

# Chapter 6

## Recommender Systems

**Luis Anido-Rifón, Juan Santos-Gago, Manuel Caeiro-Rodríguez, Manuel Fernández-Iglesias, Rubén Míguez-Pérez, Agustín Cañas-Rodríguez, Víctor Alonso-Rorís, Javier García-Alonso, Roberto Pérez-Rodríguez, Miguel Gómez-Carballa, Marcos Mouriño-García, Mario Manso-Vázquez, and Martín Llamas-Nistal**

**Abstract** The purpose of this chapter is to describe a software system that allows for discovering non-traditional education resources such as software applications, events or people who may participate as experts in some Learning Activity. Selecting the more suitable educational resources to create learning activities in the classroom may be a challenging task for teachers in primary and secondary education because of the large amount of existing educational resources. The iTEC Scenario Development Environment (SDE), is a software application aimed at offering supporting services in the form of suggestions or recommendations oriented to assist teachers in their decision-making when selecting the most appropriate elements to deploy learning activities in a particular school. The recommender is based on an ontology that was developed in a collaborative way by a multi-disciplinary team of experts. Its data set is fed not only from entries that come from registrations made by human users—using tools from the iTEC Cloud—but also from software agents that perform web scraping, that is, automatic enrichment of the semantic data with additional information that come from web sources that are external to the project. Therefore, the recommender system takes into account contextual factors when calculating the relevance of every resource. The SDE defines an API that allows third-party clients to integrate its functionalities. This chapter presents two success stories that have benefited from the SDE to enhance educational authoring tools with semantic web-based recommendations.

---

L. Anido-Rifón (✉)

Telematics Engineering Department, ETSI Telecommunication, University of Vigo, Vigo, Spain  
e-mail: [lanido@det.uvigo.es](mailto:lanido@det.uvigo.es)

J. Santos-Gago • M. Caeiro-Rodríguez • M. Fernández-Iglesias • R. Míguez-Pérez  
A. Cañas-Rodríguez • V. Alonso-Rorís • J. García-Alonso • R. Pérez-Rodríguez  
M. Gómez-Carballa • M. Mouriño-García • M. Manso-Vázquez • M. Llamas-Nistal  
University of Vigo (ES), Pontevedra, Spain  
e-mail: [jsgago@det.uvigo.es](mailto:jsgago@det.uvigo.es); [mcaeiro@det.uvigo.es](mailto:mcaeiro@det.uvigo.es); [manolo@det.uvigo.es](mailto:manolo@det.uvigo.es); [rmiguez@det.uvigo.es](mailto:rmiguez@det.uvigo.es); [agustincanas@det.uvigo.es](mailto:agustincanas@det.uvigo.es); [valonso@gist.uvigo.es](mailto:valonso@gist.uvigo.es); [jgarcia@gist.uvigo.es](mailto:jgarcia@gist.uvigo.es); [roberto.perez@gist.uvigo.es](mailto:roberto.perez@gist.uvigo.es); [miguelgomez@det.uvigo.es](mailto:miguelgomez@det.uvigo.es); [marcosmourino@gmail.com](mailto:marcosmourino@gmail.com); [mario.manso@uvigo.es](mailto:mario.manso@uvigo.es); [martin@det.uvigo.es](mailto:martin@det.uvigo.es)

**Keywords** Recommender systems • Multi-criteria decision analysis • Ontology • Information enrichment

## Introduction

In the current panorama of educational practice in primary and secondary education across Europe we find that technology is increasingly present in the classroom. On the one hand, we have government programs that provide classrooms with a technological infrastructure. For instance, the Abalar<sup>1</sup> project, financed by the Galician Ministry of Education provides classrooms with an interactive digital whiteboard, Wi-Fi Internet connection, and a laptop per student, in which a Linux distribution comes already installed and ready to be used. On the other hand, students themselves, usually have mobile devices—such as smartphones and tablets—and carry them everywhere, including the classroom.

In addition to hardware resources, nowadays we find an enormous amount of free software resources, ready to be used in the educational practice. Besides stand-alone applications, we can use many applications in the cloud, both from personal computers and mobile devices. Complete suites as that of Google<sup>2</sup> are freely available with zero cost, ready to be used in educational practice (Herrick 2009; Patterson 2007).

But the resources that may be used in educational practice are not limited to hardware and software. Many everyday events, especially cultural events, may have an educational value. As Redding (1997) states:

*Stimulating the child's desire to discover, to think through new situations and to vigorously exchange opinions, is fostered also by family visits to libraries, museums, zoos, historical sites and cultural events.*

We might think, for instance, of events such as theatre performance and lectures that may be very relevant to illustrate some points of the curriculum, and that can certainly be used in educational practice. If there is a free performance of Hamlet in our city, why do not use it as a resource for the subject of literature, especially if Shakespeare is in the curriculum? In a similar way, experts on particular topics are the best people to explain certain concepts. A doctoral student who is carrying out their Ph.D. in the area of genetic research might be very inspiring for secondary education students during their biology class.

This was the context for the work of the iTEC project which we report here. It contributed to the conception of the classroom of the future, in which technology is complemented with innovative pedagogical approaches, which entail a high degree of dynamism in educational practice. Thus, iTEC promotes an educational practice

---

<sup>1</sup> <http://www.edu.xunta.es/espazoAbalar/>

<sup>2</sup> <http://www.google.com/enterprise/apps/education/>

in which students interact in small projects which include participation in events, speeches with experts, with all of this seasoned by the use of technology.

In taking a step along the path toward iTEC's objective we were confronted by an initial difficulty: how do we select the technologies, events, and experts that will take part in an educational experience? Firstly, there is no central directory of technologies, events, and people at an European level, in such a way that a teacher may make searches in it. And, secondly, were it to exist, the difficulty of selecting between an enormous number of technologies, events, and experts would be very considerable.

In iTEC, a series of directories were developed in which technologies can be registered, as well as events and experts, which form part of the iTEC Cloud (see Chap. 4). Thus, the Composer (Simon et al. 2013) includes a directory for hardware and software technologies; the People and Events Directory (Van Assche 2012), as its name suggests, enables users to register educational events as well as experts in some knowledge area; and the Widget Store (Griffiths et al. 2012) is a repository of widgets ready to be used in the educational practice. Section "The iTEC cloud" briefly explains the components of the iTEC Cloud.

In order to solve the problem of selection from a large number of technologies, events, and experts, the iTEC project proposes the SDE, which is conceived as an artificial intelligence agent that uses Semantic Web data, and that has among its objectives to act as a recommender. Section "Background" provides some background about recommender systems. Thus, during their planning, a teacher may use the recommendations that come from the SDE in choosing the most appropriate technologies, events, and experts, as discussed in section "The SDE". In order to conceptualise the elements that contribute to educational practice an ontology was conceived, and its final version was the result of several iterations of revisions by Control Boards made up of experts in the domain and knowledge engineers. We present a brief overview of its main concepts.

The AI agent provides an API that enables client applications to integrate its recommendations. These client applications are editors that support teachers in designing their educational practice. So far, two client applications have successfully integrated recommendations from the SDE. These are: the Composer, which is part of the iTEC Cloud, see Chap. 4; and AREA see Caeiro-Rodríguez et al. (2013), which is part of a project that counts with public financing from Galician regional government. These two successful cases are discussed in section "Client Applications That Integrate SDE Recommendations".

To date, we have conducted three experiments to evaluate the SDE with teachers as end-users of this application. The first was on 6th June 2013 in Santiago de Compostela (Spain), with a focus group composed of teachers of primary and secondary education. The second took place on 18th June 2013 in Bolton (England), with end users. The third took place on 29th and 30th October 2013 in Oulu (Finland). Sections "Evaluation" and "Conclusions and Lessons Learned" discuss these experiments, and provide some conclusions and lessons learned.

## Background

As Ricci et al. (2011) state:

*Recommender Systems are software tools and techniques providing suggestions for items to be of use. The suggestions provided are aimed at supporting their users in various decision-making processes, such as what items to buy, what music to listen to, or what news to read.*

Traditionally, users of recommendation systems provide ratings for some of the items, and the system uses these ratings for the items not yet assessed (Resnick and Varian 1997). This approach is fairly flexible insofar as the output parameters are concerned, but is limited if we consider the input information available, as it does not consider, among other things, systems basing their recommendations on objective information about the items to be recommended. For our present concerns, we may apply the term *recommender* to any system offering personalized recommendations or guiding the user in a personalized way, selecting the most useful services from a variable-sized collection (Burke 2002).

Indeed, the main differences between a recommender and a search engine (or an information retrieval system) are related to the level of interest or utility of the retrieved items (recommendations). Recommendations had a clear social attractiveness even before the emergence of the information society, and they became basic building blocks of new online applications, mainly for electronic commerce and digital leisure services. Recommendation algorithms use techniques from Artificial Intelligence, Data Mining, Statistics or Marketing, among many others. Traditionally, according to the methods and algorithms used, recommendation systems are classified as: Content-based recommenders (Pazzani and Billsus 2007), Collaborative filtering recommenders (Schafer et al. 2007) and, combining both approaches, Hybrid recommender systems (Burke 2002). This classification is a very generic one and it is strongly tied to the interaction of a user with a recommender system, i.e. their preferences on the items to be recommended and their relationships to other users.

In spite of the above classification being the most frequent in the literature, it is for us preferable to focus on a classification which pays particular attention to the sources of data which the system relies on, as well as the use that the information receives. Following this approach, Burke (2002) distinguishes between five types of recommenders:

- Collaborative recommendation  
The most familiar, most widely implemented and most mature. These systems aggregate ratings or recommendations of objects, recognize commonalities between users on the basis of their ratings, and generate new recommendations based on inter-user comparisons.
- Demographic  
These recommenders categorize the user based on personal attributes and make recommendations based on demographic classes.
- Content-based  
These recommenders define their objects of interest by their associated features. These systems learn a profile of the user's interest based on the features present in objects the user has rated.

- **Utility-based**  
These recommenders make suggestions based on a computation of the utility of each object for the user. In these systems the central problem is how to create a utility function for each user.
- **Knowledge-based**  
These recommenders attempt to suggest objects based on inferences about a user's needs and preferences. Their approaches are distinguished in that they have functional knowledge: they have knowledge about how a particular item meets a particular user need, and can therefore reason about the relationship between a need and a possible recommendation.

Having established a definition and classification of recommender systems that is adequate for our proposal, we highlight three conceptual approaches that we have taken into account when developing our proposal: multi-criteria recommender systems, context-aware recommender systems and semantic recommenders. Those approaches are transversal to the types of recommenders previously presented and they try, respectively, to establish mechanisms for defining a utility function that takes into consideration several factors, to consider the context where a recommendation is produced, and to improve knowledge representation using semantic technologies. Below, we go deeper into each one of these.

### ***Multi-criteria Recommender Systems***

In traditional recommender systems, the utility function considers only one criteria, typically a global evaluation of resources or a valuation from the user. Depending on the systems under consideration, the utility function may be a valid approach though it is rather limited, since the utility of a given element for a particular user may depend on multiple factors. Taking this into consideration, in the past few years the study of multi-criteria recommender systems has increased (Lakiotaki et al. 2008, 2011; Plantié et al. 2005). Multiple Criteria Decision Analysis (MCDA) is a very mature and active research area (Figueira et al. 2005). It focuses on studying methods and management processes in systems with multiple conflicting criteria in order to identifying the best possible solution from a set of available alternatives. Starting from research and theories from that area, (Adomavicius and Tuzhilin 2010; Lakiotaki et al. 2011; Liu et al. 2011) propose approaching the problem of recommendations as one of MCDA, following the methodology that was developed by Roy (1996) for modelling these kinds of problems.

### ***Semantic Recommender Systems***

The term semantic recommender system is normally used when, in a traditional recommender, we use semantic web technologies in order to represent and process information of users and/or elements with high level descriptions. According to this

definition, we might think of content or knowledge based systems; nevertheless, semantic technologies are also used for collaborative recommender systems (e.g. Martín-Vicente et al. 2012; Shambour and Lu 2011).

## ***Context-Aware Recommender Systems***

Context is a very broad concept that has been studied across different research disciplines, including computer science, cognitive science or organizational sciences, among others. Looking for a formal definition, it can be stated that context is a set of circumstances that form the setting for an event, statement or idea, and in terms of which it can be fully understood (Oxford English Dictionary 2014).

## **The iTEC Ontology**

In order to develop a software system based on semantic techniques such as the SDE, it is necessary to define a Semantic Model which makes explicit the existing knowledge about the Universe of Discourse. This model, together with the information gathered by the system from the iTEC Back-end Registry and other possible external data sources, makes up the Knowledge Base of the SDE. The process of semantic modelling is a complex task that has led to different methodological approaches. Presently there is no standard methodology commonly used by knowledge engineers, although there are proposals with a relatively high degree of maturity.

In our case, we have adopted a methodological approach strongly based on Methontology (Fernández-López et al. 1997). We selected this methodology because it is one of the most mature and most widely used, and it is the best suited to our purpose. However, in order to adapt it to our specific needs taking into account our experience in software application development (Gago 2007), we decided to simplify and reshape some aspects of it taking into account aspects of other methodologies such as DILIGENT (Pinto et al. 2004; Uschold and King 1995; Noy and McGuinness 2001), and UPON (De Nicola et al. 2005).

One of the main advantages of semantic technologies is their support for knowledge reuse. Indeed, reuse of widely accepted terms and conceptualizations is included among the good practice guidelines for ontology design, extending or refining them when needed. Thus, in iTEC we followed this design principle by reusing those terms, properties and rules from conceptualizations that were strictly needed to capture knowledge about our universe of discourse. The objective of this approach is to have a manageable TBox, where only the knowledge strictly needed for the correct operation of the semantic applications to be developed is defined, in our case the iTEC SDE. With this approach we can guarantee the usability and efficiency of these applications. Besides, the clarity of the generated models is improved because only the terms, relations and rules from the base ontologies relevant to the terms and/or rules defined in our Semantic Model are taken into account. For exam-

ple, we have reused and included in this model most of the FOAF (People characterization), VCard RDF (characterization of the contact information of an individual or institution) and Organization Vocabulary (characterization of groups and institutions, and the relations between an individual and a group) ontologies due to their overall relevance to our application domain, but we have omitted some concepts lacking the mentioned relevance.

The parts of the semantic model that deal with technologies, events, and experts are briefly described below. The Universe of Discourse is, obviously, much wider; and certain parts of the semantic model characterise learning activities, their requirements, the educational context (e.g. students' language, age range), and many other things.<sup>3</sup>

### ***Tools Characterisation***

The SDE also facilitates the technical localisation of a learning story for a given school. Taking into account the functional requirements of learning stories, the system assesses the degree of feasibility of the learning activities in a school according to the tools available there. Thus, the semantic model needs to characterize the set of technological tools available in a school, that is, its technical setting, together with the distinct features of these tools (e.g., technical specifications, functionalities, supported languages, etc.). This enables both technical localisation, and the generation of recommendations on tools during planning. This information group collects all concepts and relations needed to model tools and technical settings, enabling eventual recommendations on tools (applications and devices) by the SDE. Figure 6.1 shows the part of the semantic model that characterises tools.

### ***Events Characterisation***

Events were also considered by the iTEC project to be relevant resources for the schools of the future. An event represents something that takes place in a given location at a given date. It includes properties such as: target audience, cost, language, place (e.g. museum, zoo) and location. Workshops, seminars, conferences and virtual meetings are examples of events that may support novel learning activities to improve the educational practice in European schools. As events are also resources, the SDE should offer recommendations on the events that best adapt to the context of a given school. Thus, event conceptualisation should be targeted to model the most relevant features of events, like the type of participants, venue, relevant dates, audience, or specific tools needed to participate. Elements identified in this information group enable a complete characterization of events, and therefore eventual recommendations on events made by the SDE. Figure 6.2 shows a diagram of the semantic model of events.

---

<sup>3</sup>The latest version of the iTEC ontology is available at: <http://itec.det.uvigo.es/itec/ontology/itec.rdf>.

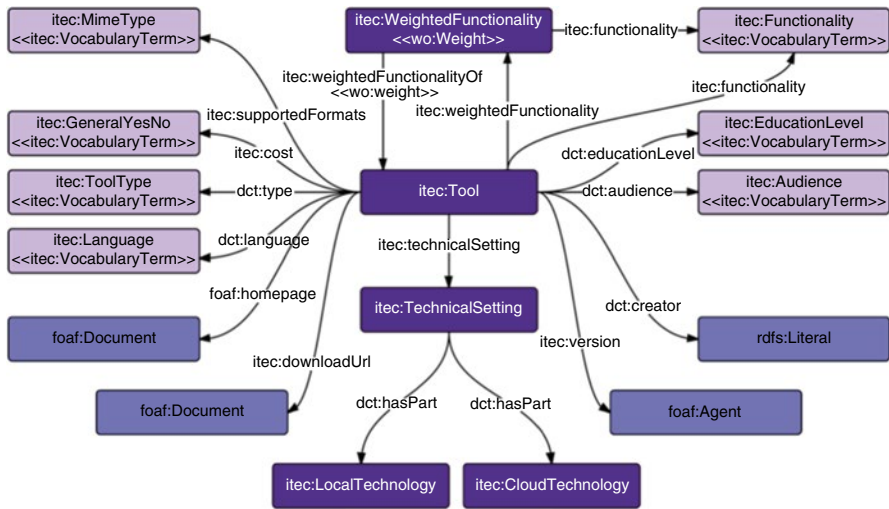


Fig. 6.1 Semantic model of tools

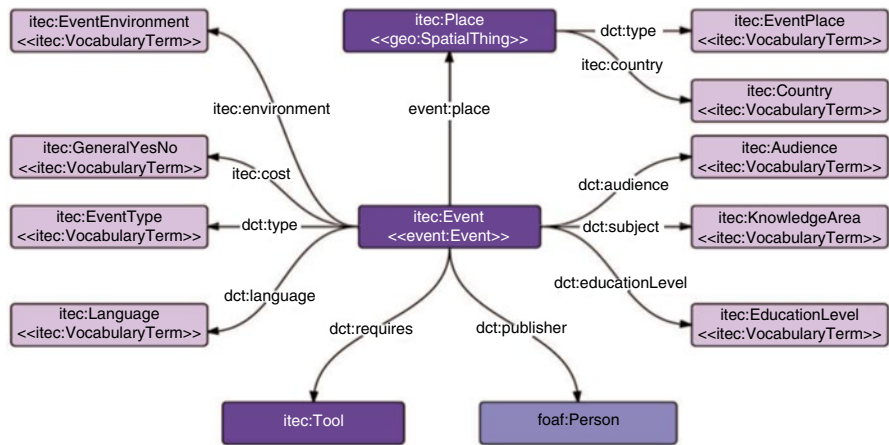


Fig. 6.2 Semantic model of an events

### People Characterisation

One of the most notable innovations of the iTEC project is that people were considered to be resources that can be utilized in a classroom to provide added value to the learning process. Besides the teacher, pupils in future classrooms may have available a rich pool of experts in several areas to provide advice and support along learning activities. According to this new vision, where people are also considered resources available to configure learning processes, the SDE supports recommendations to



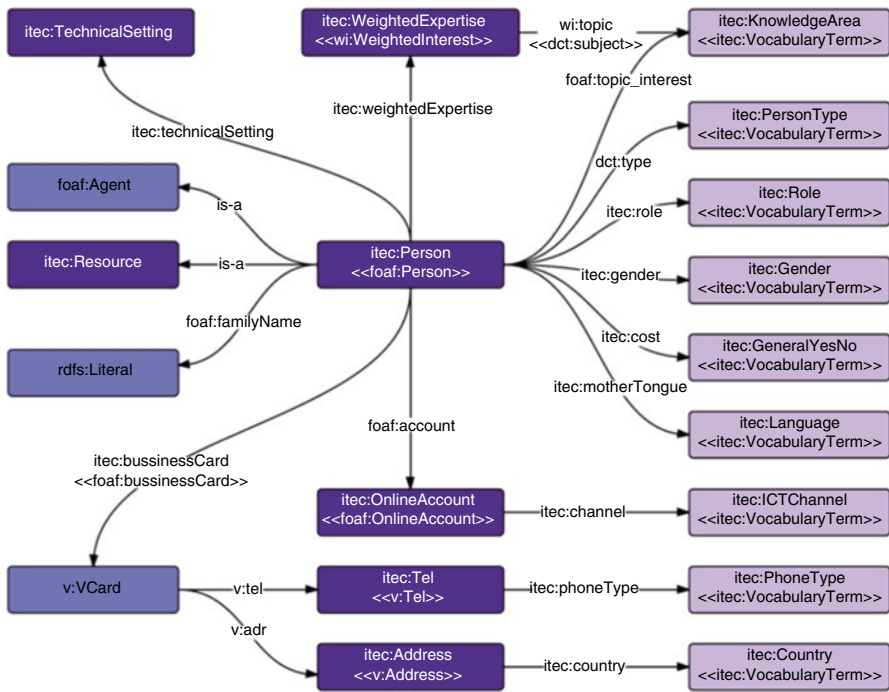


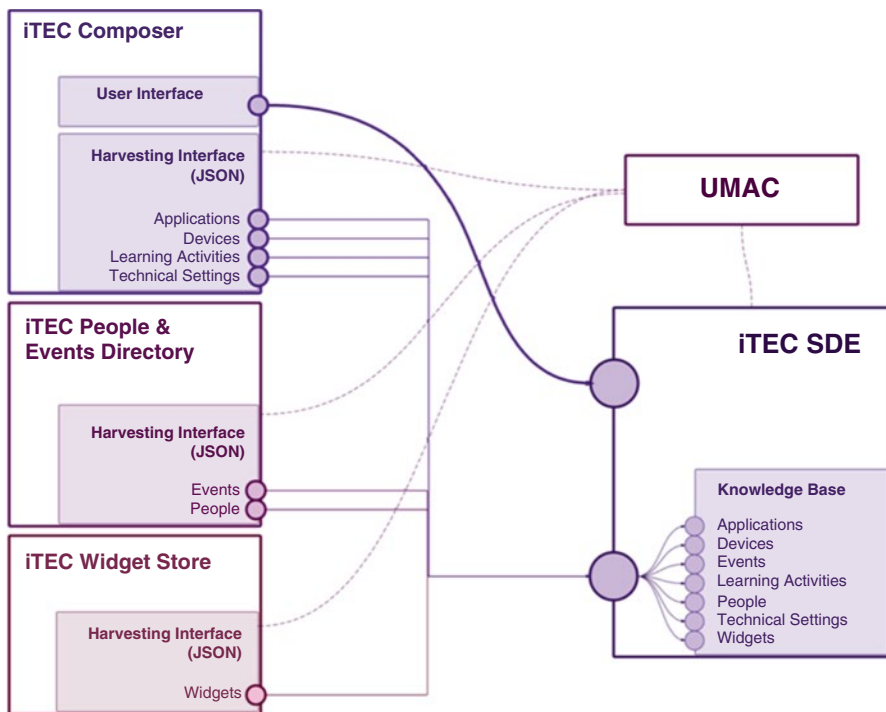
Fig. 6.3 Semantic model of a persons

teachers on the experts most suitable to enrich a given educational activity, taking into account the specific conditions at the school. Thus, the characterization of people goes beyond state-of-the-art people description, and includes all the skills, expertise and context relating to an individual relevant to educational scenarios (e.g., fluency in a given language, degree of knowledge of a particular subject, communication tools at his/her disposal, affiliation). This information group collects all the concepts and relations needed to enable the modelling of people in this context, and serves as the foundation for the recommendations that are eventually provided by the SDE. Figure 6.3 shows a diagram of the semantic model of a person.

## The iTEC Cloud

The iTEC Educational Cloud (see Fig. 4.2) is defined as the collection of systems and applications, the SDE among them, offering the functionalities developed within the iTEC project. As it can be seen in Fig. 6.4 the iTEC SDE relates to the rest of the systems in the iTEC Cloud according to three different models:

- Information harvesting. The implementation of SDE functionalities relies on data provided by other systems in the iTEC Cloud. More specifically, data registered



**Fig. 6.4** The iTEC cloud architecture from an SDE perspective

with the iTEC Composer on tools (applications and devices), learning activities and technical settings, data stored in the iTEC P&E Directory on people and events, and data registered with the iTEC Wookiee Widget Server on widget descriptions. The SDE needs to access these systems to collect data and keep its KB updated.

- Access to SDE functionalities. Access to the services offered by the SDE (technical localisation and resource planning services) is performed from the iTEC Composer through a specific Web Service API.<sup>4</sup>
- UMAC authentication. All interactions among the several systems in the iTEC Cloud, SDE’s information harvesting and access to the services provided by the SDE from the Composer in particular, together with all user interactions, has to be authenticated and authorized by the UMAC.

<sup>4</sup>A digital version of a guide of the API is available at <http://itec.det.uvigo.es/itec-sde/apidoc/index.html>

## The SDE

Traditional recommenders take into account two kinds of entities: users, and elements that make up the space of things to recommend. Context-aware recommenders follow a multi-dimensional model, instead of the traditional bi-dimensional model. The recommender integrated in iTEC does not consider the user as the main factor to take into account when generating recommendations, but rather takes the educational context as the most relevant factor. Thus, the utility function is defined in the following way:

$$f : \text{Items} \times \text{Content} \rightarrow \text{Rating} \quad (6.1)$$

In the *Items* dimension, we consider three kinds of elements—technologies, both hardware and software; events; and experts. Each one of these kinds of elements has different metadata: technologies are characterised, among other things, by their functionalities and languages of the user interface; events have space-time metadata, besides their topic; and experts are characterised, among other things, by their area of expertise. This diversity entails a multi-criteria approach, and the consideration of several factors. Each partial utility function follows a different approach—content-based, collaborative-based, or hybrid—that depends on the nature of those factors. Multiple Criteria Decision Analysis (MCDA) provides techniques and methods targeted to support the selection of the best alternative in systems where multiple criteria conflict and compete with each other. In recent years, contributions have been made in a number of different fields (Plantié et al. 2005; Lakiotaki et al. 2008; Matsatsinis et al. 2007; Manouselis and Matsatsinis 2001).

## The Learning Context

The recommender builds on a semantic model designed by iTEC partners over several iterations of Control Board revisions, and captures knowledge of the domain. The learning context is one of the key abstractions in the domain, and it includes concepts such as: the technologies that are disposable in a particular classroom; the characteristics of the target students; and space-time considerations.

## Recommendation Process

The recommendation process produces a list of recommended items—technologies, events, experts—that can be used during the performance of a learning activity in a particular context. Thus, taking the characterisation of a learning activity and its context as inputs, the recommender goes through the items in its Knowledge Base and fetches the fittest items. This process has three stages: pre-processing,

filtering an ordering of results by their relevance. All the stages are important though the ordering algorithm (relevance calculation) is the one that has most impact on the results.

In the pre-processing stage, the requirements of a given activity—the generic description of the kind of resources needed—are composed with those from the context, thus forming an integrated set of factors that have to be taken into account when calculating the relevance of resources.

In the filtering stage, some candidates are selected from the Knowledge Base, thus restricting the final number of resources whose relevance is going to be calculated. Due to the impact of this stage in the results, there are three configurable running modes:

- **Strict:** only resources that comply strictly with the requirements of the learning activity are selected.
- **Permissive:** in addition to the resources selected in the point above, this mode includes those resources with incomplete/black properties. Thus, it does not discard those resources that are not perfectly defined.
- **No filtering:** in this mode there is no filtering stage. This mode is especially useful in testing/depuration, as well as in scenarios with a low number of available resources.

Once a subset of valid resources has been obtained, the next stage consists of calculating the degree of relevance for each resource, while taking into account the requirements of the activity and the context. The heterogeneous nature of the resources and its complex description forced us to follow a rigorous strategy in order to obtain a satisfactory utility function. We followed an approach inspired by multi criteria recommender systems, which uses analysis techniques from the field of MCDA. Specifically, we followed the general methodology proposed by Roy (1996). We set (6.2) as the mechanism for calculating the relevance of resources, where  $f_i$  represents the marginal utility function for a given factor and  $w_i$  the weight that such a factor will have in the final value of relevance.

$$\sum_{i=0}^n w_i \cdot f_i \quad (6.2)$$

Below, we detail the process that we followed for selecting the factors and their associated weights. Rodríguez et al. (2013) go further into the decisions made in each of the stages of the followed methodology.

### ***Selection and Weighting of Factors***

Both the selection and weighting of factors that are taken into account in the recommendation process have been driven by iTEC Control Boards: a group of experts that collaborated in the project and that included people with technological and

pedagogical expertise. Fifty-three experts from different institutions participated in this process.

- **Selection:** we generated a document including a description of the general recommendation strategy, as well as the data model of every type of resource, with a collection of all the factors that a priori might play a role in the recommendation process. For each factor, the document included a thorough description of its meaning. After a productive discussion, with more than 100 written commentaries on the idoneity of the factors, we obtained the set of selected factors.
- **Weighting:** the experts rated the impact that each one of the factors should have in the calculation of the relevance of resources. The following tables summarise the factors that were selected by the Control Boards with their associated weights. Rodríguez et al. (2013) describe the weighting of factors in further detail. Tables 6.1, 6.2 and 6.3 shows selected factors and their weighting.

## Enrichment of Semantic Knowledge Base

The process of recommending educational resources depends on complete, thorough and up to date information being available on the knowledge base. In the end, the maintenance of information in the system is a responsibility of the community of system users. In the case of the iTEC Cloud, this community consisted primarily of teachers and technical and pedagogical coordinators registered on the platform. In many cases, these teachers lacked the appropriate knowledge and the time required to provide accurate and complete information on each of the resources catalogued (e.g., when teachers entered a new expert in the people directory, they were neither expected to be aware of all the areas of expertise of the individual

**Table 6.1** Selected factors and associated weights for resources

Factor ( $f_i$ )	Description	Weight ( $w_i$ )
Functionality	Functionality offered by a tool to a given degree	0.1307
Language	Language(s) supported by the tool's user interface	0.1031
Type	Type of the tool (i.e. application or device)	0.1011
Shell	Ranks tools according to their running environment	0.0976
Age	Prioritizes tools having as their explicitly specified audience one of the audiences specified for the context	0.0976
Cost	Prioritizes tools having no usage cost within a specified school (or context)	0.0970
Rating	Community popularity	0.0916
Technology	Discriminates whether a school already has a given tool	0.0916
Competences	References the technical expertise of a teacher	0.0883
Education level	Prioritizes tools which are explicitly targeted at an educational level among those defined for the activity	0.0979

**Table 6.2** Selected factors and associated weights for a resource of type person

Factor ( $f_i$ )	Description	Weight ( $w_i$ )
Language	Prioritizes people having as their mother tongue the language in which an activity is carried out	0.1359
Expertise	Reflects the expertise of a person in a given subject	0.1343
Experience	Considers previous experience of a person, according to the learning activities already carried out by this person	0.1238
Communication	Takes into account the communication tools a person participating in a learning activity has available	0.1186
Reliability	Indicates the degree of trust that the community, as a whole, has in the person to be selected	0.1119
Organization	Prioritizes persons belonging to the same organization as the learning activity creator	0.0998
Rating	Indicates the degree of popularity of a person	0.0984
Geographical	Indicates the degree of geographical proximity of the person to the location of the school	0.0915
Personal relations	Considers existing relations between the relations learning activity creator and the people who may participate in it	0.0856

**Table 6.3** Selected factors and associated weights for a resource of type event activity

Factor ( $f_i$ )	Description	Weight ( $w_i$ )
Subject	Used to rate an event according to the event thematic area(s)	0.1574
Required tools	Identifies online events that can be accessed when using some of the available tools	0.1444
Cost	Prioritizes free events	0.1385
Geographical	Degree of geographical proximity of an event to the location of the school where the activity is performed	0.1238
Rating	Popularity	0.1186
Organization	Relevance of the event's organizer	0.1186
Audience	Prioritizes events having as their explicit audience one of the audiences specified for the context	0.0995
Education level	Prioritizes events being explicitly targeted at an educational level among those defined for the activity	0.0995

being included, nor had the time needed to try to find out what those areas might be). Any such shortcomings in the information held lead to reductions in the quality of the recommendations provided by the system.

To try to alleviate part of this burden to end users, when developing the SDE support was included to enrich the information available in the KB transparently to other iTEC systems by leveraging the information freely available on the Web. The enrichment of the information available on the KB is performed through an enrichment module that analyses external sources and extracts relevant information to complement descriptions of educational resources already on the KB, which in turn were obtained from the information available in the collection of repositories on the iTEC Cloud. Many sources of information are available on the Web in several

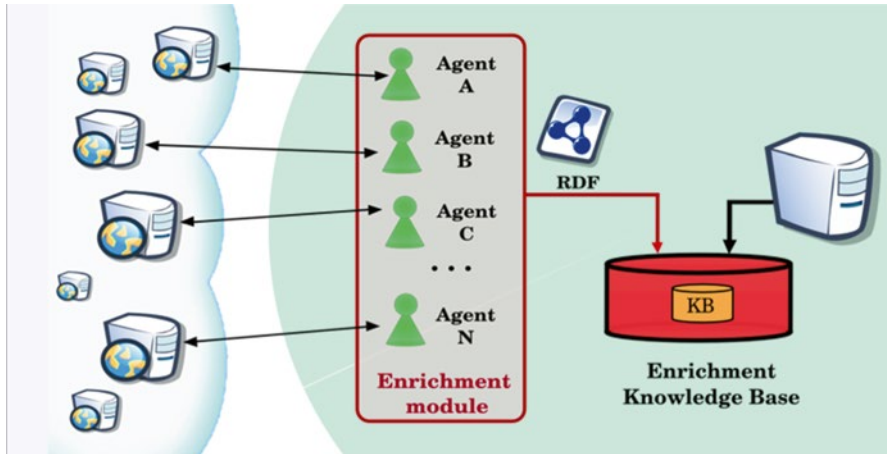


Fig. 6.5 The enrichment process

contexts that catalogue and describe in detail the information available for many entities and resources, including entities related to the resources handled in iTEC. For example, in the case of tools there are software application catalogues, which contain accurate descriptions developed by experts and endorsed by a large community of users.

In the case of the SDE, the enrichment process is carried out by a module composed of a set of smart independent agents that extract specific information from external sources (see Fig. 6.5), process it, and insert it into the KB in a way which is transparent to the rest of the system. Thus, the information available is eventually enhanced, and consequently users receive recommendations on educational resources of a better quality than those obtained solely from the information provided exclusively by the users themselves. It should be noted that in the early stages of deployment of a system lacking an enrichment module, when cataloguers have not yet entered enough information, the recommender is unable to provide quality recommendations. That is, it requires a significant initial effort from users to enter information on resources before appropriate recommendations can be offered. The extent of this effort may compromise the success of any platform. However, by the introduction of enrichment it is possible to mitigate this cold-start situation (Maltz and Ehrlich 1995) and provide available information on resources more quickly, thus considerably reducing the initial effort required from cataloguers.

Record Linkage (Winkler 1999) is one of the pillars of our enrichment algorithm. In the case of external sources publishing their information using RDF (i.e., semantic sources, as they use a form of information representation specifically targeted to preserve the meaning of statements) there are tools available (e.g., SILK (Volz et al. 2009)) that automate Record Linkage. In the case of non-semantic web sources, a specific wrapper agent has to be developed (Ferrara et al. 2011). A wrapper is an agent that extracts information from a source and transforms it to a particular

information structure, RDF in our case. The design and development complexity of these wrappers, and thus their robustness and reliability, will be ultimately determined by the type of information structure with which they have to deal. In this way, highly structured data, such as XML documents, require wrappers of lower complexity than those required to process data sources expressing their information in a semi-structured way, such as HTML documents. We provide in the next section an overall description of the tasks performed by the SDE to enrich the information initially available on the SDE's KB.

### *Overall Description*

The overall procedure that eventually leads to the enrichment of the information initially available on the SDE KB can be conceptually decomposed into a series of stages:

#### **Source Localization and Definition of Information Extraction Patterns**

The process is initiated by a domain expert who analyses the sources available in the Web to identify the most relevant ones. In other words, the sources sought are those containing useful information to complement the information available on the KB. Once the most appropriate sources have been identified, the corresponding extraction pattern is defined. This pattern is implemented by a wrapper. This piece of software determines which data and structures should be extracted, together with the operations required to extract that information and, if necessary, its transformation into RDF. The wrapper utilizes a different extraction mechanism depending on the language used to represent the information in each source (e.g., automated tools like SILK, GRDDL transformations (Connolly 2007)).

#### **Record Linkage and Retrieval of Resource Descriptions**

The next task consists of detecting the correspondence between data records in the external source and entities to be enriched, and on retrieving the information available in those records. In conceptual terms, to complete this task the following activities need to be performed:

- **Source location:** The location of relevant records in the external source can be performed directly in the case of sources providing internal searching mechanisms to final users (e.g., through SPARQL Endpoints (Prud'hommeaux and Seaborne 2008), API methods or Web content search support). These mechanisms are fairly common in most relevant sources, as these sources host large amounts of information that would be difficult to exploit without search support, and they reduce the overall complexity of the linkage process. Using the appropriate



searching service, and by means of key-based queries, it is possible to retrieve the resources related to the entity to be enriched (e.g., using an individual's name, it is possible to recover the list of individuals registered with the external source having a similar name).

- **Extraction of characterization information:** From search results, and using the previously defined extraction pattern, information characterizing each record is retrieved. Records returned by the search process usually provide limited information, including only the details required to identify each object. In addition, they usually include a key or path to recover the complete description of each object. The information extracted is structured according to the language used by the source, so it has to be translated into RDF to be further processed by the wrapper. According to the granularity desired for the detection of false positives, two strategies are possible: (1) to recover at this point all the information available for each retrieved record to have as much information as possible for filtering; or (2) to perform filtering immediately (as described below) and, once duplicate records or false positives have been discarded, to recover all the information corresponding to the remaining valid records. The first strategy facilitates a more accurate filtering process as richer information is available, whereas the second strategy is more efficient, as the number of queries required and the amount of information managed can be dramatically reduced.
- **Filtering of false positives:** For information enrichment to be correct, we need record linkage to be exact, that is, resources deemed as equal should actually be representations of the same object. As a consequence, on some occasions it is necessary to internally filter out the resources retrieved after searching the external source to discard similar but not equal objects. For instance, when we look for a specific individual in a social network, we may obtain references to individuals with similar names (e.g., Mary Smith, Maria Smith). In these occasions, a syntactic comparison is launched on the list of retrieved resources, using in our case the Jaro heuristic (Jaro 1995). This is a simple record linkage mechanism.

In cases where the source does not provide a searching service, all records available will be considered candidate results. This implies that all descriptions will be extracted from the web to be further filtered for false positives. Thus, in a context where the only objective is to enrich the information available about a local resource, an external source not providing searching support would be of little use, as enrichment would be highly inefficient in terms of time and resources required. However, if the aim includes completing the knowledge base with new, previously non-existent records, this option can be considered.

### Adaptation to the SDE Model

Data extracted follows a vocabulary defined by the managers of the external source. These vocabularies are not directly understandable by our system, which defines its own terminology through specific data models. As a consequence, extracted

information cannot be directly utilized in the recommender's inference processes. Because of this, information obtained from external sources is adapted to the SDE's data model. This translation is specific for each source and each type of educational resource to be enriched.

### **Knowledge Base Insertion**

Finally, processed information is entered in the KB to enrich the corresponding resources. This insertion process triggers several internal inference processes to obtain new information from the heuristic rules defined in the Semantic Model, and to pre-compute most of the factors needed for relevance estimation by the recommendation algorithms implemented by the SDE.

Wrappers developed according to the process described above may be periodically launched on the selected external sources. This facilitates the continuous availability of updated data without requiring additional efforts from the user community.

The generic processes described in this section are intended to enrich the information from the resource descriptions already stored on the SDE's KB. However, these same processes can be used to add new entities or non-existent records, such as new software applications that could be used in a Learning Activity that had not been yet registered by teachers because they do not belong to any technical setting in any school. That is, they also support the population of the KB with educational resources that have not been previously introduced by human cataloguers. This process will hereafter be referred as population. To do this, instead of searching for records at each external source that refer to the same resource in the KB, we will try to find all records that may serve as iTEC resources. For example, in the case of educational events, we will search events with agendas reflecting an educational or cultural event and use them to populate the KB.

This strategy is feasible for resources that, due to their characteristics and to their public nature, may be freely entered in the KB without the system detecting any difference between this automatically entered information and the resources manually inserted by cataloguers. In any case, it is always necessary to consider the treatment to be given to this data in relation to their private or public nature.

### ***Experiments Using the Enrichment Module***

We conducted experiments that dealt with the enrichment of technologies, events, and experts. For the sake of brevity, we detail here only the results of the enrichment of experts. You can see the results of enrichment events and technologies in Anido et al. (2013). The results obtained by applying the enrichment process to complete the descriptions of educational resources of type People are fairly satisfactory taking into account the initial data available. The SDE's KB included an initial list of

**Table 6.4** Preliminary results of enriching the knowledge base of experts

	# of experts	14
Initial KB	Average RDF triples per expert	28
	Total RDF triples	389
	# of enriched experts	8
	Enrichment %	~57 %
	Average RDF triples per enriched expert	190
	# of new contact accounts	7
Enriched KB	# of new expert tags	112
	# of new localizations	7
	# of new languages	12
	# of new person-languages relations	3
	Total RDF triples (enriching)	1519

14 experts associated to the iTEC project. The descriptions of these experts were used as the input of the enrichment process described above. Eventually, we have established Record Linkage relations with eight records in external sources, which refer to exactly eight different experts (cf. Table 6.4). Therefore, almost 60 % of the initial records were enriched. Analysing in further detail the enrichment process, 1519 new RDF triplets were generated, corresponding to an average of 190 triplets per expert. Most of these triplets refer to articles and other publications. Regarding the most relevant properties to the recommender, we obtained: 7 new contact accounts to facilitate communication with the corresponding experts; 112 new tags enabling the inference of new abilities and skills; 7 postal addresses that may be used to infer the geographical area of influence on an expert; 12 new evidences on language skills for 3 experts, which may be used by the recommender to propose experts according to the communication language defined for an educational.

## Client Applications That Integrate SDE Recommendations

To date, the services offered by the SDE have been successfully integrated in two different client applications.<sup>5</sup> The first, Composer (Simon et al. 2013), is the application for creating and configuring learning activities that was created in the scope of iTEC Cloud. The second, AREA (Caeiro-Rodríguez et al. 2013), is an application that includes facilities to create learning plans, and it integrates the SDE's recommendations to configure the learning activities inside learning plans.

<sup>5</sup> Apart from an ad-hoc front-end that was developed for a pre-testing with participants (Anido Rifon et al. 2012).

## ***Composer***

As mentioned above, the iTEC Composer is the iTEC's proposal to provide support to the identification of the most suitable Tools and Resources for Learning Activities. The iTEC SDE provides additional features for the iTEC Composer. Indeed, while the iTEC Composer facilitates the production of a learning plan providing access to available Tools and Resources needed to satisfy the requirements of one or several Learning Activities, the iTEC SDE analyses the actual requirements of a Learning Activity to offer recommendations on Tools and Resources satisfying these requirements according to the specific context where activities will be developed.

The iTEC Composer is an autonomous entity that may also provide basic support to the production of learning plans independently of the recommendations provided by the SDE. The first step when generating a learning plan is to provide two key elements: (1) the Learning Activities that will be eventually included in the learning plan and (2) the Learning Context, that is, the set of parameters characterizing the context where the learning experience will eventually take place (e.g., Technical Setting, language, learning subject). Then, the teacher may use the iTEC Composer to navigate across the collection of available Tools and Resources to select the most suitable to the learning plan. Additionally, the Composer may utilize the SDE to provide personalized recommendations according to the requirements included in each Learning Activity.

## ***AREA***

iTEC initiated a collaboration line with the TELGalicia<sup>6</sup> research network, whose objective is to facilitate pedagogical and technological innovation in primary and secondary education in the northwest of Spain. Given the compatibility between the objectives of iTEC and TELGalicia, a collaboration with that network was initiated that had among its outcomes the adaptation of a web application named AREA in which the services offered by the SDE were integrated together with initial content available on the SDE's KB. AREA is basically a social Web 2.0 application that facilitates access to primary and secondary teachers to innovative educational proposals. AREA provides resources and tools for authoring, exploration and social curation for teachers to design their own lesson plans. Once a lesson plan has been completed in the classroom, AREA also provides structures for teachers (and also students in those cases where teachers find it convenient) to document their experiences in a similar way as it can be done with a blog, but according to the activity structure defined in the lesson plan.

One important aspect of SDE testing was that users were able to obtain recommendations on the most appropriate resources for learning stories/learning activities through. For each activity, users could consult the requirements and perform resource selection.

---

<sup>6</sup>[www.redetelgalicia.com](http://www.redetelgalicia.com)

## Evaluation

At the time of writing this chapter, three testing sessions with end users have been completed. The first session with Galician primary and secondary education teachers, the second session consisted of a workshop with iTEC end users in the UK, and the third session consisted of a workshop in Oulu (Finland), also with iTEC end users.

A session was organized on 6th June 2013 in Santiago de Compostela (Spain) with 15 Galician primary and secondary education teachers. This session included the introduction of AREA and the integrated SDE recommendation features. Then, there was an open discussion about the questionnaire, that was created as part of iTEC's evaluation plan (Haldane and Lewin 2011), with a special emphasis on possible barriers and enablers, and on the suitability of the SDE for their needs. On 18th June 2013 a demonstration and testing session of the technologies developed in iTEC took place in Bolton (UK) with 25 teachers. As part of this, the SDE was presented in a workshop, and participants assessed the tool by means of a questionnaire. The SDE was evaluated in a similar way in the session in Finland.

On average, participants on the evaluations think that recommendations on non-traditional educational resources may foster innovation in the classroom. Teachers agree with the vision that new technologies may be very useful in teaching-learning environments, but one hindrance towards the realisation of that vision is the difficulty of knowing what technologies are most adequate for whom. Overall, participants think that recommendations from the SDE is one step forward towards filling the gap between existent, suitable, and useful technologies and being aware of their existence.

## Conclusions and Lessons Learned

This chapter has described a recommender system for non-traditional educational resources—tools, people, events—that is based on semantic technologies and that was developed in the scope of the iTEC project, whose main findings are described in this book. As the main contributions of our research we can highlight the following ones.

We defined a semantic model that characterises the universe of discourse that the recommender uses, and that is also the basis for the definition of a common language shared between the different iTEC working packages. This semantic model was implemented as an ontology, which constitutes the core of the intelligence of the recommender. The scope of the ontology developed is very broad, as it models concepts such as learning activities, contexts, technologies, events, people, and many other elements that are specific to the educational area.

The recommender system which we have described provides recommendations for technologies, events, and people (e.g. experts). This constitutes an innovative approach, at least in the area of recommender systems applied to education. Besides, the recommendation strategy is based on the learning context, rather than on students' and teachers' preferences.

The recommender's API is publicly available, and it is ready to be consumed from client applications that want to make use of recommendations. We have described how two client applications (Composer and AREA) successfully integrate SDE's recommendations. Using AREA as a front end, we tested the SDE with final users, in three experiences with teachers in Santiago de Compostela, Bolton and Oulu, and the first results were positive.

After 4 years working in this system we can point some lessons learned. First of all, the increasing number of open resources available in the web is a huge unexplored source for resources beyond content. Many applications and resources not explicitly designed to be used for education can be actually applied to that purpose. The original objective of integrating some repositories within the SDE—i.e. the Widget Store or de People and Events Directory—was not enough to provide teachers with a sufficient number of alternatives. This issue was overcome thanks to the use of enrichment techniques allowing to easily integrate external sources.

On the other hand, traditional semantic web technologies, including the academic design of ontologies and the development of recommendations algorithms based on them, are not agile enough to adapt to the community of content and application developers. Therefore a less strict approach, based for instance, on the use of soft ontologies is required.

Finally, when resources coming from different sources are to be integrated to provide recommendations to users based on whatever criteria, an extra effort is needed to appropriately classify those resources. Again, pre-design ontologies may not work for many cases. In the light of this we suggest research into Machine Learning techniques whose application to the automatic classification of educational resources may contribute to the field of automatic metadata generation.

**Open Access** This chapter is distributed under the terms of the Creative Commons Attribution Noncommercial License, which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

## References

- Adomavicius G, Tuzhilin A (2010) Context-aware recommender systems. In: Recommender systems handbook: a complete guide for research scientists and practitioners. <http://ids.csom.umn.edu/faculty/gedas/NSFCareer/CARS-chapter-2010.pdf>. Accessed 12 Mar 2015
- Anido Rifón L et al (2012) iTEC—wp 10 d10.2—support for implementing iTEC engaging scenarios v2
- Anido L et al (2013) iTEC—wp 10 d10.3—support for implementing iTEC engaging scenarios v3
- Burke R (2002) Hybrid recommender systems: survey and experiments. *User Model User-Adapt Inter* 12:331–370. Available via <http://www.springerlink.com/index/N881136032U8K111.pdf>. Accessed 12 Mar 2015
- Caeiro-Rodríguez M et al (2013) AREA: a social curation platform for open educational resources and lesson plans. In: Proceedings—frontiers in education conference, FIE, pp 795–801
- Connolly D (2007) Gleaning resource descriptions from dialects of languages (GRDDL). W3C Recommendation. <http://www.w3.org/TR/grddl/>. Accessed 12 Mar 2015

- De Nicola A, Missikoff M, Navigli R (2005) A proposal for a unified process for ontology building: UPON, vol 3588, Database and expert systems applications (Lecture notes in computer science). Springer, Heidelberg, pp 655–664
- Fernández-López M, Gómez-Pérez A, Juristo N (1997) Methontology: from ontological art towards ontological engineering. Assessment SS-97-06, pp 33–40. <http://www.cpgei.cefetpr.br/~tacla/Onto/Artigos/MethontologyFromOntologicalArt.pdf> <http://oa.upm.es/5484/>. Accessed 12 Mar 2015
- Ferrara E, Fiumara G, Baumgartner R (2011) Web data extraction, application and techniques: a survey
- Figureira J, Greco S, Ehrgott M (2005) Multiple criteria decision analysis: state of the art surveys. Springer, New York
- Gago JMS (2007) Contribución a los sistemas de intermediación en el ámbito del aprendizaje electrónico utilizando tecnologías semánticas. University of Vigo
- Griffiths D et al (2012) The Wookie Widget Server: a case study of piecemeal integration of tools and services. J Univ Comput Sci 18:1432–1453. Available via [http://www.jucs.org/jucs\\_18\\_11/the\\_wookie\\_widget\\_server](http://www.jucs.org/jucs_18_11/the_wookie_widget_server). Accessed 12 Mar 2015
- Haldane M, Lewin C (2011) WP5: revised evaluation plan. iTEC
- Herrick DR (2009) Google this! Using Google apps for collaboration and productivity. In: Proceedings of the 37th annual ACM SIGUCCS fall conference, pp 55–64
- Jaro MA (1995) Probabilistic linkage of large public health data files. Stat Med 14:491–498. Available via <http://www.ncbi.nlm.nih.gov/pubmed/7792443>
- Lakiotaki K, Tsafarakis S, Matsatsinis N (2008) UTA-Rec: a recommender system based on multiple criteria analysis. In: Proceedings of the 2008 ACM conference on recommender systems, pp 219–226
- Lakiotaki K, Matsatsinis NF, Tsoukiàs A (2011) Multicriteria user modeling in recommender systems. IEEE Intell Syst 26:64–76
- Liu L, Mehandjiev N, Xu DL (2011) Multi-criteria service recommendation based on user criteria preferences. In: Proceedings of the 5th ACM conference on recommender systems—RecSys’11, p 77. <http://dl.acm.org/citation.cfm?doi=2043932.2043950>. Accessed 12 Mar 2015
- Maltz D, Ehrlich K (1995) Pointing the way: active collaborative filtering. In: Proceedings of the SIGCHI conference on human factors in computing systems, pp 202–209
- Manouselis N, Matsatsinis NF (2001) Introducing a multi-agent, multi-criteria methodology for modeling electronic consumers behavior: the case of internet radio, vol 21, Lecture notes in computer science. Springer, Heidelberg, pp 190–195
- Martín-Vicente MI et al (2012) Semantic inference of user’s reputation and expertise to improve collaborative recommendations. Expert Syst Appl 39:8248–8258
- Matsatsinis NF, Lakiotaki K, Delias P (2007) A system based on multiple criteria analysis for scientific paper recommendation. In: Proceedings of the 11th Panhellenic conference on informatics
- Noy N, McGuinness D (2001) Ontology development 101: a guide to creating your first ontology. Development 32:1–25. <http://www.ksl.stanford.edu/people/dlm/papers/ontology-tutorial-noy-mcguinness-abstract.html>. Accessed 12 Mar 2015
- Oxford English Dictionary (2014) Oxford English dictionary online. Oxford English dictionary, 2010. <http://dictionary.oed.com/>. Accessed 12 Mar 2015
- Patterson TC (2007) Google Earth as a (not just) Geography education tool. J Geogr 106:145–152
- Pazzani MJ, Billsus D (2007) Content-based recommendation systems. Adapt Web 4321:325–341. Available via <http://link.springer.com/10.1007/978-3-540-72079-9>. Accessed 12 Mar 2015
- Pinto HS, Staab S, Tempich C (2004) DILIGENT: towards a fine-grained methodology for DIstributed, Loosely-controlled and evolvInG Engineering of oNTologies. In: 16th European conference on artificial intelligence—ECAI, pp 393–397
- Plantié M, Montmain J, Dray G (2005) Movies recommenders systems: automation of the information and evaluation phases in a multi-criteria decision-making process. In: Proceedings of the 16th international conference on database and expert systems applications, pp 633–644

- Prud'hommeaux E, Seaborne A (2008) SPARQL query language for RDF. W3C recommendation, 2009. W3C, pp 1–106. <http://www.w3.org/TR/rdf-sparql-query/>. Accessed 12 Mar 2015
- Redding S (1997) Parents and learning. Educational practices series. International Academy of Education. <http://www.ibe.unesco.org/publications/EducationalPracticesSeriesPdf/prac02e.pdf>. Accessed 12 Mar 2015
- Resnick P, Varian HR (1997) Recommender systems. *Commun ACM* 40:56–58
- Ricci F, Rokach L, Shapira B (2011) Introduction to recommender systems handbook. [http://dx.doi.org/10.1007/978-0-387-85820-3\\_1](http://dx.doi.org/10.1007/978-0-387-85820-3_1). Accessed 12 Mar 2015
- Rodríguez AC et al (2013) Providing event recommendations in educational scenarios. In: *Advances in intelligent systems and computing*, pp 91–98
- Roy B (1996) *Multicriteria methodology for decision aiding. Nonconvex optimization and its applications*. Springer, Heidelberg
- Schafer J et al (2007) Collaborative filtering recommender systems, vol 4321, *The adaptive web (Lecture notes in computer science)*. Springer, Berlin, pp 291–324. <http://www.springerlink.com/content/t87386742n752843>. Accessed 12 Mar 2015
- Shambour Q, Lu J (2011) A hybrid multi-criteria semantic-enhanced collaborative filtering approach for personalized recommendations. In: *Proceedings—2011 IEEE/WIC/ACM international conference on web intelligence, WI 2011*, pp 71–78
- Simon B et al (2013) Applying the widget paradigm to learning design: towards a new level of user adoption, vol 8095, *Scaling up learning for sustained impact (Lecture notes in computer science)*. Springer, Heidelberg, pp 520–525
- Uschold M, King M (1995) Towards a methodology for building ontologies. *Methodology* 80: 275–280
- Van Assche F (2012) iTEC—wp 9 d9.2—release of the directory. iTEC, pp 3–44
- Volz J et al (2009) Silk—a link discovery framework for the web of data. In: *CEUR workshop proceedings*
- Winkler WE (1999) The state of record linkage and current research problems. Statistical Research Division US Census Bureau, pp 1–15. <http://www.census.gov/srd/papers/pdf/rr99-04.pdf>. Accessed 12 Mar 2015