D3.3.6 Report on the current state of ASPL

Coordinator: **Martin Dzbor (OU)**

Other authors: **Dnyanesh Rajpathak (OU)**
Sylvain Dehors (INRIA)
Joerg Diederich (L3S)

Abstract.
EU-IST Network of Excellence (NoE) IST-2004-507482 KWEB
Deliverable D3.3.6 (WP3.3)

The intention of this deliverable is to describe the implications from the last year's evaluation study of ASPL-v1 and how we addressed the identified issues. We briefly summarize the ASPL functionality and the outcomes of the evaluation, and proceed with analyzing the learning task we set for the ASPL in the previous deliverables. In particular, we use the analysis to highlight how the identified gaps and shortcomings could be addressed so that the benefits and added value of the Semantic Web approach in designing the ASPL services becomes clearer and more explicit.

Document Identifier:	KWEB/2006/D3.3.6/v1.0
Class Deliverable:	KWEB EU-IST-2004-507482
Version:	v1.0
Date:	January 31, 2007
State:	Final
Distribution:	Public

Knowledge Web Consortium

This document is part of a research project funded by the IST Programme of the Commission of the European Communities as project number IST-2004-507482 with the following partners:

University of Innsbruck (UIBK) – Coordinator
Institute of Computer Science,
Technikerstrasse 13
A-6020 Innsbruck
Austria
Contact person: Dieter Fensel
E-mail address: dieter.fensel@uibk.ac.at

France Telecom (FT)
4 Rue du Clos Courtel
35512 Cesson Sévigné
France. PO Box 91226
Contact person: Alain Leger
E-mail address:
alain.leger@rd.francetelecom.com

Free University of Bozen-Bolzano (FUB)
Piazza Domenicani 3
39100 Bolzano
Italy
Contact person: Enrico Franconi
E-mail address: franconi@inf.unibz.it

Centre for Research and Technology Hellas / Informatics and Telematics Institute (ITI-CERTH)
1st km Thermi – Panorama road
57001 Thermi-Thessaloniki
Greece. Po Box 361
Contact person: Michael G. Strintzis
E-mail address: strintzi@iti.gr

National University of Ireland Galway (NUIG)
National University of Ireland
Science and Technology Building
University Road
Galway
Ireland
Contact person: Christoph Bussler
E-mail address: chris.bussler@deri.ie

Universidad Politécnica de Madrid (UPM)
Campus de Montegancedo sn
28660 Boadilla del Monte
Spain
Contact person: Asunción Gómez Pérez
E-mail address: asun@fi.upm.es

École Polythechnique Fédérale de Lausanne (EPFL)
Computer Science Department
Swiss Federal Institute of Technology
IN (Ecublens), CH-1015 Lausanne.
Switzerland
Contact person: Boi Faltings
E-mail address: boi.faltings@epfl.ch

Freie Universität Berlin (FU Berlin)
Takustrasse, 9
14195 Berlin
Germany
Contact person: Robert Tolksdorf
E-mail address: tolk@inf.fu-berlin.de

Institut National de Recherche en Informatique et en Automatique (INRIA)
ZIRST – 655 avenue de l'Europe – Montbonnot Saint Martin
38334 Saint-Ismier
France
Contact person: Jérôme Euzenat
E-mail address: Jerome.Euzenat@inrialpes.fr

Learning Lab Lower Saxony (L3S)
Expo Plaza 1
30539 Hannover
Germany
Contact person: Wolfgang Nejdl
E-mail address: nejdl@learninglab.de

The Open University (OU)
Knowledge Media Institute
The Open University
Milton Keynes, MK7 6AA
United Kingdom.
Contact person: Enrico Motta
E-mail address: e.motta@open.ac.uk

University of Karlsruhe (UKARL)
Institut für Angewandte Informatik und Formale Beschreibungsverfahren – AIFB
Universität Karlsruhe
D-76128 Karlsruhe
Germany
Contact person: Rudi Studer
E-mail address: studer@aifb.uni-karlsruhe.de

University of Liverpool (UniLiv)
Chadwick Building, Peach Street
Liverpool, L69 7ZF
United Kingdom
Contact person: Michael Wooldridge
E-mail address: M.J.Wooldridge@csc.liv.ac.uk

University of Manchester (UoM)
Room 2.32. Kilburn Building, Department of Computer Science, University of Manchester, Oxford Road
Manchester, M13 9PL
United Kingdom
Contact person: Carole Goble
E-mail address: carole@cs.man.ac.uk

University of Sheffield (USFD)
Regent Court, 211 Portobello street
Sheffield, S1 4DP
United Kingdom
Contact person: Hamish Cunningham
E-mail address: hamish@dcs.shef.ac.uk

University of Trento (UniTn)
Via Sommarive 14
38050 Trento
Italy
Contact person: Fausto Giunchiglia
E-mail address: fausto@dit.unitn.it

Vrije Universiteit Amsterdam (VUA)
De Boelelaan 1081a
1081HV. Amsterdam
The Netherlands
Contact person: Frank van Harmelen
E-mail address: Frank.van.Harmelen@cs.vu.nl

Vrije Universiteit Brussel (VUB)
Pleinlaan 2, Building G10
1050 Brussels
Belgium
Contact person: Robert Meersman
E-mail address: robert.meersman@vub.ac.be

Work package participants

The following partners have taken an active part in the work leading to the elaboration of this document, even if they might not have directly contributed writing parts of this document:

OU
L3S
INRIA

We gratefully acknowledge their assistance contribution to the discussion regarding ASPL scenarios, functionality and tool usability.

We also acknowledge the effort of Marco Ronchetti from University of Trento, who kindly assisted us with reviewing an earlier version of this deliverable.

Changes

Version	Date	Change Author	Changes
0.1	01-12-2006	Martin Dzbor, Dnyanesh Rajpathak	First draft of the report
0.2	11-12-2006	Martin Dzbor, Sylvain Dehors	Update on QBLS & ontology acquisition meth.
0.3	12-12-2006	Martin Dzbor, Dnyanesh Rajpathak	Notes on future evaluation
0.4	18-12-2006	Martin Dzbor	Implementation, exec. Summary, abstract
0.5	23-12-2006	Joerg Diederich	Update on FacetedDBLP, GrowBag, etc.
0.6	08-01-2007	Martin Dzbor	Cleanup, styles – sent for internal review
0.7	17-01-2007	Marco Ronchetti	Internal review, QA
0.8	19-01-2007	Dnyanesh Rajpathak	Clarification of evaluation plans section
1.0	02-02-2007	Martin Dzbor	QA comments, corrections, submission

Executive Summary

The intention of this deliverable is to describe the implications from the last year's evaluation study of ASPL-v1 and how we addressed the identified issues. We briefly summarize the ASPL functionality and the outcomes of the evaluation, and proceed with analyzing the learning task we set for the ASPL in the previous deliverables. In particular, we use the analysis to highlight how the identified gaps and shortcomings could be addressed so that the benefits and added value of the Semantic Web approach in designing the ASPL services becomes clearer and more explicit.

In particular we discuss and clarify the position with respect to the amendments made to the services supporting query refinement/expansion on the retrieved data (see section about "*Improved sensemaking support…*") Another amendment where we believe benefits will be more visible is described in section "*Exploratory navigation through…*", where we illustrate how automatically harvested correlations relations could be used to navigate the space of query results.

Additionally, we present our new service about question-based learning that was implemented since the evaluation of the old version of ASPL. This service is now fully integrated in the framework. In a similar role, we discuss other recommendation and data processing services, which are candidates for the inclusion into ASPL in the next period.

To complement the implementation work on ASPL services, the document briefly addresses and discusses a problem-centric approach to constructing ontologies for learning and educational support, and finally, we also highlight some initial steps towards formulating a more principled method for evaluating the current version of ASPL for its effectiveness (rather than its rate of recall).

Contents

1 Overview of ASPL and modifications

The goal of this work package is to provide a delivery platform for the educational content that is (a) stored in REASE[1], i.e. a portal repository where learning resources can be uploaded and annotated by their authors, and (b) available widely on the Web, e.g. in the form of scientific publications, communities of practice, etc.

In deliverable D3.3.3 we reported on the first version of that delivery platform, which was referred to as an advanced semantic platform for learning (ASPL). The first phase of the platform development concluded in 2005 by evaluating the application built on top of the platform. The purpose of the evaluation was *formative*; i.e. we intended to identify the gaps in the current platform, which would help us to focus on and elaborate specific strengths of our approach. In the remainder of the lifetime of the NoE we concentrate on augmenting this proof of concept ASPL-v1.

In line with this, the intention of this deliverable is to describe the augmented prototype of the advanced semantic platform for learning (ASPL-v2). Rather than merely describing the new version, we also briefly mention the rationale for re-engineering the application in particular ways and directions. While we expect future versions of ASPL to have increased functionality, ASPL is drawing upon the Magpie infrastructure — a prototype framework for semantic browsing and for rapidly developing applications involving semantic web browsing, which has been developed at the Knowledge Media Institute at the Open University. In the report we follow up on a brief overview of the Magpie (which has been published in more details elsewhere) and describe the changes to the architecture of ASPL-v1 before looking at several separate aspects of the platform-level work, which complements the actual demo application.

In particular, we summarize our work on improving ontologies and their acquisition for the purposes of being used as a part of an educational Semantic Web application. Furthermore, we sketch how the new version of the application might be assessed and evaluated to consider its effectiveness or its fit for purpose. We conclude with future plans for additional functionality and for the evaluation of ASPL.

1.1 Generic ASPL in a nutshell

ASPL intends to support the user in interpreting texts related to Semantic Web Studies. This version of ASPL includes the Magpie semantic browser framework, which was chosen in order to manage the costs of developing ASPL and balancing efficiency of the application development with an effective balance between research and implementation work. Magpie has been designed at OU to serve as a generic platform on which more sophisticated and specialized infrastructures and applications can be built.

ASPL a Magpie application is available as a plug-in for a number of browsers. It operates by making use of domain ontologies to dynamically annotate texts. Users can make use of the web services which have been associated with classes in the domain

---

[1] REASE is one of the outcomes of the project's educational area, it stands for Repository of the European Association for the Semantic Web Education, and is available at http://rease.semanticweb.org

ontology to access a range of relevant resources and activities. ASPL interacts with the user using the highlighting of entities and concepts in web pages. These lexical keywords are derived and serialized from a domain ontology.

1.2 Modifications to ASPL (version 1 → version 2)

ASPL has been evaluated by a group of users last year, and as a result of the evaluation we started investigating several means to address the identified shortcomings of the first version of the system. In this reporting period we have worked on these particular aspects and extensions of the ASPL:

Improving navigational strategies: In the first prototype we were restricted to a few, search-centric services (e.g. CiteSeer, REASE and ACM Portal). While the content of these services is useful to the purpose of ASPL, the services were considered too rigid and restrictive. Hence, we extended them so that the user can now browse in the found results (e.g. publications or authors) and explore different aspects of the problem space. Further examples of this improvement are in section "*Services supporting exploration and situated learning*".

Transitions between multiple viewpoints and sub-parts of the problem space: This activity reflects another aspect where version 1 lacked sufficient flexibility. In learning tasks it is common to explore problem spaces from different perspectives, but this basic strategy has not been implemented in ASPL-v1. Hence, algorithms for mining statistic correlations (and hence, potential semantic relationships) were improved and more tightly integrated into the ASPL application. This has led to services enabling the user to preview automatically harvested relations between people, research topics and technologies. More on this is provided in section "*Services supporting exploration and situated learning*".

More commitment to pedagogic aspects: ASPL-v1 had only limited implementation of standard pedagogic techniques. As mentioned above, it was centred on search and item retrieval/recall, rather than any synthesis or analysis. To remedy this, we elaborated one pedagogic strategy – question-based learning, and incorporated it more tightly into the existing ASPL framework. The updated QBLS service now extends and complements previously static definitions and explanations of key terms. More details on this work can be found in section "*Pedagogic services*".

2 Implications from evaluating ASPL in a learning task

In this section we describe the rationale for ASPL evolution and illustrate how we used the results from a formal evaluation of version 1 to re-design the set of functionalities provided to users. The second version of ASPL introduces more opportunities to interpret results provided by a non-semantic web mining tool. These, in turn, are used to introduce additional means for semantics-assisted exploration to the portfolio of strategies; incl. performing lateral steps and deepening the original query. We start more broadly with setting and explaining a generic issue of supporting learning with semantic web techniques, and identify desirable extensions to ASPL to address the issue. Next sections then offer more specific information on functionality and implementation of the identified gap-plugs.

2.1 Semantic web view on education

Education, like many other disciplines, aims to take advantage of Web technology to provide learning resources speedily and easily, and to tailor them to the specific needs of a learner. However, education has always relied on a strong interpretative component. In addition to recalling knowledge from knowledge bases, searching document repositories or retrieving from information warehouses, education requires also analysis and synthesis – both on the level of individual learners and at group level. Interpretation, in general, it comprises the ability to link otherwise independent information sources, to make statements about these sources, and to make inferences from the available knowledge. Above all, education is a highly social, interactive activity, which centers on the learners and expects them to actively participate in their education.

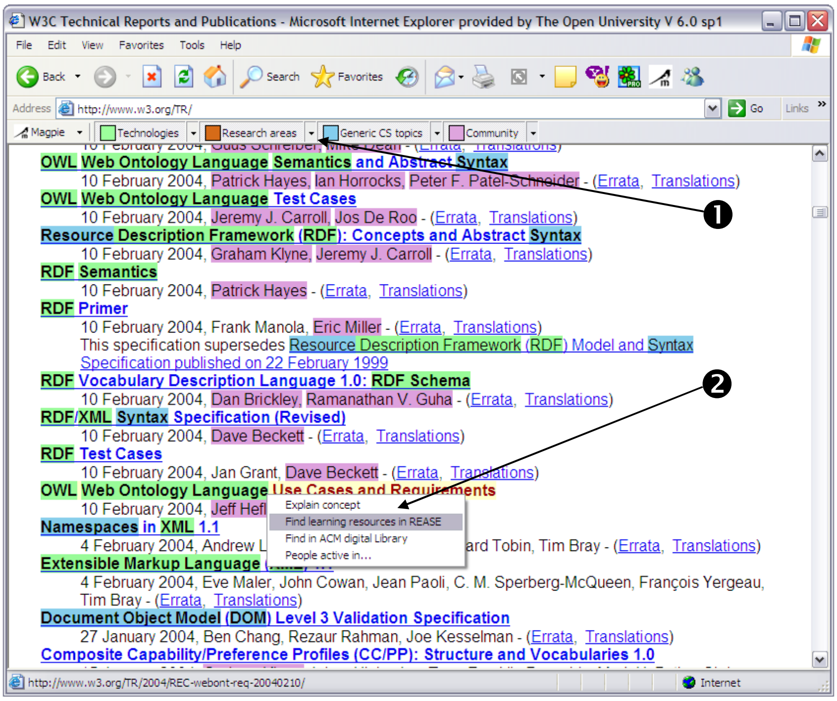
The size of databases and other repositories of resources that are suitable for learning is no longer the greatest obstacle in harnessing the Web in educational practice. There are efficient storage and data access technologies; rich information is no longer constrained by the bandwidth. The major obstacle, however, has changed remarkably little over the decades of learning systems evolution. Already in the 1940s Bush (Bush 1945) pointed out that there was already more information published than it was possible for humans to process. This matters because in learning the pieces need not only to be retrieved, but more importantly, *related* one to another and applied in new situations.

Similar remarks were made 20 years after Vannevar Bush by educational psychologists. In his report, Bloom (Bloom 1965) recognized that cognitive processing of information goes well beyond its recall and pointed out that more advanced cognitive processes, such as synthesis, analysis and judgment lead to longer-lasting knowledge. Crucially, these processes share one feature: they are *relational,* i.e., they consist of creating associations between separate pieces of information.

Although humans rely on associations to make sense of information, the challenge of supporting this associative thinking is far from resolved. Much of the current Web, and in many cases also applications of semantic web technologies, subscribes to what can be labeled as the “Web is for searching” paradigm. This has proven its worth, but there are other, pedagogically more interesting strategies. For example, in (Lieberman, Fry et al. 2001) the authors present Letitia – a tool enabling Web users to explore what lies several ‘hops’ beyond a particular web page, thus “blending browsing and searching”.

The value of exploration was also emphasized in (Eisenstadt, Price et al. 1983), where the authors point out that since (i) it is hard to formally capture all subtleties of a learning task in a tutoring system, and (ii) learner modeling is always only approximate, tutoring systems tend to be over-constrained closed worlds. In contrast with this approach, the authors in (Eisenstadt, Price et al. 1983) argue that it is more valuable for the learner to *see what can be done with a given chunk of knowledge*, rather than merely following a prescribed workflow for a learning task.

We summarize some positions arising from this research in (Dzbor, Stutt et al. 2007). In this position paper we outline a view that extends the use of semantic technologies as a means of providing learning services that are owned and created by learning communities. This, to some extent, contrasts with the work on applying these technologies to learning, which concentrated on providing novel means of accessing and making use of learning objects. We argue that this is too narrow a view: semantic technologies will make it possible to develop a range of educational Semantic Web services, such as interpretation, structure-visualization, support for argumentation, content customization, exploration, and so on.



OWL Web Ontology Language Semantics and Abstract Syntax
10 February 2004, Patrick Hayes, Ian Horrocks, Peter F. Patel-Schneider - (Errata, Translations)
OWL Web Ontology Language Test Cases
10 February 2004, Jeremy J. Carroll, Jos De Roo - (Errata, Translations)
Resource Description Framework (RDF): Concepts and Abstract Syntax
10 February 2004, Graham Klyne, Jeremy J. Carroll - (Errata, Translations)
RDF Semantics
10 February 2004, Patrick Hayes - (Errata, Translations)
RDF Primer
10 February 2004, Frank Manola, Eric Miller - (Errata, Translations)
This specification supersedes Resource Description Framework (RDF) Model and Syntax Specification published on 22 February 1999
RDF Vocabulary Description Language 1.0: RDF Schema
10 February 2004, Dan Brickley, Ramanathan V. Guha - (Errata, Translations)
RDF/XML Syntax Specification (Revised)
10 February 2004, Dave Beckett - (Errata, Translations)
RDF Test Cases
10 February 2004, Jan Grant, Dave Beckett - (Errata, Translations)
OWL Web Ontology Language Use Cases and Requirements
10 February 2004, Jeff Hefl
Namespaces in XML 1.1
4 February 2004, Andrew L ard Tobin, Tim Bray - (Errata, Translations)
Extensible Markup Language
4 February 2004, Eve Maler, John Cowan, Jean Paoli, C. M. Sperberg-McQueen, François Yergeau, Tim Bray - (Errata, Translations)
Document Object Model (DOM) Level 3 Validation Specification
27 January 2004, Ben Chang, Rezaur Rahman, Joe Kesselman - (Errata, Translations)
Composite Capability/Preference Profiles (CC/PP): Structure and Vocabularies 1.0

Fig. 1. A screenshot showing a Magpie-enhanced web browser and a web page annotated using the lexicon derived for the Semantic Web domain; pointer ❶ shows a user-selected ontology with several abstract categories of identifiable concepts (highlighted in different colours), and pointer ❷ shows a sample menu with semantic services associated with a particular category of concepts.

2.2 Setting the stage: ASPL-v1

A snapshot of one web page annotated using the ASPL-based application developed for the domain of Semantic Web as a scientific discipline is shown in Fig. 1. The ASPL experimental system was bootstrapped with automatically populated ontologies about the Semantic Web research areas and the Semantic Web community. The former ontology includes, among other things, various technical and scientific terms a student may encounter when learning about the Semantic Web. The latter comprises the scientists and people whose work is relevant to the Semantic Web.

Some categories in ASPL, such as 'Community' and 'Research areas' (violet and orange labels in Fig. 1) were automatically populated using CORDER – a tool for automated mining and capturing of entities from text corpora (Zhu, Goncalves et al. 2005). CORDER relies on a Named Entity Recognizer (Hirschman and Chinchor 1997) to identify typed entities in web pages and, by performing large-scale crawling on the web, it is able to provide accurate information about the statistical co-occurrences of these entities, using a sophisticated algorithm that takes into account several factors beyond mere co-occurrence; e.g., the distance between items in a web page. As a result, it is able to derive automatically very accurate information about, for example, who are the key people involved in a particular research area.

The ontological categories in ASPL-v1 were linked to a small set of semantic services aimed at supporting the task of preparing material for writing a critical literature review on Semantic Web topics. The services in ASPL-v1 included standard bibliographic information search in CiteSeer, aggregation of publications from ACM digital libraries, resource retrieval from REASE repository of learning materials, and term explanation. The services were more *search-* and *retrieval-focused* than inferential. This initial constraint has been opted for to prevent confounding the study with our views on how participants should carry out the task. In other words, in ASPL-v1 publications were found and displayed, with no further guidance on what to do next.

2.3 Implications of evaluating ASPL-v1

ASPL-v1 has been evaluated last year and the outcomes were reported in the previous period, therefore, we are not drawing on those outcomes in order to summarize the re-engineering requirements on the ASPL prototype. Going straight to the implications, we note that the lack of improved performance has been attributed to the quality of the semantic services implemented in version 1. As could be expected, skilled users of Web search engines did not find much value in annotated web pages for the sole purpose of retrieving results similar to those provided by a search engine.

The outlier 'Site 3' comprised students with practical experience in writing literature reviews. They seemed to have relied more frequently on the semantic annotations and retrievals through Magpie services. The annotations seem to have helped them to choose a small number of key references among retrievals. The outlier group spent more time on navigating through the resources in order to ascertain the quality, and they used a wider range of navigation strategies (e.g., via authors, via topics, via related topics, or through a repository of learning resources).

These observations can be qualitatively generalized – our participants went beyond mere information retrieval, whether semantic or non-semantic. They wanted to ensure the one at hand was the right resource – e.g. by reading its abstract, citations or co-authors. Thus they showed some preference for exploratory navigation, which was however, not available in the ASPL-v1 system. For example, most services ended with a flat list of records, with little opportunity to explore these results further (other than re-formulating the query in a search engine).

People also wanted to obtain help with formulating appropriate queries to give them 'partial findings' and process these at a later stage. Interestingly, very few people followed a good practice of literature reviews – considering cross-domain semantic relations. Normally, these are helpful if retrieval in (say) technologies yield too few or too many items. In such cases another category (e.g. authors) may give an alternative view on the domain and improve precision.

To summarize, people who managed to replicate aspects of the exploratory navigation paradigm seem to have coped with the task better. The challenge for our re-design of the ASPL system was thus to facilitate more of such exploration and embed it into standard services. In the next section we elaborate the theoretical principles of this idea.

3 ASPL re-engineering and updated services

Following the summary of the ASPL-v1 performance and identification of broad requirements on the new version, we return to the task we set for it and use it to elaborate the re-engineering focus. The task is essentially about an argument comprising different publications and previously we conceptualized it into sub-tasks it involves. One such conceptualization is shown in Fig. 2, where tasks are also indexed in terms of Bloom's learning activities (Bloom 1965), such as recall, comprehension, etc.

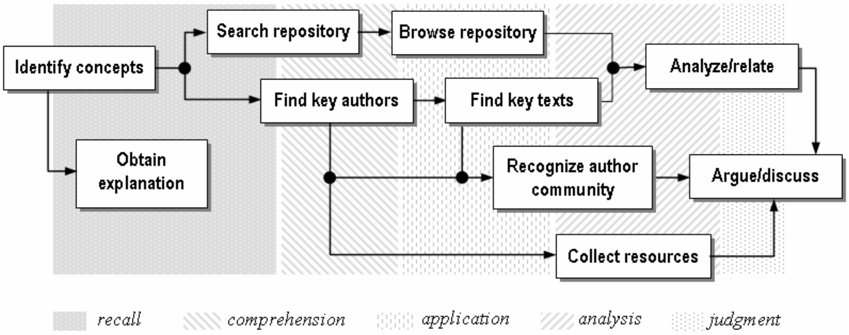


Fig. 2. Conceptual decomposition of the learning task (preparing literature review)

3.1 Conceptual analysis of the learning task

Although Fig. 2 contains paths linking activities, these should not be seen as a workflow in a traditional sense. The edges in fact reflect shallow conceptual dependencies – as in, for example, before analyzing resources, one (usually) has to find and access them. But one may not need to carry out all the sub-tasks to reach a judgment.

As our participants' behaviour in the evaluation study of ASPL-v1 showed, there may be many configurations of this conceptual task model. We call these configurations sense-making paths and show two (of many possible) examples in Table 1.

Table 1. Sample alternative paths to accomplish and make sense of the learning task

Alternative path 1	
1.	decide on the topic of your review (e.g. ontology mapping)
2.	gather authors active in the topic/area
3.	for each author identify their co-authorship community
4.	assess the citation impact of these co-author communities
5.	summarize those that are frequently cited
Alternative path 2	
1.	decide on the topic of your review (e.g. ontology mapping)
2.	identify other 'names' of this topic or similar research areas
3.	for each area find/retrieve a sample cluster of papers
4.	assess the overlaps and similarities of these clusters
5.	compare different views on the topic in question

The path composition depends on the learner's experience with the domain and his/her approach to similar situations in the past. Thus, the sense-making path, along which the original prescriptive task can be organized, is one dimension for formalizing exploratory navigation. This form of exploration corresponds with the capability to refine a query; e.g. in our case it might help to reduce the number of retrieved papers if (say) teams of co-authors can be explored as separate clusters accessible from the original query.

In the examples in Table 1, it is particularly step #2 where the exploration seems at its strongest. The two alternative paths differ in how the review task is interpreted – is it more about mentioning the prominent approaches or is it about comparing alternative approaches? Is it more author- or topic-driven? Both interpretations are valid; they are both configurations of the same underlying model that accomplish the same task.

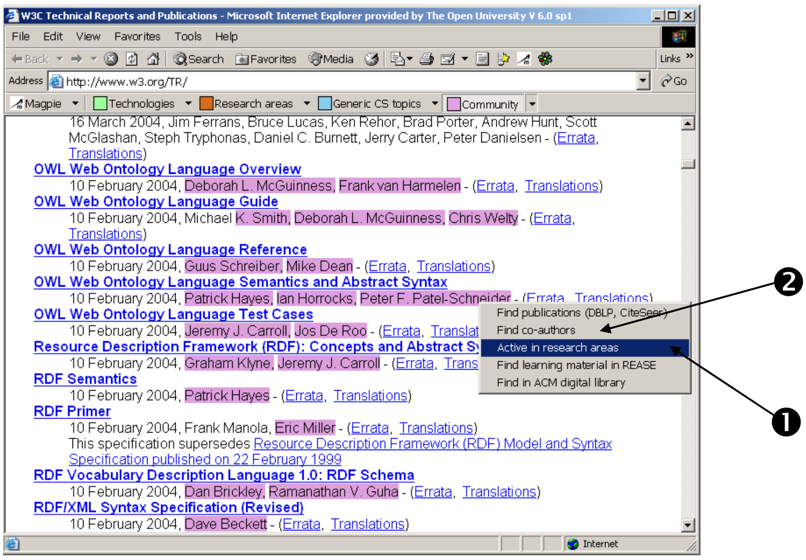
In practice the choice between the sense-making paths is *not random*. In our study we found that people tended to start with a literal interpretation of the task. If the task asked them to review "the usage of RDF in the Semantic Web", this led them use "RDF" and "Semantic Web" as keywords to search in digital collections. Many users then took whatever was retrieved from the library and listed it. Those users who achieved better quality, recognized that the search output was too coarse-grained (i.e. too many items retrieved with no apparent ranking). To reduce their frustration they could be advised to use alternative strategies: (i) filter out a few authors and follow that trail by retrieving more papers from them, or (ii) look at topics conceptually close to the ambiguous "Semantic Web" and "RDF" and use them to expand the query.

We conceptually distinguish the choice between two alternative sense-making paths as a result of *applying pedagogic knowledge*. This knowledge enables the learner to perform a 'lateral' step from one sense-making strategy to another; e.g. from a topic-driven to the author-driven one. This dimension helps with framing an open task and it is orthogonal to the task decomposition into a workflow and path composition. In practice, from their past experience or formal training, the students *know at which moment to switch* to a new or different sense-making path. The challenge for us is to facilitate this choice using Semantic Web technologies.

The user's interaction with learning resources using the Semantic Web could, therefore, be seen as navigation in a multi-dimensional space. Each dimension triggers a specific outcome; it has specific strengths and weaknesses, and is (in essence) *exploratory*. Some exploratory steps are about deepening knowledge in one domain (this is conceptually close to mSpace (Schraefel, Karam et al. 2003) and its faceted elaboration a browsing query). Other steps are 'lateral'; i.e. moving from one view of the domain to another (similar to the notion of horizontal navigation (Brusilovsky and Rizzo 2002) across distinct sub-spaces of a larger problem space). Thus, our study led to the conceptual differentiation of exploratory strategies and their formalization, so that they could be realized in the re-designed ASPL-v2.

3.2 Sense-making and pedagogic dimensions

As mentioned earlier, several authors argued the need for supporting exploration on the Web (Bush 1945; Eisenstadt, Price et al. 1983; Carr, Bechhofer et al. 2001; Lieberman, Fry et al. 2001; Brusilovsky and Rizzo 2002), so this is not a very novel conclusion from the evaluation study. What the Magpie-based application brings to the state of the art is an open, extensible framework and the capability to facilitate exploration by combining a few relatively specialized services. The services used in our study addressed fairly simple needs; e.g. retrieving publications from a digital library. The services in ASPL-v1 were not conceived as sense-making exploratory activities though. The application lacked the capability to dynamically facilitate different paths through the task.



16 March 2004, Jim Ferrans, Bruce Lucas, Ken Rehor, Brad Porter, Andrew Hunt, Scott McGlashan, Steph Tryphonas, Daniel C. Burnett, Jerry Carter, Peter Danielsen - (Errata, Translations)

OWL Web Ontology Language Overview
10 February 2004, Deborah L. McGuinness, Frank van Harmelen - (Errata, Translations)

OWL Web Ontology Language Guide
10 February 2004, Michael K. Smith, Deborah L. McGuinness, Chris Welty - (Errata, Translations)

OWL Web Ontology Language Reference
10 February 2004, Guus Schreiber, Mike Dean - (Errata, Translations)

OWL Web Ontology Language Semantics and Abstract Syntax
10 February 2004, Patrick Hayes, Ian Horrocks, Peter F. Patel-Schneider - (Errata, Translations)

OWL Web Ontology Language Test Cases
10 February 2004, Jeremy J. Carroll, Jos De Roo - (Errata, Translat

Resource Description Framework (RDF): Concepts and Abstract Sy
10 February 2004, Graham Klyne, Jeremy J. Carroll - (Errata, Trans

RDF Semantics
10 February 2004, Patrick Hayes - (Errata, Translations)

RDF Primer
10 February 2004, Frank Manola, Eric Miller - (Errata, Translations)
This specification supersedes Resource Description Framework (RDF) Model and Syntax Specification published on 22 February 1999

RDF Vocabulary Description Language 1.0: RDF Schema
10 February 2004, Dan Brickley, Ramanathan V. Guha - (Errata, Translations)

RDF/XML Syntax Specification (Revised)
10 February 2004, Dave Beckett - (Errata, Translations)

Fig. 3. Services in ASPL-v2: ❶ a lateral move from one view of the domain (people) to another (topics), and ❷ supporting 'deepening' for researchers.

For example, by identifying such concepts as authors, research areas and domain primitives, and associating services with them, ASPL/Magpie provides means to get started with the task. The re-designed semantic platform for learning (ASPL-v2) does not limit its users to a few simple services, but enriches these 'gateway services' in deepening and lateral dimensions, as mentioned in the previous section. Fig. 3 shows the services which ASPL-v2 makes dynamically available when known semantic web researchers are encountered in a web page. These includes ways of pursuing a 'deepening' path, from one researcher to his/her co-authors (marker ❷), and discovering the research areas likely to be relevant to a particular person (marker ❶).

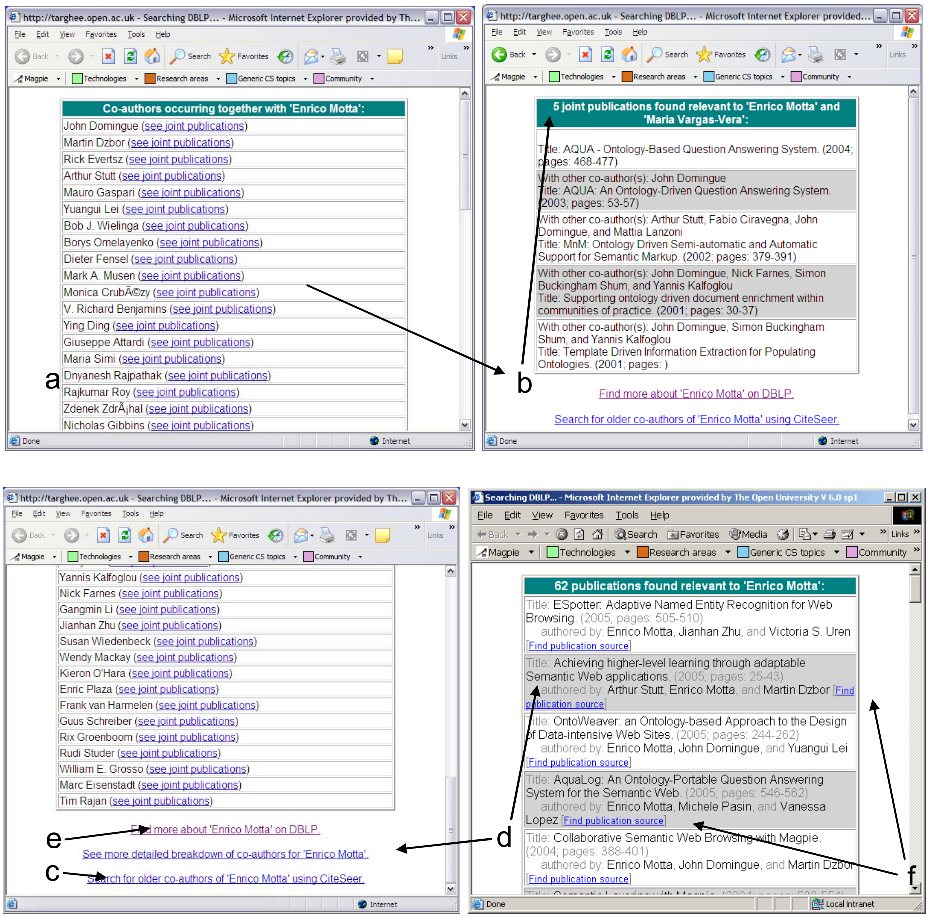
3.2.1 Improved sense-making support by query deepening

The *sense-making dimension* was implemented as a refinement of the response gained by invoking a particular service. On the cognitive level, what these refinements aimed at was to support 'spotlight navigation' (Collins, Mulholland et al. 2005) – continuously narrowing a path towards something that can be further analyzed. This path narrowing strategy is not hierarchic; it lacks the prescriptive nature of hierarchies.

For example, if the learner requests a list of co-authors publishing with Enrico Motta (see Fig. 4a), the ASPL-v2 sense-making support now offers to refine the initial list of co-authors in several directions:

1. narrow Enrico's publications to those *co-authored with specific people* (Fig. 4b)
2. look at Enrico's *co-authors over time*; e.g. older vs. newer clusters (Fig. 4c)
3. view co-authorship in terms of *publication clusters* or individuals (Fig. 4d)

4. *explore the provenance* and hence credibility of the co-author list (Fig. 4e)
5. formulate the query for *finding the source* of a particular publication (see Fig. 4f)



Co-authors occurring together with 'Enrico Motta':

John Domingue (see joint publications)
Martin Dzbor (see joint publications)
Rick Evertsz (see joint publications)
Arthur Stutt (see joint publications)
Mauro Gaspari (see joint publications)
Yuangui Lei (see joint publications)
Bob J. Wielinga (see joint publications)
Borys Omelayenko (see joint publications)
Dieter Fensel (see joint publications)
Mark A. Musen (see joint publications)
Monica CrubÃ©zy (see joint publications)
V. Richard Benjamins (see joint publications)
Ying Ding (see joint publications)
Giuseppe Attardi (see joint publications)
Maria Simi (see joint publications)
Dnyanesh Rajpathak (see joint publications)
Rajkumar Roy (see joint publications)
Zdenek ZdrÃ¡hal (see joint publications)
Nicholas Gibbins (see joint publications)

5 joint publications found relevant to 'Enrico Motta' and 'Maria Vargas-Vera':

Title: AQUA - Ontology-Based Question Answering System. (2004; pages: 468-477)

With other co-author(s): John Domingue
Title: AQUA: An Ontology-Driven Question Answering System. (2003; pages: 53-57)

With other co-author(s): Arthur Stutt, Fabio Ciravegna, John Domingue, and Mattia Lanzoni
Title: MnM: Ontology Driven Semi-automatic and Automatic Support for Semantic Markup. (2002; pages: 379-391)

With other co-author(s): John Domingue, Nick Farnes, Simon Buckingham Shum, and Yannis Kalfoglou
Title: Supporting ontology driven document enrichment within communities of practice. (2001; pages: 30-37)

With other co-author(s): John Domingue, Simon Buckingham Shum, and Yannis Kalfoglou
Title: Template Driven Information Extraction for Populating Ontologies. (2001; pages:)

Find more about 'Enrico Motta' on DBLP.

Search for older co-authors of 'Enrico Motta' using CiteSeer.

Yannis Kalfoglou (see joint publications)
Nick Farnes (see joint publications)
Gangmin Li (see joint publications)
Jianhan Zhu (see joint publications)
Susan Wiedenbeck (see joint publications)
Wendy Mackay (see joint publications)
Kieron O'Hara (see joint publications)
Enric Plaza (see joint publications)
Frank van Harmelen (see joint publications)
Guus Schreiber (see joint publications)
Rix Groenboom (see joint publications)
Rudi Studer (see joint publications)
William E. Grosso (see joint publications)
Marc Eisenstadt (see joint publications)
Tim Rajan (see joint publications)

Find more about 'Enrico Motta' on DBLP.

See more detailed breakdown of co-authors for 'Enrico Motta'.

Search for older co-authors of 'Enrico Motta' using CiteSeer.

62 publications found relevant to 'Enrico Motta':

Title: ESpotter: Adaptive Named Entity Recognition for Web Browsing. (2005; pages: 505-510)
authored by: Enrico Motta, Jianhan Zhu, and Victoria S. Uren
[Find publication source]

Title: Achieving higher-level learning through adaptable Semantic Web applications. (2005; pages: 25-43)
authored by: Arthur Stutt, Enrico Motta, and Martin Dzbor [Find publication source]

Title: OntoWeaver: an Ontology-based Approach to the Design of Data-intensive Web Sites. (2005; pages: 244-262)
authored by: Enrico Motta, John Domingue, and Yuangui Lei
[Find publication source]

Title: AquaLog: An Ontology-Portable Question Answering System for the Semantic Web. (2005; pages: 546-562)
authored by: Enrico Motta, Michele Pasin, and Vanessa Lopez [Find publication source]

Title: Collaborative Semantic Web Browsing with Magpie. (2004; pages: 388-401)
authored by: Enrico Motta, John Domingue, and Martin Dzbor
[Find publication source]

Fig. 4. A sequence of screens with different forms of 'path narrowing' for sensemaking

For example, assume the learner starts with retrieving publications using the keyword "Semantic Web", and gets thousands of references assembled purely on the basis of their publication time. In ASPL-v1, the collection used was ACM Portal[2], which offered 40,000+ hits in response to this query. What could be offered to the user at the point of receiving the results are alternative approaches to the problem:

1. try expanding or changing "Semantic Web" e.g. with related topics
2. try retrieving related authors for "Semantic Web" and search for the publications of those towards the top of the list

The above functionalities provide mechanisms for expanding the original query and re-directing the exploration into a different region of the domain.

Implementation

The service discussed above extends the old "Find publications" option that was previously based on CiteSeer and to a lesser extent on DBLP. In the re-engineered version, we placed a stronger focus on the content of the SQL database in DBLP;

---

[2] http://portal.acm.org; both 'Guide' and 'Digital Library' showed a similar performance

mainly because this is updated more frequently than CiteSeer and also contains validation of the harvested knowledge of publications.

Originally, the DBLP and CiteSeer queries supported plain keyword-based search; e.g. finding publications from a given author or with a given keyword in the title. The opportunities for deepening and amending the original query still come from the same SQL content, but are pre-computed using several patterns that are frequently used in the literature review task. These patterns include e.g. clustering of authors based on their co-occurrence, but also on more sophisticated correlation criteria: by institution, by journal or conference series, by year, etc. (for sample patterns see Fig. 4a-d).

Benefits

This enables the ASPL to draw slightly more analytic conclusions from what was previously only a flat list of SQL records satisfying a given query. The main benefit of including these pre-computed patterns is in dividing the retrieval of relevant items into a series of sub-steps enabling the user to learn more about the domain: how did it change over past X years, what are the communities of co-authors and thus potentially replaceable publication authors/keywords, etc.

Another benefit seems to be in a better management of the query results. Where in the past we merely supported a very shallow pagination of results; we can now present much larger quantities of data by taking a more “faceted approach” to the result presentation. A typical example of this kind would be presenting a list of publications retrieved for one particular researcher (say, *R*) clustered by co-authorship pairs *[R, R’]*, which reduces the complexity of result presentation up to three-, four-times.

Among the next steps for improving the faceted clustering we would like to go further towards theme-based clustering; i.e. for a large result set present it organized into themes and sub-themes. However, this pattern needs more pre-computing and interpretation than merely DBLP records; to some extent this already relies on the document content, so we need to explore our possibilities here.

3.2.2 Exploratory navigation through harvested associations

Intelligent Tutoring Systems (Frasson, Gauthier et al. 1992) and many other tools supporting learners are active applications in a sense that they lead the learner through their repositories. Yet they largely operate on a closed set of resources and tend to use manually defined abstract relationships between concepts and resources (Eisenstadt, Price et al. 1983). The most popular link of this kind is ‘requires’ – as in “*Study of ontology-based annotation requires knowledge of RDF*.” Applying the ‘requires’ link transitively, it is possible to compute user paths through the resources and ensure each user follows a prescribed learning task. However, manual annotations are not scalable; they assume one path fits all user needs. Unfortunately, as the number of links increases, this approach reduces the feasibility of this type of systems.

Rather than tying the learner into one specific learning task, we see learning tasks as an optional element in a semantic system supporting the learners. A learning task can be achieved by following several, often very distinct, *paths through the space* of (learning) resources. It is nearly impossible to formalize any one of these paths as the ideal execution of the learning task. Instead, different paths can be triggered by *associations* that happen to be useful at a particular moment. The relationship between tasks and paths is many to many – one task can be achieved by following several paths, and one particular path or its section may participate in several tasks.

The notion of exploratory user interaction with learning resources has been around for some time. In the pedagogical domain, for example, Laurillard (Laurillard 2002) characterizes education and learning as a conversation of the learner with the problem and available resources – a conversation that facilitates exploration of the relevant knowledge space.

In educational hypertext, Brusilovsky et al discuss the benefits of what they call horizontal (non-hierarchical) navigation in digital textbooks (Brusilovsky and Rizzo 2002). They notice lack of support for the horizontal links because this mode of navigation is more resource-intensive than standard classification into a vertical taxonomy. Moreover, vertical, content-based links represent the order, the plan; the horizontal, associative links are often *serendipitous* and almost always *subjective* to a learner (at least to some extent).

Implementation

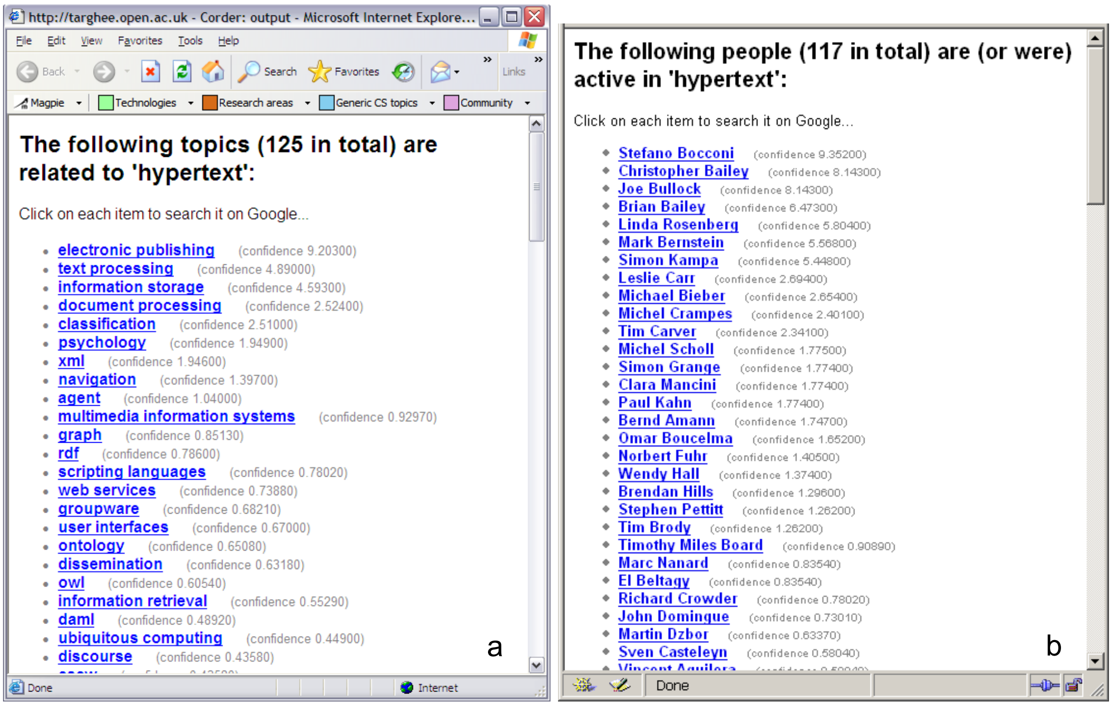
The exploratory service in ASPL relies on the data we automatically harvested from the Web using our CORDER technology (Zhu, Goncalves et al. 2005). CORDER works by starting with a seed of terms and constraining URI-s (if any are required), and follows a standard web crawling strategy to harvest new links to process. For each page/URI it uses a sliding window of a variable width to find other terms and entities co-occurring with the seed items. The window size and correlation algorithm are both parameters that can be set by the user or data manager to customize the data acquisition process.

Once data is harvested, it is stored locally in cache databases using pairs of items and their respective correlation coefficients. In the context of our ASPL service we mine for people (as named entities), research topics, technologies, and organizations. Correlations are calculated using the occurrence distance of the seed term and other terms from the same domain within a specified sliding window. Using domain ontologies one can then interepret the correlations between e.g. two people using relationships like “*has similar interests*” or “*collaborates with*”, and between a person and a research topic using e.g. “*does research in*” or “*publishes on*”. Some examples of the results from querying the CORDER-harvested data are shown in Fig. 5.

For instance, in the examples shown in Fig. 5 we assumed the learner wanted to retrieve publications using the keyword “hypertext”. In the previous version of ASPL one would get thousands of references assembled purely on the basis of containing given keyword. In ASPL-v1, the collection used in this manner was e.g. ACM Portal, which offered 40,000+ hits in response to such a query. What could be offered to the user at the point of receiving the results and prior to their presentation are alternative approaches to the problem:

- try expanding or changing “hypertext” e.g. with related topics, or
- try retrieving related authors for “hypertext” and search for the publications of those towards the top of the list

The two alternative lateral modifications of the original retrieval query as depicted in Fig. 5 are as follows: On the left we see a loosely related list of themes related to “hypertext”, while on the right is an orthogonal list based on authors active in “hypertext”. The content of Fig. 5a can be (semantically) interpreted as: *Continue exploring main topic “hypertext” in the context of sub-topic (e.g.) “navigation”*. Using a similar interpretation, one can also carry out more analytic or synthetic tasks, such as for example: *Compare the outcomes of “navigation” vs. “text processing” contexts of the main theme “hypertext”*.



The following topics (125 in total) are related to 'hypertext':

Click on each item to search it on Google...

- electronic publishing (confidence 9.20300)
- text processing (confidence 4.89000)
- information storage (confidence 4.59300)
- document processing (confidence 2.52400)
- classification (confidence 2.51000)
- psychology (confidence 1.94900)
- xml (confidence 1.94600)
- navigation (confidence 1.39700)
- agent (confidence 1.04000)
- multimedia information systems (confidence 0.92970)
- graph (confidence 0.85130)
- rdf (confidence 0.78600)
- scripting languages (confidence 0.78020)
- web services (confidence 0.73880)
- groupware (confidence 0.68210)
- user interfaces (confidence 0.67000)
- ontology (confidence 0.65080)
- dissemination (confidence 0.63180)
- owl (confidence 0.60540)
- information retrieval (confidence 0.55290)
- daml (confidence 0.48920)
- ubiquitous computing (confidence 0.44900)
- discourse (confidence 0.43580)

The following people (117 in total) are (or were) active in 'hypertext':

Click on each item to search it on Google...

- Stefano Bocconi (confidence 9.35200)
- Christopher Bailey (confidence 8.14300)
- Joe Bullock (confidence 8.14300)
- Brian Bailey (confidence 6.47300)
- Linda Rosenberg (confidence 5.80400)
- Mark Bernstein (confidence 5.56800)
- Simon Kampa (confidence 5.44800)
- Leslie Carr (confidence 2.69400)
- Michael Bieber (confidence 2.65400)
- Michel Crampes (confidence 2.40100)
- Tim Carver (confidence 2.34100)
- Michel Scholl (confidence 1.77500)
- Simon Grange (confidence 1.77400)
- Clara Mancini (confidence 1.77400)
- Paul Kahn (confidence 1.77400)
- Bernd Amann (confidence 1.74700)
- Omar Boucelma (confidence 1.65200)
- Norbert Fuhr (confidence 1.40500)
- Wendy Hall (confidence 1.37400)
- Brendan Hills (confidence 1.29600)
- Stephen Pettitt (confidence 1.26200)
- Tim Brody (confidence 1.26200)
- Timothy Miles Board (confidence 0.90890)
- Marc Nanard (confidence 0.83540)
- El Beltagy (confidence 0.83540)
- Richard Crowder (confidence 0.78020)
- John Domingue (confidence 0.73010)
- Martin Dzbor (confidence 0.63370)
- Sven Casteleyn (confidence 0.58040)

Fig. 5. Two examples of lateral navigation from one original topic-based query: (a) showing a step from ‘*hypertext*’ topic to related sub-topics, and (b) showing a step from ‘*hypertext*’ as a research topic to the authors ‘*publishing on*’ this topic

Similarly, by switching from topic-based navigation to authors, one actually reformulates the original retrieval task using semantically close entities, which may, in turn, either broaden or narrow the outcomes – depending on the user needs and the situation.

This particular set of functionalities is already available as a fully-fledged web service; it also has a web frontend for interacting with users, so this service is already, to some extent, satisfying one of the weaknesses of the old ASPL – its over-reliance and embedding in the web browser. Both functionalities shown in Fig. 5 are accessible in a simple web-based script code that takes care of invoking the actual web service, and hence, they could be included in other web-based applications (possibly REASE and other).

Benefits

In spite of the recognition of the need to open up hypermedia (Carr, Bechhofer et al. 2001) and to enable the creation of horizontal links (Brusilovsky and Rizzo 2002), one large obstacle has always been the problem of specifying meaning. In the standard Web, this is normally implicit inside the mind of the resource provider, and even when it is spelled out it needs to be discovered in the text surrounding the link.

Semantic Web technologies take a complementary stance by assuming that each relation or association can be ‘clarified’ by committing to a particular semantic interpretation. Once an explicit semantic commitment (or annotation) is made, the relation acquires a specific meaning, which in turn enables distinguishing between (say) a person being *active in* an area and one area being *similar to* another. In short, the Semantic Web extends the notion of associative navigation by the fact that associations are not only established, but more importantly, interpreted.

The idea of bootstrapping the semantic web through ‘non-authoritative’ associations can also be found in recent work on ontologies and social software. For example, Mika (Mika 2005) formally explores the role of the emerging associative component in the designed or engineered ontologies or taxonomies. One of his conclusions is that emergent semantics may be ad-hoc, subjective and incomplete, but it complements more formal, yet largely static ontological vocabularies. In contrast with Mika’s approach, ASPL-v2 uses an ontology to *give semantics to untyped associations*.

The advantage of this approach is that the application is scalable and maintainable and does not suffer from the classic knowledge acquisition bottleneck. The disadvantage is that the validity of the information cannot be guaranteed. However we do not regard this as a bug and indeed our view is that, as the semantic web grows in scale and applications begin to make use of large amounts of information derived from different sources, they will necessarily be forced to be able to perform in the face of variable data quality. This scenario is obviously in stark contrast with that of classic knowledge-based systems, which by and large tended to operate on closed domains, where data quality could be easily guaranteed.

4 New ASPL services

Today ontologies and taxonomies are the driving force behind organizing large document collections and digital libraries. Moreover, if exploited for navigational searches, they often allow for a better usability of large collections then mere keyword search (Hyvönen, Saarela et al. 2003). In this section we show how ontologies and (simpler structures like) concept hierarchies can be used to improve navigation and query amendment in a complex problem space, and also to construct ad-hoc learning narratives from simpler data and document chunks.

4.1 FacetedDBLP++

To complement the work about navigational exploration, L3S has provided two new prototype ASPL services: The FacetedDBLP++ browser, which utilizes topic hierarchies created automatically with the Semantic GrowBag approach, and the tag-based profile tool.

Faceted search uses taxonomies to structure the information space that a user can explore. But what facets should be chosen to classify a result set? Here, working on the metadata (e.g., title or author keywords) representing a corpora of documents is far more promising than directly working on full-texts that are difficult to parse and not always available. The Semantic GrowBag approach uses taggings on real-world corpora to automatically derive community-specific and time-sensitive topic facets. These facets reflect the main relations of a chosen topic with respect to its semantic environment and the topic's evolution over time. We demonstrate its practical applicability using the DBLP collection of computer science documents and present FacetedDBLP++, a faceted search engine that allows the efficient and dynamic topical grouping of results and a query expansion with semantically related keywords.

4.1.1 Principle of FacetedDBLP++ service

The faceted search paradigm (Anagnostopoulos, Broder et al. 2005; Hearst 2006) allows arranging results of Web search or document retrieval in a way that all results can be clustered according to several topics, which, in turn, reflect different views or different focal points on the data or documents. For many applications fully-fledged and expensive ontologies to drive faceting are not really needed. Especially when it comes to searching and distinguishing between topics in a structured way, somewhat simpler concept hierarchies have proved to be sufficient (Käki 2005).

A major challenge is to automatically derive topic facets tailored for specific domains or communities. We advocate using tags in large corpora of documents related to the domain; for instance, digital bibliography collections[3] or digital libraries[4]. Often these corpora do not need expensive language processing on full-texts, and are already tagged with different metadata. In fact, tagging documents and creating 'folksonomies' e.g., from author-given keywords in research papers, topics of Web pages (cf. http://del.icio.us/), or annotations of non-textual documents like images (cf. http://www.flickr.com/), has already become commonplace. Based on our Semantic Growbag technology (Diederich, Thaden et al. 2006) we designed a *faceted browser*

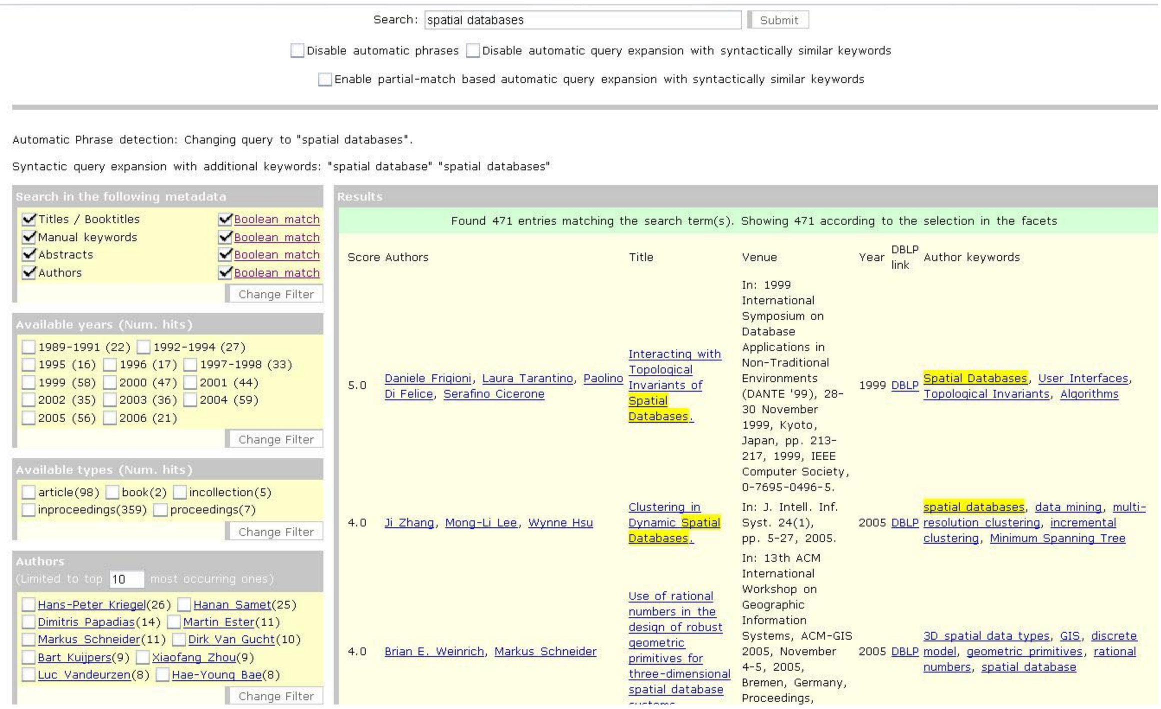
---

[3] Such as DBLP (http://dblp.uni-trier.de/) for the computer science

[4] Such as Medline Database (http://medline.cos.com/) for the area of medicine

for DBLP that allows querying research papers based on automatically created facets like topic, authors or publication year.

The Semantic GrowBag algorithm grows domain-specific concept hierarchies with respect to items from any bag of tags extracted from suitable document corpora. The first step in creating a suitable facet for some topic is to identify all relevant related concepts to a query keyword and encourage or discourage the creation of simple subsumption relations by higher order co-occurrences of tags. Investigating co-occurrences of all tags in the corpus' documents to find an initial set of related topics, we then use a biased PageRank to efficiently identify the most important topics and their relations for a given community and within a given time span. Hence, even current trends in certain topic areas can be reflected. Cross-checking the PageRank scores of related concepts, GrowBag also derives subsumption relationships with a certain confidence score for all identified concepts. The details of the algorithm are presented in (Diederich, Thaden et al. 2006).


Search: spatial databases

☐ Disable automatic phrases ☐ Disable automatic query expansion with syntactically similar keywords

☐ Enable partial-match based automatic query expansion with syntactically similar keywords

Automatic Phrase detection: Changing query to "spatial databases".

Syntactic query expansion with additional keywords: "spatial database" "spatial databases"

Search in the following metadata

☑ Titles / Booktitles ☑ Boolean match
☑ Manual keywords ☑ Boolean match
☑ Abstracts ☑ Boolean match
☑ Authors ☑ Boolean match

Available years (Num. hits)

☐ 1989-1991 (22) ☐ 1992-1994 (27)
☐ 1995 (16) ☐ 1996 (17) ☐ 1997-1998 (33)
☐ 1999 (58) ☐ 2000 (47) ☐ 2001 (44)
☐ 2002 (35) ☐ 2003 (36) ☐ 2004 (59)
☐ 2005 (56) ☐ 2006 (21)

Available types (Num. hits)

☐ article(98) ☐ book(2) ☐ incollection(5)
☐ inproceedings(359) ☐ proceedings(7)

Authors
(Limited to top 10 most occurring ones)

☐ Hans-Peter Kriegel(26) ☐ Hanan Samet(25)
☐ Dimitris Papadias(14) ☐ Martin Ester(11)
☐ Markus Schneider(11) ☐ Dirk Van Gucht(10)
☐ Bart Kuijpers(9) ☐ Xiaofang Zhou(9)
☐ Luc Vandeurzen(8) ☐ Hae-Young Bae(8)

Results

Found 471 entries matching the search term(s). Showing 471 according to the selection in the facets

Score	Authors	Title	Venue	Year	DBLP link	Author keywords
5.0	Daniele Frigioni, Laura Tarantino, Paolino Di Felice, Serafino Cicerone	Interacting with Topological Invariants of Spatial Databases.	In: 1999 International Symposium on Database Applications in Non-Traditional Environments (DANTE '99), 28-30 November 1999, Kyoto, Japan, pp. 213-217, 1999, IEEE Computer Society, 0-7695-0496-5.	1999	DBLP	Spatial Databases, User Interfaces, Topological Invariants, Algorithms
4.0	Ji Zhang, Mong-Li Lee, Wynne Hsu	Clustering in Dynamic Spatial Databases.	In: J. Intell. Inf. Syst. 24(1), pp. 5-27, 2005.	2005	DBLP	spatial databases, data mining, multi-resolution clustering, incremental clustering, Minimum Spanning Tree
4.0	Brian E. Weinrich, Markus Schneider	Use of rational numbers in the design of robust geometric primitives for three-dimensional spatial database	In: 13th ACM International Workshop on Geographic Information Systems, ACM-GIS 2005, November 4-5, 2005, Bremen, Germany, Proceedings,	2005	DBLP	3D spatial data types, GIS, discrete model, geometric primitives, rational numbers, spatial database

Fig. 6. The FacetedDBLP++ demonstrator

As a corpus, we took the computer science publications listed in DBLP enriched by matching author keywords and abstracts from ACM DL, IEEE DL and the Springer Archive. We extracted the associated tags using acronym replacement and Porter-stemming to clean the data and removed rare tags occurring less than 5 times. Our DBLP++ collection comprises rich bibliographic information of about 800,000 documents, of which 85,000 are described by a set of about 460.000 tags

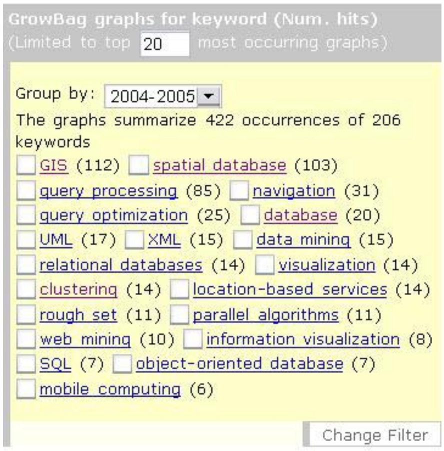
4.1.2 The FacetedDBLP++ demonstrator

Let's assume we are interested in spatial databases and, hence, type 'spatial databases' keyword into the search box. The basic appearance of the result page is shown in Fig. 6. Besides the result frame, it features facets for publications' year, article type, authors, etc.

The first visible feature is the *syntactic query expansion*. As keywords have been created by the authors of the publications, they naturally undergo some variations in spelling or abbreviations. As an example, the keyword ‘recommender system’ may be written as ‘recommendation system’. When processing the authors’ keywords, we identify similar keywords and assign the same keyword identifier to publications being tagged with all variants. Different spellings are now used to automatically expand the query with all variants (in this example ‘spatial database’ and ‘spatial databases’). However, our approach does not require a time-consuming stemming of all abstracts and titles. Instead we only stem the given keywords. This approach works well, if queries on concepts that are actually given by some author keywords prevail. Using the DBLP++ data as of September 2006, our syntactically expanded query for ‘spatial databases’ can find 471 results.

Generally results are ranked according to a score, which is the cumulative number of matches of all query terms in the title, the abstract (if available), the keywords (if available) and the author names. In the example, the first result has one hit in the title, one in the keywords, and three in the abstract, thus a score of 5. This result set, however, is still too large for manual inspection. The problem is to find a more specific query with fewer results. Therefore, to get an overview of ‘spatial databases’, FacetedDBLP++ offers a *topical facet* providing a list of ‘most frequent co-occurring keywords’. For our query, however, this list contains about 200 keywords. The main purpose of a GrowBag graph here is to compress the list of co-occurring keywords by grouping semantically related topics. In our example, we are only interested in the latest developments and ask for a GrowBag facet on spatial databases for the period 2004/2005. GrowBag graphs have two main characteristics:

1. They comprise latent related keywords for a given input using higher-order co-occurrences between keywords and thus include a certain degree of transitivity in the computation.
2. They provide subsumption relations between keywords, which makes it possible to distinguish more general or more specific concepts for a given keyword.



GrowBag graphs for keyword (Num. hits)
(Limited to top 20 most occurring graphs)

Group by: 2004-2005
The graphs summarize 422 occurrences of 206 keywords
☐ GIS (112) ☐ spatial database (103)
☐ query processing (85) ☐ navigation (31)
☐ query optimization (25) ☐ database (20)
☐ UML (17) ☐ XML (15) ☐ data mining (15)
☐ relational databases (14) ☐ visualization (14)
☐ clustering (14) ☐ location-based services (14)
☐ rough set (11) ☐ parallel algorithms (11)
☐ web mining (10) ☐ information visualization (8)
☐ SQL (7) ☐ object-oriented database (7)
☐ mobile computing (6)

Fig. 7. The topic facet of GrowBag graphs

Because they are automatically generated, the subsumption relations have a somewhat less strict semantics compared to manually crafted classifications or thesauri

(Diederich, Thaden et al. 2006). Fig. 7 depicts a sample GrowBag graph for the keyword ‘GIS’ (Geographic Information Systems); more complex graphs can be viewed using the web interface of our demonstrator. The principle for reducing the size of a ‘most frequently co-occurring keywords’ facet is to delete those keywords that occur in the GrowBag graph of a more general keyword; i.e. they are subsumed by the latent neighborhood of that keyword. This allows reducing the number of relevant keywords in our example by about 75%. Furthermore, in the graphs with each keyword we show (in brackets) the number of occurrences actually subsumed. In this way, the importance of a GrowBag graph to a specific topic can easily be inferred (also the keywords shown in the demonstrator’s topical facet are ranked by that number).

For instance, ‘Geographic information systems (GIS)’ is a keyword, whose GrowBag graph comprises even more keyword occurrences than the graph for the original query `spatial databases’. Hence, we may select this graph to extend our search and semantically filter out all documents that are not relevant to GIS (i.e. contain keyword subsumed by this graph). As a result, the ranking of relevant documents changes, since publications tagged with keywords from the GrowBag graph of ‘GIS’ get a boost in the score and the relevant document set is reduced to 85 documents. If on the other hand, we had decided that our interest is rather in the area of ‘location-based services’, we would have ended up with a slim set of 11 relevant documents concerned with typical location-based topics in spatial databases like k nearest neighbour searches, routing and road networks, or moving objects.


Keywords selected by growbag:
☑ GIS ☑ remote sensing
☑ computational intelligence ☑ web services
☑ spatial database ☑ spatial join ☑ spatial data
☑ satellite images ☑ visual query languages
☑ DEM ☑ crisis management

Fig. 8. The topical keywords subsumed by ‘GIS’

Having selected a suitable GrowBag graph for filtering, another facet is displayed that comprises all the chosen graph’s keywords. This facet allows the explicit exclusion of irrelevant keywords and, thus, refines the search again. This feature is especially useful in the case of large GrowBag graphs. In our example, having chosen the ‘GIS’ graph e.g., the keyword `Web services’ could be deselected, if we do not want papers with a connection between web services and spatial databases.

4.1.3 Discussion and comparison

FacetedDBLP++ also support the functionality of browsing co-authors (over time), provenance (by linking to DBLP), or related publications clusters. While this functionality was already integrated into ASPL-v1, it was rather flat. Now it is presented to the user in a different, richer way using the re-designed FacetedDBLP++ service. Moreover, FacetedDBLP++ provides additional navigational features such as filtering along ‘topic clusters’ to find sub-communities within a given topic. As described in the example above, users can easily find all documents related to ‘Geographic Information Systems (GIS)’ related to the original query ‘spatial databases’ and they will also find that ‘spatial databases’ are a subtopic of ‘GIS’ as more documents match to the related keywords of ‘GIS’ (112) than to ‘spatial databases' (103) even though all documents in the result set have been selected because they match 'spatial databases' in a standard IR-based search engine.

The demonstrator is currently available at:
http://demo.l3s.uni-hannover.de/FacetedDBLP and shall be included as a fully-accessible ASPL service at the next revision of the system.

4.2 Pedagogic services

New service that has been designed by INRIA with the pedagogic objectives in mind is QBLS. QBLS stands for "Question Based Learning System", and its primary aim is to support students in problem-based learning when they are performing "question-based tasks". Such tasks typically lead to the situation when a participant needs some additional information. The QBLS system provides a way to access potentially relevant resources based on the conceptual navigation paradigm.

The resources offered by QBLS may have been created on purpose or (preferably) reused from another web-based repository. A long-term goal in the domain of e-learning is to automate the reuse of existing resources coming from different sources, and to combine them in a single application. In QBLS, we support the reuse of existing heterogeneous resources through a semantic annotation and a subsequent conversion to HTML format. The annotation mechanism is semi-automated, and relies on several ontologies (domain, pedagogy and document).

We use the expressiveness and inference mechanisms now available among the semantic web technologies to express and manipulate resource annotations and also the resources themselves in a single application. We identified four practical research questions where the application of Semantic Web brings interesting results (the questions are discussed more in depth in Table 2):

1. Which generic architecture to choose for an intelligent learning system?
2. How to perform learning resource selection based on domain knowledge?
3. How to perform pedagogical reasoning?
4. How can a system be linked to other learning tools over the web?

Table 2. e-Learning issues related to finding answers in the context of Semantic Web

Architecture	Various tools are now available (web servers, semantic search engines, etc.). What would be a simple, generic and scalable architecture for future applications?
Resource selection	A review of the analysis of knowledge models for adaptive and intelligent learning systems, clearly shows that semantic web formalisms possess enough power to express those models. Can we apply Information Retrieval (IR) techniques from the Semantic Web to perform learning resource selection?
Pedagogical reasoning	A major interest of learning systems is the ability to support a given pedagogical strategy. How reasoning with OWL and rules can effectively be operated in this context?
Interoperability	One of the most striking interests of Semantic Web is the interoperability offered between systems. Can this also be achieved between learning systems?

4.2.1 QBLS architecture

The chosen engine for QBLS is the Corese search engine. Corese processes knowledge about the learning resources expressed through RDF annotations and RDFS ontologies. It also handles features of OWL-lite (e.g. transitivity). Queries to the engine are expressed in the SPARQL and answers are either in RDF or in the XML. Corese is internally based on conceptual graphs and needs to perform a translation from RDF(S) to conceptual graphs. The projection algorithm is also used to answer semantic queries. Corese integrates an inference engine based on forward-chaining production rules, expressed in a language based on SPARQL triples. The proposed architecture is generic and could apply to any "semantic search engine" taking ontologies and annotations (in OWL and RDF) and answering semantic queries (e.g. in SPARQL).

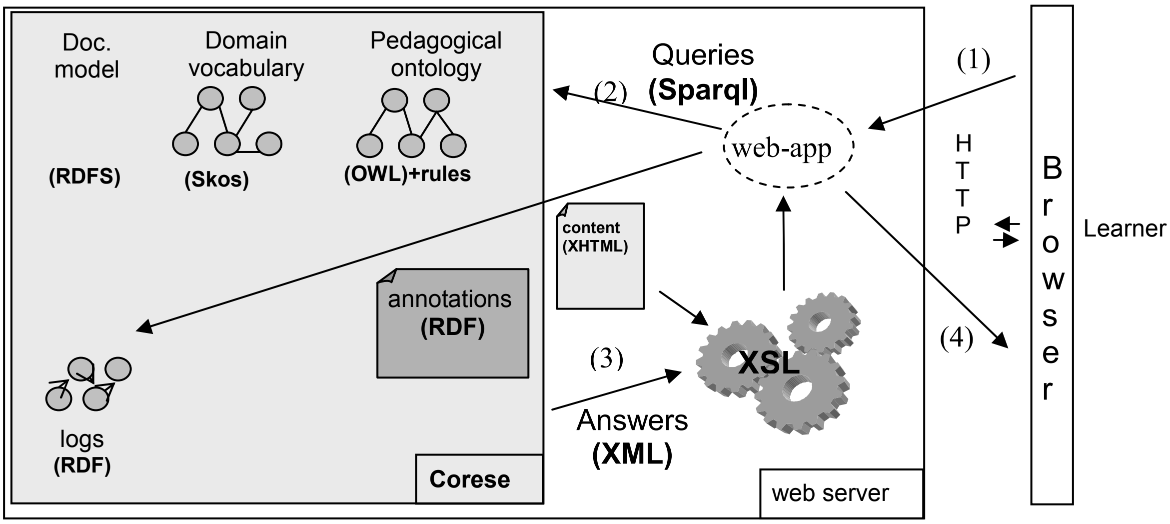


Fig. 9. Architecture of the QBLS service

The QBLS system is developed as an industry-standard web-application on Tomcat. The application accesses an instance of the Corese semantic search engine. HTTP requests from the user are answered through JSP pages and servlets that help to build the answer. The learning resources managed by the system are XHTML pages, stored on the server. This standard format allows the system to access and perform adaptation on the content dynamically. XSL transformations are used to construct the interface from both the resources and the results of the engine. XSL stylesheets are purely static and do not perform any hard coded reasoning. All the inferences are performed by the semantic search engine. The architecture of the system is summarized in Fig. 9.

The architecture shows how reusing an existing a semantic web tool like Corese reduces the amount of work necessary to deploy such an application. The amount of code written to integrate the elements shown Fig. 9 and to build a dynamic interface does not exceed a thousand lines of code (mostly XSL and JSP). Of course, other elements for generating annotations automatically, upload new courses, etc. were necessary as in any application, but the core functionalities are implemented within very simple components.

The QBLS system can be reused as it is for other similar applications, but it might also be tailored very easily to answer other needs. Handling new types of navigations or educational features is a matter of work on the models, queries and associated interfaces. Because of this scalable approach, we were able to develop several instances of the QBLS system for different learning domains. Currently, it has been deployed as a service in the modular ASPL platform.

4.2.2 How QBLS works

RDF annotations link resources to domain and pedagogical concepts. Concepts are defined in various ontologies: the domain, the pedagogical aspects and the document model. The system keeps track of the user activity by generating RDF triples that are added to the annotation graph base handled by Corese. We distinguish two ways of representing and using domain knowledge in a "semantic" learning system (but note that both identified approaches can be supported Semantic Web technologies):

- using a pre-defined OWL domain ontology,
- using a less structured domain vocabulary in SKOS

Pedagogical knowledge is expressed in a different ontology than the domain. This knowledge is used by the system to guide the learner from a pedagogical point of view. The following example, taken form QBLS, involves pedagogical knowledge: When querying for resources related to a domain concept, several answers often match the query. Answers are presented in a dedicated interface (see Fig. 11 showing the result of a search on the concept of "Knowledge Management" in ASPL). Two resources were found: a definition and a fact. The pedagogical knowledge expressed in the ontology defines the "Definition" and "Fact" as concepts. Their first advantage, typical in learning systems, is their associations in the interface with navigation links to help learners better identify the role of a resource. Here they appear as tab headings for selecting an appropriate resource while navigating.

An "intelligent" system can exploit this information further and plan a coherent path among the resources. This is done in QBLS by ordering the resources in a pedagogically sound way. In controlled experiments, we have observed that roughly 50% of the time, students visit the proposed resources in their order of presentation, from left to right. Thus, the ordering is crucial, and determines half of the learning path, the other half being decided by the learner upon the indications given by the interface. Using the inference capabilities of the Corese search engine, results are sorted according to this "role", or pedagogical concept the resources instantiate. Roles are organized in the pedagogical ontology into a classification hierarchy.

To help deciding automatically on this order, pedagogical statements are added by an expert teacher, using RDF triples. For example the following pedagogical expertise: "Fundamental resources are prior to auxiliary resources" is expressed by the following RDF statement: "`edu:Fundamental <edu:priorTo> edu:Auxiliary`". From a small set of such statements, and using forward -chaining rules, the semantic search engine completes the RDF graph to create priority relations between all roles. The resulting graph, Fig. 10, formed by the nodes of the ontology linked with priority relations is a directed graph without cycles (provided that the expert did not define explicit cycles).

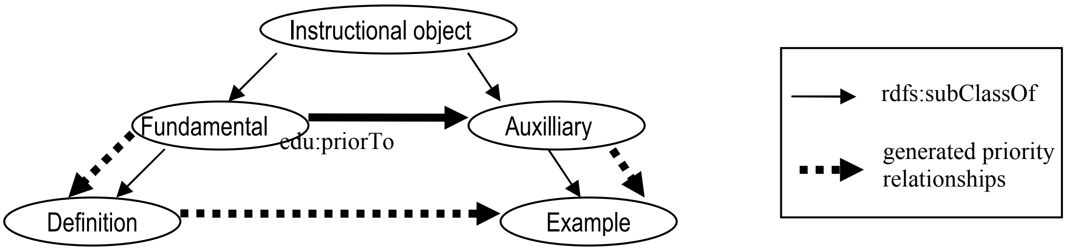
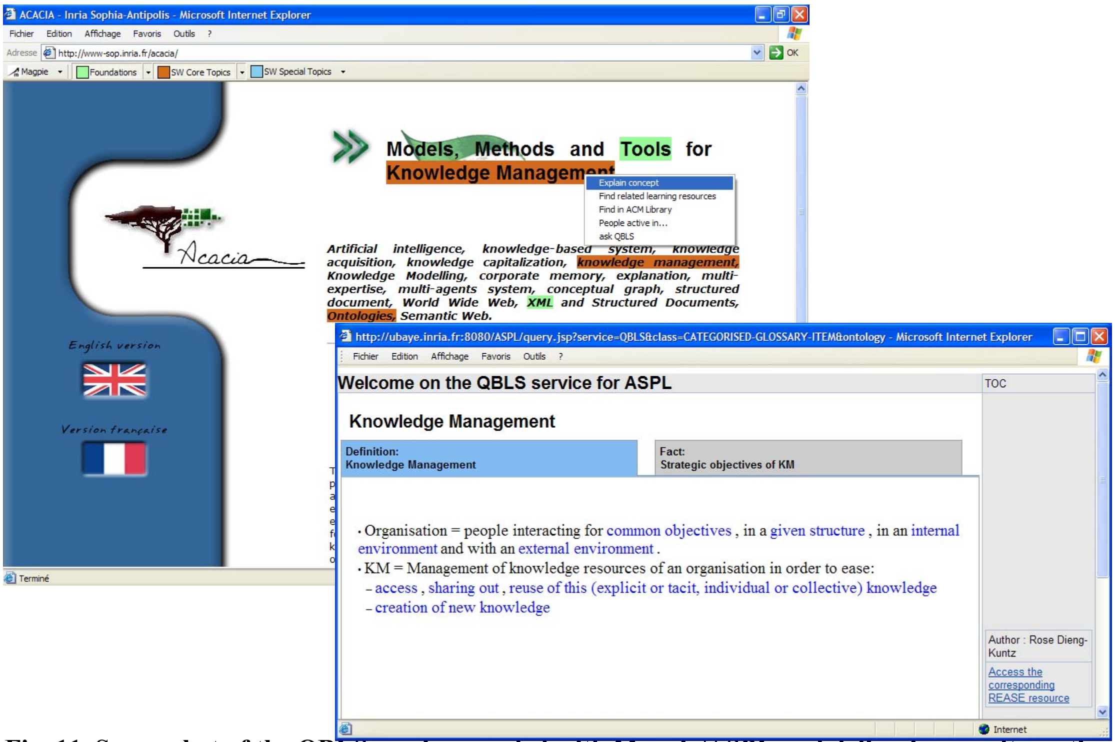


Fig. 10. Excerpt of the pedagogical ontology completed with priority links and indexes

Through this example of applying rules to propagate pedagogical relations and using them to compute an index on the ontology, we show how generic inference mechanisms can rely on pedagogical information to suggest coherent paths to learners. This information is contained in a formal ontology completed with rules.

4.2.3 QBLS in ASPL

In ASPL, learners browse web content with the Magpie plug-in and access the QBLS service for spotted domain concepts. The domain ontology is shared between the different tools, including magpie. Fig. 11 shows a screenshot of a web page where the terms relevant to semantic web studies are highlighted. For each concept (here "Knowledge Management") a contextual menu proposes a range of available services. Selecting the QBLS service opens the second window where resources associated to this concept can be browsed. Resources are reused form PowerPoint slides provided by the members of Knowledge Web and annotated with the concepts of the Semantic Web hierarchy and a pedagogical ontology.



Models, Methods and Tools for Knowledge Management

Artificial intelligence, knowledge-based system, knowledge acquisition, knowledge capitalization, knowledge management, Knowledge Modelling, corporate memory, explanation, multi-expertise, multi-agents system, conceptual graph, structured document, World Wide Web, XML and Structured Documents, Ontologies, Semantic Web.

Welcome on the QBLS service for ASPL

Knowledge Management

Definition:
Knowledge Management

Fact:
Strategic objectives of KM

- Organisation = people interacting for common objectives , in a given structure , in an internal environment and with an external environment .
- KM = Management of knowledge resources of an organisation in order to ease:
  - access , sharing out , reuse of this (explicit or tacit, individual or collective) knowledge
  - creation of new knowledge

Author : Rose Dieng-Kuntz

Access the corresponding REASE resource

Fig. 11. Screenshot of the QBLS service coupled with Magpie/ASPL and delivering an alternative (or additional) explanatory information about topic "knowledge management"

5 Improving ASPL ontology – methodological view

Another area where we attempted to further our understanding of the problem domain concerns the acquisition of a set of ontologies suitable for pedagogic purposes. Here we started by recognizing that learning is a complex and multifaceted activity, of which two broad types can be distinguished (Laurillard 2002): the "natural" learning and more artificial, "academic" learning. The former occurs in the everyday situations, and often corresponds to the acquisition of not only *passive knowledge* about a domain but also an *active knowing* (Cook and Brown 1999). This is particularly visible in foreign languages (Willis 1996) – when we learn English, we actually go and speak it in various situations; we don't merely learn about the language and its grammar.

The latter type of learning often occurs in formal education, in schools, and particularly in universities. Academic learning is essentially about abstractions; that is, it is not directly tied to reality. It mostly works with and refines various formal descriptions of what happens in the real world. Indeed, what teachers try to convey to learners are alternative accounts of everyday experiences, reusable descriptions of already known phenomena (for example in physics, formulae). Of course, these second-order descriptions are (not surprisingly) often hard to grasp in their abstract nature.

With a bit of over-simplification, the former learning is situated in a *real* situation, whereas the latter is more about *designed* problems and situations. The academic education is not always as straightforward as the simplified view may imply, and it often suffers from methodological limitations. For many years, the leading pedagogic strategy was *instructional* (Noddings 1998), where a learner's mind was conceived as some sort of repository. This empty store was susceptible to being filled with knowledge, as an unformed piece of clay waiting to be modeled.

This view undermines the learning process' outcomes, for its inherent dualistic standpoint that considers knowledge as an entity that is entirely different and independent of the learner (Noddings 1998). The learner, instead, was reduced to an objective and impersonal carrier of some static and abstract knowledge. In 1960-s a systematic opposition arose against this view by proposing a framework that takes into account the active participation of the learner in the knowledge creating process. For example, *constructivism* exemplifies the range of assumptions used to model this notion of the learning process, and consequently the engineering of the educational design. As Bruner says (Bruner 1966):

> "We teach a subject not to produce little living libraries on that subject,
> but rather to get a student to think mathematically for himself, to
> consider matters as an historian does, to take part in the process of
> knowledge-getting. Knowing is a process not a product".

More recently, other frameworks attempted to reconcile the 'user entity' with the 'knowledge entity'. Examples such those in (Cook and Brown 1999; Nonaka, Toyama et al. 2000) can be summarized under the heading of *situated cognition* (Brown, Collins et al. 1989). Learning is always *framed* in a situation, where the problem and its context are defining each other (Dzbor and Zdrahal 2002). It is actually correct to say that the solution is spelt out in terms of the context, and the problem is solved thanks to the entities a specific situation makes available. According to the authors, the situation *affords* a certain kind of knowledge production. This is true not only for the natural learning, but also for the academic learning. It is therefore crucial to contrive a setting for the learning (of both skills and abstractions) where what is taught does not appear in isolation (by definition), but in concomitance with the situation it originates from.

Through employing a task-based methodology to model the existing activities in a particular domain, we can provide a meaningful breakdown of the knowledge needed for teaching the domain *itself*. By this we mean that domain comprises not only abstract descriptors *about* it, but also includes an *active element* of practicing the domain, doing it. We present an early form of a model that offers a unique value to the educational designers; it reflects the *situated* style of the human cognition and the authentic activity through which learners create knowledge.

5.1 Task-oriented model

The formalization of tasks and subtasks in a particular domain through ontologies and other knowledge representation techniques is a well-established and well-researched discipline. For example, in (Motta 1997) or (Schreiber, Wielinga et al. 1993) ontologies of reusable task-method structures are presented and then employed to model a classic problem solving situation – e.g. parametric design. The task-based methodology has also been used in the relatively new field of the Semantic Web.

More specifically, we highlight the research brought forward by Mizoguchi and others (Chen, Hayashi et al. 1998; Aroyo and Mizoguchi 2003) on applying tasks in the field of eLearning. This research has lead to the formalization of pedagogical strategies and related actions within an ontology compounded of educational tasks and sub-tasks. A seminal work in (Mizoguchi and Bourdeau 2000) defines the new "Instructional Design" paradigm as the evolution of Intelligent Tutoring Systems and Interactive Learning Environments; that is, as a "process by which learning events can be defined or described, independently of their instructivist or constructivist orientation". Ontological engineering is clearly seen as the bridging technology between the different pedagogical approaches, which are represented at the task level in the form of activity-related concepts.

Task-based methodology is commonly used in many domains, including eLearning. For example, in (Carro, Pulido et al. 1999) the notion of task is used to specify the right sequence of learning resources the application has to present, according to their instructional value and the author's learning design. This can be pedagogically sound but by focusing on the meta level of structuring a course it still does not sufficiently bridge the gap between a natural, situated learning and the artificial, academic one. In other words, although the authors recognize that the "process of learning is more complex than navigating between different pages and reading what is written on them", they do not try to frame the learning activity within the context of a problem-based situation.

We are instead using the task-based methodology to tie the objects of a specific domain to their effective usage, that is, to their *raison d'etre*. The context of existence and usage of an object is defined as the activity occurring within a particular domain the object can be used for, can drive, inform or otherwise influence, and the motivation of this choice of the usage. This view attempts to embed the objects from a specific domain not necessarily to the pedagogic tasks and activities but to the tasks and activities that are typical for the domain in question and are embedded (possibly in many) situations.

Only through the explicit recognition of this network of an object usage, the domain knowledge can be treated as a truly situated cognition. The core section of our position centers on the notion of 'usage semantics' – which can be distinguished from e.g. descriptive semantics of the knowledge acquisition methodologies. In other words, the value of task-based methodology applied to constructing domain ontology is in recognizing the fact that it may be hard to agree upon an acceptable definition of a concept.

With respect to a widely known model in (Stojanovic, Staab et al. 2001), our approach proposes that beyond being used effectively to describe the *structure* (defined as the layout of a "set of learning materials in a learning course") and the *context* (defined as "the form the topic is presented") of the educational design, a task-based methodology can be also employed to represent the actual *content* of the course taught (namely, "what the learning material is about"). In order to do so, the following fundamental concepts are recognized in the resulting model:

- A *problem* or *problematic situation*: this can be defined as an issue or situation that is so-far unresolved, from the perspective of a particular learner. A problem in this generic sense looks for a solution and/or for further specification (Dzbor and Zdrahal 2002).
- A *solution*: this can be seen as an outcome (or a set of multiple outcomes) that to some extent resolves the issue, the conflict or the situation identified by the learner as the problem.
- A *generated problem*: defined as a specific outcome of breaking down the problem at the input into several sub-problems – with respect to a particular approach. Different approaches afford the learner to perceive different sub-problems or the same sub-problems may emerge at different times. Two types of generated problems can be identified in our model:
  - An *in-domain sub-problem* … defined as a problem or a part thereof that according to a particular approach is amenable by the domain's methods, techniques and other conceptual apparatus.
  - An *out-domain sub-problem* … defined as a problem or a part thereof that has been considered to be amenable by methods and concepts of a different domain.
- Finally, an *approach* … this can be defined as a deliberately taken perspective on the situation in which the learner is and in which s/he perceives a problem. When used in a particular situation, an approach becomes the frame that allows further processing and analysis of the problem and the synthesis of the solution.

This conceptual framework is expressed graphically in Fig. 12. Accordingly, a learner starts with a problem or an unresolved issue within a real situation – for instance, how to cope with knowledge being distributed across the Web. This problem is often subjective to a specific learner and his or her perspective. Thus, the approach a learner takes to some extent affects what one considers 'a problem' and what is just 'a background'. In general, this input problem is only a starting point for the domain exploration and subsequent conceptualization based on the usage semantics.

With a situation at hand, a learner then makes a conscious choice of (or defines) an approach, a perspective on this situation. The above situation – coping with dispersed knowledge – can be analyzed in terms of social networks, in terms of distributed systems, in terms of linking chunks of distributed knowledge, etc. Unlike other work using the task-based methodology, what constitutes our approach is not prescribed. An approach can be seen as a set of assumptions that in some way reflect learner's understanding of a particular domain. Thus, an approach is a kind of open toolkit that gives the learners some pre-existing vocabularies, ontologies and techniques, and enables them to construct a conceptual model of their situation by means of concept usage semantics. This 'usage' may be as simple as finding a tool solving a particular issue or as complex as dissecting an issue and translating it into a domain-, or better approach-specific language – as in the scenario in section 3.

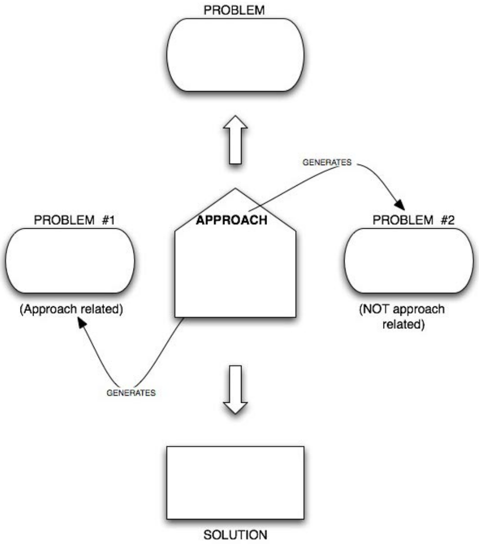


Fig. 12. Generic task model for conceptualizing a the domain by situating it into a set of recognized problems, approaches and solutions

The recursive instantiation of the model from Fig. 12 within a specific, chosen domain, as we show in the next section, is capable of iteratively generating all the concepts constituting the network of key problems in the field, their solutions and the set of "boundary" problems, which can be seen as the bridges to other disciplines and research areas. By definition (of the notion 'approach'), the completeness of the conceptual model of the domain has to be judged from the perspective of *that specific approach*. Different approach may generate different conceptual networks.

5.2 Case example: Semantic Web Studies

One of the main drivers of our effort here is to complement the existing topic-centered framework of REASE by more problem-centered learning. Currently, REASE contains in the range of 100 learning resources. These are on a fairly high level of granularity, as most consist of full tutorials, seminar talks and exercises. The core facility on REASE[5] that allows semantic structuring and browsing is the topic ontology that hierarchically organizes the topics in the Semantic Web. The only meaningful usage of this conceptual model is to comprehend an elementary linking between resources that can be inferred through the abstract is-a relationship.

This ontology, however useful in organizing a repository, carries little pedagogical value. Simply, when one wants to understand notion of 'OWL semantics' it is not sufficient to know that this belongs to the category of (say) 'Logical Foundations of the Semantic Web'. Ontology designed to catalogue resources does not reflect the usage semantics of various notions the learner may want to acquire. From the perspective of our 'usage semantics', one wants to learn how does 'OWL semantics' work, what does it afford, what are its limits, where can it be applied, etc.

5.2.1 Principles and the case scenario

Our case scenario therefore aims to instantiate the generic problem-based model from Fig. 12 in the specific field of Semantic Web *Studies*. Rather than aiming to design yet another ontology for the domain, we believe our model will enable us to complement the existing ontologies (incl. the topic catalogue in REASE) and to draw new connections between already existing materials. Doing so, a learner would be able to contextualize domain entities by their *usage*, or utility value in tackling specific issues and problems. This would enable him or her to draw connections between the problems, concepts and solutions, other than an artificial, imposed top-down classification of coarse-grained resources into one topic hierarchy.

As showed in Fig. 13, we take an existing application developed at KMi, ASPL/Magpie, and try to frame it using our model from Fig. 12. The basic question that drives this kind of contextualization has this form: "If there is X, what does solve or tries to solve it?" The question aims at revealing the 'usage semantics' of an entity (as explained earlier). In general, since we do not often have complete solutions but only attempts or partial ones, the aim of the above question may become to reduce the complexity of an unknown problem/situation by breaking it down into smaller (and more easily understandable) components.

The starting point (situation or problem) is totally arbitrary. Not only that, it is actually irrelevant, since the iteration of the model would eventually "spit out" all the fundamental problems and solution in the field (of course, constrained by our particular chosen approach, e.g. Semantic Web as distributed knowledge representation systems).

For example, the problem a tool such as Magpie (and in general, any other semantic browser) addresses is the dispersion of potential learning resources on the Web and their recollection on-the-fly depending on the learner's context of usage. We define the Semantic Web approach to frame this problem as comprising a model of the relevant domain knowledge, relying on ontology and performing a web search with that ontology as a filter. The problem is thus broken down into (at least) two broad sub-problems:

---

[5] REASE is one of the outcomes of the project's educational area, it stands for Repository of the European Association for the Semantic Web Education, and is available at http://rease.semanticweb.org

(i) how to model a domain and
(ii) how to implement an information extraction technique using an ontology.

The first one is considered again as a problem relevant to the Semantic Web studies domain (so it will iteratively move to the upper box generating a new instantiation of the model). The second one is a "border" problem, outside the scope of our teaching (again, this scoping depends on how the approach has been chosen). Eventually, the purpose of the iterations is to identify all these "border" problems, which of course can be linked to the problematic issues arising in or tackled by other domains (Brusilovsky and Rizzo 2002).

5.2.2 Discussion of the proposed case

It is important to notice how the formulation of the approach drives the whole characterization of problems as Semantic Web related or not. Even if reaching a general agreement over what constitutes a Semantic Web problem will be extremely difficult (if not impossible), we decided to take the following standpoint. We defined the specific approach taken in Fig. 13 as one based on distributed knowledge representation systems, as opposed to other technologies (e.g. databases or folksonomies).

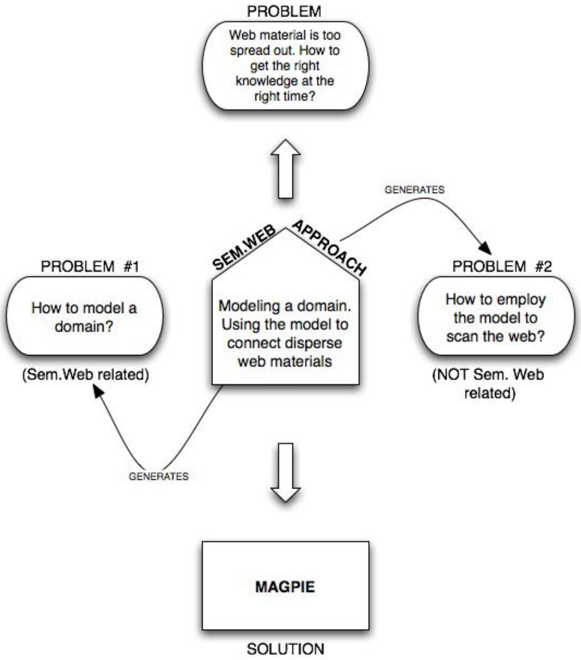


Fig. 13. Specific model conceptualizing the domain of Semantic Web Studies starting by situating one particular problem into a set of sub-problems and a technological solution

This choice is of course totally arbitrary and subject to criticism. But since the overall aim of this framework is to dissect a domain for teaching purposes, and not to define or acquire once and for all the essence of the Semantic Web (or any other domain), we do not discuss this choice any deeper at this stage of our investigation, but note it as a potentially interesting follow-up topic. As a teacher inevitably has to take a specific set of assumptions in order to explain and explore with his/her students an object domain, we also have defined the Semantic Web approach in a specific way in order to shed light on related "teaching materials". It would be interesting to see how an alternative definition would organize the problem space according to different principles.

The key aspect we would like to highlight instead is the methodology chosen: the problem-based model is being experimented with, in order to produce the "situated context" of the entities forming a domain. In the long run, as our understanding evolves, the definition of the approach (that is, the specific view of what the Semantic Web is) will be probably revised and refined. However, our shaping the approach differently shall not affect the strategic choice that attempts to relate domain objects to their usage semantics.

Our focus shall not therefore be on producing an ontology that *defines* classes and sub-classes, but on carving out the relations between entities in the domain. An interesting outcome would be, for example, the usage of a hierarchical classification (such as the one in REASE) *within* the model: since Magpie is classified as a semantic browser, also other similar applications can be retrieved as examples of solutions to the same problem. In other words, if Magpie is presented to learners in the context of the problem of remote and disconnected web materials, automatically another semantic browser can be treated as relevant since it also fits in the actual context. On the other hand, these other 'solutions' may shed new light on the existing problem.

We believe this view is compliant with a known pedagogic paradigm that interprets learning as a process of situated cognition. In this way, knowledge chunks and their conceptual abstractions are not considered isolated, but in the context specified by their *Problem-Approach-Solution* space. We have also introduced an early and ongoing work aiming at instantiating this generic model for the situations involving teaching the students to practice the Semantic Web. An example from the field describes our methodology and gives a glimpse of the future direction of this research.

More specifically, the key challenge this task-centered model has to tackle includes the problem of reproducing what typically happens during "natural learning" also in the more abstract, descriptive kind of learning. For example, we see one potential benefit in tying the abstract concepts from the latter kind of learning back to the applicable natural situations (i.e. problems). This application of natural strategies in more formal types of learning is partially supported by various cognitive studies that show domain experts often seeing their problems through associations directly linked to their problem space and environment, rather than relying on any formal representation in prescribed classification schemas (Brown, Collins et al. 1989; Cook and Brown 1999; Nonaka, Toyama et al. 2000).

Disclaimer: This section is loosely based on a paper from M. Pasin and M. Dzbor titled "*A Task Based Approach to Support Situated Learning for the Semantic Web*" and published at the SW-EL workshop collocated with Conf. on Adaptive Hypertext 2006

6 Towards methodology for ASPL evaluation

In terms of evaluating tools like ASPL we would like to differentiate between:

- *User independent evaluation of the tool* and
- *User-based evaluation of the tool*

The purpose of the former type, i.e. user-independent evaluation, can be summarized in the following three facets/steps:

  - Firstly, *the core functionalities associated with the tool need to be determined*. The core functionalities allow us to describe the scope and main features of the tool. Moreover, the scope of the tool is useful to determine what the tool is expected (by its designer) to perform while the different features associated with the tool specify what engineering activities can be achieved by the user while using the tool.
  - Secondly, once the scope of the tool is determined, then it is important to *determine what evaluation measures are needed* for evaluating the core functionalities of the tool. These evaluation measures allow us to check whether the core functionalities of the tool perform different engineering activities correctly and efficiently for which they are engineered in the tool.
  - Thirdly, once the evaluation criteria are determined for *the family of tools, which are performing similar engineering action(s)*, e.g. data categorization, then the family of tools can be compared and contrasted against the evaluation matrix. The results obtained after performing the evaluation study can be used to determine the performance against these 'benchmark' tools.

After performing user independent evaluation of the tool, the next stage is to perform the evaluation of the tool in use by different users. This would allow analyzing the performance of the tool when used by its intended users in a particular setting or task. Users may vary in terms of their experience, academic background, and/or knowledge of the domain:

  - Typically, here the first aim of the evaluation study is to hand over the tool to the users with a set of tasks, each allocated a fix duration within which the users are expected to perform the given tasks.
  - Having users completed given tasks, the performance of the tool such as speed of carrying out a task, accuracy (i.e. tool efficiency), relevance of the results (i.e. tool effectiveness), etc. can be analyzed.
  - The evaluation results obtained through this stage of the study would, in turn, allow identifying specific strengths and weaknesses of the tool in a particular activity, and this might be done in comparison against other tools addressing similar issues.

Both types of evaluation studies mentioned above can be performed at different stages of the software development process in order to analyze the performance of the tool. Our objective in this section is to start describing the evaluation of ASPL, viewed as a suite of techniques, which carry out specific tasks such as the data categorization, information retrieval, and identification of correlations among the terms by taking the textual data as an input. First, we describe the core functionalities that may occur in connection with such primitives, techniques. Secondly, having identified the core functionalities, we consider how these functionalities might be evaluated.

6.1 Data categorization functionality

The main aim of data categorization is to perform systematic classification of the terms that occur in a document (in our case, named entities and phrases appearing in a web page) by taking into account the way they are associated with each other. For example, in ASPL the categories include various topics deemed relevant to teaching and learning about Semantic Web. In such a situation, the tool may (i) recognize such entities as e.g. *RDF*, *XML*, or *ontology mapping*, and (ii) categorize the former two as (say) *Semantic Web technologies* and the latter as a *research theme*. Obviously, from this perspective, the choice of categories depends on the domain, or on the focus, etc.

The following typical evaluation criteria may apply to this functionality of the ASPL system. Alternatively, we can see these criteria as sufficiently informative *if* we decide to treat the ASPL as a system that can be used purely for data categorization:

1) *The quality and applicability of the different choices of categories used by the tool to categorize the data*:
    a) The main objective of this statement is to find out whether the tool provides a comprehensive coverage of *a particular topic, theme*, *or domain* in terms of available classification categories. This is important as it influences the perception of tool's effectiveness in a specific task/domain. If used appropriately, it can be used to distinguish, for example, between the data coming from the same domain or from another, heterogeneous domain.

    b) It may also be important to assess whether the tool takes into account relevance of the categorized terms when putting them in a specific category. For example, if the text comprises different aspects and entities associated with the Semantic Web, such as the languages used or the people working in this field, then the *categorization* function should show consistency and classify terms that intuitively viewed by the users as related in the same category (e.g. *RDF* and *OWL* are both intuitively viewed as languages for the Semantic Web).

2) *Different modes of data categorization used by the tool to categorize the data:*
    a) Here we may distinguish e.g. *visual categorization* where the main aim is to use e.g. different colour schemes or font sizes to differentiate between categories depending on the way these terms are related with each other.

        i. One can observe whether the visual categorization is *self-explanatory*. In other words, it may be important to assess that the categorizing schemes/views provide an indication about the sub-domains to which different terms belong.

    ii. Similarly as mentioned in point 1b above, terms that are considered related or somehow similar in a given domain should have a similar visual appearance (e.g. in our example, both *RDF* and *OWL* would be highlighted in the same colour). The visual clue is used here as a factor to discriminate between different categories, and hence its efficiency and effectiveness can be assessed from this perspective.

    iii. Another aspect to evaluate is whether the visual categorization is performed correctly. In other words, if the terms like *Magpie* occur in a document, which is about the web-based tool, then it is important that the tool categorizes such a term by using the same colour as any other tools that may be mentioned in this context, and that the tool does not consider the term *Magpie* as a bird (to use an extreme example).

    iv. Having performed visual categorization, one can assess whether the colour schema used by the tool is not ambiguous. In other words, our aim could be to measure the granularity of the categorization; i.e. how many visual schemes are used by the tool (and the user), and consequently, how efficient these clues are for the user to discriminate between different categories.

    v. Finally, one can assess whether the tool maintains consistency in using a specific visual clue for the data categorization in heterogeneous documents. For instance, if the tool makes use of the blue colour clue to categorize occurrences of such terms as *XML* or *RDF* in one web page, then the same clue shall be used e.g. for terms like *Java* or *Lisp* occurring in a different document to indicate that there is some conceptual connection between them.

b) Another way to discriminate data may be a *slot-based categorization*. This functionality typically makes use of various tree structures to categorize data elements. In this case, our aim may be to evaluate the following criteria particularly relevant to the slot-based categorization:

    i. As above, we can assess whether the different nodes of the classification structure used by the tool to categorize the data elements are comprehensive enough. In other words, it is useful to evaluate whether the different nodes used by the tool provide an adequate coverage and topographic structure to a particular domain. The adequateness can be seen as an appropriate depth and/or breadth of the tree, a number of children per node, tree balance, etc.

    ii. The categories in the structure may be assessed in terms of their compliance with the general naming conventions and/or taxonomies of the domain to reduce the level of ambiguity. For example, avoid hard-coding category structure into names, e.g. 'Researchers' and not 'People-Researchers', or topical category names that should ideally be singular, e.g. 'Semantic Web technology'.

iii. Another aim may be to assess that the different categories used by the tool actually make the browsing or navigation any easier or more effective. This is particularly interesting from the end user's viewpoint because it represents the most tangible benefit from using the semantic tool to browse through documents and articles relatively straightforwardly.

iv. With respect to the categorization structures, it is useful to evaluate whether the siblings in the structure are associated with other members correctly. This would make the data categorization consistent and it also helps users to retrieve the data that are more closely related to the queried term(s).

v. The categories can be devised in such a way that the same term does not appear in more than one category. For example, when categorizing the term *RDF*, this should not appear both among "languages for the Semantic Web" and among (say) "research topics" – mainly to avoid a potential ambiguity. Furthermore, it may also be interesting to consider other aspects of structural categorization; e.g. to see if it is not cyclic in nature or logically incoherent, etc.

c) After visual and structure based approaches, the categorization may also be *ontology-driven*, where an ontology can be used as a basis for categorization. In this case, the tool subscribes to the underlying ontological structure and data in order to categorize the data.

i. Here, the evaluation criterion for the data categorization performed by the tool may assess whether the underlying ontology and the knowledge base is sufficiently comprehensive to facilitate precise and semantically sound categorization.

ii. Furthermore, in addition to ontology precision, it may be useful to assess the degree of recall; i.e. how many terms occurring in the text are actually correctly processed and found relevant to a given ontology.

iii. The time taken by the tool to perform data categorization can be also an evaluation criterion. In other words, if the tool performs the data categorization correctly but it takes a long time, then the users may prefer to use an alternative tool, which performs the same task faster but maybe less precisely.

iv. Another possibility is to evaluate whether the tool allows modification in the underlying ontology structure or its content by adding new concepts or re-assigning instances. This is particularly interesting when the existing ontology fails to provide necessary level of detail, precision, or degree of recall.

    v. Another aspect that may be assessed at this point is the tool's *interoperability* and possible combination with other tools and datasets that may have been developed for handling heterogeneous domains. The main purpose of this evaluation strategy is to see whether the tool handles ontologies developed by using different standards or by different authors.

  d) One should also not overlook the usability criterion of the tool, which may need to be evaluated with an aim to see if different users can use the tool with a particular level of expertise.

    i. The tool may be evaluated on its scalability to give an indication about how it handles large data sets. Some of the important evaluation criteria closely related to the scalability of the tool include its performance, speed, consistency or the accuracy of categorization – usually in terms of gains or degradation.

    ii. If the tool performs the data categorization then the reusability may be an issue to assess. Here, one can evaluate: a) whether the tool can be reused over heterogeneous domains, or b) the amount of reengineering effort needed to reuse the tool in other domains. If major engineering efforts are needed for using the tool in other domains, this may be seen as a particular weakness.

6.2 Dictionary and glossary functionality

The main aim of this functional viewpoint is to provide an explanation or definition for the annotated data elements, such as concepts, named entities, etc. For instance, if the tool annotates the concepts, such as *RDF*, *XML* or *OWL*, then the glossary services can be used to provide an explanation of these terms or their examples. Below we describe different evaluation criteria that can be used to evaluate this aspect of the ASPL tool:

1) One can evaluate whether an explanation provided by the dictionary-like service is *relevant and conceptually correct* with respect to the domain.

  a. For example, users may have different motivation when performing, e.g., a scholarly review of the domain as opposed to defining the core concepts. In the former case, it is more appropriate to provide a comparative explanation of the annotated terms; in the latter case, one may opt for a single more detailed and descriptive explanation.

2) Alternatively, one may evaluate the level of detail provided by the dictionary-like services while explaining or defining the annotated terms.

    a. From this perspective, one can consider whether an explanation provided by the dictionary-like services takes into account varying degrees of expertise of the tool users.

    b. It may be important for the tool to support users who are comparatively new to the domain, and for such users simplified technical information needs to be provided or customized.

    c. On the other hand, for the experienced users the tool needs to provide information on a more appropriate technical level to achieve the same benefits; so again, a degree of customization or value added can be observed and studied.

6.3 Information retrieval functionality

The main aim of this view of the ASPL is to retrieve relevant materials from the database or another repository based on a keyword-based query submitted by the user. This functionality can be used to retrieve the internal material, e.g. from REASE in case of ASPL; in this situation, the target audience are typically the students accessing the learning and teaching materials for their coursework. Alternatively, the tool may retrieve various types of external publications from a third-party repository on the Web (such as CiteSeer or DBLP). The target audience of this service consists of both the students as well as relatively expert researchers in respective domains. The information retrieval functionality of the tool can thus be evaluated by using the following criteria:

1) Evaluation criteria for the information retrieval *from internal repositories*:
    a. One may assess whether the documents retrieved by the tool are indeed learning and/or teaching materials as opposed to e.g. flyers or abstracts or general presentations.

    b. One may obviously consider checking the precision of the retrieved documents; i.e. whether the material retrieved by the tool is pedagogically relevant to the query submitted by the user.

2) Evaluation criteria for the information retrieval *from the external repositories*:
    a. Similarly as for the internal materials, one may want to assess that the resources retrieved by the tool are actually publications; e.g. workshop, conference, or journal publications.

    b. Furthermore, it is also possible to evaluate the quality of the retrieved materials to see an extent to which they are related to the submitted query.

   c. Another criterion in this scope might comprise aspects of detail; i.e. did the tool retrieve metadata *about* a publication or was it the publication itself?

3) More generic evaluation criteria for the information retrieval include:

   a. Evaluation of the tool to assess whether the information retrieval activity is performed in a reasonable time interval, e.g. in real time or what is the acceptable delay otherwise.

   b. Once the relevant documents are retrieved and displayed, it may be also useful to assess whether the tool provided references to other related or similar materials (e.g. in a form of a neighbourhood) that may be associated with the retrieved items.

   c. While performing the information retrieval activity based on a keyword-based query, it may be interesting to assess if the tool can identify some similarity patterns between the queries of different formats. For instance, if a user is looking for the documents, which are about '*Problem Solving Methods*', then it may be desirable to identify semantically close variants of the submitted query such as '*PSM*', '*PSMs*', or syntactic variants such as '*Problem-Solving-Methods*', which essentially refer to the same topic.

6.4 Identification of correlations

The main aim of this functionality in ASPL would be to establish a relationship between the keywords in order to retrieve the more meaningful data set. This can be sub-divided into the following four categories:

1) A relationship can be established between two separate data terms in order to perform more meaningful search and retrieve the correct set of documents.

   a. If a user makes search in a document set that contains heterogeneous topics, and submits phrase "john smith" as a query, this may be associated with different domains, and hence have a different semantic connotation. In a leisure domain, "john smith" may lead to retrieving items and concepts associated to breweries, beers, etc. For the Semantic Web domain, the same keyword may have different associations, including "john smith" as an author, an algorithm name, etc.

   b. It may be interesting to evaluate the appropriateness of interpreting the association; e.g. seeking "john smith" may mean finding other items of the same type (say, researcher) close to the original term. Alternatively, it may mean finding materials citing this original term or describing it. This can obviously lead to different associative networks, and consequently to different explanations, retrievals, categorizations, etc.

c. The threshold limit can also be tested to assess the statistical significance of correlations. In other words, if we have a collection of terms representing two separate domains, say ‘people-in-organization’ (PIO) and ‘research-topics’ (RT), then the relationship among them can be established, i.e. PIO + RT => Cor, and only correlations between the terms classified a-priori in given domains can be considered as a valid solution set.

d. Furthermore, if the context of search are documents (say) about ‘description logic’, then one needs to take into account more than mere classification. For instance, in this example, ‘$\text{PIO}_{\text{enrico-franconi}}$ + $\text{RT}_{\text{description-logics}}$’ tend to have more chances of achieving the set threshold in comparison with ‘$\text{PIO}_{\text{stefan-rüger}}$ + $\text{RT}_{\text{description-logics}}$’. As a result, it makes more sense for the tool to consider the former association as more likely candidate while retrieving the documents, which are about description logics.

e. Having filtered out the statistically significant correlations, the user then can make use of such a combination in order to retrieve more meaningful set of documents. This allows us to assess the utility of correlations, e.g. in terms of query expansion, problem space navigation, problem or solution re-formulation, etc.

7 Future work

Work reported in this deliverable focused mainly on the learner's interaction with resources on the Semantic Web. We highlighted the fact that for the purposes of learning these interactions are more than mere annotation, retrieval and subsequent browsing of semantic metadata. In order to apply semantic knowledge, the re-designed version of ASPL used an exploratory approach to interacting with distributed learning resources. Specifically we implemented two distinct modes of exploratory learning: (i) convergent, 'spotlight-style' (Collins, Mulholland et al. 2005) browsing of semantically enriched resources, and (ii) divergent, 'serendipitous' browsing into an open web space (Brusilovsky and Rizzo 2002).

Applying Semantic Web to construct multiple exploratory paths and attending to different aspects of the exploration, rather than to the individual nodes of the semantically enriched space, has several side effects. For instance, from the user experience viewpoint, the application becomes more flexible. A semantically enriched application does not confine its user to one specific activity or role. Another side effect is the dynamics of the semantic application. Ontology-driven solutions are often brittle; often based on closed worlds that enable reasoning solely about the known concepts. Linking the association discovery to the presentation overcomes this brittleness, and also avoids the knowledge acquisition bottleneck.

One of the outstanding tasks is to assess whether the re-engineering actually worked. In other words, in the remaining period of the project, we shall focus more resources on formulating a useful strategy and method to evaluate such a complex application as ASPL. We intend to further develop the evaluation method that was sketched in the previous section, with a particular emphasis on assessing the effectiveness of the ASPL system and its services with respect to supporting the learning task at hand.

As described in the section discussing the evaluation methodology, the current approach to evaluating ASPL is more theoretically driven; i.e. we are exploring the grounding of evaluation of the tools like ASPL. In the future, one of our main aims is to apply this theoretical foundation to a set of real-life applications. This aim is, however, conditioned by concluding the ASPL application – i.e. deciding on a sub-set of services that would be fully enabled and later evaluated. As can be seen in the report, there is some redundancy in the services scope; e.g. the query deepening techniques and FacetedDBLP++ are somewhat overlapping. This is acceptable, as both services have started from slightly different assumptions. Nevertheless, for the purpose of evaluation we may decide to use only one of them – depending on what particular task we decide to pursue.

Once the ASPL application is finalized, we can start instantiating the methodology we hinted at in the previous section. This is important as we need to design a plausible evaluation strategy that actually provides a useful set of conclusion. For example, comparing the precision of two services for visualizing DBLP yields entirely different outcomes as assessing the effectiveness and educational added value of only one of them (say FacetedDBLP++). As the choice of evaluation strategies strongly influences the identified strengths and weaknesses (as we found in the previous study of ASPL-v1), we want to spend more time analyzing the implications on the conceptual and theoretical level.

Yet another possibility we are currently considering would be to evaluate ASPL with respect to other, similar tools from the same category. This would be very informative – it would enable us to compare and contrast the performance of the different tools on a particular activity. While this benchmarking and comparative analysis is useful and desirable, it major disadvantage is to find the appropriate competitors so that we compare like with like. In the past, Magpie and ASPL have been compared with more specialized information retrieval and named entity recognition tools, but this did not really give us much added knowledge for e.g. educational scenarios.

Furthermore, ASPL features several rather different capabilities ranging from data and document retrieval to problem space navigation and to query expansion, amendment and re-formulation. So far, majority of Semantic Web tools focus on partial functions compared to ASPL; they are more specialized and optimized for a single task. On the contrary, ASPL has been conceived as a flexible framework addressing different stages of a fairly complex learning task (in particular, gathering date for literature review). The ASPL tasks are far less well defined and more open than mere named entity recognition or document retrieval. Hence, these aspects need to be taken in account in the remaining time of the project – both to inform the evaluation and to drive the selection of appropriate services in the final revision of the ASPL system.

8 References

Anagnostopoulos, A., A. Broder, et al. (2005). Sampling search-engine results. 14th Intl. WWW Conf., Japan.

Aroyo, L. and R. Mizoguchi (2003). Process-aware Authoring of Web-based Educational Systems. Proc. of the 1st International Workshop of Semantic Web for Web-based Learning, Austria.

Bloom, B. S. (1965). A Taxonomy of Educational Objectives Handbook 1: Cognitive Domain. New York, US.

Brown, J. S., A. Collins, et al. (1989). "Situated Cognition and the Culture of Learning." Educational Researcher **18**(1): 32-42.

Bruner, J. (1966). Toward a Theory of Instruction. Cambridge, MA, Harvard University Press.

Brusilovsky, P. and R. Rizzo (2002). "Map-Based Horizontal Navigation in Educational Hypertext." Journal of Digital Information **3**(1): 156.

Bush, V. (1945). "As we may think." The Atlantic Monthly **176**(1 (July)): 101-108.

Carr, L., S. Bechhofer, et al. (2001). Conceptual Linking: Ontology-based Open Hypermedia. 10th Intl. WWW Conf., Hong-Kong.

Carro, R. M., E. Pulido, et al. (1999). TANGOW: Task-based Adaptive learNer Guidance On the WWW. 2nd Workshop on Adaptive Systems and User Modeling on the WWW, Toronto, Canada.

Chen, W., Y. Hayashi, et al. (1998). An Ontology-based Intelligent Authoring Tool. Proc. of the 6th International Conference on Computers in Education, Japan.

Collins, T., P. Mulholland, et al. (2005). Semantic Browsing of Digital Collections. Proc. of the 4th Intl. Semantic Web Conf., Ireland.

Cook, S. D. N. and J. S. Brown (1999). "Bridging Epistemologies: The Generative Dance Between Organizational Knowledge and Organizational Knowing." Organization Science **10**(4): 381-400.

Diederich, J., U. Thaden, et al. (2006). The Semantic GrowBag - Automatically Creating Directed Tag Graphs. Hannover, Germnay, L3S Research Center.

Diederich, J., U. Thaden, et al. (2006). The Semantic GrowBag Demonstrator for Automatically Organizing Topic Facets. SIGIR Workshop on Faceted Search, Seattle, US.

Dzbor, M., A. Stutt, et al. (2007). "Representations for semantic learning webs: Semantic Web technology in learning support." Journal of Computer Assisted Learning **23**(1):69-82 (*in press*).

Dzbor, M. and Z. Zdrahal (2002). Design as interactions between problem framing and problem solving. 15th European Conference on AI (ECAI), Lyon, France.

Eisenstadt, M., B. A. Price, et al. (1983). "Software Visualization As A Pedagogical Tool." Instructional Science **21**: 335-365.

Frasson, C., G. Gauthier, et al. (1992). Intelligent Tutoring Systems. Intl. Conf. on Intelligent Tutoring Systems (ITS), Springer-Verlag, Berlin.

Hearst, M. (2006). "Clustering versus faceted categories for information exploration." Communications of the ACM **49**(4): 59-61.

Hirschman, L. and N. Chinchor (1997). Named Entity Task Definition. 7th Message Understanding Conf. (MUC-7).

Hyvönen, E., S. Saarela, et al. (2003). Ontogator: Combining View- and Ontology-Based Search with Semantic Browsing. XML Finland 2003: Open Standards, XML, and the Public Sector, Finland.

Käki, M. (2005). Findex: search result categories help users when document ranking fails. Conf. on Human Factors in Computing Systems, Oregon, US.

Laurillard, D. (2002). Rethinking University Teaching: A Conversational Framework for the Effective Use of Learning Technologies. London, UK, RoutledgeFarmer.

Lieberman, H., C. Fry, et al. (2001). "Exploring the web with reconnaissance Agents." Comm. of the ACM **44**(8): 69-75.

Mika, P. (2005). Ontologies Are Us: A Unified Model of Social Networks and Semantics. 4th Intl. Semantic Web Conf., Ireland.

Mizoguchi, R. and J. Bourdeau (2000). "Using Ontological Engineering to Overcome AI-ED Problems." International Journal of Artificial Intelligence in Education **11**(2): 107-121.

Motta, E. (1997). Reusable Components for Knowledge Modelling. The Netherlands, IOS Press.

Noddings, N. (1998). Philosophy of education. Boulder, Colorado, Westview Press.

Nonaka, I., R. Toyama, et al. (2000). "SECI, Ba and Leadership: a Unified Model of Dynamic Knowledge Creation." Long Range Planning **33**.

Schraefel, m. c., M. Karam, et al. (2003). mSpace: interaction design for user-determined, adaptable domain exploration in hypermedia. Workshop on Adaptive Hypermedia and Adaptive Web Based Systems, UK.

Schreiber, A. T., B. J. Wielinga, et al. (1993). KADS: A Principled Approach to Knowledge-Based System Development. London, Academic Press.

Stojanovic, L., S. Staab, et al. (2001). eLearning based on the Semantic Web. Proc. of WebNet2001, World Conference on the WWW and Internet, Florida, US.

Willis, J. (1996). A Framework for Task-Based Learning. Harlow, Addison Wesley Longman Ltd.

Zhu, J., A. L. Goncalves, et al. (2005). Mining Web Data for Competency Management. Proc. of the IEEE/WIC/ACM Conf. on Web Intelligence, France.