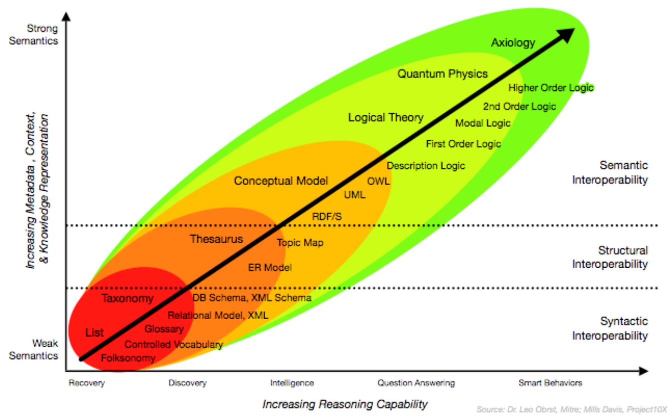

Figure 1 A Panorama of Semantic Representation Levels (source: Leo Obrst, Mitre and Mills Davis, Project 10X)

Today, Semantic Web applications are at the stage of conceptual models described in terms of ontologies. An ontology is defined as "a formal explicit specification of a shared conceptualization" (Gruber, 1993). Ontologies describe a specific knowledge domain by using a vocabulary of classes and relationships to describe the related entities. The main component of an ontology, called statement or triple is formed by two concepts and a relation between them. By combining statements, the ontology becomes a directed graph of concepts and relations, which can be further enriched with constrains and rules. The main Semantic Web languages for ontology representation are Resource Description Frameworks (RDF), RDF Schema and Web Ontology Language (OWL), while for the query of ontologies the main query language is SPARQL.

More and more ontologies are made available on the Web. Domain ontologies in basically any area of knowledge, from arts to science to commerce, have been developed and validated by experts, scholars and associations. For example, in science, the *BBC Wildlife ontology*[1] defines terms for describing the names and ranking of biological taxa, including phyla, families, and species, as well as providing support for describing their habitats, conservation status, etc. The *SNOMED ontology*[2] is

[1] BBC Wildlife ontology: https://www.bbc.co.uk/ontologies/wo
[2] SNOMED: http://www.snomed.org

a huge multilingual vocabulary in medicine. In 2011 it already grouped 311,000 concepts structured in taxonomies of diseases or health problems and symptoms. These concepts are linked together by more than a million links or relations for example describing the physiological location or the possible causes of a health problem. *The GoodRelations vocabulary*[3] for ecommerce is used to annotate Web pages describing a commercial entity in terms of its products or services, its location, commercial agents, delivery modes, prices, licensing conditions, guaranties, etc.

Although domain ontologies offer a great opportunity, the whole power of the Semantic Web is reached when these domain ontologies are open, freely accesible and connected to each other creating a cloud of linked open data (ONTOTEXT, 2017). The Linked Open Data (LOD) community, which started in 2007 with just a few open datasets published under Linked Data principles, has become a large space containing more than 1200 datasets[4]. There is great potential in this inter-linked data space; applications can benefit from a set of data in different domains, continuously updated and virtually unlimited. Currently, the main domains represented in the LOD are geography, government, life science, linguistics, media, publications and social networking.

The term "Knowledge Graph" (KG) has been recently used to refer to graph-based knowledge representations such as the ones promoted by the Semantic Web community with the RDF standard. According to Paulheim (2017), the term Knowledge Graph was coined by Google in 2012, and is used to refer to Semantic Web knowledge bases that describe real world entities and their interrelations through ontological schemas, that allow for potentially interrelating arbitrary entities with each other and that cover various topical domains. Although most RDF datasets published under the Linked Open Data initiative meet the first criteria, most of them do not cover different domains of knowledge. OpenCyc, Freebase, Wikidata, DBpedia and YAGO are identified as the main open Knowledge graphs on the Semantic Web. The most popular of them, DBpedia has more than 4.800.000 instances and 170.000.000 statements.

Figure 2 shows the concept "Artificial Neural Network" and some of its existing links in the DBpedia Knowledge Graph. The concept belongs to two categories which are part of one of DBpedia hierarchical structures, the one induced by the relation "broader". Concepts in a knowledge graph are also linked via non-hierarchical properties that express other semantic relationships as is the case of the "Machine Learning" concept in the example.

1.2 The automatic manipulation of the Semantic Web knowledge

Semantic Web ontologies and knowledge graphs represent knowledge that is available, structured, up to date, validated by experts and shared by communities and that can be exploited by humans for several tasks such as understanding and learning from a domain or organizing domain related resources when designing a course.

But, Berners Lee's vision of the functional Semantic Web goes beyond the human use of the ontologies. It includes the automated reasoning achieved by "software agents that collect Web content from diverse sources, process the information" and

[3] GoodRelations ontology: http://wiki.goodrelations-vocabulary.org/Quickstart
[4] Linked Open Data Cloud: https://lod-cloud.net/

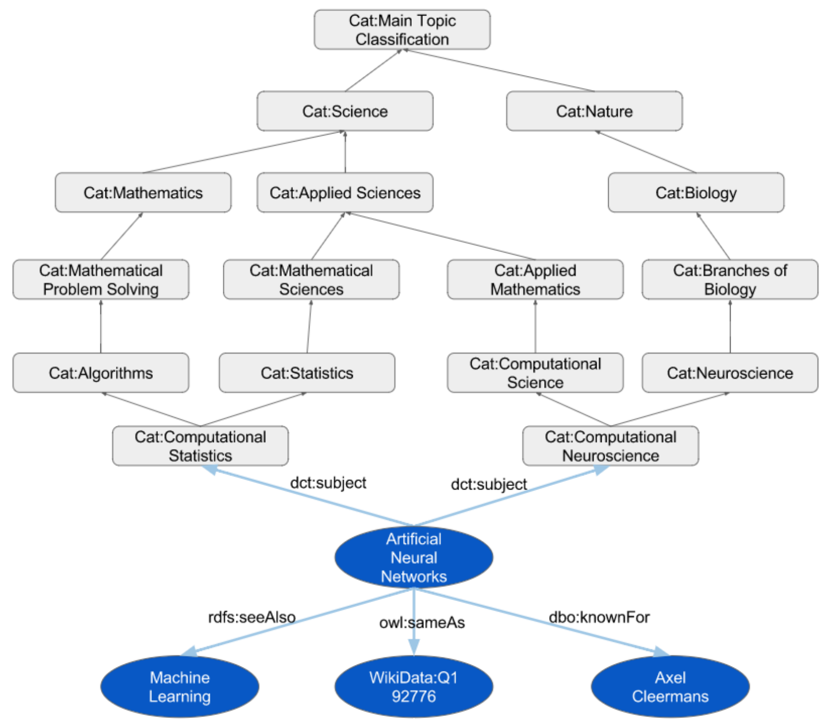

Figure 2 Excerpt of the background knowledge present in DBpedia for the concept "Artificial Neural Network" (arrows without label correspond to the predicate skos:broader).

either "carry out sophisticated tasks for users" or are Web services which "exchange the results with other programs" (Berners-Lee et al., 2001).

In this introduction, we presented the Semantic Web, we explained the main elements of this technology and we gave an overview of its current state. We showed the tip of the iceberg of the great potential of this technology for Higher Education by presenting several domain ontologies as well as the hugh amount of knowledge available in the linked open data cloud.

In the next section we explore further this potential by analysing benefits of using Semantic Web to enhance Higher Education. We reflect on these benefits from the perspective of the instructional designers, the teachers and the students.

In the following sections, we present our own innovative Semantic Web-based applications to enhance the learning and teaching in Higher Education. In section "Semantic Web tools for Learning Environments Design", we present two main applications that we developed at (Blinded: Institution 1) to produce more flexible, powerful, elearning and knowledge management environments. Section "Semantic Web tools for lifelong autonomous learning" presents the Semantic Web based set of tools that we developed at (Blinded: Institution 2) to automatically offer autonomous life-long learners adequate learning sequences.

2 Semantic Web opportunities to enhance Higher Education

In this section we discuss some of the ways in which the Semantic Web helps enhance Higher Education. After giving an overview of the way Information and Communication Technologies and in particular Artificial Intelligence have supported the evolution of education, we present four different levels in which the Semantic Web offers an opportunity to enhance current Higher Education: 1) Competency and knowledge based design of programs and courses: from the perspective of instructional designers and teachers to help them create competency based pedagogically solid programs and courses, populate them with pertinent resources and maintain them, 2) Personalized learning experience: courses built on top of competency and knowledge ontologies can also be designed for personalized learning, as student previous knowledge and competencies can be taken into account for adapting activities and recommending resources 3) Autonomous and lifelong learning: beyond structured courses, learners look for the Web in search of information to complete their learning process and the Semantic Web can help find and organize Web documents for a given learning goal and 4) Learning communities: the semantic annotating of resources and products of a learning community can help trace its evolution in terms o interest and connect it with other related communities creating a global community of learning communities.

2.1 Technology Enhanced Learning and the Semantic Web

The use of AI and the Semantic Web to support learning environments has evolved with the shift in the emphasis on various learning theories as well as with technological advances both in software engineering and in AI and Semantic Web. The literature identifies four major learning theories. For Behaviorism (Skinner, 1954) learning is to acquire transmitted knowledge from reliable sources. For Cognitivism (Newell and Simon, 1972) learning is processing information using oneself cognitive processes. For Constructivism (Le Moigne, 2001; Piaget, 1936), learning is to construct mental images of reality, not only through perception but by the actions of the learner. Finally, for Socio-constructivism (Vygotsky et al., 1978), learning is to construct meaning through social interactions and communication.

In the early ages of Computer-Based Learning, learning environments took generally the form of programmed instruction based on behaviorism. The learner would acquire small chunks of static content provided in sequences followed by multiple choice questions. The system would give the same feedback to all learners according to typical answers. These drill and practice systems were later enhanced to include a basic learner model as to personalize entrance points and complexity of exercices while still being alligned with behaviorism. Critiques of this programmed approach have encourage education scientists and learning environment designers to adopt more and more the three other learning theories. Mainly starting in the eighties and the nineties, this algorithmic approach moves towards more heuristic systems such as educational games, simulators and Discovery Learning Environments (Papert, 1980) that promote cognitivism and constructivism. Later on Collaborative Learning Environments (Goodsell, 1992) provided concrete applications of Socio-constructivism.

The area of Intelligent Computer-Aided instruction started with the use of expert systems as learning laboratories alligned with the constructivism (R.T. et al., 1988) while the rise of Intelligent Tutoring Systems tackled the issues of student model and personalization pushing the drill and practice systems to a higher level (Anderson and Pelletier, 2008; Wenger, 1987). All these systems were based on the symbolic AI paradigm and on closed, tailored graph or rule-based knowledge representations.

In the last decades, several advances both in educational models and in technology boost the possibilities to enhance education and in particular Higher Education. On the technology side, software engineering advances in model-driven and ontology-driven architectures as well as in service oriented frameworks, open the door to programmable learning portals and to the creation and easy adaptation of customizable Learning Management Systems. In the area of AI, the connectionist paradigm exploded with the success of machine learning algorithms. In this context, the rise of the Semantic Web (Baader et al., 2003; Berners-Lee et al., 2001) and the following emphasis on the Web of Linked Open Data (Allemang and Hendler, 2011) offer the possibility to develop semantic service oriented platforms (Carbonaro, 2020), intelligent-computer aided learning environment based on open knowledge representations (García et al., 2013) as well as semantic based personalization and recommendation (Figueroa et al., 2015), eventually enhanced with machine learning techniques (see section 4).

From the point of view of education, Millenial and Z generation students are challenging the classical model of teaching inside the class forcing Higher Education Institutions to innovate with models such as blearing, eLearning and Massive Open Online Courses (MOOC), to design competency based courses and curricula, to integrate information and communication technologies such as Web search engines and social networks in their learning activities and to help each learner find his/her own learning path. Here again, the Semantic Web offers great possibilities to address these challenges (Devedzic, 2004). In the rest of the section, we will dig deeper into these possibilities.

2.2 Using ontologies to plan courses and programs in Higher Education

It is the goal of every professor or educator to help their students develop more elaborated representation of their knowledge domain. Traditionally, professors preparing a course for Higher education will use a set of textbooks and articles, together with some search on the internet, in order to lie out a table of content for their courses, basically a hierarchical list of subjects. Within their interaction with students, they will present controlled vocabularies, glossaries and taxonomies, and sometimes a concept map of the knowledge domain. As shown in figure 1, ontologies provide stronger semantics and relationships between concepts within a knowledge domain. The Web of linked open data emphasizes also links between different knowledge domains. Domain ontologies such as the one presented in section 1 in biology, commerce and medicine, are the result of a careful analysis of a domain, sometime by large teams of scholars, so they are good candidates to prepare a Higher Education course or program (Quezada-Sarmiento et al., 2020; Tapia-Leon et al., 2019). If a professor cannot find an existing suitable ontology on the web, a good practice is to model one, using some ontology modeling tool. Ontologies have also been used as input to learning activities proposed to students, in order for them to explore the learning domain (see section 3). Finally, a big challenge of program and course designer is sustainability and the possibility of reusing learning resources. Since the work of Dietze et al. (2013), the Semantic Web based Learning community has been active in proposing ways to connect learning resources and Semantic Web, not only when designing courses but also for semi-automatic and automatic resources discovery, personalization and recommendation.

Besides ontologies for a subject domain's knowledge, many other uses of ontologies have been developed for Higher Education[5]. The *Bowlogna ontology* defines terms of a standard schema for European universities involved in the Bologna Reform aiming to ensure comparability in the standards and quality of higher-education qualifications. The *Academic Institution Internal Structure Ontology (AIISO)* provides a schema to describe the internal organizational structure of an academic institution; it is designed to work together with the *AIISO-Roles ontology* describing roles people play within an institution, and also the *Resource List ontology* describing academic courses and collections of references: such as reading lists, bookmarks, bibliographies. *SPAR, the Semantic Publishing and Referencing Ontologies*[6] offers the possibility to describe and relate bibliographic entities such as books and journal articles, to reference citations to the component parts of documents, and to various aspects of the scholarly publication process.

2.3 Semantic-based personalization design in Higher Education courses

Personalization refers to instruction that is paced to learning needs, tailored to learning preferences, and tailored to the specific interests of different learners. In an environment that is fully personalized, the learning objectives and content as well as the method and pace may all vary so personalization encompasses "differentiation and individualization"[7]. The advent of Massive Open Online Course (MOOC) has rendered personalization even more necessary than before. The same course can be followed by thousands of learners in various parts of the world, all with different background, knowledge and culture, which makes it difficult, if not impossible, to provide an efficient one-size-fits-all learning environment. The large number and the diversity of learners preclude providing close guidance by human tutors, as in distance learning university online courses. Commercially mature recommender systems have been introduced during recent years in popular e-commerce web sites such as Amazon or eBay with similar personalization intents. Yet, according to (Adomavicius and Tuzhilin, 2005), new developments must "include, among others, the improved modeling of users and items, and incorporation of the contextual information into the recommendation process".

Semantic Web technologies have been widely used to personalize learning experiences and provide recommendations (Bahmani et al., 2012; Baseera and Srinath, 2014; Brut and Sèdes, 2010; Jevsikova et al., 2017; Tarus et al., 2018), see also our own work presented in sections 3 and 4). One solution for personalization is to build an individualized learner path composed of resources selected according to a learner's learning objectives. Another solution is based on group collaboration in a learner's community sharing a predefined core scenario that can evolve and dynamically update learners' profiles, recommend adapted resources and activities

[5] Linked Education: https://linkededucation.wordpress.com/data-models/schemas/
[6] SPAR : https://opencitations.wordpress.com/2010/10/14/introducing-the-semantic-publishing-and-referencing-spar-ontologies/
[7] These concepts are defined in the 2010 US Education Technology Plan.

and adapt the initial scenario to individual profiles. Of course, both approaches can be combined in many ways.

In methods for personalization such as OWL-OLM (Denaux et al., 2005) or Personal Reader (Dolog et al., 2003), taxonomies or lightweight ontologies, have been used to tag resources with knowledge references in order to recommend resources to users. We can go a step further, complementing knowledge selected in a domain ontology with mastery levels, generic skills and performance levels references. Referencing resources with a set of concepts from a simple taxonomy of subjects would ignore the richer structures of OWL-DL ontologies. Furthermore, to state that a person has to “know” a concept does not say what exactly the person is able to do with the concept. For example, stating that someone “knows a certain device” may mean competencies ranging from “being able to describe its structure”, to “being able to recognize its malfunction”, to “being able to repair it”. Also, it is very different if a person is able to diagnose or repair the device in a familiar or novel situation, or with or without help; these are examples of performance indicators or criteria adding precision to a generic skill.

A competency model that extends a domain ontology is central for the personalization process enacted by one or more designers. This is the approach taken by our work in BLINDED:Institution1 and presented in section 3.

- In the first step of the personalization process, a specific competency model or profile expanding the domain ontology is built and resources are tagged with entry and target competencies selected in the model.
- Second, learners must be assessed for their actual competencies to enable comparison with entry and target competencies of the tagged resources. This operation can be done in various ways. The learner can self-assess his actual competencies using a self-assessment tool pre-configured with the list of target competencies provided by the design team, so he/she can enter performance levels for each, thus defining his own competencies. Automatic assessment can be achieved by building semi-automatically a Q&A test based on the list of the target competencies. Peer-assessment can be done by parameterizing a tool that selects a number of peer-assessors to evaluate a learner’s document for competency levels.
- Thirdly, a learner clustering operation can combine the various competency assessment results using a clustering policy specified by the designers. This policy will include the relative weights assigned to the various evaluation methods, a specified number of learner subgroups to be created and the criteria used to assign individual learners to a subgroup. It will create automatically the lists of subgroup members and notify each learner on his membership, his peers’ average competencies in his subgroup and the list of resources and activities recommended to the group. If massive historical data exists from a course’s previous deliveries, it becomes also possible to apply learning analytics techniques to create the subgroups.
- Fourthly, once a variant of a scenario is set for a team or subgroup or learners, the system can proceed to individual recommendation to learners by comparing each learner’s actual competencies with those of a critical module, activities or resources in the course scenario.

This process for recommendation and personalization is another important application of Semantic Web technology. It is based on a competency model, which extends domain ontologies with skills and performance levels, serving as a common semantic referential for activities, resources and learners in the learning scenario.

2.4 Help students navigate the “right” Web

Today students are digital citizen whose main source of information both for personal and learning tasks is the Web (Currie et al., 2010; Nadzir, 2015; Nikolopoulou and Gialamas, 2011). If the Web contains lots of useful information, the number of fake and inaccurate documents and data has been growing steadily in the last decades. Higher education students turn to general search engines such as Google or Yahoo as the first source of information and knowledge for a learning goal or research task, not always recovering relevant, up to date or even true information. While learning object repositories have tried to address this problem by curating and annotating learning resources, they cannot compete with the impressive amount of information on the rest of the Web and only a small number of students use these repositories as their first source of information. And yet searching for information has become a central activity in learning, to the point that the recent research on ”searching as learning” is focused on understanding and enhancing the learning processes that take place when searching online (Ghosh et al., 2018; Moraes et al., 2018; Tibau et al., 2018).

Providing Semantic Web based search tools like the ones presented in section 4 can help students navigate the Web the “right” way. These tools do not only select web resources based on a knowledge representation of the concepts addressed by these resources, but can also limit the document corpus of search to consider only curated documents (i.e. the ones referred to by google scholar, by TED or by conference paper repositories such as CORE). Moreover, other than finding the ”right” resources, to fulfill a learning goal, learners have to study them in an adequate order. Here again, the identification of prerequisite relations in the Semantic Web can help organize the recovered resources, as will be shown in section 4.

2.5 Semantic Web support for collaboration and learning communities

The Semantic Web technologies have a great potential to integrate and enrich social networks. This evolution towards the Social Semantic Web supports collaboration for learning in cMOOCs or other kinds of online learning environment based on community work, in schools or in organizations (Da Costa et al., 2017; Tiropanis et al., 2009). Semantic technologies can help learning communities by linking documents, data and applications involved in a variety of situations thus breaking up the silo effect in the Social Web. Conversely, online communities using Social Web software produce massive data and information that can be processed by Learning Analytic techniques to extract new knowledge for the Semantic Web, in order to create more intelligent applications than the ones available today.

3 Semantic Web tools for resource management, knowledge/ competency referencing, search and recommendation.

In section 2.1, we have underline the evolution in the support for learning theories, from Behaviorism to Cognitivism, Constructivism and Socio-constructivism.

In parallel, learning environments have moved from programed instruction software, to Intelligent tutoring systems (Anderson and Pelletier, 2008; Wenger, 1987), Discovery learning environments (Papert, 1980) and Collaborative Learning Environments (Goodsell, 1992).

The initial integration of AI in learning environment were at first based on Symbolic AI technologies, but at the turn of the century, new concepts had emerged from web-based software engineering such as programmable learning portals, service oriented frameworks, model-driven and ontology-driven architectures. The rise of the Semantic Web (Baader et al., 2003; Berners-Lee et al., 2001) and the following emphasis on the Web of Linked Open Data (Allemang and Hendler, 2011), have deeply influenced our work in order to design more flexible, powerful, yet user-friendly elearning environments.

In this section, we discuss two main applications of Semantic Web technologies (Domingue et al., 2011) developed at the (Blinded: Institution 1). They illustrate the use of the Semantic Web on two levels: for platform services execution and for knowledge and competency resource referencing, search and recommendation. More importantly, they provide tools to learning designers to implement cognitivist, constructivist and socio-constructivist learning environments.

3.1 Overview of TELOS, an ontology-based platform

TELOS, Technology Enhanced Learning Operating System (BLINDED, 2007) developed at (Blinded: Institution 1) is a Semantic Web system based on an explicit technical ontology structuring the objects to be processed by the system, acting as its executable blueprint. The execution of the services of TELOS is realized through queries to the system's technical ontology. Multi-actor scenarios provide the central learning environment aggregation mechanism grouping the actors, the operations they perform and the resources that they use or produce. The multi-actor scenario editor and its execution engine are a central piece of TELOS (Davies et al., 2003; Kleppe et al., 2003; Tetlow et al., 2006). They provide support for the design of cognitivist, constructivist, as well of socio-constructivist learning environments. They also provide support to upper level environments for the design of learning environments.

Figure 3 displays the user TELOS desktop interface in a Web browser, with three main tools open: the Resource Manager, the Scenario Editor and the Task Manager. The GMOT OWL Editor, a competency profile editor and other tools can also be launched from the TELOS desktop.

The TELOS Resource Manager serves to integrate and manage the resources that actors use or produce in TELOS, including the learners, individually or in teams, and also facilitators, that is professors, tutors, content experts or designers, whatever their role in the learning environment. All kinds of resources are classified into classes in the technical ontology embedded in the system to guide the execution of its services in a way adapted to the specificity of each class. For resources of type Scenario, the View and Modify functions open the scenario editor shown on the second window of figure 3, while the Run option starts the inference engine that will execute the scenario and present it to its users in the Task Manager. For resources of type TELOS Users, the View and Modify functions open a User Browser in order to view or enter personal information such as e-mails, photo, portfolio, etc. This tool is linked to an ePortfolio presenting the actual competencies of a user and some evidence (tests, productions. . .) of their acquisition. Information about users and other resources are available during the scenario execution processes. Software components are stored as Operations in the resource manager. Selecting such resources launch them during scenario execution to provide a variety of web services.

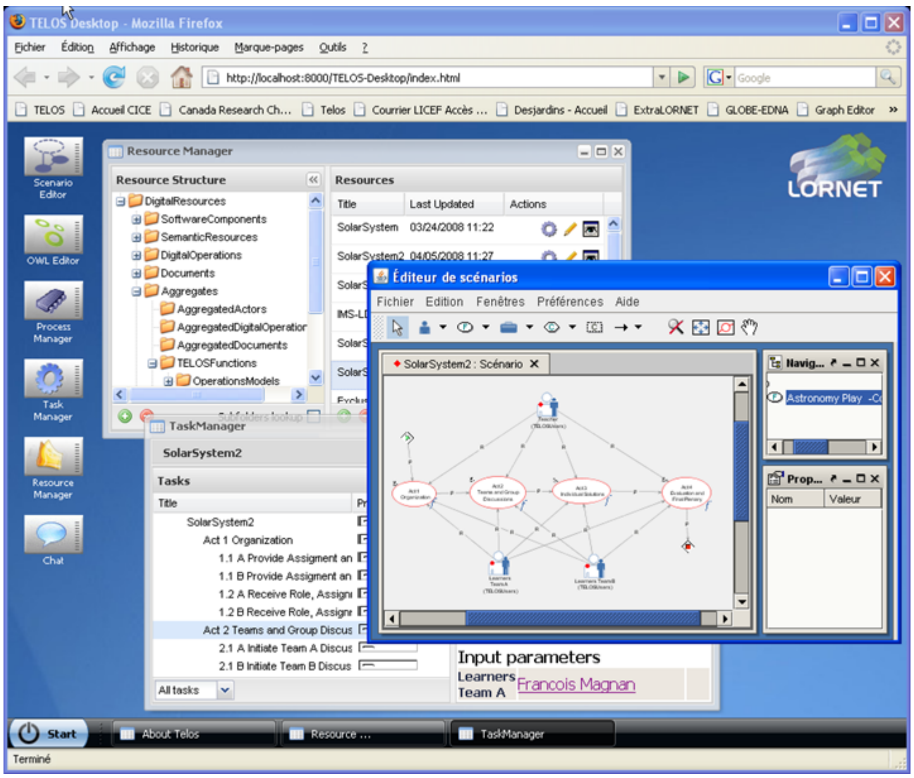

Figure 3 TELOS desktop with three important tools open.

The TELOS scenario Editor (BLINDED, 2010) enables the construction of visual workflow graphs using the MOT visual language where concept symbols represent all kinds of resources: documents, tools, semantic resources, scenarios, actors, activities, operaions, and data. MOT procedure symbols represent Functions that are aggregates of resources achieving together part of a scenario. Functions can be decomposed into “smaller” functions at any depth down to Activities enacted by humans, or Operations performed automatically by the system. MOT actor symbols represent users, groups, roles or software agents, all seen as control objects that rule the activities using and producing resources as planned by the scenario model. The condition symbols (not shown on figure 3) represent control elements inserted within the basic flow to decide on the next functions, activities or operations to be executed.

The TELOS scenario language is a high-level visual programming language. This generic language is designed for various levels of TELOS users for example an instructional designer building a course in higher education or a technologist building an upper level design process. Two Ph.D. doctoral students have used the TELOS scenario editor in this way to build new instructional design processes within their successful doctoral work. The first one has built a process to supports a course designer that has to deal with cultural diversity in a learners' group (BLINDED, 2013). The second one has built a design process for the integration of personalization features in a MOOC learning environment (BLINDED, 2017a).

The TELOS Task Manager shown on figure 3 is used by learners and facilitators to interact with some scenarios at run time. It will automatically open for its users when a scenario selected in the Resource Manager is launch. The task manager, guided by the technical ontology, presents adapted interfaces potentially different for every participant in the scenario. For example, if so designed, the teacher will see all the documents, the progress task bars of all the learners involved in the scenario, while learners will see only the tasks they are involved in and only the documents or tools they are supposed to use or produce. This flexibility of the system is made possible through its Semantic Web design and execution.

3.2 Example of a social-constructivist scenario

The MOT scenario graph on figure 3 is presented on figure 4 to offer a more detailed example of a socio-constructivist scenario designed using the TELOS scenario editor. It is a sequence of four activities in which three actors participate: a teacher and two teams of learners. In Act 1, learners are provided with a list of target knowledge and competencies about planets of the solar system and they are organized in two teams, each member and the composition of each team being registered in the resource manager. In Act 2, the learners are provided with communication tools and they proceed with team and group discussions. In Act 3, they are to build individual solutions to some problems about planets. In Act 4, they will participate in a group evaluation and plenary forum.

On figure 4, we see two other screen shots, first a sub-model of Act 2, and second, a sub-model for the Team A discussion. In Act 2, each team is provided with different planet properties to explore, together with the objectives for the discussion. After a certain time set in the scenario, both teams will join in a group discussion forum to prepare their work for Act 3.

The third screen shot shows how the Team A discussion is designed. A chat operation has been selected that has four roles: participant, moderator, manager and observer. The teacher is assigned to all 4 roles and each learner in team A is assigned as a participant. During the Team A discussion, the teacher is enabled to stop the exchanges (not shown on the figure) and provide some help to the team members if needed. The Team B discussion proceed in a similar way but with other planet properties to analyze.

Other choices could have been made at design time, providing more complete liberty to learners or, on the contrary closer supervision by the teacher, and very different collaboration patterns could have be chosen. When the pedagogical choices have been made in the scenario editor, the environment designers need to tell TELOS what each icon in the graphic scenario represents in terms of its technical ontology, so that the system is able to process it. This is called the icon's execution semantic. An interface in the scenario editor provides a service to designers

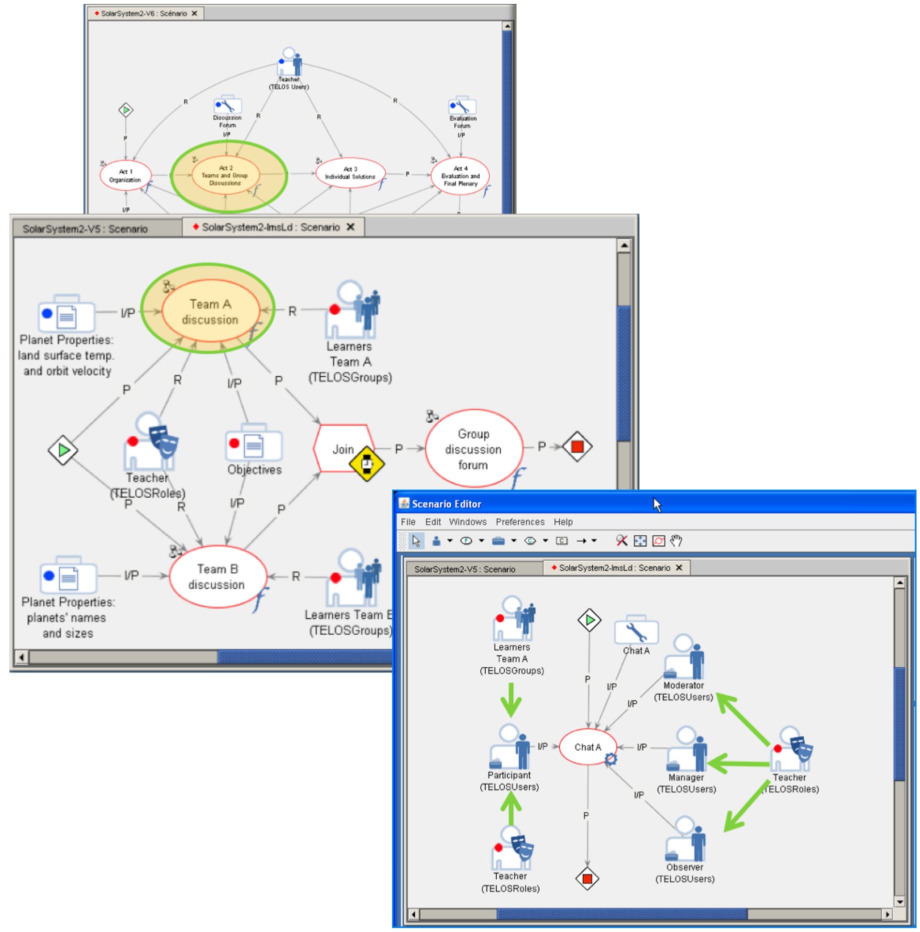

Figure 4 A socio-constructivist scenario in TELOS.

in order to set the semantic properties for each resource in the scenario. Using it, the designer can tells the system that a certain icon is an activity that has to be shown in the task manager at run time that another icon is a user or a group, or that still another icon is a document of a certain type. For example, a PowerPoint presentation (or a text, or a video) stored in the Resource manager has to be associated to the planet property icons in the Act 2 scenario, in order to be opened at runtime from the Resource Manager. The teacher icon will be associated to an individual actor described in the Resource Manager, and the group icons for the teams of learners will be associated each to a list of precise learners, previously entered by the scenario designer in the resource manager.

These operations involve important features of a Semantic Web system like TELOS that must provide all the capacities needed to design collaborative learning environments. In our example scenario, it enables the system to display in the task manager at run time to Team A members only the activities and resources assigned to Team A. These are different from those that Team B learners will see in their task manager interface. Without this capacity, a very important pedagogical property of the scenario would be lost if the two team would see the same resources. The teacher on the other hand will see both and can view the task bars that will help him monitor the learners' progress.

Another benefit from the ontology in TELOS is the learner and instructor improved efficiency in collaborative multi-role scenarios. Instead of losing too much time explaining verbally to each actor his or her role, the visual flow of activities and resources are transparent and situated globally, to clarify both learners' roles and facilitator roles as well. This is what a graphic scenario can provide as long as it is supported by a general ontology that provides the capacity, whichever role is selected for a user.

3.3 Referencing resources with knowledge and competencies

Another important use of Semantic web technology is the referencing of resources using domain ontologies augmented by a competency referential. Semantic referencing of resources aims at a number of important goals in TELOS: to inform learner and designers of the knowledge and competency a resource embeds; to enable search methods for resources based on their knowledge and competency properties; to inform recommendation agents so they can assist users performing certain activities and recommend them resources suitable for their actual knowledge and competencies.

The resource referencing process proceeds in three steps.

- First a domain ontology is selected in the Web of linked open data or is built directly by the designer using the TELOS ontology editor.
- Second, using the TELOS competency editor, a competency structure (Devedzic, 2006) is built, associating skills and performance levels to some knowledge elements in the domain ontology.
- Third, a resource referencing tool is provided in TELOS, to select resources in the Resource manager or in a scenario, and to associate to each resource a knowledge reference selected in the domain ontology and a competency selected in the competency structure.

Once this is done, these semantic references can be consulted in the Resource manager as properties of a resource. At design time, they can be viewed or edited from within the scenario editor. At run time, the resource semantic properties can be consulted in the Task manager.

Using this model, components of a scenario (actors, activities, resources) can be referenced using comparable competencies, based on the same domain ontology and competency structure. Resources and activities or tasks in a scenario are referenced by two sets of competencies, one for prerequisite competencies, and the other, for target competencies (i.e. learning objectives). Learners are referenced by their actual competencies which are regularly updated as learners consult resources or achieved activities in the scenario, gaining the target competencies referenced in these activities or resources.

3.4 Competency-based search and recommendation

To enable competency-based search for resources and provide recommendations to users, we need to be able to compare any two competencies $C1 = (K1, S1, P1)$ and $C2 = (K2, S2, P2)$, where K is the knowledge part, S is the skill's level and P is the performance's level. The system can evaluate the semantic proximity or nearness between $C1$ and $C2$, based on the respective positions of their knowledge parts in the domain ontology graph and on the values of the skill and performance levels.

We have developed such a competency comparison algorithm (BLINDED, 2012). A recommendation agent using this algorithm can be inserted at any point in a scenario. It can evaluate if a user's actual competency is very near, near, or far from the prerequisite or target competencies of a resource, in order to recommend using this resource or not. The agent can also compare the actual competencies of a user with other user's competencies to recommend pairing them in a team for certain activities. It can also evaluate if a competency is stronger or weaker than another one according to the levels of its skill and performance parts, or it can determine if the competency is more specific or more general according to the positions of the corresponding knowledge components in the ontology.

The knowledge and competency comparison algorithm is also used in TELOS to enable ontology-based search for resources (documents, tasks, users, etc.). We have implemented three different search methods. The first one is a simple keyword search using a knowledge or competency identifier; it can only retrieve resources with exactly the same identifiers in their properties. The second method queries the system with a knowledge or competency identifier in order to retrieve semantically near resources. The third method will also use this algorithm with a resource as the query; it will retrieve all the resources semantically near to this resource.

These recommendation operations are possible because TELOS is a Semantic Web system based on a technical ontology to which a domain ontology extended by a competency structure can be added and associated to users, activities and resources in a scenario or in resources stored in the Resource manager.

3.5 Resource management on the web of linked open data

After a decade of research and practice in the field of Open Educational Resources (OER) repositories, a number of limitations to their larger use have appeared. Many of these limitations require new approaches. In (Blinded: Institution 1), our tools for resource management evolved from a first-generation tool called PALOMA (BLINDED, 2004) based on traditional Learning Object Metatada using relational database technology, to the use of ontology-based annotations within the TELOS system. Finally, we moved to a tool, COMÈTE (BLINDED, 2015), using semantic technologies for the Web of linked open data (Heath and Bizer, 2011).

This second-generation, is being used in the colleges of Quebec for educational resource referencing and search. The biggest technical problem that was solved was the lack of interoperability between the various resource repositories that exist throughout the world. Some repositories use a proprietary schema (such as in TELOS), while a standardization effort has been made towards the Dublin Core or the Learning Object Metadata (LOM) standards. But LOM repositories use a diversity of application profiles, each different controlled vocabularies, so there are important difficulties to search for resources across various repositories, each described with different metadata schema and local terms. For example, a search in France or Spain to find resources for a course at a certain education level will not necessarily find resources in repositories in other countries where this education level does not exist or is labeled differently. Another problem is the identity coherence of authors or organizations linked to a resource that are referenced differently from one repository to another. Finally, in a knowledge domain like education or economy, not the same terms are used in different repositories for concept with the same or related meaning.

The solution to these interoperability problems is to reference resources using Semantic web technologies. The ISO/IEC 19788 standard (ISO-MLR, 2013) proposes an RDFS referencing schema where repositories can all be queried using the SPARQL Query language (SPARQL, 2013). This proposal is intended to provide optimal compatibility with both DC and the LOM repositories while preventing the proliferation of non-interoperable application profiles. Most important, it enables search within the Linked Open Data graph (LOD, 2019), using widely used data sets like DBpedia, Foaf, Geobase, etc. in order to reference resources using any LOD vocabulary including DC and LOM.

The COMÉTE Resource manager relies on these principles. It allows harvesting educational resources that constitute the heritage of an organization whatever metadata schema they use. It integrates the descriptions of the harvested resources in an homogeneous graph of RDF triple illustrated on Figure 5. Shown on the figure are interlinked icons for resources, authors or contributors, organizations they belong to and vocabulary/ontology references for the content of the resources.

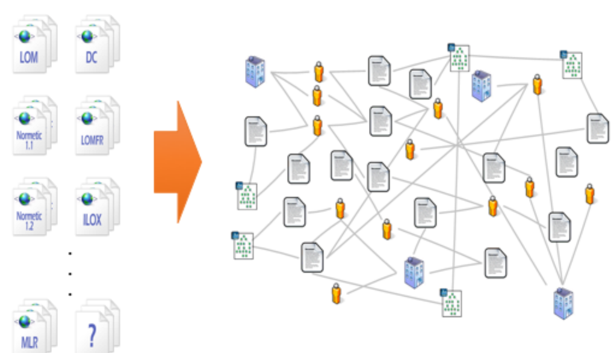

Figure 5 Integrating metadata records in RDF graphs

By various techniques, the system tries to maximize the inner coherence of the graph. Its Identity module takes care of the importation of identities representing persons or organizations, making sure each stays unique. The Vocabulary module implements the management of vocabularies, thesauri, and ontologies and manages correspondences between concepts and properties in ontologies from various sources. A useful example of alignment is the mapping between different school-level taxonomies of different countries to ensure the interoperability of resources between national repositories.

The use of Semantic Web technologies overcomes the interoperability problems between repositories. There is no need to stick to a one-size specification like the LOM with lots of fields to be filled and that compete with other norm. The resources are reference with their knowledge content and various other properties. The Web of linked data breaks the silo effect and offers one big repository where all kinds of repositories car be harvested and searched.

3.6 Educational benefit of Semantic web systems and tools

We now conclude this section by presenting some of the benefits that can be expected from Semantic web systems and tools.

- *Global Systemic View.* The technical ontology of TELOS can be seen as a kind of Virtual Campus model. It provides a global view to support the cohesion of the activities, from the upper level where an institution can create a global workflow to coordinate its major processes, to the lower levels where a professor designs scenario-based learning environments.
- *Extended set of actors.* Compared to the commercial learning and content management systems (LCMS), TELOS semantic design enables any set of actors and provides multi-actor coordination. As we have seen, this is essential, for example in discovery collaborative learning environments.
- *Visible scenarios and workflows.* Learning scenarios can be consulted in a visible graph. Links between resources, activities and actors are understandable rapidly, including their knowledge and competency references. Each user taking an actor's role can visually see the context of the activities he has to perform, what resources to use, what outcomes to produce.
- *Resource reusability* is a goal pursued by many advocates of learning object repositories, but it is not easy to achieve. Using ontologies to annotate each resource within the same linked open data (LOD) framework breaks down the silo effect and brings solutions to many reusability problems. For example the integration into COMÉTE's RDF graph of a proprietary resource repository lacking standardized metadata reference, makes it interoperable with other LOD vocabularies.
- *Focus on education.* The proposed approach has the potential to reduce the technology noise often present in eLearning applications when too much time is devoted to solving pure technology problems instead of focusing on education problems.
- *Fidelity from requirements to Code.* Capturing the main use cases and conceptual architecture concepts in an ontology driven system improves its fidelity to users' requirements. For example, the ontology-driven architecture of TELOS has facilitated its evolution when new concepts (e.g. competency) needed to be integrated in the system.

4 Semantic Web tools for lifelong autonomous learning

4.1 The challenges of lifelong autonomous learning

New generations are challenging the paradigms of education, and Higher Education is no exception. More and more adults are becoming lifelong learners. Even while following formal academic programs, this new generation is expanding its knowledge through knowledge acquisition processes outside of the classroom, using technologies as simple as a generic search engine (Kurt, 2018; Tsai et al., 2012).

The paradigm of lifelong learning supported by technology is redefining the classic tutor-oriented approach, moving from closed models in which objectives, content and sequence are predetermined, towards more open and self-directed scenarios. In the new era of learning, people take on learning processes that are not necessarily carried out under the guidance of a tutor (Brookfield, 2009; Morris, 2019; Selvi, 2009). In this format, learners are in charge of searching for, selecting and organizing the most appropriate resources to accomplish their learning goals. Learning resources are any type of resources relevant to a learning goal, and the main source of these resources is the Web. Indeed, even though most documents are not annotated as educational on the Web (as Learning Objects), they are still being used as learning resources. In the last years, the Web has taken a leading role as a provider of learning resources: Slideshare presentations, Youtube videos, question/answers in forums, blog posts and news articles are examples of Web resources that have been used for learning purposes. Resource-based learning (RBL) is the name given to the form of self-regulated learning that uses learning resources found on the Web (Hannafin and Hill, 2007; Helen, 2014). In lifelong learning, learners are continually pushed into an RBL scenario. In such a setting, self-directed learners face the challenge of effectively searching and selecting content that might not be educationally annotated, as well as organizing this content in a pedagogically sound way.

The selection challenge arises because learners do not always have enough literacy skills to do an effective and selective search (Hill, 2012). There is a great volume of resources available on the Web that can overwhelm learners, especially in unfamiliar domains. Learning resources on the Web are generally not well indexed because of the lack of metadata annotation. As a result, the answer of traditional search engines may not be adequate to the learning goal (Changuel et al., 2015). If selecting adequate resources is difficult for a self-directed learner, it is even more difficult for him/her to be able to organize the retrieved learning contents in a useful way for his/her learning purpose (Philipp Scholl and Steinmetz, 2007). Unlike traditional approaches, learners do not have the assistance of an expert to structure and organize the selected resources. And this is a main handicap as the order in which the knowledge is acquired is crucial for an effective learning because complex concepts require the understanding of more basic ones (Talukdar and Cohen, 2012). Indeed, this principle is based on a pedagogical theory called the Elaboration Theory. The Elaboration theory is a cognitivist model for instructional design, proposed by Charles Reigeluth and his colleagues in the late 1970s (Reigeluth, 1979; Reigeluth and Darwazeh, 1982). It provides a macro prescriptive framework for selecting, sequencing, synthesizing and summarizing the content. The key principle of the elaboration theory is that the content being taught should be organized starting from the simplest ones and then increasing the order of complexity by following the prerequisite relationships among involving concepts (Elsayed, 2015; Reigeluth and Darwazeh, 1982). Despite having been proposed more than forty years ago, this theory is still relevant. In fact, most MOOC courses contents are organized in a sequence following this principle, as has been validated in our evaluation.

4.2 A semantic web based approach for supporting lifelong learning

At (Blinded: Institution 2), we addressed the problem of supporting autonomous lifelong learning and proposed and developed semantic web based automatic strategies to select and organize resources using the rich open knowledge in the Linked

Open Data (LOD). The LOD provides a variety of structured knowledge bases freely accessible on the Web and interconnected to each other. As a result, there is a big amount of semantic aware machine-readable data available in the so-called LOD cloud that could be exploited to build more intelligent systems or to improve the effectiveness of existing algorithms (Musto et al., 2017). From this huge space of data, Knowledge Graphs (KGs) are particularly important since they concentrate the knowledge of multiple domains, and in general they specify a large number of interrelationships between concepts (Paulheim, 2017). Given such a constantly increasing amount of published and connected knowledge in Knowledge Graphs, we assume that the processes of selection and organization can be more effective, automatic and generalizable to multiple domains. Based on this assumption, we developed an automatic process that, starting with a learning goal stated by an autonomous learner, selects web documents and organizes them in a sequence that respects the Elaboration Theory. The whole process takes advantage of the background knowledge found in LOD Knowledge Graphs of the semantic web.

The main idea of our proposal is that the learner states his/her learning goal in natural language and is given as output a sequence of selected learning resources, appropriate for his/her learning goal. This sequence includes all possible learning paths to study the selected resources coherent with the elaboration theory, thus covering more simple concepts first. The learning resources are selected from a corpus of documents, which can be as wide as the whole web or as narrow as a particular learning objects repository such as videos in the Coursera MOOC platform or academic papers from the CORE database.

In a nutshell, the proposed solution follows three steps: first we annotate semantically, that is in terms of concepts and relations of a referential knowledge graph, both the learning goal and the resources of the considered corpus and based on this annotation we select the most relevant resources for the intended learning goal; then we identify prerequisite relations between concepts in the knowledge graph, and finally, using both the semantic representation of the selected resources and the prerequisite relations, we organize these resources in a sequence coherent with the Elaboration Theory, which means that resources dealing with prerequisite concepts are placed in the sequences before the ones that require these concepts.

Figure 6 presents the three components responsible for these three steps. As inputs, there is a set of resources (documents) and a learning goal expressed in natural language. The global output is a pedagogically sound sequence of resources proposed to attain the learning goal.

Each of the components of our approach has been developed and evaluated independently in different authors' previous works. The semantic representation was presented and evaluated in BLINDED (2017b,c, 2018a). The core concepts identification strategy using semantic representations was evaluated in BLINDED (2018b). The prerequisite concept identification is presented and evaluated in detail in BLINDED (2019a,b). The sequence strategies are presented in the doctoral thesis BLINDED (2019c). In the following sections, we present the most important aspects of each of the components.

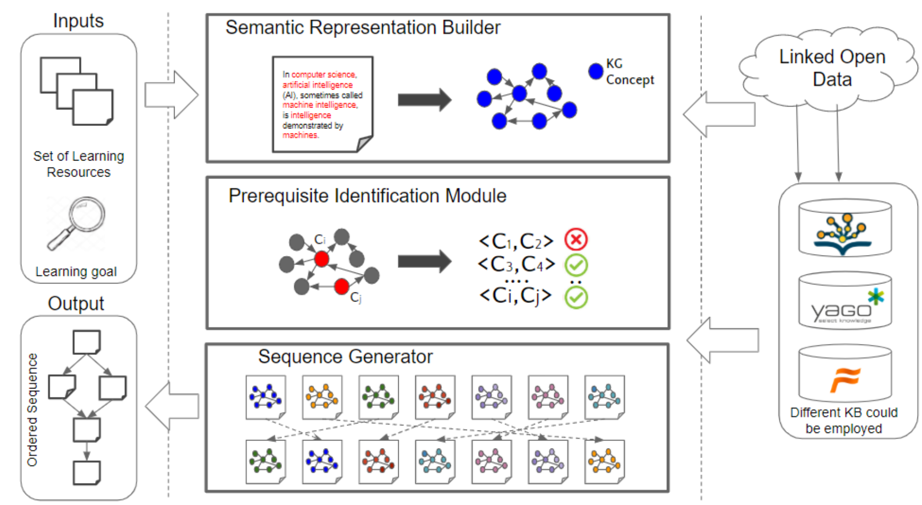

Figure 6 Systems components

4.3 Building the semantic representation of a document and finding its core concepts

Our first component is the semantic representation builder. The goal of this component is to identify the core concepts addressed by each resource.

Several previous works have addressed the challenge of automatically determining the main concepts of a resource or, the other way around, the "coreness" or importance of a concept in a resource (BLINDED, 2018a; Farhat et al., 2015; Foster et al., 2012; Jebali and Farhat, 2013; Roy et al., 2008; Sultan et al., 2014). Some of these systems use machine learning techniques based on text-based features while other take advantage of close, domain specific, expert built ontologies to guide the identification of concepts. The work proposed by Krieger et al. (2015) is, to the best of our knowledge, the only one using LOD knowledge graphs to identify the relevance of a resource for a learning context. However, they do not identify the core concepts of the resource, which for us is a critical information for selecting and organizing resources.

We developed two different paths to identify the core concepts addressed by each resource. The first path is the construction of a graph that represents the semantics of the resource, that is a structure that presents the concepts addressed in the resource, their relations and the importance of each concept in the resource. We call this structure the semantic representation of the resource. The second path uses machine learning techniques enriched with web semantic features extracted from the semantic representation of the resource to determine the "coreness" of a concept in a resource. Figure 7 presents the modules developed for both paths of this component.

4.3.1 Construction of a weighted directed graph of Knowledge Graphs concepts and relations to represent the semantics of a resource

The first path is the construction of the semantic representation of each resource. The semantic representation of a resource r_i is a directed graph G_i having as nodes concepts and as edges relations between concept. Both concepts and relations in

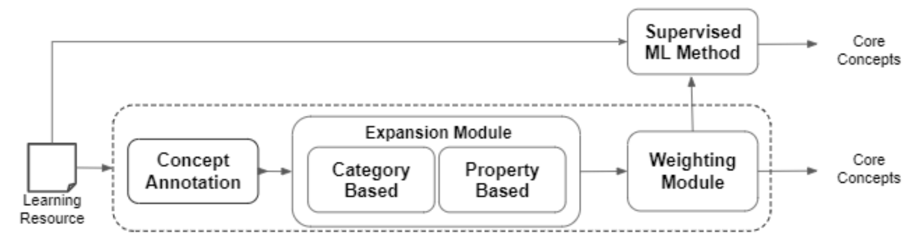

Figure 7 Core concepts identification paths

the representation correspond to entities found in the referential knowledge graph. Throughout the project, we used DBpedia as the referential knowledge graph; we will refer to it simply as "knowledge graph". The nodes and edges of the semantic representation have associated weights which correspond to their importance in the represented resource.

A Knowledge Graph consists of a set of concepts/entities C and literals L that are interrelated through a set of properties/predicates P. Under an RDF model, KG data consists of a set of statements $S \subset C \times P \times (C \cup L)$. Each $s \in S$ is a triplet composed of a subject, a predicate, and an object/literal. Considering the above, our semantic representation follows Definition 1.

Definition 1 *The semantic representation of a learning resource r_i is a directed weighted graph G_i. The set of nodes $N_i = \{c_1, c_2, ..., c_k\}$ are entities/concepts belonging to the space of a KG ($c_j \in C$). The node weight $w(r_i, c)$ denotes how relevant the node c is for the learning resource. A connection edge e between two nodes (c_a, c_b) represents the existence of at least one statement s in the KG that links both concepts. The weight of the edge $w(r_i, e)$ denotes how strong is this linkage between the considered concepts.*

The concept annotation module looks for concept mentions in the text (i.e. annotations) and links them to concepts in the Knowledge Graph. Entity linking and word sense disambiguation services such as DBpedia Spotlight[8], Aylien[9], or Babelfy[10] can be employed for this task. The result of this step is a list of concept identifiers (URI). As was stated by Waitelonis et al. (2015), incomplete or incorrect annotation may occur. To mitigate these problems and to enhance the representation, we propose an expansion and a weighting module.

Expansion is used to enrich the representation with concepts that are not explicitly mentioned in the text or were not identified by the annotation service. We expand the set of annotations (i.e. concepts found in the text) by following the taxonomical and the property links of each mentioned concept, in the Knowledge Graph. A taxonomical link connects a concept with its categories or connects a category with a broader one and a property link connects two concepts which are in a relation, for instance the concepts of space and speed.

[8] https://github.com/dbpedia-spotlight/dbpedia-spotlight

[9] https://aylien.com

[10] http://babelfy.org

Finally, the importance of each concept in the semantic representation and thus in the related resource, is evaluated via different weighting functions that take advantage of the graph-based structure of the representation. Three weighting functions were defined: concept frequency, semantic connectivity score and centrality measures. For each weighting strategy, a ranked list of concepts is generated.

As is done for the candidate resources, this whole process is also followed with the text stating the intended learning goal, for which a weighted graph and core concepts are also extracted. As an example of this first path, let's consider the following text defining the notion of Class in the domain of computer programming[11]:

"*In object-oriented programming, a class is an extensible program-code-template for creating objects, providing initial values for state (member variables) and implementations of behavior (member functions or methods). In many languages, the class name is used as the name for the class (the template itself), the name for the default constructor of the class (a subroutine that creates objects), and as the type of objects generated by instantiating the class; these distinct concepts are easily conflated.*"

After the concept annotation step, the following set of concepts are retrieved : "Constructor (object-oriented programming)", "Member variable", "Method (computer programming)", "Object (computer science)" and "Subroutine". Even though the text is the definition of the KG concept "Class (computer programming)", this very concept is not returned by the annotation service used. However, the concept is added via property-based expansion to the representation since it is connected to several recovered concepts. In the same way, the expansion module adds concepts and categories such as "Object lifetime", "Variable (computer science)" and "Category: Programming language topics". The final graph representation G_i , as shown in Figure 8, is then built using as nodes the resulting set of concepts after the expansion process. For the edge conformation, property paths between every pair of nodes in the Knowledge Graph are returned and analyzed via SPARQL queries. Basically, if a property path between two concepts is found, a directed edge in G_i is created following the direction of the connection in the Knowledge Graph. Following previous experimental results, we limit the search to property paths of length less than or equal to 2 (BLINDED, 2018c).

A normalized score that considers the number of property paths found is used as edge weight $w(r_i, e)$. In Figure 7, red nodes are annotations while green nodes are concepts incorporated via the expansion module. The thickness of the edge indicates how strong is the link between two concepts via $w(r_i, e)$. The size of the node is proportional to the node weight $w(r_i, c)$ and it is used as an indicator of its importance in the representation. $w(r_i, c)$ is obtained in the final weighting module.

4.3.2 Machine learning techniques based on semantic features to find the core concepts of a resource

Several researches on the identification of the importance of a concept in a document use supervised machine learning approaches (Sultan et al., 2014). In these systems, the features considered are text-based feature such as the relative position of a concept in a sentence or whether a concept appears in the title of the

[11] https://en.wikipedia.org/wiki/Class_(computer_programming)

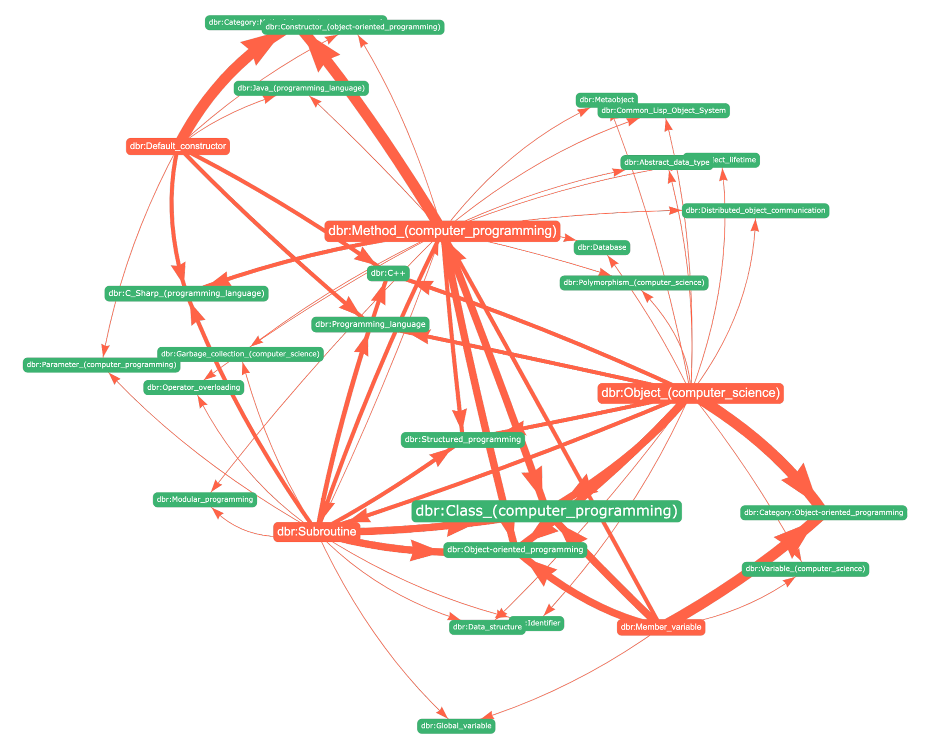

Figure 8 Semantic representation example

document. As them, we decided to try a supervised machine learning approach to identify the coreness of a concept in a document. The main innovation of our approach is that we enhanced the set of features to include features related to the semantic representation of the documents, thus extracted from the previous generated semantic representation. More precisely, given a pair concept-learning resource (c, r_i), we assign the "coreness" score via a supervised regression model which we called $CoCoDisK_{superv}$. For each concept - learning resource pair (c, r_i) - we calculate four types of features: text-based, graph-based, semantic similarity-based and complexity-based features. We only consider concepts in the semantic representation G_i of the resource r_i.

$CoCoDisK_{superv}$ requires a training process that involves the selection and configuration of the regression model. In order to train the model, we used the [CCI] dataset that was proposed in BLINDED (2018b). [CCI] is a dataset built for the identification of core concepts in learning resources. In [CCI] the coreness of a concept was evaluated via expert human knowledge and in order for a concept to be included as a core concept, at least two experts had to annotate it in their respective evaluations. On average each one of the 96 learning resources was annotated by at least 3 experts and contains on average 3.8 core concept annotations (374 core concepts in total). For each pair "concept-learning resource" the number of experts that identified it as a core concept is assigned as the output value. Since in [CCI] there are at least three annotations, the output range goes from 0 (no expert identifies it as a core concept) to 15 (all 3 experts identified it as a core concept) with equal intervals for intermediate values. The assignment of the output value is based on the simple idea that the more experts have identified it, the more likely it is to be a core concept and the higher their score. The video transcripts of the [CCI] learning resources are used as input for the semantic representation construction process. On top of the semantic representation the set of proposed features are extracted. We trained two ensemble methods widely used for regression tasks: Gradient Boosting regression (GBR) and Random Forest Regressor (RFR).

4.3.3 Evaluation

To evaluate our two proposals, a group of expert annotated 192 MOOC video lectures in three different languages: English, Spanish and French. The final dataset contains 419 annotations of core concepts of the video lectures. As baselines, we used SWAT[12] and TextRazor[13], two state of the art systems from the related area of salient entity analysis (Ponza et al., 2017). This area analyses important entities in documents where an entity is a Wikipedia page which can easily be related to a DBPedia concept. In general, supervised machine learning based approaches give better results than our unsupervised approach – the generation of a weighted semantic graph. Our supervised approach overcomes all baselines. It is worth noticing that all systems perform better with English resources than with Spanish and French ones, probably due to the low performance of the tagging services such as DBpedia spotlight in these languages. Finally, while supervised approaches give better results than our unsupervised proposal, this unsupervised approach does not require a training step, thus it might be preferred when time is an issue or not training step is feasible.

4.4 Identifying prerequisite relations between concepts using the LOD

The automatic identification of prerequisite relationships between concepts has been identified as one of the cornerstones for modern, large-scale online educational applications (Gasparetti et al., 2017; Pan et al., 2017; Talukdar and Cohen, 2012). Prerequisite relations exist as a natural dependency between concepts in cognitive processes when people learn, organize, apply, and generate knowledge (Laurence and Margolis, 1999). Recently, there has been a growing interest in automatic approaches for identifying prerequisites (Liang et al., 2015; Pan et al., 2017) and their applications in the organization of learning resources (BLINDED, 2018c), and automatic reading list generation (Fabbri et al., 2018). Most of these approaches take advantage of natural language processing techniques and machine learning strategies to extract latent connections among concepts in large document corpora to find prerequisite dependencies. Unlike previous approaches, our proposal uses open knowledge graphs as the main source to identify prerequisite dependencies. Thus, given a target concept c, the goal is to identify its prerequisites in the Knowledge Graph's concept space. As a Knowledge Graph can have millions of concepts, a first step consists of retrieving the candidate set of concepts that might eventually be prerequisites of the concept c. This candidate set is built with two types of analysis. The first type looks at the hierarchical links and includes all concepts that share a common category with c, while the second type traverses the non-hierarchical links

[12] https://services.d4science.org/web/tagme/swat-api
[13] https://www.textrazor.com

relating c with other concepts in the knowledge graph and includes all concepts found through non-hierarchical path of a given lengths.

This first candidate set is then reduced through a pruning process. Our concept pruning strategy is based on a simple measure that analyzes references between concepts. This measure, called RefD, was proposed by Liang et al. (2015) and is originally calculated to evaluate the degree to which a concept ca requires a concept c_b as a prerequisite. The main notion behind RefD is that if most related concepts of c_a refer to c_b but few related concepts of c_b refer to c_a, then c_b is more likely to be a prerequisite of c_a. In our approach, RefD was slightly modified to be applied to knowledge graphs. The modified version of RefD, called SemRefD, takes into account semantic paths in the knowledge graph to indicate the likelihood of one concept being a prerequisite of another (BLINDED, 2019b).

Once a candidate set is obtained the prerequisite relation between all possible pairs of concepts of the final candidate set and the target concept is evaluated. The evaluation is carried out by means of a supervised model proposed by BLINDED (2018c).

4.4.1 Evaluation

We used DBpedia 2016-10 as Knowledge Graph to implement both the search and the pruning. The hierarchical structure of a concept was drawn from categories in DBpedia categorical system. For all our experiments, we set $\beta = 0.5$ and $l_{max} = 2$ following previous experimental results presented in (BLINDED, 2017d). To evaluate our approach, we selected 15 target concepts in the domains of computer science and mathematics. For each concept, the entire process was carried out: initial sets of prerequisite candidates were obtained, which were then pruned using the SemRefD function. For a concept in an initial candidate set to be included in the final set, the value of SemRefD had to exceed a threshold value (teta). In the final step, all the concepts in each set were evaluated via a supervised model to validate the prerequisites relations with the target concept. Considering the complete set of target concepts and all possible candidate sets, a total of 582 different concepts pairs were evaluated by the supervised model. The candidate concepts identified as prerequisites constituted the output of the complete process. We built a ground truth to evaluate our proposal in terms of precision, false positives and true positives: 3 master students with computer science and mathematics backgrounds evaluated the 582 different concept pairs. Details of the evaluation are given in (BLINDED, 2019a).

The evaluation showed that the initial candidate sets should include concepts related to the target concepts through non hierarchical links of path 1, rather than concepts related through hierarchical links. Also, using the hierarchical version of SemRefD for the pruning step led in most cases to a high number of false positives and consequently to the lowest values of precision, while the non hierarchical version of SemRefD led to the lowest false positives values and the highest true positives and precision values The final precision obtained using the above functions varies between 83% and 92.9% and depends mostly on the configurable parameter (θ) of the pruning method. A theta value of 0.2 seems to be appropriate to regulate the trade off between the precision and the number of correct prerequisite concepts identified. Finally, is it worth noticing that the process relies heavily on the chosen Knowledge Graph, in our case DBpedia. Thus, new trending concepts with poor representation in the graph might give bad results.

4.5 Automatic generation of pedagogically sound sequences of resources to achieve a learning goal

Our last module is the sequence generator. This module orchestrates the prerequisite relationships calculations and the semantic representations of the learning resources, to produce a pedagogically sound (coherent with the Elaboration Theory) sequence of learning resources.

The problem of generating a sequence can be understood as that of generating a partial order over a finite set of elements. The problem in our case is that the ordering is based on pairwise prerequisite relations. This implies that the ordering directive does not directly relate the elements (i.e. learning resources), but concepts in their representation. Additionally, it is possible to extend the prerequisite relationship between concepts to precedence relations between learning resources. The sequencing strategies presented in this section are, therefore, based on prerequisite relations among concepts and precedence relations among resources.

We designed and implemented three different strategies (BLINDED, 2019c). The first strategy, called Simple Ranking method (SRM) relies on the idea of ranking each learning resource r_i. If the number of learning resources addressing concepts that are identified as prerequisites of the concepts found in r_i if high, r_i will be ranked high. This strategy is simple but it can be expensive as for every pair of resources, every pair of core concepts are compared pairwise. To reduce the number of prerequisite calculations, we proposed a second strategy called Bridge Ensemble Method (BEM). This method is based on the assumption that when a new concept is introduced in a learning resource, the resource is "expected" to mention its most important prerequisite concepts. The prerequisite concepts are therefore "bridges" towards the explanation of more complex concepts. Bridge concepts are introduced in an earlier learning resource but re-appear in a later learning resource when some new concepts are introduced. This implies that bridge concepts are the common concepts among learning resources.

In Figure 9, three learning resources (L_1; L_2; L_3) are presented. Each learning resource is represented by three concepts. The arrows indicate the existence of a prerequisite relationship between the concepts (e.g. c_3 is a prerequisite of c_4 and c_5). Under the assumption that to introduce a new concept its prerequisite concepts must be mentioned, there must be resources that share common concepts in their representation, "bridge" concept. In the figure, c_3 is a "bridge" concept between L_1 and L_2. The bridge concepts provide valuable information about the precedence relationship between pair of resources and, as a consequence, they are a good estimator of the appropriate ranking of the resource in the sequence. Consider the Venn diagram between L_1 and L_2. The prerequisite analysis between the bridge and non-bridge concepts shows a common direction of the prerequisite relation. In this case, it is clear that it is more appropriate to assign to L_1 a ranking higher than L_2. Now consider the Venn diagram between L_1 and L_3 in which the bridge concept is c1. The prerequisite analysis between the bridge and non-bridge concepts are calculated again. Unlike the previous case, there is no clear direction of the prerequisite relationship which is indicative that L_1 and L_3 could have similar rankings. Finally, the Venn diagram between L_2 and L_3 shows that there are no bridge concepts so their analysis does not really contribute to the ranking process. With this strategy, the number of prerequisite calculations is considerably reduced.

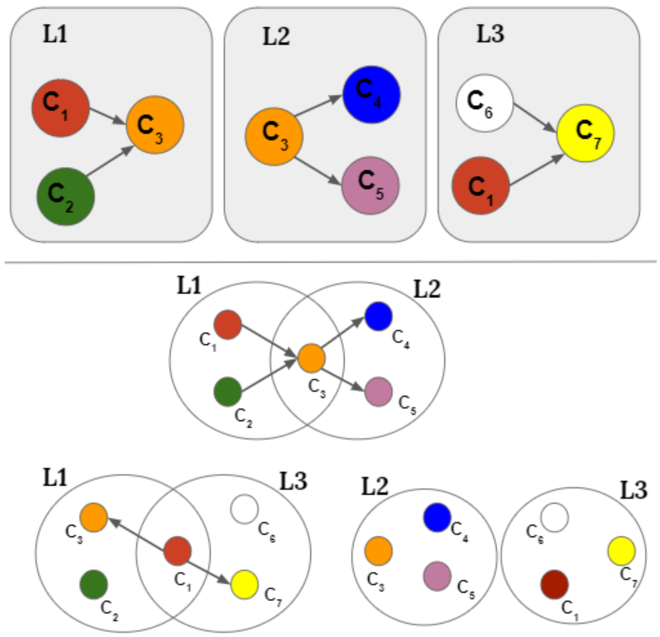

Figure 9 Bridge concept principle. Arrows indicate a prerequisite relation between concepts. c_3 and c_1 are bridge concepts. There are no bridge concepts between L_2 and L_3.

Our last strategy is called Unit Discovery Strategy (SUD). This strategy is based on the fact that courses are usually structured in units and the learning resources inside a unit share common concepts. To mimic this organization strategy, we designed an automatic strategy to group resources with similar semantic representations into units based on an agglomerative hierarchical clustering (AGC) algorithm (Gulagiz, 2017; Steinbach et al., 2004). After grouping the learning resources into units, the preceding strategies (SRM) or (BEM) are applied both to order the units treating them as resources and to order the resources inside each unit.

4.5.1 Evaluation

To evaluate our proposal, we generated a data set with resources (transcripts of videolectures and text documents) from 50 massive open online courses (MOOC) from Coursera platform. We selected 26 courses in computer science, 9 in electrical engineering, 4 in mathematics, 4 in physics and chemistry and 7 course from other different areas such as music, astronomy, and management. To validate the selection of courses validated by students and experts, the ratings given by students and the reviews of the community platforms MOOC-list [14] and Classcentral [15] were used. The minimum characteristics of a course in order to be included in the dataset

[14] https://www.mooc-list.com

[15] https://www.classcentral.com/

were to have at least 50 ratings (adding those of all the aforementioned platforms) with a total score of at least 4/5. Most courses were in English but we also had some in French and in Spanish. Resources labeled as optional in the courses were not included. The order of study of the learning resources suggested by the professor who designed the MOOC was used as the correct sequence to evaluate against our results. We obtained a total of 50 sequences with 3983 learning resources. On average, the length of the sequences was almost 80 learning resources, with a minimum of 29 and a maximum of 207.

The problem of comparing two sequences is known in the literature as the comparison of ranked lists. Among the different metrics employed for this task the Kendall Tau rank distance and the Spearman's footrule are current standards (Fagin et al., 2003; Kumar and Vassilvitskii, 2010). The Kendall tau rank distance is a metric that counts the number of pairwise disagreements between two ranking/ordered lists. There is a disagreement between two resources r_i and r_j in the sequences if in one of the sequences r_i appears before r_j but not in the other. Kendall tau will be equal to 0 if the two lists are identical and equal to 1 if one list is the reverse of the other. The Spearman's footrule looks for every resource r_i, its position in both sequences and calculates the absolute distance between these positions. The Spearman's footrule distance is the sum of these distances, normalized with the maximum possible distance.

We found four works in the literature referring to the problem of the automatic sequencing of learning resources (Changuel et al., 2015; Gasparetti et al., 2017; Shen et al., 2015; Siehndel et al., 2014). Among them, the closest to our work are the sequencing strategy proposed by Changuel et al. (2015) and the work of Gasparetti et al. Gasparetti et al. (2017). As neither the code nor detailed information was provided in these cases, we could not replicate their results so we opted to implement the proposals of Sheng et al. (Shen et al., 2015) and Siehndel et al. (Siehndel et al., 2014) and use them as baselines. In (Shen et al., 2015), learning resources are represented as a weighted bag of words in which the weight is given by a semantic measure calculated on WordNet. The strategy proposed in (Shen et al., 2015) requires the annotation of the learning resources with Wikipedia concepts. We used TextRazor as annotation service. Each learning resource is represented by a set of features that are extracted from the respective set of Wikipedia concepts pages. In all cases, our three strategies, even in their worst cases (worst calibration of the parameters) overcome the selected baselines and the difference is statistically significant ($\rho < 0.05$) using a two tailed t-test over Kendall Tau measure.

We find that our approach is appropriate for the generation of learning sequences. Nevertheless, the algorithms and strategies proposed can be used to create applications in other areas different from the one that motivates us. The proposed semantic representation and the sequencing based on prerequisite relationships serve to create search engines focused on learning processes. Our organizational strategy can serve as a more appropriate ranking method when looking for resources to learn. Our proposed strategies can also serve to build tools to support course designer processes. Both the selected resources and the sequence could facilitate the course design process by reducing the search and organization time.

4.6 Potential and limits of a Semantic Web approach to automatically organize Web resources in learning sequences for autonomous learners

We first summarize the benefits we expect from semantic web-based tools such as these to automatically support autonomous learning:

- *Personalization*: By autonomous learning we mean that the learner is in charge of his/her learning process. Be it a worker needing to learn new competences or a student in a Higher Education institution who needs or wants to get a deeper knowledge on a particular issue, the tools presented in this section allow him/her to define a personal specific learning goal and to obtain the sequence of resources to attain it.
- *Scalability*: One main goal of all work done recently in automating processes for intelligent search of resources, identification of prerequisites and sequencing and organizing resources recovered after a search, is to be able to attend a large number of autonomous learners with different learning goals, without having to design, with the help of experts, a particular learning sequence. As was shown with our evaluations: by enriching the process with semantics using knowledge graphs, one can outperform state of the art solutions based only on machine learning techniques.
- *Quality and Evolution*: as the linked open data movement continues to increase and new ontologies are linked to this cloud, we can expect even better quality in the results provided by this kind of tools. Moreover, as these knowledge graphs are open and constantly debugged by the concerned community, the knowledge they model may be considered the state of the art in each discipline and thus reliable as the backbone of the selection and organization of learning resources intended for someone to appropriate the domain.
- *Interdisciplinarity*: Today big world challenges are interdisciplinary by nature and the new generation of Higher Education learners should be prepared to tackle them, which is challenging the domain silos of traditional institutions. Semantic web tools based, not on isolated domain specific ontologies but, on the broad interconnected interdisciplinary open linked data cloud, can provide a coherent sequence of learning resources to deal with any learning goal of a particular learner, be it in a specific discipline or related to various domains.
- *Pedagogical foundation*: The high coincidence between our automatically generated learning sequences and the sequences proposed by current highly successful MOOC suggests that the Elaboration Theory is a pedagogically sound framework for structuring learning resources, which validates the pedagogical foundations of our proposal.

The results presented here show the great potential of knowledge graph-based Semantic Web applications in Higher Education. But these new opportunities still have some limitations. Our evaluation showed that the quality of the results was not even across disciplines and languages. The generated sequences for courses in English very more accurate than the sequences for French and Spanish courses, which could be explained by the fact that knowledge graphs in English are much larger than in other languages. Thanks to the steady growth of the LOD, we might hope that this problem will be solved in the next future. On the other hand, the generated learning sequences were more accurate for courses in which the appropriation of some concepts highly depends on prerequisites concepts, like the courses we used in computer science or mathematics. On the contrary, the sequence we generated for courses where the organization of resources does not seem to be guided by prerequisite relations was very different from the original one. This was the case with the course “Internet Giants: The Law and Economics of Media Platforms”. Although this might look as bad news, probably this means that the order of study of resources in this kind of courses does not affect the learning process. Finally, learning resources used as recapitulation or introduction were difficult to place in the sequence.

On a broader perspective, despite the very good results obtained and the benefits underlined, a learning experience is more than an adequate sequence of relevant learning resources. Taking into account the autonomous learner learning style and preferences as well as pedagogical strategies other than the Elaboration Theory should be a future challenge, if we expect really interesting learning experiences to be automatically generated and supported.

5 Conclusion

In the last ten years, the Semantic Web technologies have gone from laboratories to mainstream applications and the knowledge being represented in ontologies and knowledge graphs has grown from a few datasets to more than one thousand highly interconnected datasets, some with millions of instances and hundred thousands of links. We hope that the reflection on the four categories of semantic web-based opportunities for Higher education, presented in section 2, serves as a road map for future researches and developments in the fertile arena of the Semantic Web based Higher Education. Our own innovations presented in sections 3 and 4 are concrete examples of the way in which Semantic Web-based applications can enhance learning and teaching processes, while the related reflection on advantages and limits should help generalize our experience and identify challenges still to address.

Availability of data and materials

The datasets used and/or analysed during the current study are available from the corresponding author on reasonable request.

Acknowledgements

Not applicable.

Funding

The work in section 3 was supported by the Canadian NSERC and the Quebec Education Ministry. The work in section 4 was partially supported by COLCIENCIAS PhD scholarship (Call 647-2014).

References

Adomavicius, G., Tuzhilin, A.: Toward the next generation of recommender systems: a survey of the state-of-the-art and possible extensions. IEEE Transactions on Knowledge and Data Engineering **17**(6), 734–749 (2005). doi:10.1109/TKDE.2005.99

Allemang, D., Hendler, J.: Semantic web for the working ontologist (second edition), Second edition edn., pp. 1–12. Morgan Kaufmann, Boston (2011). doi:10.1016/B978-0-12-385965-5.10001-9

Anderson, J.R., Pelletier, R.: A development system for model-tracing tutors. In: Proceedings of the International Conference of the Learning Sciences. (2008)

Baader, F., Calvanese, D., McGuinness, D.L., Nardi, D., Patel-Schneider, P.F. (eds.): The Description Logic Handbook: Theory, Implementation, and Applications. Cambridge University Press, New York, NY, USA (2003)

Bahmani, A., Sedigh, S., Hurson, A.: Ontology-based recommendation algorithms for personalized education. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) **7447 LNCS**(PART 2), 111–120 (2012). cited By 3

Baseera, Srinath: Design and development of a recommender system for e-learning modules. Journal of Computer Science **10**(5), 720–722 (2014). doi:10.3844/jcssp.2014.720.722. cited By 3

Berners-Lee, T., Hendler, J., Lassila, O.: The semantic web. Scientific American **284**(5), 34–43 (2001)

BLINDED: BLINDED, pp. 254–268. Routledge/Falmer, ??? (2004)

BLINDED: In: Pierre, S. (ed.) BLINDED, pp. 79–109. Springer, London (2007)
BLINDED: Blinded. Technology ,Instruction, Cognition and Learning **7**(3-4), 263–286 (2010)
BLINDED: Blinded. (2012)
BLINDED: Blinded. In: Proceedings of Culturally-aware Technology-enhanced Learning (CulTEL), Paphos, Cyprus, September 17 (2013)
BLINDED: Blinded (2015)
BLINDED: Blinded (2017a)
BLINDED: Blinded. In: Proceedings of the International Conference on Web Intelligence. WI '17, pp. 1246–1252. ACM, New York, NY, USA (2017b). doi:10.1145/3106426.3109440. http://doi.acm.org/10.1145/3106426.3109440
BLINDED: Blinded. In: Solano, A., Ordoñez, H. (eds.) Advances in Computing, pp. 3–16. Springer, Cham (2017c)
BLINDED: Blinded. In: Różewski, P., Lange, C. (eds.) pp. 23–37. Springer, Cham (2017d)
BLINDED: BLINDED. In: Serrano C., J.E., Martínez-Santos, J.C. (eds.) pp. 177–191. Springer, Cham (2018a)
BLINDED: Blinded. In: Ichise, R., Lecue, F., Kawamura, T., Zhao, D., Muggleton, S., Kozaki, K. (eds.) (2018b)
BLINDED: Blinded, pp. 198–205 (2018c)
BLINDED: Blinded, vol. 2161-377X, pp. 332–336 (2019a)
BLINDED: Blinded. Smart Learning Environments **6**(1), 21 (2019b). doi:10.1186/s40561-019-0104-3
BLINDED: Blinded. PhD thesis (2019c)
Brookfield, S.D.: In: Maclean, R., Wilson, D. (eds.) Self-Directed Learning, pp. 2615–2627. Springer, Dordrecht (2009)
Brut, M., Sèdes, F.: Ontology-based Solution for Personalized Recommendations in E-learning Systems: Methodological Aspects and Evaluation Criterias, pp. 469–471 (2010). doi:10.1109/ICALT.2010.136. cited By 6
Carbonaro, A.: Concept integration to develop next generation of technology-enhanced learning systems. In: Rehm, M., Saldien, J., Manca, S. (eds.) Project and Design Literacy as Cornerstones of Smart Education, pp. 121–129. Springer, Singapore (2020)
Castells, P.: La web semántica. Technical report, Escuela Politécnica Superior Universidad Autónoma de Madrid (2003)
Changuel, S., Labroche, N., Bouchon-Meunier, B.: Resources sequencing using automatic prerequisite–outcome annotation. ACM Trans. Intell. Syst. Technol. **6**(1), 6–1630 (2015). doi:10.1145/2505349
Currie, L., Devlin, F., Emde, J., Graves, K.: Undergraduate search strategies and evaluation criteria: Searching for credible sources. New Library World **111**(3-4), 113–124 (2010). doi:10.1108/03074801011027628. cited By 44
Da Costa, A.F., De Carvalho, C.L., Ferreira, D.J.: A Collaborative Learning Environment Based on the Semantic Web Principles, pp. 256–260 (2017). cited By 0
Davies, J., Fensel, D., van Harmelen, F.: Towards the Semantic Web: Ontology-driven Knowledge Management. Wiley, ??? (2003). https://books.google.com.co/books?id=kREOBAAAQBAJ
Denaux, R., Aroyo, L., Dimitrova, V.: An approach for ontology-based elicitation of user models to enable personalization on the semantic web. In: Special Interest Tracks and Posters of the 14th International Conference on World Wide Web. WWW '05, pp. 1170–1171. ACM, New York, NY, USA (2005). doi:10.1145/1062745.1062923. http://doi.acm.org/10.1145/1062745.1062923
Devedzic, V.: Education and the semantic web. International Journal of Artificial Intelligence in Education **14**(2), 165–191 (2004). cited By 81
Devedzic, V.: Semantic Web and Education, p. 261. Springer, Boston, MA (2006). doi:10.1007/978-0-387-35417-0_1. https://doi.org/10.1007/978-0-387-35417-0_1
Dietze, S., Sanchez-Alonso, S., Ebner, H., Yu, H.Q., Giordano, D., Marenzi, I., Nunes, B.P.: Interlinking educational resources and the web of data: A survey of challenges and approaches. Program **47**(1), 60–91 (2013). doi:10.1108/00330331211296312
Dolog, P., Henze, N., Nejdl, W., Sintek, M.: Towards the adaptive semantic web. In: Bry, F., Henze, N., Małuszyński, J. (eds.) Principles and Practice of Semantic Web Reasoning, pp. 51–68. Springer, Berlin, Heidelberg (2003)
Domingue, J., Fensel, D., Hendler, J.A. (eds.): Handbook of Semantic Web Technologies. Springer, ??? (2011). doi:10.1007/978-3-540-92913-0. https://doi.org/10.1007/978-3-540-92913-0
Elsayed, A.M.: Effectiveness of using elaboration theory in teaching mathematics to develop academic achievement and critical thinking for primary students in oman. International Journal of Humanities and Cultural Studies **2**(3) (2015)
Fabbri, A.R., Li, I., Trairatvorakul, P., He, Y., Ting, W.T., Tung, R., Westerfield, C., Radev, D.R.: Tutorialbank: A manually-collected corpus for prerequisite chains, survey extraction and resource recommendation. In: ACL 2018, Melbourne, Australia, July 15-20, 2018, pp. 611–620 (2018)
Fagin, R., Kumar, R., Sivakumar, D.: Comparing top k lists. In: Proceedings of the Fourteenth Annual ACM-SIAM Symposium on Discrete Algorithms. SODA '03, pp. 28–36. Society for Industrial and Applied Mathematics, Philadelphia, PA, USA (2003). http://dl.acm.org/citation.cfm?id=644108.644113
Farhat, R., Jebali, B., Jemni, M.: Ontology based semantic metadata extraction system for learning objects. In: Chen, G., Kumar, V., Kinshuk, Huang, R., Kong, S.C. (eds.) Emerging Issues in Smart Learning, pp. 247–250. Springer, Berlin, Heidelberg (2015). doi:10.1007/978-3-662-44188-6_34
Figueroa, C., Vagliano, I., Rocha, O.R., Morisio, M.: A systematic literature review of linked data-based recommender systems. Concurrency Computation **27**(17), 4659–4684 (2015). doi:10.1002/cpe.3449. cited By 29
Foster, J.M., Sultan, M.A., Devaul, H., Okoye, I., Sumner, T.: Identifying core concepts in educational resources. In: Proceedings of the 12th ACM/IEEE-CS Joint Conference on Digital Libraries. JCDL '12, pp. 35–42. ACM, New York, NY, USA (2012). doi:10.1145/2232817.2232827
García, I., Benavides, C., Alaiz Moreton, H., Alonso, A.: A study of the use of ontologies for building computer-aided control engineering self-learning educational software. Journal of Science Education and Technology **22** (2013). doi:10.1007/s10956-012-9416-6
Gasparetti, F., Medio, C.D., Limongelli, C., Sciarrone, F., Temperini, M.: Prerequisites between learning objects: Automatic extraction based on a machine learning approach. Telematics and Informatics, (2017)

Ghosh, S., Rath, M., Shah, C.: Searching as learning: Exploring search behavior and learning outcomes in learning-related tasks. In: Proceedings of the 2018 Conference on Human Information Interaction & Retrieval. CHIIR '18, pp. 22–31. ACM, New York, NY, USA (2018). doi:10.1145/3176349.3176386. http://doi.acm.org/10.1145/3176349.3176386

Goodsell, A.S.: Collaborative Learning: a Sourcebook for Higher Education. National Center on Postsecondary Teaching, Learning, and Assessment, ??? (1992). https://books.google.com.co/books?id=GYVYAAAAYAAJ

Gruber, T.R.: A translation approach to portable ontology specifications. Knowl. Acquis. **5**(2), 199–220 (1993). doi:10.1006/knac.1993.1008

Gulagiz, F.K.: Comparison of hierarchical and non-hierarchical clustering algorithms. International Journal of Computer Engineering and Information Technology **9**(1) (2017)

Hannafin, M.J., Hill, J.R.: Resource-Based Learning. In: Handbook of Research on Educational Communications and Technology, 3rd edn. (2007). Chap. 40

Heath, T., Bizer, C.: Linked data: Evolving the web into a global data space. Synthesis Lectures on the Semantic Web: Theory and Technology **1**(1), 1–136 (2011). doi:10.2200/S00334ED1V01Y201102WBE001

Helen, M.: Personalized recommender systems for resource-based learning hybrid graph-based recommender systems for folksonomies. PhD thesis, Technischen Universitat Darmstadt (2014)

Hill, J.R.: In: Seel, N.M. (ed.) Resource-Based Learning, pp. 2850–2852. Springer, Boston, MA (2012). doi:10.1007/978-1-4419-1428-6_449. https://doi.org/10.1007/978-1-4419-1428-6_449

ISO-MLR: ISO-IED 19788 Information Technoogy – Learning, Education and Training – Metatda for Learning Resources Multipart Standard. ISO-IED, (2013). ISO-IED

Jebali, B., Farhat, R.: Ontology-based semantic metadata extraction approach. In: 2013 International Conference on Electrical Engineering and Software Applications, pp. 1–5 (2013). doi:10.1109/ICEESA.2013.6578408

Jevsikova, T., Berniukevicius, A., Kurilovas, E.: Application of resource description framework to personalise learning: Systematic review and methodology. Informatics in Education **16**(1), 61–82 (2017)

Kleppe, A.G., Warmer, J., Warmer, J.B., Bast, W.: MDA Explained: The Model Driven Architecture : Practice and Promise. Object technology. Addison-Wesley, ??? (2003)

Krieger, K., Schneider, J., Nywelt, C., Rösner, D.: Creating semantic fingerprints for web documents. In: Proceedings of the 5th International Conference on Web Intelligence, Mining and Semantics. WIMS '15, pp. 11–1116 (2015). doi:10.1145/2797115.2797132

Kumar, R., Vassilvitskii, S.: Generalized distances between rankings. In: Proceedings of the 19th International Conference on World Wide Web. WWW '10, pp. 571–580. ACM, New York, NY, USA (2010)

Kurt, A., Bulent, G.: Analysis of students' online information searching strategies, exposure to internet information pollution and cognitive absorption levels based on various variables. Malaysian Online Journal of Educational Technology **6**, 18–29 (2018)

Laurence, S., Margolis, E.: Concepts and cognitive science. In: Margolis, E., Laurence, S. (eds.) Concepts: Core Readings, pp. 3–81. MIT Press, ??? (1999)

Le Moigne, J.-L.: Pourquoi je suis un constructiviste non repentant. Revue du MAUSS **17**(1), 197–223 (2001). doi:10.3917/rdm.017.0197

Liang, C., Wu, Z., Huang, W., Giles, C.L.: Measuring prerequisite relations among concepts. In: Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing. Association for Computational Linguistics, Lisbon, Portugal (2015)

LOD: The Linking Open Data Cloud Diagram. (2019). http://lod-cloud.net, last consulted June 2019

Moraes, F., Putra, S.R., Hauff, C.: Contrasting search as a learning activity with instructor-designed learning. In: Proceedings of the 27th ACM International Conference on Information and Knowledge Management. CIKM '18, pp. 167–176. ACM, New York, NY, USA (2018). doi:10.1145/3269206.3271676. http://doi.acm.org/10.1145/3269206.3271676

Morris, T.H.: Self-directed learning: A fundamental competence in a rapidly changing world. International Review of Education **65**(4), 633–653 (2019). doi:10.1007/s11159-019-09793-2

Musto, C., Lops, P., de Gemmis, M., Semeraro, G.: Semantics-aware recommender systems exploiting linked open data and graph-based features. Knowledge-Based Systems (2017). doi:10.1016/j.knosys.2017.08.015

Nadzir, M.M.: Identifying the information-seeking behaviours among school of computing undergraduate students. Journal of Theoretical and Applied Information Technology **74**(2), 149–154 (2015). cited By 2

Newell, A., Simon, H.A.: Human Problem Solving. Englewood Cliffs : Prentice-hall, ??? (1972)

Nikolopoulou, K., Gialamas, V.: Undergraduate students' information search practices. Themes in Science and Technology Education **4**(1), 21–32 (2011)

ONTOTEXT: What Are Linked Data and Linked Open Data? (2017). https://ontotext.com/knowledgehub/fundamentals/linked-data-linked-open-data/ , last consulted June 2019

Pan, L., Li, C., Li, J., Tang, J.: Prerequisite relation learning for concepts in moocs. In: Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics, pp. 1447–1456 (2017). doi:10.18653/v1/P17-1133

Papert, S.: Mindstorms: Children, Computers, and Powerful Ideas. Basic Books, Inc., New York, NY, USA (1980)

Paulheim, H.: Knowledge graph refinement: A survey of approaches and evaluation methods. Semantic Web **8**(3), 489–508 (2017). doi:10.3233/SW-160218

Philipp Scholl, C.R. Doreen Mann, Steinmetz, R.: Support of Acquisition and Organization of Knowledge Artifacts in Informal Learning Contexts. In: In European Distance and E-Learning Network EDEN - Book of Abstracts (2007)

Piaget, J.: Le problème biologique de l'intelligence, (un des nombreux documents disponibles sur le site de la fondation jean piaget) (1936)

Ponza, M., Ferragina, P., Piccinno, F.: Document aboutness via sophisticated syntactic and semantic features. In: Frasincar, F., Ittoo, A., Nguyen, L.M., Métais, E. (eds.) Natural Language Processing and Information Systems, pp. 441–453. Springer, Cham (2017)

Quezada-Sarmiento, P.A., Enciso, L., Conde, L., Mayorga-Diaz, M.P., Guaigua-Vizcaino, M.E., Hernandez, W.,

Washizaki, H.: Body of knowledge model and linked data applied in development of higher education curriculum. In: Arai, K., Kapoor, S. (eds.) Advances in Computer Vision, pp. 758–773. Springer, Cham (2020)
Reigeluth, C.M.: In search of a better way to organize instruction: The elaboration theory. Journal of instructional development **2**(3), 8–15 (1979). doi:10.1007/BF02984374
Reigeluth, C.M., Darwazeh, A.: The elaboration theory's procedure for designing instruction. Journal of instructional development **5**(3), 22 (1982). doi:10.1007/BF02905492
Roy, D., Sarkar, S., Ghose, S.: Automatic extraction of pedagogic metadata from learning content. Int. J. Artif. Intell. Ed. **18**(2), 97–118 (2008)
R.T., W., E.T., W., R.D., D.: Theories of learning in intelligent computer-aided instruction. In: Estes, J. (ed.) The Fifth International Conference on Technology and Education. CEP Consultants, vol. 1, pp. 86–99 (1988)
Selvi, K.: In: Tymieniecka, A.-T. (ed.) Lifelong Learning and Self-Actualization, pp. 51–66. Springer, Dordrecht (2009)
Shen, S., Lee, H., Li, S., Zue, V., Lee, L.: Structuring lectures in massive open online courses (moocs) for efficient learning by linking similar sections and predicting prerequisites. In: INTERSPEECH 2015, 16th Annual Conference of the International Speech Communication Association, Dresden, Germany, September 6-10, 2015, pp. 1363–1367 (2015)
Siehndel, P., Kawase, R., Nunes, B.P., Herder, E.: Towards Automatic Building of Learning Pathways. In: Proceedings of the 10th International Conference on Web Information Systems and Technologies, pp. 270–277 (2014). doi:10.5220/0004837602700277
Skinner, B.F.: The science of learning and the art of teaching. Harvard Educational Review **24**, 86–97 (1954)
SPARQL: SPARQL 1.1 Query Language, W3C Recommendation,. http://www.w3.org/TR/sparql11-query/
Steinbach, M., Ertöz, L., Kumar, V.: In: Wille, L.T. (ed.) The Challenges of Clustering High Dimensional Data, pp. 273–309. Springer, Berlin, Heidelberg (2004)
Sultan, M.A., Bethard, S., Sumner, T.: Towards automatic identification of core concepts in educational resources. In: Proceedings of the 14th ACM/IEEE-CS Joint Conference on Digital Libraries. JCDL '14, pp. 379–388. IEEE Press, Piscataway, NJ, USA (2014)
Talukdar, P.P., Cohen, W.W.: Crowdsourced comprehension: Predicting prerequisite structure in wikipedia. In: Proceedings of the Seventh Workshop on Building Educational Applications Using NLP, pp. 307–315. Association for Computational Linguistics, Stroudsburg, PA, USA (2012). http://dl.acm.org/citation.cfm?id=2390384.2390423
Tapia-Leon, M., Aveiga, C., Chicaiza, J., Suárez-Figueroa, M.C.: Ontological model for the semantic description of syllabuses. In: Proceedings of the 9th International Conference on Information Communication and Management. ICICM 2019, pp. 175–180. ACM, New York, NY, USA (2019). doi:10.1145/3357419.3357442. http://doi.acm.org/10.1145/3357419.3357442
Tarus, J.K., Niu, Z., Mustafa, G.: Knowledge-based recommendation: a review of ontology-based recommender systems for e-learning. Artificial Intelligence Review **50**(1), 21–48 (2018). doi:10.1007/s10462-017-9539-5
Tetlow, P., Pan, J.Z., Oberle, D., Wallace, E., Uschold, M., Kendall, E.: Ontology driven architectures and potential uses of the semantic web in systems and software engineering. W3C Working Draft Working Group Note 2006/02/11, W3C (March 2006). http://www.w3.org/2001/sw/BestPractices/SE/ODA/
Tibau, M., W. M. Siqueira, S., Pereira Nunes, B., Bortoluzzi, M., Marenzi, I., Kemkes, P.: Investigating users' decision-making process while searching online and their shortcuts towards understanding. In: Hancke, G., Spaniol, M., Osathanunkul, K., Unankard, S., Klamma, R. (eds.) Advances in Web-Based Learning – ICWL 2018, pp. 54–64. Springer, Cham (2018)
Tiropanis, T., Davis, H., Millard, D., Weal, M.: Semantic technologies for learning and teaching in the web 2.0 era. IEEE Intelligent Systems **24**(6), 49–53 (2009). doi:10.1109/MIS.2009.121. cited By 36
Tsai, M.-J., Liang, J.-C., Hou, H.-T., Tsai, C.-C.: University students' online information searching strategies in different search contexts. Australasian Journal of Educational Technology **28**(5) (2012). doi:10.14742/ajet.822
Vygotsky, L.S., Cole, M., John-Steiner, V., Scribner, S., Souberman, E.: Mind in Society: Development of Higher Psychological Processes. Harvard University Press, ??? (1978)
Waitelonis, J., Exeler, C., Sack, H.: Enabled generalized vector space model to improve document retrieval. In: Proceedings of the Third NLP&DBpedia Workshop (NLP & DBpedia 2015) Co-located with the 14th International Semantic Web Conference 2015 (ISWC 2015), Bethlehem, Pennsylvania, USA, October 11, 2015., pp. 33–44 (2015). http://ceur-ws.org/Vol-1581/paper4.pdf
Wenger, E.: Artificial Intelligence and Tutoring Systems: Computational and Cognitive Approaches to the Communication of Knowledge. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA (1987)