ITEC - WP 10

D10.4 - SUPPORT FOR IMPLEMENTING ITEC ENGAGING SCENARIOS V4

“This document has been created in the context of the ITEC project. All information is provided “as is” and no guarantee or warranty is given that the information is fit for any particular purpose. The user thereof uses the information at its sole risk and liability. The document reflects solely the views of its authors. The European Commission is not liable for any use that may be made of the information contained therein.”

<table>
<thead>
<tr>
<th>CONTRACT NO</th>
<th>2577566</th>
</tr>
</thead>
<tbody>
<tr>
<td>DATE</td>
<td>13/08/2014</td>
</tr>
<tr>
<td>ABSTRACT</td>
<td>This document corresponds to Deliverable D10.4 “Support for Implementing iTEC Engaging Scenarios v4”. The outcomes of the activities performed by WP10 along the fourth year were intended to: (i) enriching the Knowledge Base of the SDE; (ii) enabling the exploration of the Knowledge Base by end-users in a sort of Semantic Search; (iii) evaluating the SDE, which included the development of a user interface.</td>
</tr>
<tr>
<td>AUTHOR, COMPANY</td>
<td>Luis Anido Rifón, Miguel Gómez Carballa, Mario Manso Vázquez, Marcos Mouriño García, Roberto Pérez Rodríguez</td>
</tr>
<tr>
<td>INTERNAL REVIEWERS</td>
<td>Will Ellis, Jean-Noël Colin</td>
</tr>
<tr>
<td>WORKPACKAGE</td>
<td>WP 10</td>
</tr>
<tr>
<td>CONFIDENTIALITY LEVEL1</td>
<td>PU</td>
</tr>
<tr>
<td>FILING CODE</td>
<td>ITEC-D10_4_20140813</td>
</tr>
</tbody>
</table>

1 PU = Public

PP = Restricted to other programme participants (including the EC services);

RE = Restricted to a group specified by the Consortium (including the EC services);

CO = Confidential, only for members of the Consortium (including the EC services).

INN - Internal only, only the members of the consortium (excluding the EC services).
### DOCUMENT HISTORY

<table>
<thead>
<tr>
<th>Version</th>
<th>Date</th>
<th>Reason of change</th>
<th>Status</th>
<th>Distribution</th>
</tr>
</thead>
<tbody>
<tr>
<td>V0.1</td>
<td>1/07/14</td>
<td>First draft</td>
<td>Draft</td>
<td>WP10 members</td>
</tr>
<tr>
<td>V0.2</td>
<td>3/07/14</td>
<td>Initial structure of the deliverable</td>
<td>Draft</td>
<td>WP10 members</td>
</tr>
<tr>
<td>V0.3</td>
<td>5/07/14</td>
<td>Report of the enrichment sources</td>
<td>Draft</td>
<td>WP10 members</td>
</tr>
<tr>
<td>V0.4</td>
<td>9/07/14</td>
<td>Report of the SDE user interface</td>
<td>Draft</td>
<td>WP10 members</td>
</tr>
<tr>
<td>V0.5</td>
<td>15/07/14</td>
<td>Formulations of Semantic Search</td>
<td>Draft</td>
<td>WP10 members</td>
</tr>
<tr>
<td>V0.6</td>
<td>17/07/14</td>
<td>Evaluation of the SDE</td>
<td>Draft</td>
<td>WP10 members</td>
</tr>
<tr>
<td>V0.7</td>
<td>23/07/14</td>
<td>Exploitation</td>
<td>Draft</td>
<td>WP10 members</td>
</tr>
<tr>
<td>V0.8</td>
<td>25/07/14</td>
<td>Conclusions</td>
<td>Draft</td>
<td>WP10 members</td>
</tr>
<tr>
<td>V0.9</td>
<td>08/08/14</td>
<td>Appendixes</td>
<td>Draft</td>
<td>WP10 members</td>
</tr>
<tr>
<td>V1.0</td>
<td>13/08/14</td>
<td>Delivered for internal review</td>
<td>Draft</td>
<td>Internal reviewers</td>
</tr>
<tr>
<td>V1.1</td>
<td>15/09/14</td>
<td>Updates after internal review</td>
<td>Final</td>
<td>Public</td>
</tr>
</tbody>
</table>
Executive summary

This document reports the main actions carried out by WP10 in the fourth year of the iTEC project. These actions fall broadly in four groups. The first one had the objective of designing and developing a Web interface for the SDE, providing support to the life cycle of Future Classroom Scenarios from their definition to their didactic planning. The second group of actions was aimed at designing and developing an information retrieval system for enabling the exploration of the Knowledge Base by end-users, following the principles of Semantic Search. The third group has to do with the enrichment of the Knowledge Base of the SDE, using advanced techniques in information extraction from Web sources. Finally, the fourth group of actions was centred at evaluating the SDE with end users, in several workshops across Europe.

WP10 has performed important actions that resulted in a user interface for the SDE. This interface was implemented as a Web application, which provides support to the entire life cycle of Future Classroom Scenarios, from their definition to their didactic planning. A lot of dedication was put on the graphical design, as well as on interaction design, so that end users may find easy to interact with the application.

The design and development of a system for enabling the exploration of the Knowledge Base by end users centred a big part of the researching efforts of WP10. The development of an information retrieval system following the principles of Semantic Search corresponds to a request by end users, who wanted to actively interact with the Knowledge Base, thus going beyond a simple passive reception of recommendations. This document reports the foundations of our approach to Semantic Search, for its theoretical formulation to the description of the software architecture that actually implements it.

The enrichment of the Knowledge Base of the SDE is a direct continuation of the research line started in the second year and continued in the third year with very promising results. In the fourth year we kept on researching in this direction, widening a lot the number of sources the SDE extracts information from, as well as the number of elements actually extracted. Both the refinements in the extraction technique as well as the figures of extracted elements are reported in this document.

The evaluation of the SDE was the objective of a series of actions by WP10. Particularly, the SDE was presented to end users in several workshops across Europe, and a final evaluation questionnaire was issued to end users for gathering results. These actions are reported in this document.

Finally, possible actions for the exploitation and continuity of works are discussed, putting a strong focus on the exploitation of the technical contributions from WP10, identifying the main technical contributions, suggesting possible improvements and future research lines, and also outlining possible models for the exploitation of the SDE.
# TABLE OF CONTENTS

Introduction .............................................................................................................................................. 8

Reminder of the context .......................................................................................................................... 8

Purpose and scope of the task ............................................................................................................... 9

Relationship with other tasks .............................................................................................................. 10

Structure of the document .................................................................................................................. 11

Impacts of the Deliverable .................................................................................................................. 12

iTEC project ......................................................................................................................................... 12

Ethical Issues ......................................................................................................................................... 12

IPR issues .............................................................................................................................................. 12

Report .................................................................................................................................................. 13

1. A USER INTERFACE FOR THE SDE ......................................................................................... 13

   1.1. Using the SDE for registering scenarios ................................................................................. 13

   1.2. Using the SDE for discovering resources ............................................................................ 14

      1.2.1. Suggesting concepts ........................................................................................................ 15

      1.2.2. Seeing concept summaries .............................................................................................. 17

      1.2.3. Detail of an educational resource ..................................................................................... 17

   1.3. Bookmarking resources ......................................................................................................... 21

   1.4. Getting recommendations .................................................................................................... 22

   1.5. Creating a Learning Story .................................................................................................... 24

      1.5.1. Alternative resources ....................................................................................................... 24

2. SEMANTIC SEARCH ON THE KNOWLEDGE BASE ............................................................... 27

   2.1. Introduction to semantic search ............................................................................................ 27

   2.2. Concept-based representation of documents ........................................................................ 27

      2.2.1. From combinations of words to concepts ....................................................................... 27

   2.3. Automatic extraction of concepts ......................................................................................... 28
2.3.1. Automatic extraction of concepts "under the hood"................................. 29

2.4. Semantic relatedness between resources .............................................. 29

2.4.1. Related resources "under the hood"............................................... 30

2.5. Suggesting resources to explore ........................................................ 31

2.6. Suggesting alternative resources ......................................................... 32

2.7. Architecture of the information retrieval system .................................. 34

2.7.1. Understanding the architecture with the help of a metaphor ............. 36

3. ENRICHMENT OF THE KNOWLEDGE BASE ...................................... 39

3.1. Resources indexed in the SDE ............................................................... 40

3.1.1. Places of interest ............................................................................ 40

3.1.2. Events ............................................................................................. 41

3.1.3. Lectures .......................................................................................... 42

3.1.4. Documentaries .................................................................................. 43

3.1.5. Courses ............................................................................................. 44

3.1.6. Articles ............................................................................................... 44

3.1.7. Blog posts ........................................................................................ 44

3.1.8. Presentations .................................................................................... 45

3.1.9. Applications ...................................................................................... 45

3.1.10. Biographies ..................................................................................... 45

3.1.11. News ................................................................................................. 46

4. EVALUATION OF THE SDE ................................................................. 47

4.1.1. Workshop in Oulu (30/31 November 2013) ....................................... 47

4.1.2. Workshop in Lisbon (17th January 2014) ......................................... 49

4.1.3. Workshop in Budapest (29 January 2014) ........................................ 50

4.1.4. Workshop in Vigo (27th February 2014) .......................................... 51

4.1.5. Workshop in Bad Hofgastein (25th March 2014) ............................. 52
4.1.6. Workshop in Vilnius (11th June 2014) ................................................................. 53
4.2. Evaluation questionnaire ................................................................................ 53
  4.2.1. Evaluation of the collections of resources in the Knowledge Base. ............... 55
  4.2.2. Evaluation of the information retrieval ....................................................... 57
  4.2.3. Evaluation of word-sense disambiguation ................................................. 58
  4.2.4. Evaluation of exploratory search ............................................................... 59
  4.2.5. Evaluation of recommendations ............................................................... 60
  4.2.6. Evaluation of editors ................................................................................ 61
  4.2.7. General evaluation .................................................................................... 62
4.3. Discussion ........................................................................................................ 63

5. REFERENCES ......................................................................................................... 65

Conclusions .............................................................................................................. 67

6. Appendix I: LIST OF ABBREVIATIONS .......................................................... 68

7. Appendix II: CONTROL BOARDS .................................................................... 69
  7.1. Document preparation .................................................................................. 69
  7.2. Results .......................................................................................................... 69
INTRODUCTION

Reminder of the context

In the past few years, technology was playing an increasingly important role in classrooms. Thus, governments, ministries of education, and entities with competences in education are launching programs that aim at providing classrooms with technological infrastructure.

The availability of technologies in the classroom opens up a world of possibilities for the teaching practice. However, in spite of the great potential of technology-provided classrooms, a big part of them remain underused. In the literature, we can find several reasons for that:

• The first one is that technology does not fit well traditional teaching approaches—those based on master classes, in which technology is perceived as a distracting element.
• The second one is that teachers lack training on how to effectively use technologies in their teaching practice. Training has a twofold dimension: pedagogical and technological.

As a step towards overcoming those problems, several research groups working together within the framework of the iTEC project have proposed a series of pedagogical approaches, which have been tested in large-scale pilot experiences throughout Europe. Those pedagogical approaches emphasise the importance of teacher training for their adequate implementation.

The cornerstone of those approaches is the Future Classroom Scenario, conceived as a small cooperative project in which students, working in groups, carry out several Learning Activities. In those Learning Activities—in addition to the contents of the curriculum—students practice 21st Century Skills, such as: working cooperatively, using technology, and speaking in public.

The final outcome of iTEC is an integrated set of tools grouped under the name of Future Classroom Toolkit, especially designed to facilitate the implementation of iTEC pedagogical approaches. The software tools created under the umbrella of iTEC—aimed at facilitating the process of defining and planning new Future Classroom Scenarios, as well as adapting existing ones—are a key part of the Future Classroom Toolkit. More concretely, iTEC has among its objectives “To build a prototype assistant for advising users how to find, select and combine resources that support the project scenarios”.

The Composer allows teachers to find Learning Activities in accordance with taxonomy of 21st Century Skills, as well as sharing new Learning Activities with their community. The Widget Store allows teachers to create and find W3C widgets that may be of interest, and that are easily embeddable in a number of educational platforms, such as Moodle and .LRN. The P&E Directory allows for registering and finding events and persons, which may be of interest for the teaching practice. A crucial part of the iTEC software tools is the Scenario Development Environment (SDE), which applies semantic technologies to provide “intelligent” advice to users when defining, planning, and adapting Future Classroom Scenarios.

2 iTEC DoW, page 11.
One of the main contributions of the WP10 has been the semantic modelling of Scenarios, Technical Settings, Resources, and Services; concreted in an ontology whose final version is described in D10.3. The other main contribution has been the development of the SDE, understanding it as a backend accessible through an API and integrated with the rest of the iTEC software tools—and especially with the Composer, which acts as a consumer of the services that the SDE provides through its API.

In spite of not being described in the DoW, the enrichment of the Knowledge Base of the SDE from external sources was already identified in D10.2 as a very promising line of research. That research was subsequently elaborated and refined in D10.3. In the fourth year, we continued working on that research line, extending dramatically the number of elements in the Knowledge Base of the SDE—and inserting new types of resources in addition to tools, people, and events.

In the fourth year, we performed works on Semantic Search aimed at enabling the exploration of the Knowledge Base of the SDE by end-users.

**Purpose and scope of the task**

The activities carried out by WP10 for three past years have generated the following outcomes:

1. A detailed Semantic Model of all the elements relevant in the processes of Localisation and Planning.
2. A Reference Architecture for the SDE that defines its design principles and main functionalities.
3. A version of the SDE that implements all the functionalities defined in the Reference Architecture, as well as a prototype of client system that consumes the SDE API with demonstration and testing purposes.

As stated in D10.3, all the tasks specified in the DoW have been fulfilled:

- **Task 10.1 (M1-M48)—Semantic Modelling of Educational Scenarios and Technical Settings.**
  - T10.1.4 (M10-M48)—Evaluation and Maintenance.
- **Task 10.2 (M1-M48)—Semantic Modelling of Resources and Services.**
  - T10.2.4 (M10-M48)—Evaluation and Maintenance.
- **Task 10.3 (M1-M48)—Scenario Localisation.**
  - T10.3.3 (M15-M48)—Requirements’ Update.
- **Task 10.4 (M1-M48)—Planning.**
  - T10.4.3 (M15-M48)—Requirements’ Update.

WP10 is a research-oriented work package, whose main research goal is to validate the appropriateness of semantic technologies for supporting users in defining, localising, and planning Future Classroom Scenarios. In the fourth year, we carried out activities in the line of continuing the enrichment of the Knowledge Base of the SDE—previous work in this regard was reported in D10.3. Those activities are part of subtasks in T10.4.3.

---

3 iTEC DoW, page 30.
• Much more sources of information for tools, persons, and events have been incorporated; and the Knowledge Base of the SDE has thus increased considerably.
• Some NTCs were interested in incorporating other types of elements in the Knowledge Base of the SDE, apart from tools, persons, and events. For this reason, the Knowledge Base was extended in order to include other elements such as courses, lectures, and documentaries.

During the fourth year—always following the recommendations from NTCs—we performed actions targeted at enabling Semantic Search on the Knowledge Base of the SDE. The iTEC DoW states that the planning process “is supported by the resource discovery services”. In order to give greater visibility to the resource discovery process we created a series of mechanisms and Web interfaces that enable users to interact with the Knowledge Base of the SDE in a sort of Semantic Search. All these actions fall within T10.4.3.

The bulk of the actions carried out by WP10 in the fourth year were targeted at evaluating the SDE with end-users. This included the development of a Web application for that purpose, which enables users to interact directly with the SDE, and supports the entire life cycle of Future Classroom Scenarios.

Relationship with other tasks

WP10 has a direct relationship with other work packages. The main ones in the fourth year are:

• WP6—The evaluation of the SDE was coordinated with WP6, which also gathered evaluation results for the SDE.
• WP7—The functionality that the SDE offers through its API is consumed from the iTEC Composer, thus offering intelligent advice to users when creating Learning Activities. Besides, the SDE accesses elements stored in the Composer through its Harvesting API, adding those elements to the Knowledge Base of the SDE.
• WP8—The SDE is able to access the Widget Store for retrieving W3C widgets and storing them in the Knowledge Base of the SDE. Therefore, the SDE acts as an access point to the widget store, and it allows for adding widgets to the didactic plans of Future Classroom Scenarios.
• WP9—The information stored in the P&E Directory is accessed by the SDE in order to insert events and people in the Knowledge Base. In this way, elements from the P&E Directory populate the Knowledge Base of the SDE and users may benefit from the semantic enhancements provided by SDE technology when trying to discover resources of that kind for their Future Classroom Scenarios.
• WP11—The results from WP10 ready to be exploited were communicated to WP11.
Structure of the document

Chapter 1 reports the work performed in order to creating a user interface that enables the evaluation of the SDE with end-users. This chapter is divided into the following sections:

- Using the SDE for registering scenarios (Section 1.1). In this section, we describe the editor of Future Classroom Scenarios that was integrated into the SDE.
- Using the SDE for discovering resources (Section 1.2). This section details the user interface that enables to discover educational resources from the Knowledge Base of the SDE, and that is based in the principles of exploratory search presented in the chapter before.
- Bookmarking resources (Section 1.3). This section explains the mechanism for bookmarking resources that is integrated in the SDE.
- Getting recommendations (Section 1.4). In this section, we describe how users may get recommendations of educational resources.
- Creating a Learning Story (Section 1.5). This section explains how to create lesson plans in the form of narratives, which combine text with resources from the Knowledge Base of the SDE. We also explain the mechanism followed for suggesting alternative educational resources—with the objective of adapting Future Classroom Scenario to realities other than those they were conceived for.

Chapter 2 reports the work performed in order to enable Semantic Search on the Knowledge Base of the SDE; which uses state-of-the-art techniques in information extraction and information retrieval. This chapter has the following sections:

- Introduction to semantic search (Section 2.1). This section introduces the notion of semantic search, and it justifies why it results an appropriate approach to enabling the interaction of users with the Knowledge Base of the SDE.
- Concept-based representation of documents (Section 2.2). This section presents the concept-based representation of documents, understanding concept as “unit of knowledge”; and it justifies why the concept-based paradigm is superior to the word-based one.
- Automatic extraction of concepts (Section 2.3). In this section, we expose the procedure followed for automatically extracting concepts from text. The procedure builds on state-of-the-art techniques on Natural Language Processing and Machine Learning.
- Semantic relatedness between resources (Section 2.4). In this section, we expose the mechanism we use for calculating the semantic relatedness between resources.
- Suggesting resources to explore (Section 2.5). This section details our approach for suggesting resources to users on the basis of their browsing history.
- Suggesting alternative resources (Section 2.6). This section describes the procedure for suggesting alternatives to a given resource. This procedure builds on the notion of semantic relatedness.
- Architecture of the information retrieval system (Section 2.7). In this section, we detail the components that play a role in enabling concept-based information retrieval, as well as interactions among them.

Chapter 3 reports the actions that were carried out with the objective of enriching the Knowledge Base of the SDE, and that are a direct continuation of those performed in the years before. It describes types and sources of the resources currently indexed in the SDE.
Chapter 4 reports the evaluations of the SDE carried out with end-users. Concretely, we report the workshops in which the SDE was tested by end-users, and also the results of a questionnaire for evaluating the SDE that was distributed among the teachers participating in the project.

**Impacts of the Deliverable**

**iTEC project**

Review ‘Risk Analysis’

There are no new risks detected during the period under report.

**Ethical Issues**

None.

**IPR issues**

The Scenario Development Environment, which is the main contribution from WP10, will be released under the Apache Software License v2.0.
REPORT

1. A USER INTERFACE FOR THE SDE

The creation of a user interface for the SDE concentrated a big part of the efforts of WP10 in the fourth year of the iTEC project. The initial motivation for that interface was that some end-users manifested their interest in interacting directly with the Knowledge Base of the SDE, since they considered it to have many possibilities to offer, way beyond the suggestion of educational resources having into account the learning context of users.

The workshops with end-users (reported in Chapter 4) that WP10 carried out this year, allowed us to obtain a great quantity of feedback on the user interface, which enabled us to develop the users interface in an iterative and incremental way. At the end of that process, the Web interface of the SDE gives provides support to the entire life-cycle of Future Classroom Scenarios: defining a Scenario, discovering educational resources, creating the didactic planning, and even adapting other existing Scenarios. The Web interface can be accessed at:

http://www.itec-sde.net

The following sections document the process of interacting with the user interface of the SDE—hereafter we call it just SDE, for simplicity. For the sake of clarity, this chapter uses a prototypical Scenario created ex profeso to illustrate the interaction of users with the user interface. The Scenario is entitled: Life in Ancient Greece.

1.1. Using the SDE for registering scenarios

The SDE provides an interface for registering Future Classroom Scenarios, so that it may be later accessible by the creator and also by other SDE users. A Scenario is conceived as a first idea or a “vision”, which must be subsequently defined in more concrete terms—that is to say, a didactic planning. The icon designed to represent the element Scenario, which can be seen in Figure 1, aims at reflecting that idea. It is a speech bubble—typical in the word of comic—that is often used to denote that some character has had an idea. In the case of a Scenario, such an idea is a narrative—denoted by the “ancient” storybook. It should also be noted that the ownership of Future Classroom Scenarios may be shared with members of a team.

The template of Future Classroom Scenario is based on the one developed by WP2. It contains:

- The title, description, and a representative image of the Future Classroom Scenario.
- The core purpose of the Scenario, that is, its aspiration and aims.
- A narrative of the Scenario, which can be either a story of the Scenario in third person—describing what happens to the characters—or a “day in life” story narrated from the point of view of a main character.
- The trends embraced by the Scenario, in accordance with the model for identifying trends from the Future Classroom Toolkit.
- Possible approaches to teaching and assessment.
- Environments, that is to say, where the Scenario is going to take place.
- Persons and roles involved in the Scenario—teachers, experts, learners, etc.
- A first draft of the activities that will be carried out during the performance of the Scenario.
• Resources that will be used during the course of the Scenario, including technologies.

![Figure 1 Template for editing a Future Classroom Scenario. In this figure, only the title and description fields are filled up](image)

### 1.2. Using the SDE for discovering resources

As we said in the introduction to this chapter, the Knowledge Base of the SDE, after having been enriched with educational resources from external websites, stores a great quantity of educational resources. With the objective of enabling the exploration of the Knowledge Base by end-users, our idea was to develop a user interface in accordance with the principles of Semantic Search. Figure 2 shows the home page of the SDE; in the top menu, under the discover menu item, the list of categories of resources is displayed.

As (Guha, McCool, & Miller, 2003)—among other authors—state, the information in the Web should, ideally, be understandable by machines. That is the main purpose of the so called Semantic Web: to offer the information in a semantic format, in such a way that other software systems may consume that information “understanding” its meaning. Regrettably, as it was already discussed in D10.3, the websites that publish information in a semantic way still are a minority, in comparison to those that publish information using semi-structured formats such as HTML. Therefore, it is necessary to process that information in order to translate it to a semantic format,
and thus be storable in the Knowledge Base of the SDE. That is to say, the approach we follow could be described—paraphrasing Francis Bacon—as “if the mountain does not come to us, we go to the mountain”.

The other line of action to make information understandable by machines consisted on the semantic annotation of educational resources in accordance with their thematic. The traditional way of performing that task is to employ human annotators (or indexers), who assign concepts to resources in accordance with some taxonomy of concepts. That approach has two main drawbacks. The first one has to do with scalability, because the task of assigning concepts to resources is very time consuming. The second one has to do with the error proneness of the annotation process; besides, several human annotators might follow several criteria.

Our approach to indexing resources in accordance with their thematic is based on using Natural Language Processing and Machine Learning techniques, in order to train a software agent so that it may create semantic annotations of educational resources.

![Figure 2 Home page of the SDE, with the categories of resources that can be discovered](image)

1.2.1. Suggesting concepts

The SDE allows for accessing the search functionalities to any user. Thus, any user can access to [www.itec-sde.net](http://www.itec-sde.net) and try the Semantic Search of educational resources.

The interaction with the Knowledge Base starts with the introduction of some search term in the search box in the SDE. Thus, following with the example proposed in the introduction we can
introduce the search query “greece”. As it can be seen in the Figure 3, “greece” may refer to several different concepts. Hence, retrieved resources may relate to any of those concepts. As a consequence, for the particular case of the query “greece”, we can see among the search results some resources that refer to, among others:

- Greece, the country.
- Greece national football team.
- Ancient Greece. This is the concept we are interested in.
- Kingdom of Greece.

The most relevant thing from the listing of educational resources is the presentation in categories. In this way, we can see places of interest about “greece”, lectures about “greece”, documentaries about “greece”, etc. This way of presenting results is very different from that of the traditional search engines, because the SDE’s categories refer to the types of educational resources, and not to their file format.

Figure 3 This image shows the list of concepts proposed for the “greece” search query
1.2.2. Seeing concept summaries

When we select a concrete concept from the list of suggested concepts, only the resources about that particular concept are displayed—all the rest are filtered. Thus, if we select the concept “Ancient Greece”, all the resources about “Greece (country)”, “Greece National Football Team”, and “Kingdom of Greece” are filtered.

Let’s see this idea with another example. Since we are trying to find educational resources about the Ancient Greece, we are going to search about its mathematics. Thus, we enter the query “Pythagoras”; then, we click on that concept, and go to the page that visualises the concept. As it can be seen in the Figure 4, the grey box at the top of the page displays information on the concept “Pythagoras”: a short paragraph describing his life and works; a representative picture of Pythagoras, a sculpture concretely; and also a list of key-value pairs that relate Pythagoras to other concepts. That list of key-value pairs allows users to explore the graph of concepts. For instance, from the concept “Pythagoras” we can navigate towards the concepts “Thales”, “Anaximander”, and “Pherecydes”, which represent the influences of Pythagoras; but we can also navigate towards “Samos” and “Metapontum”, which are, respectively, the places where Pythagoras was born and died.

This way of interacting with the Knowledge Base described in the paragraph above is called Exploratory Search through the Space of Concepts, since from a particular concept we can keep on exploring through other related concepts.

After the description of the concept “Pythagoras”, the SDE displays a list of the educational resources that have been annotated with the concept “Pythagoras”—as shown in Figure 5. We can see, among other categories and resources:

- Artworks, such as “Pythagoras Emerging from the Underworld” by Salvator Rosa, and “Pythagoreans Celebrate Sunrise” by Fyodor Bronnikov.
- Places of interest, including the Pythagorion Archaeological Museum.
- Lectures, such as “Music of the (p)Spheres” by Alessandro Panconessi from the Sapienza University of Rome.
- Documentaries, such as the one entitled “Genious: Pythagoras”, and also “The Story of Maths”.
- Blog Posts, such as “Double Integral of Pythagoras over rectangular region”.
- News reports; for instance, “Why we all love numbers”.
- Open Educational Resources, such as the one on Pythagora’s Theorem from the LRE.

1.2.3. Detail of an educational resource

When a user chooses an educational resource from the search results, the system displays a summary of that educational resource—which includes a brief extract from its textual content, a representative image (in case it is available), plus metadata of the educational resource. Besides that, the system shows the tags that have been automatically generated, and that correspond with the bag-of-concepts representation of the resource—Figure 6 shows the detail of a particular resource. Each of those tags is a navigable link, in such a way that clicking on a tag the user can navigate to a concept summary—and to the relevant resources for that concept.
The behaviour described above is called exploratory search through the space of tags. This kind of exploratory search works seamlessly in conjunction to the exploratory search through the space of concepts.

The exploratory search through the space of tags is inspired in social tagging sites such as Del.icio.us, Flicker, or even Twitter and Instagram. For instance, in Del.icio.us, users assign tags to the resources that they incorporate into the platform; and those tags are navigable—clicking on a particular tag, other resources that were assigned the same tag are retrieved. In Flickr, where users assign descriptive tags to pictures, the same mechanism is applied. In all those systems that implement social tagging the exploration of new documents through tags is natural and convenient.
Figure 5 Search results for the concept “Pythagoras”, organised by categories
A similar approach is used in repositories of educational resources such as Klascement, in which users assign tags to the educational resources that they aggregate to the repository, and those tags may be used for exploring related educational resources. In other repositories, professional indexers perform the task of tagging educational resources.

Our approach to assigning tags presents some advantages over the “manual” generation: it is less time-consuming; it has not any problems with synonymy and polysemy; and it is very consistent—it does not depend on personal preferences when assigning tags.

The detail of an educational resource also includes the visualisation of resources that are related to it, as shown in Figure 6 and Figure 7. In those figures, we can see resources related to The Odyssey, which include other books—such as The Iliad—and also other kinds of resources, for instance, the archaeological site of Troy.
1.3. Bookmarking resources

Our approach to exploratory search has to do with the idea that users interact with the Knowledge Base of the SDE, navigating through the space of concepts, using the semantic annotations in educational resources for navigating through other concepts, and navigating towards the resources related to a given one. The final objective of that exploration is the gathering of educational resources that may be later inserted in the didactic planning of a Future Classroom Scenario.

As a mechanism for gathering educational resources, we implemented the functionality of bookmarking them, in a similar way to that of a Web browser. Thus, from the page for visualising a resource, users can bookmark the resource that is being currently displayed. Besides, in the personal space of users, they can access “My bookmarks”, which is a page that displays all the resources they have bookmarked insofar, organised in accordance with their categories.

In the Figure 8 we can see how the SDE organises the bookmarked resources. For each one of the categories the SDE shows a thumbnail with the title plus a representative picture of each bookmarked resource. In this screen, users can also un-bookmark resources. Following with our example of the Future Classroom Scenario entitled “Life in Ancient Greece”, the figure shows some bookmarked educational resources, gathered in an exploratory search:
1.4. Getting recommendations

The interactions that the user performs with the SDE—concretely, visualising resources and bookmarking them—gives precious information that enables the SDE to infer user’s interests. We
can say that visualising an educational resource shows that the user has some interest in it, and that bookmarking it shows that the user has even more interest.

The semantic annotations—or tags—of educational resources are key for defining the interests of users. Thus, when a user bookmarks a resource that contains the concepts “archaeology” and “mycenae”, the SDE infers that those are concepts of interest for the user, and it will recommend to the user other educational resources that have those semantic annotations.

Recommendations are presented to users after logging into the system, and they are organised in accordance with the categories of educational resources. Those recommendations may serve as a personalised starting point for the exploratory search of educational resources.

Figure 9 shows recommended educational resources for a particular user. These recommendations are created in function of some interests (past interactions with the Knowledge Base), such as: archaeology, astronomy, and classic culture.
1.5. Creating a Learning Story

In the SDE, we integrated an editor devoted to the creation of Learning Story to support lesson planning. The approach of that editor is combining resources and free text in the form of a narrative. In this way, a teacher may combine Learning Activities, events, and applications, in a sort of “story”.

Thus, the concept of narrative or “story” is central in this proposal. This is a key distinctive feature from other planning alternatives, which choose a more structured approach.

The icon designed for illustrating our narrative approach tries to represent an ancient storybook. As we can see in the Figure, each Learning Activity included in the Learning Story is marked with an icon that denotes the beginning of a new “chapter”.

The bar at the bottom of the page allows for selecting the type of element that we want to incorporate into the narrative: free text, Learning Activity, application, event, etc. We wanted the editor to be as easy-to-use as possible; hence, the interface works in a drag-and-drop way, enabling a convenient sorting of elements of the Learning Story.

Figure 10 shows a narrative with three elements: free text, a Learning Activity, and a documentary. The user is dragging the introductory paragraph. In the bottom of the editor we can see the bar for adding new elements to the narrative.

1.5.1. Alternative resources

One of the objectives of the SDE is to provide functionalities that enable the adaptation of Learning Stories to contexts different from those they were initially conceived for. One of the key mechanisms for going towards that objective is that of suggesting alternative resources to a given one. The main idea here is that when a user visualises an already planned Learning Story—with all its educational resources already inserted in it—and wants to adopt it, the SDE provides alternatives to those resources that the user does not have at disposal at that moment. For instance: an alternative to an event that has already taken place, an alternative to a museum that is in another city, and alternative to a book that is written in another language, etc.

Alternatives are accessible when the user is visualising a Learning Story created by another user. There, in each box representing an educational resource, the user can click on the “three arrows” icon and then a pop-up with alternative resources is displayed.

Figure 11 shows the alternatives to Google Drive. The set of alternatives is composed of Cloud applications that can be used instead of Google Drive, that is, applications that can be described with a similar set of concepts.
Figure 10 This is the editor for the creation of Learning Stories.
Figure 11 Alternatives to Google Drive.
2. SEMANTIC SEARCH ON THE KNOWLEDGE BASE

2.1. Introduction to semantic search

(Guha et al., 2003) differentiate between two types of user interactions with search engines: navigational searches and exploratory searches. In navigational searches, users provide the search engine with a phrase or combination of words that must be present in the retrieved documents—but that particular combination does not necessarily denote a concept. According to the authors, in navigational searches, users make use of the search engine for “navigating” towards a particular document. Conversely, exploratory searches are from a very different kind: the phrase or set of words that users provide to the search engine does denote a concept—with the objective of gathering information or doing research about it. In an exploratory search, there is not a particular document that the user knows a priori and tries to retrieve; the search engine is used as an aid for exploring a space of documents.

(Verbert et al., 2012) reflect about the difficulty in expressing an educational requirement through keywords in a traditional search engine (such as Google):

> It is difficult to express a specific learning requirement through keywords. For example, search engines do a poor job when a learner needs content about ‘relativity theory’, oriented to high school level, with a duration of about 30 minutes.

We share that view and propose exploratory search as a first step in the direction of overcoming the difficulties pointed above. In an exploratory search, the search terms denote a concept—understanding concept as a unit of knowledge (Huang, 2011). Retrieving documents in which a specific concept is relevant presents serious difficulties, which have to do with synonymy and polysemy:

- As spotted by (Egozi, Markovitch, & Gabrilovich, 2011), when the search terms are different from the terms used in a particular document, that document will not be retrieved. That is to say, synonymy negatively affects recall.
- Besides, polysemy negatively affects precision. When a user is searching for documents about planet Mercury, the retrieved documents that are about the chemical element called mercury have to be seen as an imprecision of the search engine.

2.2. Concept-based representation of documents

2.2.1. From combinations of words to concepts

In traditional information retrieval systems, documents are represented in accordance with the bag-of-words paradigm. Let’s suppose that we have two text documents:

- Document A: “mercury is a chemical element”.
- Document B: “Mercury is a planet of the solar system”.

On the basis of the two documents above, we can build the following dictionary: “mercury”: 1; “is”: 2; “a”: 3; “chemical”: 4; “element”: 5; “planet”: 6; “of”: 7; “the”: 8; “solar”: 9; “system”: 10.
Following with this representation, documents are represented as vectors—with as many dimensions as entries in the above dictionary:

- Document A = [1, 1, 1, 1, 1, 0, 0, 0, 0, 0]
- Document B = [1, 1, 1, 0, 0, 1, 1, 1, 1, 1]

In this way, when a user introduces the term “mercury” the search engine checks the representation vectors and returns documents A and B—being document A completely irrelevant for a user that is looking for information on planet Mercury.

Creating a dictionary of concepts (Sahlgren & Coster, 2004) in the following way would solve the above problems: “mercury (element)”: 1; “chemical element”: 2; “mercury (planet)”: 3; “planet (celestial body)”: 4; “solar system”: 5. Thus, in accordance with this dictionary, documents would be represented as:

- Document A = [1, 1, 0, 0, 0]
- Document B = [0, 0, 1, 1, 1]

Therefore, it is clear that a concept-based representation of documents will help to solve the problems of synonymy and polysemy, and it is thus the cornerstone of the exploratory search strategy implemented in the system.

2.3. Automatic extraction of concepts

To know what concepts are referenced in a text document is a non-trivial task. Traditional approaches to this problem are based on using humans for generating the so-called semantic annotations (Reeve & Han, 2005), which are chosen from a set of possible annotations called taxonomy.

The principal problem with “manual” annotations is that it does not scale well. The generation of manual annotations is very time-consuming and error-prone. In order to overcoming that, some authors have proposed the automatic extraction of concepts from text documents, using techniques from the fields of Natural Language Processing and Machine Learning (Gabrilovich & Markovitch, 2007; D. Milne & Witten, 2008a). Following that approach, given a text document, a software component “extracts” the concepts that particular document refers to.

In the literature, we can find several examples of automatic extraction of concepts (Mendes, Jakob, Garcia-Silva, & Bizer, 2011; Mihalcea & Csomai, 2007; D. Milne & Witten, 2013a). (Gabrilovich & Markovitch, 2007) propose a method called Explicit Semantic Analysis (ESA), by which any text document is represented as a weighted vector of concepts from Wikipedia. The authors chose Wikipedia because it is the greatest repository of knowledge in the entire Internet. Evaluation of the effectiveness of this approach was performed by comparing its performance in the task of computing semantic relatedness between text documents to the traditional bag-of-words approach, concluding that ESA gives better results than the traditional method.

In a similar line, (D. N. Milne, Witten, & Nichols, 2007) represent documents on the basis of concepts from Wikipedia, to be used in the Koru search engine. According to the authors, Wikipedia is very suitable for this task, because it contains a big number of terms, with relationships between terms that have been added manually and are domain-independent.
Other systems build on resources different from Wikipedia for their concept-extraction task. Thus, (Boubekeur & Azzoug, 2013) derive the concepts for representing documents from linguistic resources such as WordNet and WordNetDomain.

### 2.3.1. Automatic extraction of concepts “under the hood”

The algorithm for automatic extraction of concepts used in the system was proposed by (D. Milne & Witten, 2008b), and it uses techniques from the fields of Natural Language Processing, Machine Learning, and Data Mining—concretely, it is based on data mining in Wikipedia. The process for extracting concepts is performed in three steps (see Figure 12), which are discussed below.

The first step is **candidate selection**. Given a text document, a document with all the n-grams—being an n-gram a contiguous sequence of n words—present in the document is generated; then, the algorithm queries a vocabulary with all the anchor texts from Wikipedia to check if some n-gram is present in the vocabulary. Thus, the most important candidates (n-grams) will be those that are used more often as anchor texts in Wikipedia.

The second step is **disambiguation**. Given a vocabulary of anchor texts, the algorithm selects the most probable target article for each candidate. The procedure is based on Machine Learning and has two factors into account:

- The commonness of the relationship between an anchor text and a target article.
- The relation with the non-ambiguous terms from the context.

As the training sequence, the articles from Wikipedia—that contain good examples of “manual” disambiguation—are used.

The third step is **link detection**, whose objective is to measure the relevance of concepts in text. To that end, a Machine Learning algorithm is used—following the approach that each article in Wikipedia is an example of what constitutes a relevant link and what does not.

### 2.4. Semantic relatedness between resources

In the detailed view of a particular resource, we can see how the system suggests to the user other resources that might be of interest, on the basis of the semantic relatedness to that resource.

In the first place, the system suggests resources from the same type of that currently being displayed. For instance, when the system is displaying the details of a particular course, the first row of related resources is composed of other resources with a similar thematic. Immediately below, the system displays related resources from other categories—documentaries, lectures, etc. Thus, the user may click on any related resource and navigate to its detailed view, which in turn will display its own related resources. This procedure for discovering content is called exploratory search through the space of related resources.
2.4.1. Related resources “under the hood”

The problem of retrieving related resources is defined as the problem of approximating the function

\[ R : \mathcal{D} \times \mathcal{D} \rightarrow [0,1] \]

where \( \mathcal{D} \) is the domain of documents, \( R(d_i, d_j) = 0 \) when document \( d_i \) is totally unrelated to document \( d_j \), and \( R(d_i, d_j) = 1 \) when the two documents are completely related.

Following the Vector Space Model (Salton, Wong, & Yang, 1975), a document \( d_i \) is represented as a vector.
composed of weights of features, being $\mathbb{F}$ the set of features, and $w_{ij}$ the weight the feature $f_i$ in the document $d_j$, which is a measure of the contribution of that feature to the semantics of the document. Thus, a document defines a point in a $|\mathbb{F}|$-dimensional space.

Defining some measure of distance between points we can calculate the degree of similarity between documents. The most common metric is the cosine similarity; thus the relatedness between documents is approximated as follows:

$$R(d_i, d_j) = \cos(d_i, d_j) = \frac{d_i \cdot d_j}{\|d_i\|_2 \cdot \|d_j\|_2}$$

Using the formula above, we have approximated the function that defines which documents are most related to a given one.

### 2.5. Suggesting resources to explore

The objective of a recommender system is to guide users in a personalised way towards interesting documents in a big space of possible options (Lops, Gemmis, & Semeraro, 2011). There are several types of recommender systems in accordance with their way of working. Among them, content-based recommender systems try to recommend documents similar to those that have interested the user in the past.

An important objective in our research is to explore the applicability of the BoC paradigm, which gave us so good results in information retrieval, to content-based recommendations. This is a not-well explored area of research, as stated by (Lops et al., 2011)

To the best of our knowledge, there are no (content-based) recommender systems able to exploit the above mentioned semantic text representations for learning profiles containing references to world facts. The positive results obtained exploiting the advanced text representations in several tasks, such as semantic relatedness, text categorization and retrieval, suggest that similar positive results could also be obtained in the recommendation task. It seems a promising area, not yet explored.

The notion of recommendation is formulated as follows. Let $\mathbb{D}$ be the domain of documents and $\mathbb{U}$ the domain of users. The problem of recommendation can be defined as the problem of approximating the function

$$\mathcal{J} : \mathbb{U} \times \mathbb{D} \rightarrow [0,1]$$

being $i(u_i, d_j)$ the interest of user $u_i$ in the document $d_j$. Then, for each user $u \in \mathbb{U}$ we want to find the document $d'$ that maximises the interest of the user

$$\forall u \in \mathbb{U}, d' \in \mathbb{D} \arg \max_{d \in \mathbb{D}} i(u, d)$$
In order to known which resources may be relevant for a particular user we need to formulate the notion of interest. In a domain of concepts \( C = \{c_1, c_2 \ldots c_{|C|}\} \) the interest of a user in a concept is modelled as the function

\[
\mathcal{I}: \mathbb{U} \times C \mapsto \mathbb{R}^+
\]

In this way, the interest of user \( u_i \) can be modelled as a weighted vector of interests in concepts

\[
\forall u_j \in \mathbb{U}, \text{interests}(u_j) = [i_c(u_j, c_1), i_c(u_j, c_2), \ldots, i_c(u_j, c_{|C|})]
\]

The calculation of the interest of a user is modelled as a machine-learning problem. Let \( A = \{a_1, a_2 \ldots a_n\} \) be the set of possible actions that a user \( u \in \mathbb{U} \) can perform on a document \( d \in D \). The informativeness of actions is modelled as

\[
Inf: A \mapsto [0,1]
\]

which tells us how much a user is interested in a document when performing an action upon that document. Thus, we can model, for instance, that bookmarking a document is more informative of the user’s interest in the document that just viewing it

\[
\text{bookmarking} \in A, \text{viewing} \in A, \text{inf(booking)} > \text{inf(viewing)}
\]

Having that into account, we define the variation in users’ interests as a result of an action as

\[
\forall i,j, \quad \Delta \text{interests} = \text{inf}(a_i) \cdot d_j
\]

which is then added to the vector of interests for a user

\[
\text{interests} = \text{interests} + \Delta \text{interests}
\]

In summary, every action that users perform on documents contributes—provides information—so that the system may learn about users’ interests.

The most simple strategy for recommending content—educational resources—is based on calculating the distance of documents to the vector of interests of users and thus retrieving the nearest ones—using some metric of distance, such as cosine similarity.

### 2.6. Suggesting alternative resources

One of the most ambitious objectives of our research is to provide support to the adaptation of Learning Stories to realities different from those they were conceived for. Let’s see that with some examples:

- A particular Learning Story includes the visit to a Natural History museum in Paris; it is adapted to be carried out in the city of Barcelona—visiting the Natural History museum of Barcelona.
- A Learning Story includes the attendance to a poetry reading that has already taken place; it could be adapted so to include the attendance to another similar event still to take place in the future.
• A Learning Story that makes use of a proprietary application could be adapted so that it uses, in its place, a free software application that provides equivalent functionalities.
• A Learning Story that uses a certain book as reference could be adapted to use a free book.

It is important to define to what Learning Stories are adapted. A Learning Story must be adapted to the context in which it is going to take place, understanding as context all those variables that have influence in the adaptation process; and that, in some way, define to what the Learning Story is going to be adapted. For instance, we can think on location, time, and language—among others. Thus, the context is defined as a set of contextual dimensions $\mathbb{K}$, which include time, location, and language.

The function alternative is defined as

$$\mathcal{A}: \mathbb{D} \times \mathbb{K} \mapsto \{\mathbb{D}\}$$

where $\mathbb{D}$ is the domain of documents, $\mathbb{K}$ is the set of contextual dimensions, and $\{\mathbb{D}\}$ is an ordered set of alternative documents.

Depending on the type of document that we want to adapt to a particular context, we must have into account a lesser or bigger number of contextual dimensions. For instance, if we want to provide alternatives to a place of interest it may be enough to have into account one contextual dimension: the location

$$\mathcal{A}: \mathbb{D} \times \mathbb{L} \mapsto \{\mathbb{D}\}$$

In this way, we can say that

$$a(d_i, l_j) = \{d_k, d_l, \ldots\}$$

are the alternative documents to the document $d_i$—being it a document that describes a place of interest—in the context that is given by the location $l_j$, and that $d_k$ is the best alternative available.

In case we want to obtain an alternative to a document that represents an event, it is necessary to take into consideration another contextual dimension: time. The function alternative will have the following formulation

$$\mathcal{A}: \mathbb{D} \times \mathbb{L} \times \mathbb{T} \mapsto \{\mathbb{D}\}$$

Thus, for a document describing an event we can say that

$$a(d_i, l_j, t_k) = \{d_l, d_m, \ldots\}$$

are the alternatives to document $d_i$, in the context described by the location $l_j$, and the time $t_k$.

The notion that a document $d_i$ constitutes an alternative to document $d_j$ builds on the notion of semantic relatedness. There are multiple measures of distance, but we use the cosine similarity

$$\mathcal{R}(d_i, d_j) = \cos(d_i, d_j) = \frac{d_i \cdot d_j}{\|d_i\|_2 \cdot \|d_j\|_2}$$
The method we follow for retrieving the list of related documents is *filtering a posteriori*, defining several filtering thresholds for each one of the contextual dimensions. Thus, the results—alternatives—will be ordered in accordance with their semantic relatedness and, after that, filtered on the basis of their belonging to a range of values for each contextual dimension. This approach enables a great flexibility, because the adaptation problem is reduced to obtaining the set of ordered similar documents and, after that, filtering it as in a conventional faceted search.

Maybe the biggest advantage of using a concept-based representation is that we can measure the semantic relatedness between documents that have been written in different languages. This is possible because a majority of concepts have a correspondence with concepts in English, hence enabling us to use those representations as vectors in concept space, and apply those strategies. More formally, we can say that for a book the function alternative takes the following form

\[ A : \mathcal{D} \times \mathcal{L}_A \mapsto \{\mathcal{D}\} \]

and the same function applies to other documents such as lectures, slideshows, etc. This enables the suggestion of alternative documents written in other languages. For instance, the SDE can suggest to a Spanish user an alternative to a book that is written in English, on the basis of the concepts that book is about.

### 2.7. Architecture of the information retrieval system

The architecture of the information retrieval system (see Figure 13) is based on several software components that cooperate between them: a Web application, a pool of Wikipedia Miner (D. Milne & Witten, 2013b) instances, an indexer, a data store, and several web scrapers.

The Web Application is based on Ruby on Rails, which is the combination of the Ruby programming language and the framework Rails, which is specially designed for developing Web applications—in fact, we could say that Rails provides a Domain-Specific Language for coding Web applications. The great community of developers that was built around Ruby on Rails in the past few years, as well as the solidity and maturity of the most part of library modules available—known as “gems”—make Ruby on Rails a very attractive option to serve as basis of the Web application module.

Ruby on Rails follows the Model-View-Controller (MVC) paradigm (Krasner & Pope, 1988). Thus, controllers receive HTTP requests and, after having queried the models, return to the browser the result of processing a particular view. Let’s see that with an example: when a user enters the following in the address bar of the browser

http://www.itec-sde.net/en/lectures/1

the lectures controller queries the models in order to retrieve the lecture with id=1 and, after that, processes the view

/lectures/show.html.erb

that is in charge of generating the HTML code with the detailed view of lecture 1, including its related resources.
In the example request discussed above, the model Lecture represents the class of a lecture that is persisted in the component called Data Store, which is a relational database. The transition from the “relational world” to the “object world” (and vice versa) is performed following the Active Record (Fowler, 2003) paradigm. The technological option for implementing the data store in the system has been a MySQL database.

Performing searches directly on the data store is very time consuming, which may be solved by maintaining an inverted document index. That index is defined as the function

\[ J: \mathbb{C} \rightarrow \{< D, [0,1] >\} \]

where \( \mathbb{C} \) is the domain of concepts, \( \mathbb{D} \) is the domain of documents, and \([0,1]\) is the range of the weight of a concept in a document. Thus, we can say that

\[ J(c_i) = \{< d_j, w_{ij}, \ldots, < d_z, w_{zj} >\} \]

being \( d_j \) and \( d_z \) documents that make reference to concept \( c_i \), and \( w_{ij}, w_{zj} \) the weights of that concept in those documents.

As technological option for implementing the indexer we opted for Apache Solr (Smiley & Pugh, 2009). With some simple adjustments, Apache Solr can be used to implement the bag-of-concepts strategy:

- Concepts are represented as a text string preceded by the symbol 
- Stemming is disabled for text strings that are preceded by the symbol 
- The tokenisation strategy for text strings that are preceded by symbol # is to leave the text string intact.
- The most important point in the strategy is to exploit the functionality of Solr known as “boosting”. We defined several fields to be indexed for each document, assigning a certain “boosting” to each field. Thus, the concepts in each document are indexed with a given “boosting”, which depends on the relevancy of concepts in particular documents.

With the above adjustments, we can leverage a very mature technology (Apache Solr) to be used as the basis of the indexer component in the system. This guarantees the stability and scalability of the indexer, as well as a blazing fast response time. Another important advantage of this approach is that, since the core of Solr is not modified, we can make use of a complete ecosystem of software components that were developed for Solr. One of those components is the library module—or “gem”—known as Sunspot, which provides a Domain Specific Language for querying Solr from Ruby on Rails applications, and that is used in the Web application.

With the goal of obtaining the weighted concept vectors that represent documents we use a pool of Wikipedia Miner instances. The rationale for needed several instances is that the process of extracting concepts has a high computational cost. In the system, we have as many Wikipedia Miner Proxy objects as instances of Wikipedia Miner. The working mechanism is based on the classical producer-consumer problem. In the system, there are several producers of semantic annotation tasks, which are certain scripts that run in background. A task is characterised by its type and its input document. Following this approach, producers add tasks to a queue, and consumers—proxies of Wikipedia Miner instances—extract them from the queue, thus maximising the throughput of Wikipedia Miner instances.
Finally, the components called “scrapers” are in charge of obtaining the textual representation of documents to be indexed, in such a way that Wikipedia Miner instances may extract their associated weighted concept vectors. Sources of documents—educational resources—include well-known portals such as Khan Academy, Coursera, and the Learning Resource Exchange.

### 2.7.1. Understanding the architecture with the help of a metaphor

The metaphor for the architecture of the search engine (illustrated in Figure 14) is as follows. Let’s imagine a huge library with the vocation of storing all the knowledge generated by the humanity. That store could be named Alexandria, like the famous library of Antiquity. To that library new books arrive every day. The staff of the library—which is very selective—buys those books and they only buy books of outstanding quality. When a new book enters the library, a member of the team of Diderot and d’Alembert—the annotators—is in charge of reading it. When he/she finishes reading it, labels are attached to the book. Each label corresponds to any entry in the last version of the Encyclopaedia. Once the book is properly labelled, they pass it to Hypatia, who takes care of creating new sheets in the index of the library, which is a huge room with archives. When a user arrives to the library, he/she ask the librarian—we could agree that Aristotle might be a good candidate for that job—about a particular topic. In case of the term is ambiguous, the librarian asks which option the user refers to, and then checks the index (in the archives) in order to retrieve books for the user. If the user considers interesting a particular book, the librarian may suggest similar books to him/her, on the basis of the topics the book is about.

**Data Store.** In a way similar to a library, which stores books on the basis of their genera, our data store keeps information on educational resources based on their type. In this way we will have some shelves for biographies, other ones for places of interest, etc. The store allows for enabling new shelves when the existing ones cannot allocate new books.

**Scrapers.** The library in our metaphor receives new books every day. These books are placed in shelves based on their genera. Each of the persons in charge of selecting and bringing the books is specialised in a particular genera. For instance, the library has among its staff an expert in biographies and an expert in travel books. In the architecture of the SDE we have several Web scrapers, each one devoted to a certain kind of educational resources—and in many cases to a certain source in particular. This approach is highly scalable, because in case of incorporating more sources, it is sufficient to allocate more machines for running the scrapers.

**Annotators.** In our metaphorical library, it is necessary to classify every book that arrives—and for that it is needed to read it. To that end, the library has a team of experts: Diderot and their colleagues; who are in charge of reading books and labelling them—creating annotations. Diderot and their team follow the approach based on the premise that annotating a book—identifying its most relevant topics—it is conceptually similar to identifying the relevant topics in an article of the Encyclopaedia. When the library hires a new member for Diderot’s team is needed to train him/her in order to learn the strategy for labelling books. To that end, a series of articles can be used for training the new member. In the architecture of the SDE we have several servers for creating semantic annotations of educational resources. This task is computationally expensive. The good news is that the approach is very scalable, because if we want to raise the volume of resources that we want to enter in the store we just have to allocate more servers to run Wikipedia Miner.
Indexer. Following with the metaphor of the library, it has a room with archivers. When Hypatia—who takes care of the index—receives the annotations of a new book, she takes care of writing sheets and archiving them. The number of entries in the archive is thus equal to the number of articles that there are in the Encyclopaedia of Diderot and d’Alembert. If the index increases we just have to allocate more archives. In the SDE architecture, since the indexing is based in Solr, this guarantees a high scalability, because Solr allows for a distributed indexing. In this way, if the index increases we just have to allocate more servers.

Figure 13 Architecture of the information retrieval system
Search application. The librarian of our metaphor takes care of several tasks. In the first place, he takes care of disambiguation. That is to say, when a user asks about an ambiguous term, he tries to disambiguate it—using a copy of the Encyclopaedia to that end. Then, he presents the most relevant books about the topic to the user, which he gets by checking the index. In the second place, when a user wants to go deeper on the topics of a particular book, the librarian suggests related books to the user. In the SDE architecture, the librarian corresponds to the application layer, which is a Web application written in Ruby on Rails. Also this part is highly scalable, because if the number of users increases, we just have to allocate more servers to run the application—to hire more librarians.

Figure 14 Metaphor of the architecture. From bottom to top: personnel insert books in the shelves of the library; encyclopaedists annotate those books putting tags on them; the indexer creates entries for those books in the index; finally, the librarian suggests interesting books to the user.
3. ENRICHMENT OF THE KNOWLEDGE BASE

In D10.2, we identified the enrichment of the Knowledge Base of the SDE as a promising research area. The first positive results of using enrichment are reported in D10.3, where three mechanisms of enrichment were discussed—in accordance with the type of source from which information is extracted. Thus, extraction can be performed: from semantic sources, which publish information in RDF; from structured sources, such as those that publish information in highly structured formats, such as XML and JSON; and, finally, from semi-structured sources, such as websites that publish information in HTML. In the fourth year of the iTEC project, we centred on refining our extraction technique from semi-structured sources, as they represent the bulk of the Web.

Figure 15 The “scrapers” are Internet spiders that crawl the Web retrieving pieces of information for specific target websites
The extraction from semi-structured sources is a mechanism that is commonly known as Web scraping. Figure 15 aims at representing graphically the scraping process. Scrapers—as they were described in Section 2.7—are Internet spiders that crawl the Web in search for target websites from which to extract information. In the figure this is represented as the scrapers “cutting” pieces of relevant information from target websites. Once those pieces of information are cut, they are added to the Knowledge Base of the SDE.

3.1. Resources indexed in the SDE

The choice of new types of resources, as well as the sources where they were extracted from, was performed in accordance with their perceived value for serving as educational resources. In the process, we follow the “beyond content” paradigm of iTEC, as stated in the DoW, and we incorporated non-traditional educational resources such as: places of interest (i.e. museums, exhibitions, architectural buildings), events, and even posts in high-quality blogs. We also decided to include more classical educational resources, such as: documentaries, lectures, and presentations, among others. It follows a description of those sources of elements.

3.1.1. Places of interest

Sites are places that may have an educational value. Those include monuments, museums, libraries, architectural sites, etc. For instance, if we search for “astronomy”, some places of interest correspond to observatories and planetariums. This information can be extracted from the Internet, from websites such as Spain is Culture http://www.spainisculture.com, belonging to the Spanish Ministry of Culture; the website of the UNESCO, where places that are part of the heritage of humanity are listed; or Trip Advisor http://www.tripadvisor.co.uk a private company that aggregates information about touristic places. An important piece of information about a place of interest is its location, which allows us for making searches having that factor into account---and in that way be able of contextualising the search results for a user.

Table 1 Sources of places of interest.

<table>
<thead>
<tr>
<th>Logo</th>
<th>URL</th>
<th>Description</th>
<th>Number of resources</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image" alt="Spain is Culture" /></td>
<td><a href="http://www.spainisculture.com">www.spainisculture.com</a></td>
<td>Website of the Spanish ministry of tourism.</td>
<td>7992</td>
</tr>
<tr>
<td><img src="image" alt="UNESCO" /></td>
<td>en.unesco.org</td>
<td>UNESCO World heritage sites.</td>
<td>983</td>
</tr>
<tr>
<td><img src="image" alt="Trip Advisor" /></td>
<td><a href="http://www.tripadvisor.com">www.tripadvisor.com</a></td>
<td>Travel website.</td>
<td>195957</td>
</tr>
</tbody>
</table>
3.1.2. Events

There are a lot of events that may have an educational value: exhibitions, concerts and recitals, festivals, shows and performances, fairs, conferences, workshops, webinars, etc. The information about events can be extracted from a large number of sources, such as Visit Portugal http://www.visitportugal.com, which belongs to the Portuguese Ministry of Tourism, the University of Oslo, the website of All Conferences www.allconferences.com/. Important information about events that have to be extracted by means of Web Scraping includes the venue, as well as start and end dates. This information is crucial for contextualising search results.

<table>
<thead>
<tr>
<th>Logo</th>
<th>URL</th>
<th>Description</th>
<th>Number of resources</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image1.png" alt="Spain" /></td>
<td><a href="http://www.spainisculture.com">www.spainisculture.com</a></td>
<td>Website of the Spanish ministry of tourism.</td>
<td>630</td>
</tr>
<tr>
<td><img src="image2.png" alt="Finland" /></td>
<td><a href="http://www.discoveringfinland.com">www.discoveringfinland.com</a></td>
<td>Finland travel guide.</td>
<td>91</td>
</tr>
<tr>
<td><img src="image3.png" alt="UNESCO" /></td>
<td>en.unesco.org</td>
<td>Events on the UNESCO Web site.</td>
<td>20</td>
</tr>
<tr>
<td><img src="image4.png" alt="Finnbay" /></td>
<td><a href="http://www.finnbay.com">www.finnbay.com</a></td>
<td>Mass media firm that distributes content about Finland.</td>
<td>19</td>
</tr>
<tr>
<td><img src="image5.png" alt="Europa" /></td>
<td><a href="http://www.openeducationeuropa.eu">www.openeducationeuropa.eu</a></td>
<td>Portal launched by the European Commission aiming at offering access to all existing European Open Educational Resources.</td>
<td>83</td>
</tr>
<tr>
<td><img src="image6.png" alt="Portugal" /></td>
<td><a href="http://www.visitportugal.com">www.visitportugal.com</a></td>
<td>The official website for Portugal as a tourist destination, developed by Turismo de Portugal (Ministry of Economy).</td>
<td>6</td>
</tr>
<tr>
<td><img src="image7.png" alt="Lisboa" /></td>
<td><a href="http://www.ulisboa.pt">www.ulisboa.pt</a></td>
<td>Website of the University of Lisboa.</td>
<td>304</td>
</tr>
<tr>
<td><img src="image8.png" alt="Oslo" /></td>
<td><a href="http://www.uio.no">www.uio.no</a></td>
<td>Website of the University of Oslo.</td>
<td>91</td>
</tr>
<tr>
<td><img src="image9.png" alt="Hungary" /></td>
<td>visit-hungary.com</td>
<td>Travel site about Hungary operated by the Hungarian Tourism Plc.</td>
<td>844</td>
</tr>
</tbody>
</table>
3.1.3. Lectures

Lectures delivered by experts constitute very valuable educational resources. Thus, the SDE indexes resources such as TED talks and academic lectures from videolectures.net.

<table>
<thead>
<tr>
<th>Logo</th>
<th>URL</th>
<th>Description</th>
<th>Number of resources</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="https://i.imgur.com/TED.png" alt="TED Logo" /></td>
<td><a href="http://www.ted.com">www.ted.com</a></td>
<td>TED is a “platform for ideas worth spreading”.</td>
<td>1690</td>
</tr>
<tr>
<td><img src="https://i.imgur.com/videolectures.png" alt="Videolectures.net Logo" /></td>
<td>videolectures.net</td>
<td>Open access educational video lectures repository.</td>
<td>108898</td>
</tr>
</tbody>
</table>
3.1.4. Documentaries

In the Internet we can find a great number of websites that act as directories of documentaries. Thus, the SDE extracts information from websites such as documentaryheaven.com or toptdocumentaryfilms.com.

<table>
<thead>
<tr>
<th>Logo</th>
<th>URL</th>
<th>Description</th>
<th>Number of resources</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="topdocumentaryfilms.com" alt="Logo" /></td>
<td>topdocumentaryfilms.com</td>
<td>Directory of documentary films.</td>
<td>2839</td>
</tr>
<tr>
<td><img src="www.ciberdocumentales.com" alt="Logo" /></td>
<td><a href="http://www.ciberdocumentales.com">www.ciberdocumentales.com</a></td>
<td>Web about online documentaries in Spanish.</td>
<td>1784</td>
</tr>
<tr>
<td><img src="documentariosvarios.wordpress.com" alt="Logo" /></td>
<td>documentariosvarios.wordpress.com</td>
<td>Directory of online documentaries in Portuguese.</td>
<td>339</td>
</tr>
<tr>
<td><img src="documentaryaddict.com" alt="Logo" /></td>
<td>documentaryaddict.com</td>
<td>Directory of free documentaries online.</td>
<td>4374</td>
</tr>
<tr>
<td><img src="documentaryheaven.com" alt="Logo" /></td>
<td>documentaryheaven.com</td>
<td>Website that lists online documentaries.</td>
<td>2700</td>
</tr>
<tr>
<td><img src="docunet.nl" alt="Logo" /></td>
<td>docunet.nl</td>
<td>Directory of documentaries in Dutch.</td>
<td>674</td>
</tr>
<tr>
<td><img src="www.belgesell.com" alt="Logo" /></td>
<td><a href="http://www.belgesell.com">www.belgesell.com</a></td>
<td>Online documentaries in Turkish.</td>
<td>1930</td>
</tr>
<tr>
<td><img src="www.documentarne.sk" alt="Logo" /></td>
<td><a href="http://www.documentarne.sk">www.documentarne.sk</a></td>
<td>Directory of documentaries in Slovak.</td>
<td>153</td>
</tr>
<tr>
<td><img src="www.magyarvagyok.com" alt="Logo" /></td>
<td><a href="http://www.magyarvagyok.com">www.magyarvagyok.com</a></td>
<td>Online documentaries in Hungarian.</td>
<td>1843</td>
</tr>
<tr>
<td><img src="www.lrt.lt" alt="Logo" /></td>
<td><a href="http://www.lrt.lt">www.lrt.lt</a></td>
<td>Directory of online documentaries in Lithuanian.</td>
<td>595</td>
</tr>
<tr>
<td><img src="www.nrk.no" alt="Logo" /></td>
<td><a href="http://www.nrk.no">www.nrk.no</a></td>
<td>Online documentaries in Norwegian.</td>
<td>5888</td>
</tr>
</tbody>
</table>
3.1.5. Courses

The SDE is not oblivious to the rise of MOOCs, so it indexes courses from a variety of sources, such as coursera.org and edx.org, among others.

<table>
<thead>
<tr>
<th>Logo</th>
<th>URL</th>
<th>Description</th>
<th>Number of resources</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="https://via.placeholder.com/150" alt="Coursera Logo" /></td>
<td><a href="http://www.coursera.org">www.coursera.org</a></td>
<td>Education platform that offers free courses through partnerships with universities and organisations worldwide.</td>
<td>624</td>
</tr>
<tr>
<td><img src="https://via.placeholder.com/150" alt="EdX Logo" /></td>
<td><a href="http://www.edx.org">www.edx.org</a></td>
<td>Non-profit online initiative created by founding partners Harvard and MIT. It offers online classes from the world’s best universities.</td>
<td>135</td>
</tr>
<tr>
<td><img src="https://via.placeholder.com/150" alt="MIT OCW Logo" /></td>
<td>ocw.mit.edu</td>
<td>MIT OpenCourseWare offers materials under a Creative Commons License.</td>
<td>2248</td>
</tr>
</tbody>
</table>

3.1.6. Articles

Articles in magazines devoted to scientific dissemination are other excellent type of educational resource. Sciencedaily.com, for instance, has several tens of thousands of articles of scientific dissemination.

<table>
<thead>
<tr>
<th>Logo</th>
<th>URL</th>
<th>Description</th>
<th>Number of resources</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="https://via.placeholder.com/150" alt="Sciencedaily Logo" /></td>
<td><a href="http://www.sciencedaily.com">www.sciencedaily.com</a></td>
<td>News website for topical science articles founded in 1995.</td>
<td>79259</td>
</tr>
</tbody>
</table>

3.1.7. Blog posts

In the current Internet, the blogosphere is a source of educational resources that we cannot afford to ignore. In every area of knowledge, experts have launched blog for publishing their reflections, and those blogs constitute an excellent source of scientific dissemination, more accessible to the general public than the traditional academic channels. The SDE gets quality blogs by checking the information in reputable blog directories, such as botw.org. These blog directories use to provide information about addresses of RSS feeds for every individual blog, and those feeds contain links to any of the particular posts.
3.1.8. Presentations

Presentations (keynotes or power points, for instance) are many times interesting educational resources. Thus, the SDE indexes thousands of presentations from websites such as slideshare.org.

<table>
<thead>
<tr>
<th>Logo</th>
<th>URL</th>
<th>Description</th>
<th>Number of resources</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="Logo.png" alt="slideshare logo" /></td>
<td><a href="http://www.slideshare.net">www.slideshare.net</a></td>
<td>Slideshare allows users to share presentations, infographics, documents, videos, and webinars.</td>
<td>18518</td>
</tr>
</tbody>
</table>

3.1.9. Applications

In every area of knowledge, there exist a lot of useful applications that can be used in teaching-learning processes.

<table>
<thead>
<tr>
<th>Logo</th>
<th>URL</th>
<th>Description</th>
<th>Number of resources</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="Logo.png" alt="alternativeeto logo" /></td>
<td><a href="http://www.alternativeto.net">www.alternativeto.net</a></td>
<td>Online directory of applications.</td>
<td>24057</td>
</tr>
</tbody>
</table>

3.1.10. Biographies

Biographies are educational resources that are particularly relevant for the study of historical figures.

<table>
<thead>
<tr>
<th>Logo</th>
<th>URL</th>
<th>Description</th>
<th>Number of resources</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="Logo.png" alt="bio logo" /></td>
<td>biography.com</td>
<td>Online directory of biographies.</td>
<td>7301</td>
</tr>
</tbody>
</table>
3.1.11. News

In every area of knowledge in which research is very active, news are a first-level source of knowledge. Therefore, the SDE indexes news from sources such as the science section of Reuters, the CNN, and the BBC, among others.

<table>
<thead>
<tr>
<th>Logo</th>
<th>URL</th>
<th>Description</th>
<th>Number of resources</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image1.png" alt="Reuters Logo" /> <img src="image2.png" alt="CNN Logo" /></td>
<td>Several URLs</td>
<td>Several online newspapers</td>
<td>12401</td>
</tr>
</tbody>
</table>
4. EVALUATION OF THE SDE

In order to design the SDE, we followed the problem-centred approach described by (Peffers, Tuunanen, Rothenberger, & Chatterjee, 2007), and that is known as Design Science Research Methodology (DSRM). We think that DSRM is particularly well suited to address the evaluation of an artefact such as the SDE.

Several workshops were conducted under the umbrella of the iTEC project. As DSRM proposes, every workshop allowed us to gather feedback for refining the design. The table shows data of those workshops. At the end of the iTEC project, an online survey was issued. The sample of the survey consisted on 20 teachers. The online questionnaire served us to gather data that enabled a more formal evaluation.

Table 2 Workshops where the SDE was presented to teachers.

<table>
<thead>
<tr>
<th>Place</th>
<th>Date</th>
<th>Participants</th>
</tr>
</thead>
<tbody>
<tr>
<td>Oulu</td>
<td>30/31 November 2013</td>
<td>21</td>
</tr>
<tr>
<td>Lisbon</td>
<td>17 January 2014</td>
<td>9</td>
</tr>
<tr>
<td>Budapest</td>
<td>29 January 2014</td>
<td>9</td>
</tr>
<tr>
<td>Vigo</td>
<td>27 February 2014</td>
<td>6</td>
</tr>
<tr>
<td>Bad Hofgastein (Austria)</td>
<td>25 March 2014</td>
<td>10</td>
</tr>
<tr>
<td>Vilnius (Lithuania)</td>
<td>11 June 2014</td>
<td>10</td>
</tr>
</tbody>
</table>

4.1.1. Workshop in Oulu (30/31 November 2013)

The workshop in Oulu (Finland) held on 30th and 31st of November 2013 was conducted by Markku Lang. During the two days of the workshop, the SDE was introduced to participants as a useful tool for supporting the life cycle of Future Classroom Scenarios. Particularly, the use of the SDE for supporting the Edukata process allowed us to get a valuable feedback about the use of the SDE by end-users.

WP10 in coordination with the Oulu team, made an important effort in the preparation of this workshop, in order to adapt the user interface of the SDE to Finnish. That adaptation was complex sometimes, since the SDE followed an approach quite centred on Knowledge Engineering; and, in order to create some complex specifications, it assumed syntax structure similar to that of English—requirement that was not fulfilled in Finnish.

According to the conductor of the workshop, Markku Lang, “First impression of the SDE is a simple and handy UI”. This allowed us for realising that we were on the right track regarding user interface design, as it was also demonstrated by the fact that teachers participating in the workshop were able to upload their elements created for the planning of Future Classroom Scenarios to the SDE.
This workshop raised an important matter: it seemed very convenient to go for a simplification of the definition of elements in the SDE, in such a way that the support of the life cycle of Future Classroom Scenarios did not depend so heavily on its specification at a knowledge level. From that time, we started working on a lightweight approach to definition and planning, which does not depend on Knowledge Engineering methods, being thus adequate to be put into practice during workshops.

Another important matter raised by this workshop was the convenience of enabling the exploration of the Knowledge Base by end users, thus going far beyond automatic recommendations of educational resources—the principal mechanism for interacting with the Knowledge Base at that time.

Finally, relying on feedback from this workshop, we started to work on the support of group work in the SDE. Since Edukata relies heavily on teamwork, teachers were interested in sharing their created elements with the other members of their teams.

Figure 16 Pictures from the workshop in Oulu (Finland) where teachers had the opportunity to interact with the SDE
4.1.2. Workshop in Lisbon (17th January 2014)

The workshop in Lisbon was organised by the DGE, especially by Fernando Rui Campos, and conducted by Roberto Pérez from UVIGO. This workshop was held in a morning and an afternoon session. A number of 9 teachers participated in this workshop, which was dedicated entirely to the SDE.

WP10 devoted an important effort in weeks before the workshop, especially in tasks aimed at enriching the Knowledge Base of the SDE: people, applications, and events. In order to evaluate rigorously the potential of the enrichment of events, WP10 performed a lot of work extracting information on events all across Portugal.

This workshop was very productive for WP10, as it allowed us to gather a lot of very valuable feedback. Many comments—gathered in an evaluation questionnaire—were very positive, and confirmed that we were on the right track with the lightweight approach to planning Learning Stories. “It also allows to find different resources and put the useful in a single place for a class plan”, reported a teacher.

Other comments for teachers were in the line of reporting bugs and possible improvements for the user interface, such as: hints in forms, so that users may know what is each field for; and also incorporating a confirmation dialog when users want to leave a form they are currently editing—so that they do not lose information by mistake. All those comments were very valuable for the development team in UVIGO.

At the end of the workshop, there was a very productive discussion, where many improvements and ideas for further developments were raised. Some teachers were particularly interested in developing a module for Moodle, as this Learning Management System is used extensively in Portugal, and the barrier of entry for the technologies underlying the SDE would be lower if they are implemented as a sort of plug-in in Moodle.

From the point of view of WP10, this workshop was very productive, and it started a tight cooperation work with the DGE, which led to the translation of the SDE to Portuguese, and the use of the SDE by Portuguese teachers in the iTEC Future Classroom Transformative pilots taking place in April-May 2014.
4.1.3. Workshop in Budapest (29 January 2014)

The workshop in Budapest had place in the framework of an iTEC meeting. This workshop was of a special kind, since participants were National Technical Coordinators from several of the countries participating in iTEC; besides, the time for evaluating the application was not so large as in the workshops before. This workshop allowed us to gather feedback of an extraordinary value, since NTCs were accustomed to working with teachers and presents innovative technologies to them.

In the first place, some participants did not like the fact that results depend so heavily on the learning context declared by users when registering to the application, and they would like a more free exploration—which enables them to perform searches on the Knowledge Base without the “search bubble” of personalisation and adaptation. Those comments motivated that WP10 decided to invest in the development of functionalities that enable the Semantic Search of educational resources in the Knowledge Base of the SDE.

In the second place, several participants were aware of the potential that semantic techniques that WP10 was researching offered for discovering educational resources, and they advocate the incorporation in the Knowledge Base of the SDE of resources from categories other than people, events, and applications. In particular, one participant considered that it would be interesting to discover scientific articles.

After this workshop, WP10 decided to extend the Knowledge Base of the SDE in order to index all categories of educational resources that could be useful when planning Future Classroom Scenarios, such as documentaries, articles, biographies, etc. Besides, WP10 decided to improve the Semantic Search of educational resources, furthering in its development and making it more visible in the application.
4.1.4. Workshop in Vigo (27th February 2014)

The workshop in Vigo had place the 27th of February 2014. This workshop was performed in the framework of a cooperation agreement between the University of Vigo (WP10 leader) and the Galician Ministry of Education. There were six teachers from across Galicia participating in this workshop.

In this workshop, we introduced the SDE to teachers who did not know it before, and we requested them to use it to define the Learning Story they were designing—this workshop was the fourth formative session on iTEC. All participants finished that task without difficulties.

We also requested them to use the SDE to find educational resources of interest for their Future Classroom Scenarios. That was the first time we were evaluating a full-fledged Semantic Search system with final users, and results were highly promising. Participants found easy the interaction
with the search functionalities. Some of them were positively surprised of the concept-based search and of the appropriateness of the results obtained, in comparison to the results that are obtained in other search engines, such as Google.

It is remarkable that these participants used the SDE also after this workshop, and it provided them support in all the life cycle of their Future Classroom Scenarios, which they finally carried out in their classrooms. Two of these participants also took part in the iTEC Future Classroom Transformative Pilots taking place in April-May 2014, and they used the SDE in that process too.

4.1.5. Workshop in Bad Hofgastein (25th March 2014)

The workshop in Bad Hofgastein (Austria) was conducted by Hermann Morgenbesser on 25th of March 2014, the Austrian NTC, and ten teachers from the Austrian iTEC group participated in it.
In the weeks before this workshop, WP10 kept working on the improvement of Semantic Search functionalities, both the search algorithms and the user interface for searching. In parallel, we also continued to work on the enrichment of the Knowledge Base of the SDE, incorporating new sources of information.

The feedback we gathered from teachers, via the conductor of the workshop, was very positive in many regards. Teachers found easy the interaction with the SDE, and they perceived that “it is very simple/comfortable in use”. That was a very important confirmation for WP10, as we were making a lot of effort in creating a user interface as friendly and intuitive as possible.

In regards to the Semantic Search functionalities, feedback was extraordinarily positive. Teachers manifested that “it motivates us to create/discover scientific content over this interface”. It demonstrated that the exploratory approach to search was good for enabling users to interact with the Knowledge Base of the SDE. In addition to the usability of the search, teachers manifested that “above this we are very impressed by the functionalities of the algorithm”. That feedback was very important for WP10, as it confirmed that we were on the right track implementing Semantic Search with the concept-based approach.

In summary, feedback from Austrian teachers was very positive and encouraging, as they perceived the SDE as a “very matured tool” and expressed they desire to continue using it in the future.

4.1.6. Workshop in Vilnius (11th June 2014)

The workshop in Vilnius was conducted by Virginija Bireniené on the 11th of June 2014, with teachers from the Lithuanian group. The workshop was about several iTEC tools, and the SDE was among them.

The cooperation with the NTC from Lithuania allowed WP10 to discover some minor usability issues in the SDE, which had to do this the didactic planning of Future Classroom Scenarios. Regarding the usefulness of the SDE, the number of elements they created in the SDE demonstrates that it served Lithuanian teachers to get support for the entire life cycle of Future Classroom Scenarios.

According to the NTC of Lithuania, “SDE was a big success”. The NTC also expressed that they wanted to keep using the SDE in the future, specially in the second cycle of the CCL project4.

4.2. Evaluation questionnaire

With the aim of obtaining some quantitative evaluation results of the SDE at the end of pilot experiences, WP10 elaborated a comprehensive evaluation questionnaire that was distributed among NTCs, so that they, in turn, may distribute the questionnaire among teachers participating in the project. Regrettably, the sample obtained (20 teachers) is not large enough for having a solid statistical value; notwithstanding, it does offer a good indication of how teachers perceive the SDE.

http://creative.eun.org/
The evaluation questionnaire was divided into several blocks, in accordance with the topics we wanted to validate. The first block consisted of questions about the collections of educational resources indexed in the SDE. This block aimed at capturing the perception of teachers on the appropriateness and completeness of resources in each category.

The second block of questions was focused on evaluating the information retrieval of educational resources. To that end, the questionnaire requested teachers to identify a topic of study, and then to use Google and the SDE to try to discover educational resources on that topic. After completing that process, teachers were asked to compare Google to the SDE in accordance with their performance in:

- **Variety.** A question tried to evaluate which system returns a bigger variety of educational resources, that is, educational resources of many types.
- **Recall.** Another question focused on elucidating which system has a bigger recall—which one return a bigger number of educational resources for the topic of choice.
- **Precision.** This one was an important point in the evaluation. Precision, in this context, was understood as the capability of returning educational resources and not other types documents.

In order to evaluate the concept-based search we designed the following process. We asked teachers to think on a search term that could have different meanings—a polysemic one. Then, we requested them to look for it in Google, using its word-sense disambiguation features. After that, we requested teachers to search for that term in the SDE, using its concept suggestion features. At the end of that process, teachers had to judge which system has a better disambiguation mechanism.

The next block of questions had as the objective of evaluating the functionalities of exploratory search. To that end, we asked teachers to evaluate it in its three modalities: by the space of concepts, by the space of tags, and by the space of related resources. We also wanted to evaluate the algorithm for calculating semantic similarity between educational resources, and thus one question was devoted to that. Finally, the last two questions were a bit more generic, as they asked teachers whether they discovered any interesting educational resource during their interaction with the Knowledge Base of the SDE or not; and we also requested teachers to evaluate the search tool as a whole.

Recommendations were evaluated in the next block of the questionnaire. To that end, we devoted a question to each one of the categories from which educational resources were recommended. Besides, we devoted a question for capturing the perception of teachers on the usefulness of recommendations by asking them whether they had discovered any interesting educational resource thanks to recommendations.

The next block was aimed at evaluating the didactic planning features for Future Classroom Scenarios integrated in the SDE. To this effect, this block requested teachers to evaluate the tools for creating Future Classroom Scenarios and rest of elements for planning. We also tried to capture the perception of teachers on the usefulness of receiving suggestions of alternative resources, in order to adapt Future Classroom Scenarios created by other teachers to their particular learning context. The last question of this block was aimed at evaluating the usefulness of bookmarking educational resources.
Finally, the last block consisted of more generic questions, such as: the evaluation of the usability of the SDE as a whole; whether they would recommend it to other teachers or not; and the general evaluation of the SDE.

4.2.1. Evaluation of the collections of resources in the Knowledge Base.

The evaluation of the appropriateness and completeness of the indexed collections was evaluated through the following question: *Do you think that the following collections of resources indexed in the SDE are appropriate and complete enough?*

The following figures show the results of that question.

![Figure 20](image.png)

**Figure 20** Appropriateness and completeness of the collection of applications indexed in the SDE

![Figure 21](image.png)

**Figure 21** Appropriateness and completeness of the collection of events indexed in the SDE

![Figure 22](image.png)

**Figure 22** Appropriateness and completeness of the collection of biographies indexed in the SDE

![Figure 23](image.png)

**Figure 23** Appropriateness and completeness of the collection of lectures indexed in the SDE
Figure 24 Appropriateness and completeness of the collection of sites indexed in the SDE

Figure 25 Appropriateness and completeness of the collection of documentaries indexed in the SDE

Figure 26 Appropriateness and completeness of the collection of courses indexed in the SDE

Figure 27 Appropriateness and completeness of the collection of articles indexed in the SDE

Figure 28 Appropriateness and completeness of the collection of OERs indexed in the SDE
4.2.2. Evaluation of the information retrieval

Firstly, you have to choose a topic that you feel like learning about it or teaching it. For instance: Napoleon, Thermodynamics, DNA, etc.

Now you have to try to gather educational resources about that particular topic. We propose to you to search for it in Google and try to gather (or at least to identify) educational resources about your topic of choice. Then, we have to look for educational resources about the topic in the SDE. You should not spend more than 2 minutes in each platform.
4.2.3. Evaluation of word-sense disambiguation

Now we propose that you choose a topic that might have different senses. Now search for your topic of choice in Google. Now look for the chosen topic in SDE.

Figure 36 Which platform do you think that disambiguates better?
4.2.4. Evaluation of exploratory search

Figure 37 How do you evaluate the navigation through the space of topics?

Figure 38 How do you evaluate the exploratory search through the keywords of particular resources

Figure 39 How do you evaluate the exploratory search through related resources?

Figure 40 Are “related resources” really related?

Figure 41 Did you discover new and interesting resources in the process?
4.2.5. Evaluation of recommendations

Figure 42 Could you give a general evaluation of the search tool?

Figure 43 How do you evaluate recommendations of sites?

Figure 44 How do you evaluate recommendations of events?

Figure 45 How do you evaluate recommendations of news?
Figure 46 How do you evaluate recommendations of other resources?

Figure 47 Did you discover any new and interesting educational resource thanks to recommendations?

4.2.6. Evaluation of editors

Figure 48 Evaluate the usability of the tool for creating Scenarios and Learning Activities.

Figure 49 Evaluate the usability of the tool for creating new Learning Stories.
Figure 50 Evaluate the “alternatives” for resources embedded in a Learning Story.

Figure 51 Are bookmarks useful?

4.2.7. General evaluation

Figure 52 Was easy for you to interact with the SDE?

Figure 53 Would you recommend it to other teachers?

Figure 54 Evaluate the SDE.
4.3. Discussion

The results of the evaluation of the SDE are, in general, very positive. The Web application is perceived by end-users as intuitive and easy-to-use; and the Semantic Search techniques usually surprise end-users because of their precision and appropriateness for retrieving information on educational resources.

An important advantage of the bag-of-concepts paradigm is that it is easily communicable to end-users, who have no problem understanding how it works. For instance, it is easy to explain how the SDE retrieves resources related to a particular one: getting those that are about the same concepts. In a similar way, end-users find easy to understand how the SDE retrieves resources in accordance with their relevance for a given concept. Overall, the concept-based paradigm is perceived as possessing a great potential.

In general, the collections of resources indexed in the SDE are perceived as complete enough for exploiting the capabilities of the SDE. The biggest problem we faced here was finding sources of educational resources in languages other than English. In this regard, evaluation results evidence that it is necessary to keep working on incorporating more sources in other languages.

One of the reflections that we make after evaluating the SDE with end-users is that they are very interested in having at disposal a platform that indexes documents from the Internet that may have an educational value. Resources such as, for instance, lectures from videolectures.net have a great educational value; and it is convenient to have at disposal a central hub, which enables to access all those resources that are scattered all across the Web.

The evaluation of the information retrieval strategies clearly evidences that the way in which the SDE displays search results—that is, classified in accordance with the type of educational resource—is more convenient that displaying information in accordance with their file format. Therefore, it is positive to choose categories because they represent types of educational resources and not formats for storing information—audio, video, etc.

Another aspect that is worth highlighting is that the point where Google is clearly perceived as outperforming the SDE is recall of educational resources. While it is true that recall is better in Google—because it indexes most part of the Internet—, precision is better in the SDE; that is to say, in each search the SDE retrieves a bigger percentage of educational resources in comparison to searches in Google.

Word-sense disambiguation was subjected to a quite exhaustive evaluation. To that end, we designed a series of steps that end-users had to follow:

1. Choosing a search term that suffers from polysemy.
2. Choosing the intended sense from a list of available options.
3. Checking whether there are false positives or false negatives in the results.

We requested to repeat those steps performing searches in Google and in the SDE. As we can see, end-users perceive the word-sense disambiguation capability in the SDE as outperforming that which Google implements.

Regarding exploratory search, it is one of the favourite features of the SDE as evaluation results evidence: it was evaluated very positively by end-users. Exploratory search in all its forms—by the
space of concepts, tags, and related resources—is perceived as a very convenient way of interacting with the Knowledge Base of the SDE.

The evaluation of recommendations shows that, as in the case of collections of resources, it is necessary to incorporate more resources in languages other than English so that, in this way, the SDE may have a wider range of options for suggesting resources to users in other languages.

In regard to the editors, end-users find easy to interact with the editors, and they consider the application to be intuitive and easy-to-use. Besides, they found useful the suggestion of alternative resources, which is perceived as an efficient way of obtaining resources for adapting existing Future Classroom Scenarios to the needs of particular contexts.
5. REFERENCES


CONCLUSIONS

This document reported the actions carried out by WP10 in the fourth year of the iTEC project. This section presents some considerations a posteriori on the research lines followed and elaborates some conclusions that may be useful for future research in the field.

In the first place, it is worth highlighting the great potential of the techniques for enriching the Knowledge Base of the SDE that WP10 has developed for the past few years. Information extraction from external sources reveals itself as a very promising approach to augmenting the knowledge at the disposal of the SDE; potentially, all the knowledge stored in the Web could be accessed and used by a knowledge-based system. At the time of writing, the SDE indexes 204932 places of interest, 5932 events, 110588 lectures, 23452 documentaries, 3007 courses, 79259 articles, 59521 blog posts, 18518 presentations, 24057 applications, 7301 biographies, and 12401 news items.

In the second place, it is remarkable the great potential of the techniques for Semantic Search that were developed during the fourth year of the iTEC project, and that enable a exploratory search, well suited for discovering educational resources that deal with particular concepts. The concept-based approach completely overcomes the problems of synonymy and polysemy, associated with traditional keyword-based searches. As it was shown in Chapter 2, these techniques were formulated in a rigorous manner. Besides, a software architecture that implements Semantic Search was described—that software architecture lays at the foundation of the information retrieval system that enables users to explore the Knowledge Base of the SDE. It is worth noting that the formal—mathematic—foundations of Semantic Search could serve as the basis for the implementation of information retrieval systems on top of existing repositories of educational resources, which could benefit enormously from all the possibilities that Semantic Search offers. Particularly, the Learning Resource Exchange is a very good candidate for the implementation of the techniques described in Chapter 2 on top of its huge collection of educational resources. In fact, WP10 developed a proof-of-concept of that idea by integrating educational resources from the LRE into the SDE; the success of that proof led the LRE subcommittee to unanimously approve the elaboration of a project proposal in the line of developing Semantic Search capabilities on top of the LRE—to be submitted to next year’s call for proposals of the Horizon 2020 programme. Besides, the positive results of the evaluation of Semantic Search with end-users reported in Chapter 4 make of Semantic Search a very promising research line.

In the third place, it is important to manifest the value of the SDE—understanding it as the backend services plus the user interface developed during the fourth year of iTEC—as a tool for supporting the entire life-cycle of Future Classroom Scenarios, from their initial definition to their final didactic planning. Here too, evaluation results are very positive, since end-users perceive the SDE as an intuitive and easy-to-use Web application. In spite of having room for improvement, the SDE is nevertheless exploitable in its current form; as it is demonstrated by a cooperation agreement that the University of Vigo signed with the Galician Ministry of Education, and that includes the exploitation of the SDE for providing support to teachers when defining and planning Future Classroom Scenarios; and it is also demonstrated by the interest shown by the Lithuanian NTC, who wants to keeps using the SDE in the framework of the CCL project, and by the Austrian NTC, in the context of Scientix.
# 6. Appendix I: LIST OF ABBREVIATIONS

<table>
<thead>
<tr>
<th>ACRONYM</th>
<th>MEANING</th>
</tr>
</thead>
<tbody>
<tr>
<td>API</td>
<td>Application Programming Interface</td>
</tr>
<tr>
<td>BoC</td>
<td>Bag of Concepts</td>
</tr>
<tr>
<td>DoW</td>
<td>Description of Work</td>
</tr>
<tr>
<td>DSRM</td>
<td>Design Science Research Methodology</td>
</tr>
<tr>
<td>HTML</td>
<td>Hypertext Mark-up Language</td>
</tr>
<tr>
<td>HTTP</td>
<td>Hypertext Transfer Protocol</td>
</tr>
<tr>
<td>iTEC</td>
<td>Innovative Technologies for an Engaging Classroom</td>
</tr>
<tr>
<td>JSON</td>
<td>JavaScript Object Notation</td>
</tr>
<tr>
<td>MOOC</td>
<td>Massive Online Course</td>
</tr>
<tr>
<td>NTC</td>
<td>National Technical Coordinator</td>
</tr>
<tr>
<td>RSS</td>
<td>Really Simple Syndication</td>
</tr>
<tr>
<td>SDE</td>
<td>Scenario Development Environment</td>
</tr>
<tr>
<td>WP</td>
<td>Working Package</td>
</tr>
</tbody>
</table>
7. Appendix II: CONTROL BOARDS

During the fourth year of the iTEC project we carried out the seventh and last Control Board iteration. The Control Board meeting was held on the 10th of July 2014 via Flashmeeting. This Control Board had a twofold purpose. In the one hand, it was aimed at gathering feedback for the evaluation of the SDE, understanding it as the integrated application that provides editors for supporting the entire life cycle of Future Classroom Scenarios—from their initial definition to their final didactic planning—plus the Semantic Search capabilities. On the other hand, this Control Board was also aimed at attempting to elucidate possible directions for future research on the SDE and, in general, on the approaches and techniques that underlay the SDE, especially those that have to do with Semantic Search.

Due to the especial character of this last Control Board iteration, we focused the participation on NTCs, since they have field experience on the use of the SDE by teachers—for defining and planning Future Classroom Scenarios, as well as for discovering educational resources that may be used in those Scenarios. Participants were Ildikó Csordás, NTC of Hungary, and Hermann Morgenbesser, NTC of Austria.

7.1. Document preparation

Given the particular purpose of this Control Board, participants did not need any additional documentation, because what really counted was to have experience using the SDE with teachers. Thus, with a detailed agenda was enough documental support for the preparation of the Control Board.

7.2. Results

Regarding the part of the Control Board that had to do with the evaluation of the SDE, results are in line with the analysis reported in Chapter 4. In general, teachers consider the SDE to be an intuitive and easy-to-use application. Besides, the resource discovery functionalities surprises teachers in a positive way, and makes pleasant the task of gathering resources for incorporating them to didactic plans.

In regards to the part of the Control Board that dealt with the identification of possible research lines for further improving the SDE, the principal line identified was the improvement of the SDE in languages other than English.