D3.3.7 Evaluation and current state of ASPL-v2

Coordinator: Martin Dzbor (OU)

Other authors: Dnyanesh Rajpathak (OU)

Abstract.
EU-IST Network of Excellence (NoE) IST-2004-507482 KWEB
Deliverable D3.3.7 (WP3.3)

Work reported in this deliverable focused mainly on the learner's interaction with resources on the Semantic Web; in particular with the semi-structured data that can be exposed to the user via domain-specific inference templates. We assessed this capability of the service-based ASPL-v2 framework in terms of assisting users with interpreting connections in the academic domain; for example, filtering leading scientists, recognizing communities of practice, or associating research topics and issues with particular publication outlets. The outcomes of a user-based study were reported, and the ASPL-v2 was found to outperform other tools – including the generic search engine aggregator Ask and semi-specialized Google Scholar.

Document Identifier:	KWEB/2006/D3.3.7/v1.0
Class Deliverable:	KWEB EU-IST-2004-507482
Version:	v1.0
Date:	November 20, 2007
State:	Draft
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

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

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

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

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

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

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

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

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

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

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

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

We gratefully acknowledge their assistance contribution to the discussion regarding ASPL scenarios, functionality and tool usability. We also acknowledge the effort of Joerg Diederich (L3S) and Jerome Euzenat (INRIA) in their roles of quality assessors.

Changes

Version	Date	Change Author	Changes
0.1	01-10-2007	Dnyanesh Rajpathak	First draft of the report
0.2	11-10-2007	Martin Dzbor, Dnyanesh Rajpathak	Update on user evaluation
0.3	19-10-2007	Martin Dzbor	Positioning analysis
0.4	29-10-2007	Martin Dzbor	Revision, findings, etc.
0.5	16-11-2007	Jerome Euzenat	QA, review and text revision
1.0	20-11-2007	Martin Dzbor	Addressing comments from QA

Executive Summary

Work reported in this deliverable focused mainly on the learner's interaction with resources on the Semantic Web; in particular with the semi-structured data that can be exposed to the user via domain-specific inference templates. We assessed this capability of the service-based ASPL-v2 framework in terms of assisting users with interpreting connections in the academic domain; for example, filtering leading scientists, recognizing communities of practice, or associating research topics and issues with particular publication outlets. The outcomes of a user-based study were reported, and the ASPL-v2 was found to outperform other tools – including the generic search engine aggregator Ask and semi-specialized Google Scholar.

The current report is set on the backstage provided by the previous report (D3.3.6), which presented and concluded with a theoretical justification of our decision to re-engineer and almost completely revise the suite of learning services, so that more interaction is offered to the user, alongside with novel, semantically driven inferences. One of the outstanding tasks in the previous report was to assess whether the re-engineering actually worked. In particular two aspects came up in this report:

(i) finalizing the design and implementation of learning services for the revised ASPL-v2 framework, and

(ii) carrying out a comparative assessment of ASPL-v2 vis-à-vis other tools that have a similar scope and may be commonly used by the users

Generally speaking, it was observed that ASPL with the new and improved learning services (as set out by point (i) above) turned out to form an appropriate tool when the users were expected to carry out analysis and synthesis on the academic data set. In other words, the outcome of point (ii) above is a two-pronged methodology we used to assess and position tools performing a certain class of tasks – in our case, search and data retrieval tools. The outcomes of this research activity are reported in the following style: First, we present the *positioning analysis* the ASPL/DBLP++ combination with respect to its nearer or more distant competitors, and second, we present the outcomes of *assessing the user-centred effectiveness* of the ASPL-v2 re-engineering, or better the effectiveness, usefulness, pluses and shortcomings of the re-engineered learning services

In a nutshell, the study has shown that the users preferred to use ASPL to perform the search in three out of five tasks. ASPL was observed to perform better than Google Scholar on three tasks that required a degree of synthetic or analytic thinking. However, we also point to some identified shortcomings, which can be addressed in the remaining life of project; in particular, the capabilities of ASPL to handle the incompleteness need to be looked at and possibly improved. If this is accomplished, the tool is likely to gain a good support within a narrow domain of academic literature and interpretation of community relationships.

Contents

1 Overview and context of the evaluations

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.

ASPL (Advanced Semantic Platform for Learning) 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 was originally designed and prototyped as a Magpie-based application (see Figure 1 for reference), and it has been available as a plug-in for a number of web 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 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 domain ontologies.

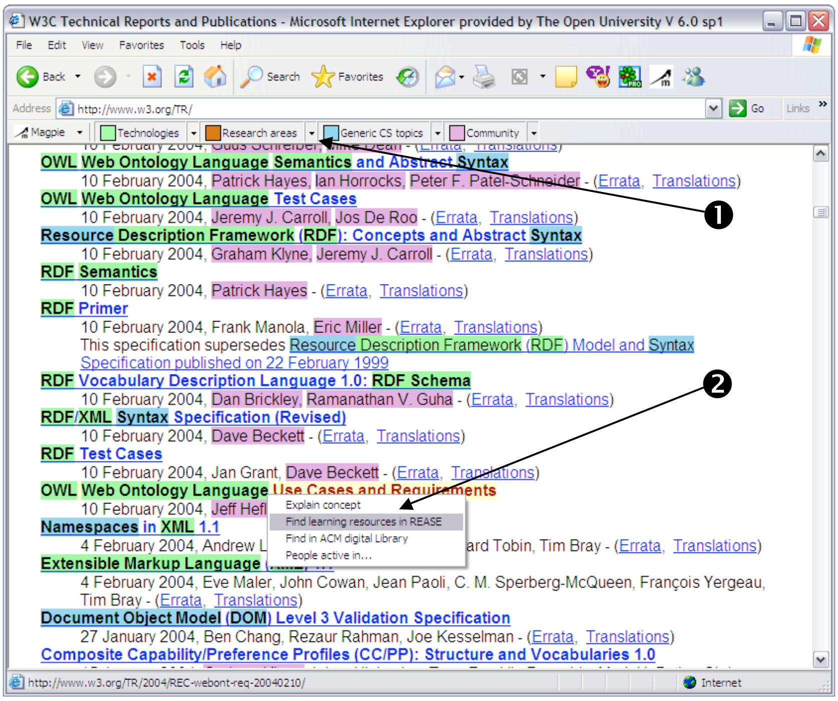


Figure 1. A screenshot showing a Magpie-enhanced web browser and a web page annotated using the lexicon derived for the Semantic Web domain; pointer ❶ shows a user-selected ontology with several abstract categories of identifiable concepts (highlighted in different colours), and pointer ❷ shows a sample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

[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

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.

The deliverable D3.3.6 presented the theoretical underpinning of the process we intended to pursue to augment the prototype of the advanced semantic platform for learning (ASPL-v2). Rather than merely describing the new version, we also mentioned the rationale for re-engineering the application in particular ways and directions. The current versions of ASPL have two-pronged functionality: (i) ASPL is still 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.; and (ii) ASPL also acts as a lightweight, web-based search tool or a front end its primary data store (the DBLP++ KB).

1.1 Modifications to ASPL learning services

In this report we get back to the key argument from the past reports that for the purposes of learning the interactions between a user/learner are more than mere annotation of web pages, retrieval and subsequent browsing of semantic metadata. In order to apply semantic knowledge, the re-designed version of ASPL supports a more exploratory approach to interacting with distributed learning resources, focusing on creating interpretative pathways rather than merely retrieving simple data.

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). Together, the two helped us to introduce support for analytic and synthetic learning tasks, and the value of our approach has been corroborated in a user-based study – majority of users liked the way ASPL-v2 helped them to navigate through the problem space in a structured way, which they could mimic and thus develop a skill in analyzing academic data.

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.

The previous report (D3.3.6) concluded with a theoretical justification of our decision to re-engineer and almost completely revise the suite of learning services, so that more interaction is offered to the user, alongside with novel, semantically driven inferences. One of the outstanding tasks in the previous report was to assess whether the re-engineering actually worked. In other words, in the past period we focused more resources on two aspects:

(i) finalizing the design and implementation of learning services for the revised ASPL-v2 framework, and
(ii) carrying out a comparative assessment of ASPL-v2 vis-à-vis other tools that have a similar scope and may be commonly used by the users

In terms of the former point, L3S has put a substantial effort into their DBLP++ tool. This tool essentially draws upon the well-known database of publications in computer science, and in its interactive form, it offers the user a rich, faceted interface to access the content of DBLP. Nonetheless, this rich user interface may be somewhat daunting, so it was decided to pursue also a parallel approach promoting a lightweight user interface via the ASPL platform.

The ASPL platform is essentially about associating web services with the concepts and instances from a particular ontology, which is of interest to the user. Thus, OU and L3S developed a suite of web service end points for the DBLP++ data set, and these were later complemented with a user-friendly front end – simple, Google-style user interface for querying the content of DBLP and also for making knowledge-level inferences and connection interpretations. In particular, the following web services were exposed from the DBLP data set:

- *Person's publications and interests* … a combination of a simple data retrieval (of publications) with an interpretative inference based on the publication keywords, Semantic Web Topic Hierarchy matches, etc.
- *Person's interests* … an interpretative inference based on the occurrence of keywords and phrases (also from Semantic Web Topic Hierarchy)
- *Person's community characteristics* … an interpretative inference based on the co-occurrence of co-authors, keywords and themes allowing generalizations from the individual nodes (researchers) to their collections (communities)
- *Person's co-authors and communities* … a combination of the retrieval function with an interpretative function as described above
- *Leading experts on topic* … an interpretative inference based on the occurrence of certain phrases within individuals' profiles combined with a statistics
- *Main publication outlets for topic* … an interpretative inference allowing the user to generalize from single nodes (publications and authors) to their collections (journals, conferences, etc.)

1.2 Report outline

ASPL has been evaluated by a group of users in late 2005, and as a result of the evaluation we implemented several improvements in its learning services to address the identified shortcomings of the first version of the system. This report largely focuses on the aspect of carrying out a comparative assessment of ASPL-v2 vis-à-vis other tools that have a similar scope and may be commonly used by the users

In particular, we start by setting out the comparative analysis part in Section 2. Then we present the findings of this positioning analysis and specific positioning diagrams in Section 3. We then dedicate Section 4 with a summary of the user-centred study, whose aim was to assess the effectiveness of the ASPL approach. We conclude with a brief overview of key findings, the key benefits of our two-pronged assessment strategy, and also present some near-future plans to address the weak aspects of the ASPL performance.

2 Feature comparison and evaluation

In this section we describe the first method for evaluating ASPL-v2 – in a direct feature comparison with several other tools that are available to the user for the purpose of searching, retrieving or otherwise accessing information on the Web. The objective of this study was to present an objective view of what functionalities and functions are actually available in the domain of information and knowledge search. Subsequently, this objective view enables us to position the ASPL-v2 against its (mostly indirect) competitors. These include more traditional ones, such as Google Scholar, but also some of the emerging more or less semantically aware tools, such as AquaLog [11].

In a nutshell, in this section we attempted to present a generic multi-dimensional space, which can be used to compare and contrast different tools with each other based on various functional aspects that are typically associated with search and information retrieval – searched content, context treatment, query formulation, query analysis, query output processing, and the use of learning mechanism. Because this multi-dimensional space is generic in nature it allows users to compare the existing tools with newly emerging ones, and, as a result, it provides a uniform basis to analyze the strengths and weaknesses of the new tool with respect to the existing ones.

This multi-dimensional space is flexible in nature; in the sense that if the users need to compare and contrast only specific functions associated with the heterogeneous tools, say, context treatment, they can safely perform such an evaluative scanning without having to consider other functions, which may not be relevant for their needs. Hence, for sake of user friendliness, we present partial projections of the multi-dimensional space as a series of bar graphs drawn around selected pairs of functional features.

Finally, this multi-dimensional positioning space also allows its users to introduce new evaluation functions, which may be beneficial for their specific purposes. We included in our scope functions that are interesting from the perspective of assessing the ASPL platform, but being careful not to bias the positioning analysis.

2.1 Determining the scope of the tools

Education, like many other disciplines, aims to take advantage of Web technologies 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, 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 learners and expects them to actively participate in the process. Let us explore here how different tools contribute to this challenge.

The evaluation criteria described here can primarily be used to determine different dimensions, which we may need to take into account in order to evaluate the search and question-answering tools. Generally speaking, tools we are interested in assessing and position in this first study take as an input a query submitted by a user. Then, they retrieve the best matching information and give it as an output, i.e. a result, which to a lesser or greater extent satisfies the requirements of a submitted query.

As a part of this evaluation focus, our main objective is to determine the functions that need to be taken into account while comparing and contrasting any two apparently heterogeneous tools with each other. For instance, some tools search only the textual data, whereas other tools specialize their search or information retrieval technique to image and audio data content. Nonetheless, it is interesting to find and describe dimensions for comparing them. In some other cases, there are tools that perform search only within a specific domain, which can be contrasted with those tools that are more generic in nature. As a result, differently scoped tools can be used to perform the search in different situations and will perform differently in these situations.

In order to compare and contrast such a set of different tools, it is important to establish evaluation dimensions, which take into account different functions that may be associated with the task of searching for information, so that a comprehensive but at the same time an unbiased evaluation between these tools can be performed.

2.2 Key evaluation features of the tools

Here, our main objective is to describe and to understand the different categories that are used in the subsequent analysis to assess and describe the scope of tools. This assessment space definition allows us to appreciate which part of this complex space these tools occupy. Generally speaking, we identify two broad categories that can be used to determine the position of the tools (these are described below in more detail):

(i) searched content and
(ii) capability of an intelligent (i.e. contextual) adaptation.

The rationale for using the "searched content" dimension is simple. As already mentioned in Section 2.1, different tools may have different focus (and hence different underlying infrastructure) in order to perform search over different data sets and data types. Under the "searched content" criterion we distinguish these sub-types:

- **Single-domain tools**
  This category of the searched content looks at the input and includes those techniques that take as an input a textual query submitted by the users, and after analyzing the query they look for the information only within a specific domain. A typical representative here is ASPL, which processes the domain of academic literature and scientists.
- **Multi-domain tools**
  This category does not explicitly subscribe to any specific domain, but is capable of performing the search in different domains. Yet, it is not fully generic, only multi-domain. A typical example is Hakia[2] that uses natural language technology to analyze the submitted query and can be used to search in medical or legal domains.
- **Generic (no-domain) tools**

[2] www.hakia.com

The last category includes generic tools that do not specialize in any particular domain, of which a typical representative is e.g. the omnipresent Google[3] that tends to be generic and as a result it can be used to perform search over heterogeneous domains.

Within the "searched content" family we can also consider the type of the searched content:

- **Raw textual data**
  Here we include all those tools that can process any text and do not make use of the text structure or any other form of topology, semantics or ordering between its constituting items.
- **Structured (but textual) data**
  In this category are those tools that still focus on text, but rather than working with documents and web pages, concentrate on structured data sets or make use of the document structure to interpret the text within the documents.
- **Media and multimedia (in principle)**
  The last category includes techniques that work with multimedia data and can, in principle, search for information in different media formats (video, sound, picture, etc.)

The second family of dimensions "contextual adaptation" also has several sub-types. Let us elaborate on the core reason for including this dimension in our analysis first. The notion of context [1, 2] when viewed in the tasks of searching and question-answering can be defined as an additional information or knowledge to modify, to analyze or to interpret the query submitted by the users in order to find the relevant results as an output [3, 4]. Context as such may be used at two different stages during the information gathering process:

- At the first stage while performing a query analysis the tools may identify **an implicit context**, which is part of the query submitted by the user. For instance, if the user submits a query, 'Publications of Author X on research topic Y?', the tool may contextualize the query and limit the scope of results to finding a set of publications in a specific discipline the user is known to work in.
- In the second category, the tools make use of an **explicit context** knowledge, which can be modelled and can be used by the tool to retrieve correct information for the submitted query. For instance, in order to find answers to the query 'Publications of Author X on research topic Y?', the tool may use ontology or another model that contains and establishes formal relationship between e.g. different research sub-topics, and thus enables some form of query expansion.

In the latter situation, we have determined three different treatments of context that can be used to determine the scope of the tools: deployment of the user specific context, context identification, capture and acquisition, and finally, the way context is modelled as a part of tool architecture. These treatments are described below in a detail.

[3] www.google.com

User specific context.
Here, our main aim is to assess how the tools make use of the keywords, phrases, or sentences that are submitted by the users as a query. In particular, we are interested in how these structural elements of the query help to formalize the context. The query, in this case, acts as the main source of context that can be used by the tool to comprehend the query. For instance, if a user submits a query such as ‘a batsman with the highest run rate in test matches’, then in such a case, a tool may make use of various keywords from the query like *batsman*, *test match*, and *highest run rate* to establish a context firstly to identify that a user is interested in finding ‘a (specific) person’ in ‘a test (cricket) match’ (i.e. a specific type of sport event), who has the ‘highest run rate’ (i.e. a statistic associated with this sport). By making use of this “query sugar” as a context knowledge a tool could retrieve the correct information, i.e. a text document with a mention of, Sir Donald Bradman with run rate average of 99.94.

Context identification.
The main aspect of the context identification or acquisition feature is firstly to analyze whether the tool is capable of distinguishing between multiple occurrences of a term or a phrase in a query submitted by the users to retrieve the necessary information. Having identified the occurrences of a term or a phrase, it is then important that the tool identifies the situations, in which such a term or a phrase appears in the data set. For example, if term ‘Magpie’ appears in the query submitted by a user, then the context identification feature may be able to distinguish that the term ‘Magpie’ can occur either in the context that is about the ‘birds’ or alternatively the same term can also occur in the context of ‘tools and technologies (for the Semantic Web)’.

Moreover, having identified the context in which a term or a phrase is specified, the tool may then be able to classify the retrieved results in appropriate clusters by taking into account additional information gleaned from the data set. Unlike in the previous situation, context appears here at the end of the query processing, i.e. to display the results in somewhat intelligent manner.

Modelling context.
As described earlier, typically the search or question-answering tools subscribe to some background (or contextual) knowledge to find the material relevant to the query submitted by the users. In some cases, formal ontologies or other formal schemas are engineered in order to represent the context associated with a specific domain completely separately from the user, the actual query, and the data set. A range of different modeling frameworks can be used; e.g. topic maps are the ISO/IEC 13250:2003 standard for representing and interchanging the knowledge with a specific emphasis to find the required information. To this end, the main evaluation criterion here is to compare and contrast how the context is modelled in systems’ architecture of these different tools.

2.3 Phases of the user interaction

In the previous section we introduced and briefly described the families of key features that we have taken into account while evaluating the capabilities of the heterogeneous search and question-answering tools. This type of evaluation is of course important because it helps to determine the overall scope of the tools. In other words, it allows the users to appreciate what type of search or services the tools are capable of performing and what purposes they fit.

In addition, however, it is equally important to evaluate how the different tools make use of different techniques to formulate a query, to process and analyze it, or the way they represent the results and present them to the users in an appropriate (user-friendly presumably) way. In the following text we describe important phases that we have focused on while evaluating the performance and functionality of the tools.

Query formulation.

The most difficult part of the search strategy is to allow users to formulate a query in a sufficiently robust yet flexible manner. This is an issue particularly for the question-answering tools because these allow the users to submit their queries in a natural language. To this end, it is important to provide a certain level of flexibility for the users to submit their queries in the ways that are suitable for them. This can be achieved in many ways; examples include making use of controlled vs free keywords or phrases, or making use of the complete natural language sentences.

While assessing the query formulation feature of the tools our aim is to consider the level of flexibility offered to the users. We are interested in learning how different tools approach this task; e.g. specific *types* of keywords or keyword modifiers that are submitted as a part of a query – as in “John Smith language:en site:.net”. Another example might be the use of languages primitives such as what, where, when, how, why – as in “Who are the members of the band that performs in Wembley in July”. Yet another possible value in this dimension might the capability of handling and interpreting the punctuation or other grammatical nuances.

Query analysis.

An important part of processing the query is to analyze it correctly – as ‘intended’ by the users. The main purpose of the query analysis process is to interpret a query so that an appropriate information can be retrieved and formalized. In this dimension we look at how tools treat this phase of searching. In some cases the analysis process may include a fine-tuning of a query in order to get the correct (or better) information. In other cases, tools may support some form of query expansion or reduction.

Generally speaking, a performance of any tool does almost invariably depend on how that tool performs query analysis. Hence, our aim here is to assess whether the tool analyzes the submitted query correctly – in terms of a correct set of data can be searched to retrieve the appropriate information. It is also useful to look at the amount of data that a tool needs to process before it returns a result. Next, this dimension also includes the appreciation of the processing speed during the analysis of the submitted queries. Finally we would also like to consider whether the tool handles fine-tuning of a query in scenarios where the user-specified query fails to retrieve the information.

Since this dimension is rather rich in content, below we describe different aspects that are associated with the query formulation process:

- ***Query expansion***. This is a process of reformulating a seed query to improve the retrieval performance of a tool [6, 7]. In the case of question-answering or web-based search, query expansion typically involves evaluating the user’s input (i.e. making sense of what words are typed in), identifying other types of relevant data, and finally expanding the search query with some relevant knowledge. The situation with the query expansion may have several shapes.

In some cases, the originally formulated query is sufficient; we see this as an empty expansion (the simplest form). In other cases the tool may need to find the synonyms to the keywords appearing in the query. Here, it is interesting to assess whether the tool interprets synonyms correctly, and also it is important to consider whether the tool is capable of relating different morphological forms of words by stemming each keyword in the search query.

Another type of query expansion may be the tool's capability to handle spelling errors automatically and then perform search by using the corrected form of a query. Finally, query expansion may be accomplished automatically (i.e. only by the tool), semi-automatically (i.e. by the tool with some minor input from the user), manually (i.e. by the user), and what form of interaction with the users is preferred.

A classic measure that can be used to describe the utility of the query expansion technique would be (the change in) recall [8]. In case of query expansion acting as a constraining feature on the search space, it is also useful to consider (the change in) precision. However, these two measures were investigated quite extensively, so we will only touch on them briefly.

- ***Query Extraction***. One can imagine that in some cases the users would like to break the confines of a keyword-based search, and instead make use of the natural language sentences. This often happens even in tools that are not inherently equipped with NLP algorithms, so it is interesting to assess how the actual query is extracted from the 'query sugar' (i.e. a sentence with possibly some less useful terms).

  In order to handle such queries the tools should (ideally) take into account both the structure and the nuances of natural language, which help them to extract information from the submitted query more meaningfully. To this end, firstly, we look at whether different tools are equipped to handle the structure and the nuances of the natural language. In cases where the tools handle the natural language we consider to what extent these tools extract the queries correctly and retrieve the correct answers.

- ***Query term co-relation***. Here, the main principle of this functional feature, when occurring in search or question-answering tasks, is to establish a relationship between the keywords submitted by the users, and to use the knowledge of relationship(s) to retrieve more meaningful results.

  For instance, if the user performs a search in a document set that consists of heterogeneous topics, and submits query with keywords 'John Smith ales' then this query may be associated with different domains, and therefore has different semantic neighbourhoods. In domain independent set of documents, the query "john smith" may be associated with breweries, beers, etc. In the

Semantic Web domain, the same query may create different associations, including "john smith" as an author, a name of algorithm, etc.

What we are interested in is whether a tool is capable of recognizing a correct correlation between the individual elements of the query. Also, we look at whether the tool manages to establish the co-relation between the query terms and how it makes use of this additional knowledge.

Query Output.
After other technical aspects mentioned earlier, such as query formulation and query analysis, which are associated with search or question-answering tools, it is also important for the tools to represent and present an output to the submitted query in such a way that it helps users in their tasks and activities. Here, our aim is to look at four key functional capabilities: ranking mechanism for the query output, interaction with the end-users, post-processing of the query output, and explanation or definition given by the tool for the (parts of an) output. These capabilities are briefly discussed below in more detail.

- ***Ranking mechanism***. Search or question and answering tools typically retrieve several results for any query submitted by the users. Typically, most relevant results tend to appear at the top of the list of retrieved results, while lower down in the list the results get less relevant. This ordering or ranking mechanism helps users to focus their attention on the most appropriate set of results quickly.

  To this end, our aim is to check whether the tools we are comparing make use of any kind of ranking mechanism to help sorting the results in some systematic manner. Moreover, our aim is also to look at the performance of the ranking mechanisms (if used at all) to see that they rank the results correctly.

- ***Interaction with the end users***. In many cases the presentation of query results happens simply by gathering an opinion on structuring the results retrieved by the tool from the users. For instance, if a user submits an open keyword query 'Magpie', which yields many hits, then the tool may consider the nature of this query and it may provide multiple connotations associated with.

  In this case, the default case might relate to 'magpie as a bird', so if the user meant something else, they may need to expand the query either by manually typing another keyword (say, 'project' or 'plugin') or choosing from the cluster – if any such are provided by the tool.

  Anyway, in such situations it may be useful to interact with the users to get an opinion about which sources of information are more relevant, to choose and/or to provide input. This preference can be used to filter out the data sets or the documents that are not relevant for the further search, but also it can be used for constructing the user profile, which in turn, might be reused to pre-structured future queries and/or to act as an implicit context.

- ***Post-processing of a query output***. The post-processing of a query output is an important function of the tools particularly for the queries, which work with mixed sets of result or data, because it helps users to interpret the results more easily and in a self-explanatory fashion. Below we describe different techniques of the query post-processing feature that may be associated with the tools that we looked at.

First, we may take into account whether the retrieved results are related to each other and then classify them in an appropriate category. For instance, if a user submits query 'technologies and people and "Semantic Web"', then for this query, a tool may retrieve results such as RDF, OWL, Tim Berners-Lee, Jim Hendler, Protégé, etc. Some tools may be capable of classifying RDF and OWL as core technologies, Protégé and Magpie as applications/tools associated with the Semantic Web, while Tim Berners-Lee and Jim Hendler are scientists associated with the Semantic Web.

Second, the classification may be accompanied by an appropriate *visual categorization*, such as different colour schemas, font sizes etc. to differentiate between different categories of query results. Here, our aim is to see whether the visual categorization used by the tool is self-explanatory and consistent. Moreover, it is also important to check whether the visual categorization used by a tool is correct and not ambiguous in nature.

A third strategy may be that a tool provides a quick summary of the retrieved results. Here it is interesting to observe if the tool highlights key terms associated with the results, and also summarizes important issues discussed in the textual description(s) associated with the results.

- ***Explanation or definition of a query output***. The main aim of assessing this function of the tools is to provide an explanation or definition for the concepts, named entities, links, etc. that appear in the retrieved results and which may have a well-defined meaning in principle. For instance, if a tool retrieves concepts, such as *RDF*, *XML* or *OWL*, as a result it may provide an explanation of these terms or their examples. Below are some scenarios that will be considered for evaluating this feature.

  For instance, we may take into account whether the retrieved results are related to each other and then classify them in an appropriate category. For instance, if a user submits query 'technologies and people and "Semantic Web"', then for this query, a tool may retrieve results such as RDF, OWL, Tim Berners-Lee, Jim Hendler, Protégé, etc. Some tools may be capable of classifying RDF and OWL as core technologies, Protégé and Magpie as applications/tools associated with the Semantic Web, while Tim Berners-Lee and Jim Hendler are scientists associated with the Semantic Web.

  One can obviously assess whether an explanation provided by the dictionary- or glossary-like service is relevant and conceptually correct with respect to the subscribed domain. In addition to quality, a complementary measure may be the level of detail provided by the tool for explaining or defining technical concepts or terms that appear in the set of results.

Learning Mechanism.

This next category of assessment criteria considers a useful feature associated with some tools, which allows them to learn when faced with new queries. Frequently, this mechanism is used to adapt quickly to the queries that show some degree of similarity or re-occurrence to the queries processed in the past [9]. Learning mechanisms help to improve the performance of a tool over the time because the tool learns to deal with the queries that were not processed before. More importantly, learning mechanisms help to improve the performance of tools for a certain type of domains because the users may provide a feedback that helps a tool in capturing and establishing an association between the submitted query and the (preferred set of) results that are retrieved for this specific query/context/user.

3 Findings from the comparative tool study

Having described the features that can be taken into account in order to objectively assess the performance and the functional scope of the tools we make use of these features to compare and contrast the performance of several specific tools.

In the subsequent section we will present a summary of findings for the following tools and applications:

- Hakia (www.hakia.com): a search engine specializing in medical, financial, scientific, and legal literature, available online as a web application
- Ask.com (www.ask.com): a generic search engine, available online
- AskMeNow (www.askmenow.com): a search engine specializing in the telecommunication domain, available as a web application
- TextDigger (www.textdigger.com): a generic search engine, available online
- Google (www.google.com): a generic (and probably the most widely used) search engine
- Blinkx (www.blinkx.com): a search tool specializing in visual media
- Aqualog (kmi.open.ac.uk/technologies/aqualog [11]): question answering tool for NLP sentences, available as an applet
- Precise ([12]): a question-answering tool for multiple domains, analyzed from literature
- Masque/SQL ([13]): NLP query interface to databases, analyzed from the publicly available literature
- ILQUA ([14]): generic question answering tool, analyzed from literature
- ASPL/DBLP++ (neon-project.org/aspl-v2): domain oriented knowledge search and interpretation engine, available via a browser plugin and online

The comparison of the above tools facilitates a practical case study on how the function-centric features discussed in the previous sections can be used to map the space within which our application operates. This type of evaluation can be treated as an objective, user independent evaluation of the tools, but later we will demonstrate how the performance of the selected tools can be evaluated by the user specific approach.

3.1 Positioning tools on searched content

With respect to this family of dimensions, we compared the tools on their coverage of one specific domain, several strictly defined domains or generic coverage of any domain, in principle. We selected the tools so that we had almost evenly represented the generic search engines as well as the domain-specific ones. Although, during the analysis we found that some domain-specific tools are in fact shells that can be fit or instantiated with an appropriate domain model to specialize them. Nevertheless, where only one such model has been developed, described and tested, we considered the tool to be specializing on a single domain – purely from pragmatic point of view.

In terms of the type of searched content, the distribution was even wider. Although many tools claimed to work with text, only some were in fact capable to process text in its different forms. Hence, we distinguished between the capability to search in the raw text (such as e.g. present in the web page files marked up using HTML) and rich formats (such as Adobe PDF, Microsoft DOC or PPT). From the tested tools only three searched in the document space in general, additional 2 supported search in the space of text documents and text files. The majority of search engines participating in the study were in fact front ends to particular databases (or in two cases, knowledge bases). These data stores usually contained a pre-modelled view of the problem space a particular tool targeted; hence, these tools are more akin to information retrieval applications rather than tools for finding information in an open space.

Although much research is devoted to finding information on multimedia content, only two tools were actually supporting this format of data in a reasonably testable manner. In a similar position are semantic data stores and knowledge bases with respect to standard databases. There are some research prototypes of semantic search and knowledge access applications, but these tended to be relatively small-scale for a comparison with larger search tools. However, this is likely to change in the near future, as shortly after concluding the study we became aware of a large-scale effort related to the production of RDF-represented data in the context of DBPedia.org and OpenLink.org initiatives – these new offerings may slightly help to perhaps distinguish applications using knowledge bases as a separate and functionally comparable category in some future studies.

Below, in Figure 2 we visually summarize the space of tools in the searched content dimension and position our ASPL/DBLP application in this space.

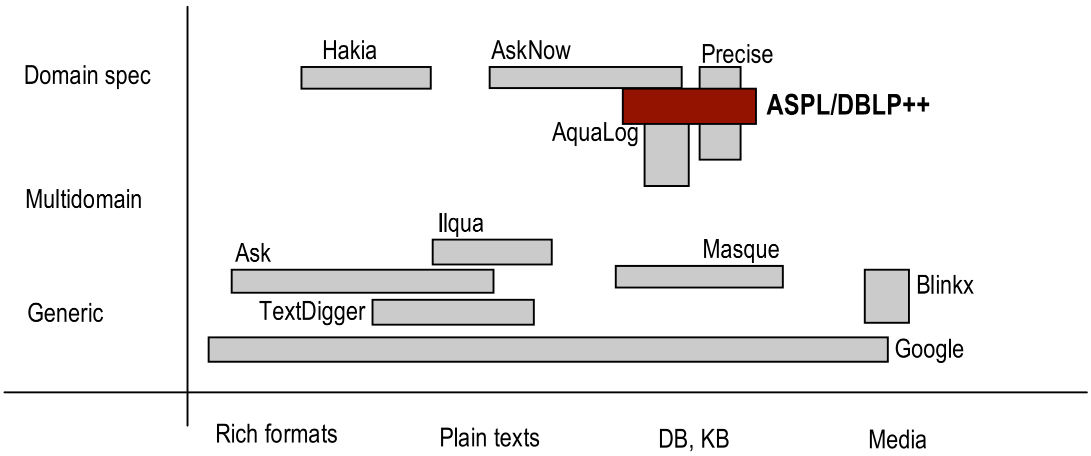


Figure 2. Distribution of tools on searched content dimensions

As shown in the figure, ASPL/DBLP++ belongs among the specialized tools that cover fairly well one specific domain, but are in principle applicable to other domains. It focuses on exposing its underlying databases *and* knowledge bases by means of a controlled set of keywords. With this respect, the closest tools are those from the question-answering domain (AquaLog, Precise, Masque). The only search tools that cover databases to some limited extent are AskNow and Google. However, neither of them makes use of knowledge bases (e.g. RDF-annotated datasets).

3.2 Positioning tools on context treatment

In the dimension of making use of context, contextual factors and using some underlying context models, the situation is varied. First, only three tools were observed to explicitly associate context with a formal ontology the user could (in principle) subscribe to. A majority of applications (7) considered context on the level of co-occurrence patterns among the query terms and/or keywords, and they treated the context as an implicit modifier of the 'core query'. Natural language technology was observed to play role in 3 tools, but only one of them relied on an ontological grounding for the interpretation of language constructs identified in the user queries.

We believe it is safe to state that the level of contextual support is better in the case of question answering tools than in search engines. This is likely related to the fact that the query for question answering systems takes as input a sentence that very closely resembles the natural language. Hence, these applications make use of rich sentence structuring in the first run and many are also linked to services like WordNet for bringing in synonyms, homonyms, etc. Since WordNet itself already contains some form of semi-formal expression of semantically related terms, these connections can be exploited as a context for the queries.

We found surprisingly little presence of user-specific context being taken in account – apart from ASPL/DBLP that treated this aspect in its academic task-specific composition of results and interpretation of some connections between the result items, only Google was to some extent exhibiting this feature (e.g. in terms of its defaulting to searching the Web 'within a given country' as opposed to the US-centric nature of many other tools). One reason for the lack of this functionality is its relation to authenticating the user in and then logging his or her actions.

First, many users may not be willing to personalize their search in such a direct manner, and second, even if they do log in, there are currently no immediate benefits observable in terms of a changed behaviour of the tools. Hence, this seems to be an area where (i) there is scope for a product, and (ii) ontologies may provide a more meaningful targeting of user context so that the user privacy is not violated, yet some additional filtering can be enabled.

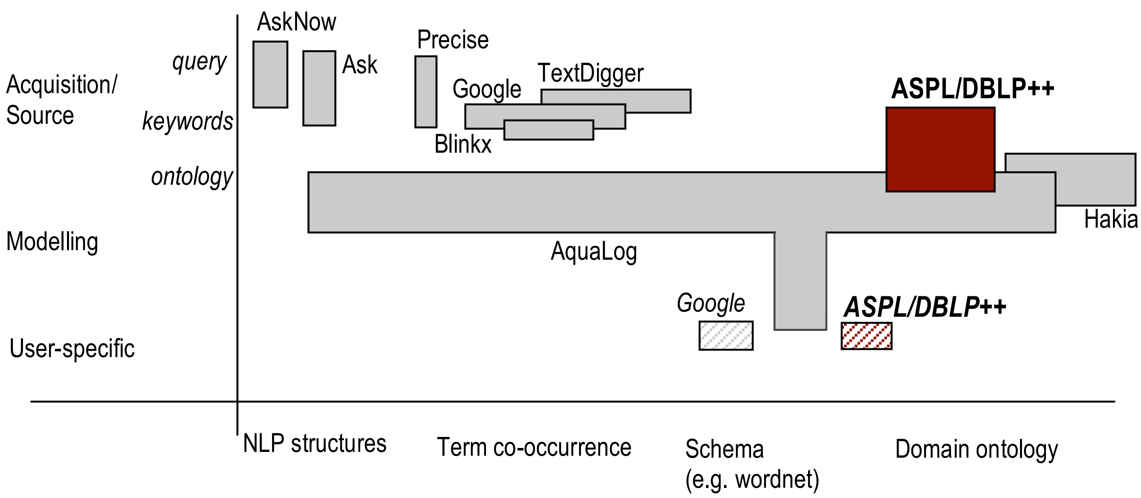


Figure 3. Distribution and positioning of tools on context treatment

In Figure 3 we show how different tools scored on the functionalities in the context-related dimensions. On the horizontal axis, the first two values (i.e. 'NLP structures' and 'Term co-occurrence') represent two common ways of treating the context implicitly; whereas the values 'Schema' and 'Ontology' are forms of explicit use of a context. As can be seen, only one tool covers both possible forms, implicit *and* explicit – AquaLog (thanks to its merger of NLP with ontological schemas) for query disambiguation. The other tools tend to sit at either end of the spectrum, with most of them treating fairly simple aspects of context in the form of keyword co-occurrence. Ontologies and more formal models of contextual factors are currently represented and used fairly sparsely.

AquaLog has also another feature where it explicitly attempts to capture at least a part of the user-specific context, when it remembers ('learns') what the user meant by his or her question – in case the user prefers one disambiguation schema over another.

In this dimension, we could not ascertain the role of context in Masque/SQL and ILQUA tools, so these two were omitted from the positioning figure. Nevertheless, in terms of positioning, ASPL/DBLP++ seems to occupy the section, which is fairly sparsely populated and covered by its larger competitors. Perhaps to improve this standing, it might be useful to consider adding some explicit context modelling capabilities and/or capabilities to acquire user-specific context. This can be done e.g. by exposing some of the data interpretation patterns and templates to the user, so that s/he can used them for filtering or biasing the queries (e.g. finding experts on a given topic *satisfying a particular user-defined condition* – e.g. on their standing in the community, etc.)

3.3 Positioning tools on query formulation

Most of the tools we surveyed in this study support a keyword-based formulation of the queries (in total 7) and most of these extend the support of primitive keywords (i.e. one-word terms such as 'service' or 'Madrid') to simple phrases (i.e. multi-term queries such as 'knowledge management' or 'baroque music'). Although many tools argue about their support for natural language (NL) questions, the majority actually enables the user to state his or her query in the semi-natural fragment or phrase.

Hence, as can be expected it was mostly the question answering tools that actually scored on this aspect. From the search engines, we acknowledged limited support to NL queries in tools like Google or Ask, which usually entails matching the entire pattern of keywords against the data set, and cannot be considered a true formulation of a query in NL.

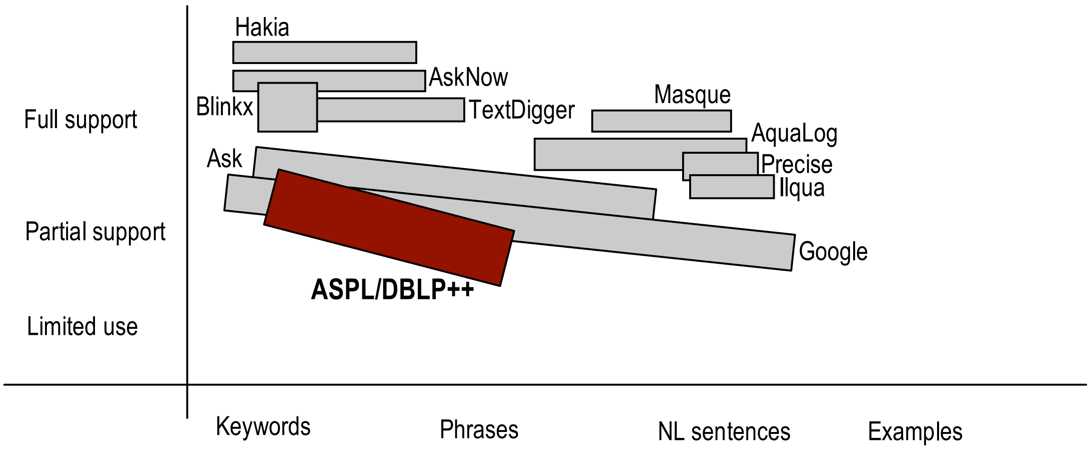


Figure 4. Positioning of tools in terms of query formulation

As shown in Figure 4, none of the tested tools supported asking questions by example. This is a fairly recent development that made some inroads in the specialist domains, but it seems not yet on a larger and more generic scale.

The position of ASPL/DBLP++ in this space is roughly aligned with standard search engines coverage – obviously with the caveat of being domain-centric and thus being able to return meaningful results to a potentially restricted set of keywords. Our tool ASPL/DBLP++ showed a mixed performance with respect to keywords and phrases. While on keywords it was fairly reliable and provided largely correct outputs, but there were some issues with the capability to process incomplete phrases.

Here, the rich part of the pair (DBLP++) showed greater flexibility as it was able to explicitly distinguish the keyword search from the phrase search – allowing the user to set this as an additional option. On the other hand, the lightweight ASPL front end did not make this differentiation, but suffered somewhat due to expecting complete phrases for some of its services (e.g. for finding publications it was necessary to provide full name, say, 'Joe Bloggs', which may be restrictive for users not remembering it). However, we should be fair and also acknowledge that the ASPL front end was merely an exposure of the underlying semantic services, which were tied to its plug-in version and that has taken care of recognizing partial names…

Nonetheless, there is a clear scope for improving ASPL positioning in this space by making its keyword- and phrase-based query formulation more flexible – perhaps by linking it to ontologies similar to those of AquaLog. This would not only make the procedure of query formulation more flexible, but it would also make use of the distributed nature of the infrastructure on which ASPL/DBLP++ relies. In other words, an incomplete query may be complemented by means of delegating the task to a specialized tool, such as SemSearch [15].

3.4 Positioning tools on query analysis

In this section we look at three items and present them on a one on one basis. First, we consider the capability of extracting the actual 'core query' from all the 'syntactic sugar' that users often submit in the initial query. Secondly, we explore the strategies used to expand and modify queries. In both cases, we set these functions against the initiative; i.e. whether the activities are done automatically or manually.

3.4.1 Positioning tools on query extraction

In the context of query extraction, we observed, in principle, three broad approaches instantiating this analytic dimension. Tools were almost evenly spread around these three values, which shows they are mutually equal alternatives, rather than evolutionary stages in the query extraction task. Each strategy corresponds roughly to the way how a given tool supports query formulation (i.e. sentences vs. keywords).

The first strategy worked solely with keywords and terms already available as a part of the user's initial query. The principle of extracting some additional 'context' was limited to keyword restructuring, re-ordering or seemed to have relied on some form of term prioritization (e.g. based on the statistical popularity of the individual segments/terms of the query). Typical representatives of this strategy are traditional search engines (Google, Ask), but also our ASPL/DBLP++ exhibits this type of behaviour. Majority of tools implement this form of query analysis in a semi-automatic or manual style – i.e. offering a modification to the user but waiting for his or her explicit 'approval'.

The second strategy relied on the use of some schemas or templates to which the user query was matched, which subsequently enabled the tools to extract the core items and distinguish them from the auxiliary ones. This capability was observable in the tools that made use of some form of structured knowledge in their infrastructure – be it ontologies (as in AquaLog or ASPL/DBLP++) or database schemas (as in Precise). Unlike the query extraction based on reshuffling the keywords (that was executed automatically but offered for an approval), the schema- or template-based techniques were usually applied automatically with no user approval opportunity.

Finally, the third strategy comprised the use of natural language processing power to identify and take advantage of the nuances of NL – to a lesser or greater extent. It is not surprising that this feature was strongly linked with the tools allowing NL input of the queries; i.e. most of the question answering tools scored in this section of the space. NL analysis was usually carried out automatically, but more advanced methods (e.g. those in AskMeNow or AquaLog) aimed to present different interpretations of the query and let the user choose (or explore). NL processing ranged from simple acknowledgement of connectors (such as non-words 'and', 'of', etc.) to more complex clause interpretation (i.e. semantic breakdown of 'who', 'which', 'how', etc. clauses of the original query). A usual intermediate step comprised some form of formal representation of the NL parsed query.

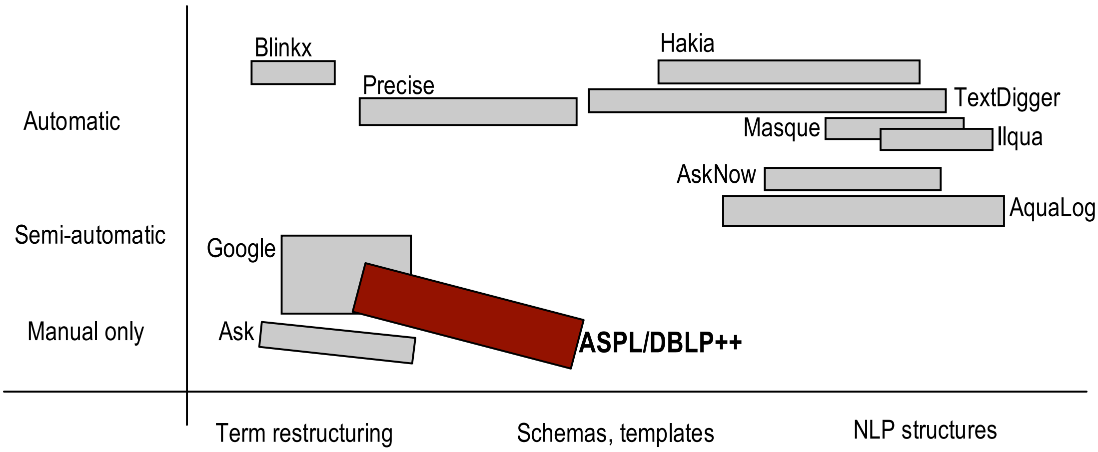


Figure 5. Distribution of tools on query extraction capabilities

As can be seen in Figure 5, there seems to be a certain degree of divide between the two polar treatments of the initiative – the tools either approach the query analysis fully automatically or rely (almost fully) on the user intervention – e.g. as in the case of Google or ASPL/DBLP++. From the pragmatic point of view, the approach used by some question answering and NL query processing tools is a good compromise: the tools has a go in the initial disambiguation of the possible senses of the query, shows these to the user as relevant interpretative paths, and the user chooses the path to interpret the query.

What makes this approach better, e.g. when compared to Google's term restructuring, is an explicit presentation of how the query has been interpreted – it is a kind of 'explanation of reasoning', an intermediate step in the query processing chain that helps the user to make sense of the extraction result (i.e. modified query).

As shown in the figure, ASPL/DBLP++ is positioned partially in this interactive, semi-automatic section of the problem solving space, which is a potential opportunity. The schemas are derived from ontologies and DB schemas, and (for known concepts) enable to extract a more appropriate (i.e. ontologically grounded) form of the used term (as in 'managing knowledge' → 'knowledge management'). Also, thanks to DBLP, the tool is partially capable of handing some variations in terms (as in 'Harmelen' → 'van Harmelen').

However, as already noted earlier, one weak point of ASPL/DBLP++ seems to be the capability to reuse ontological and database schemas to complete a partially formulated query. Especially, the lightweight ASPL front end relies on full names being queried (in the context of 'People' tab). If only surnames are given (which is fairly common in user queries), the tool shows gaps in terms of (not) proposing any possible completions. Since this feature is fairly common in other tools, it should be addressed in the near-term revisions of the ASPL functionality.

3.4.2 Positioning tools on query expansion

Query expansion is closely related to what has been discussed in the previous section; however, after a careful analysis of the tools we decided to treat it as a separate dimension for structuring our positioning space. The reason is that whereas the query extraction is more about playing with the existing query, the process of expansion is more about adding new information (e.g. terms, modifiers, etc.) to the original query. As such, query expansion is much more correlated with the capability to support some form of contextual adaptation than with the capability to comprehend the core of the query (as in the extraction-type of analysis).

In term of query expansion, we differentiated between the syntactic, structural and morphological modifications on one hand, and the semantic modifications on the other hand. The former strategy seemed to have used more statistically biased techniques (such as typical term correlations or popular term correlations) to make suggestions in terms of correcting or amending the original queries.

The latter strategy showed an attempt to make use of some background conceptual structures (e.g. concept maps or ontologies) to make suggestions that expanded the query. Thus, in our opinion, this strategy is more representative of what can be truly considered as query expansion, and, from our suite of tested tools, mostly those applications working with the contextually rich queries or media were observed as implementing query expansion. Typical representatives of the 'rich media' style include the tools that process the underlying texts using NLP techniques (e.g. the TextDigger algorithms); whereas the 'rich query' camp is typified by the NL equipped tools (i.e. those allowing the queries to be asked in NL or near NL).

In Figure 6, we show the summary of how different tools are distributed in terms of supporting query expansion. As can be seen, the majority of them are in the syntactic and statistical section of the space. Two tools that truly seem to support query expansion are AquaLog and TextDigger; however, we also included into this section those tools that make partial use of automatic term association based on their underlying schemas (rather than statistics) – these are shown using dashed bars on the right of the figure.

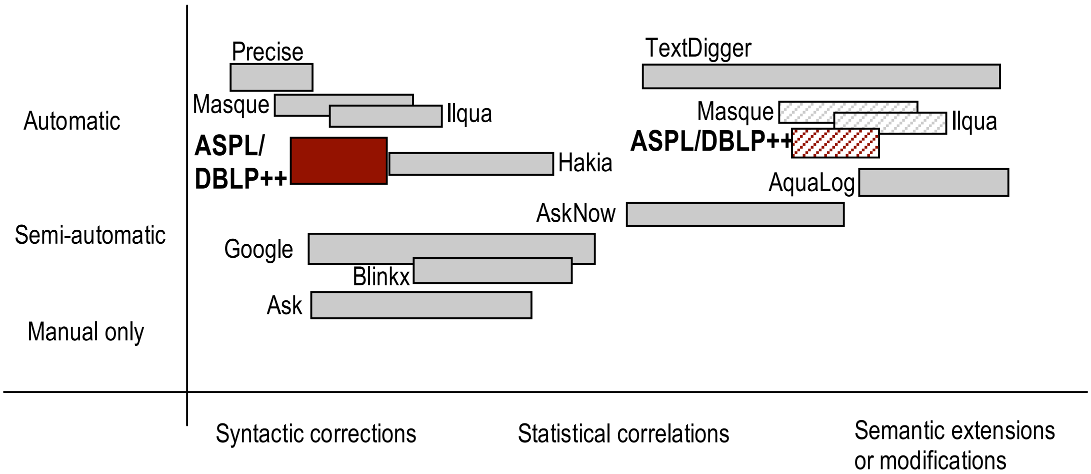


Figure 6. Positioning of tools in terms of query expansion functionality

Examples of this tentative support for semantic query expansion ranges from Masque's use of 'completion schemas', which enable it to propose (forms of) the terms that are expected to appear in the query. ILQUA has a dedication expansion component, which considers correlations – however, between the underlying concepts rather than directly between the terms of the query. ASPL/DBLP++ has some implicit support for expansion, e.g. in terms of faceting (DBLP++ core capability) or in terms of term-concept mappings of ASPL's ontological infrastructure. However, as we already mentioned in the previous discussion, it might be desirable to somewhat merge these two approaches and make the coupled ASPL/DBLP++ more consistent with respect to term completion (e.g. 'Smith' → 'Barry Smith' or 'James E. Smith' or 'Kate A. Smith', etc.)

3.5 Positioning tools on query output processing

In this section we look at three dimensions and their mutual, i.e. one to one interactions. First, we consider the capability of ranking or ordering the results of a user query using one of the common post-processing strategies. Secondly, we explore the status in terms of an explanatory power of different tools with respect to post-processing methods.

3.5.1 Positioning tools on query output ranking

First, we looked at the capabilities of the tools with respect to supporting some form of ranking the results using some explicit criteria. Among the ranking approaches we saw two broad types: the support for relevance-based ordering and the support for popularity/statistics based ordering. However, there were also tools where no obvious ranking schema could be recognized.

Most of the tools make an attempt on relevance-driven presentation of the results, but some of the larger-scale tools seem to rely more on the statistical models than the semantic relevance (a typical example is the popularity-driven PageRank algorithm underlying Google). The distribution of initiative is fairly evenly divided between fully automated approach allowing no user intervention whatsoever, automated with a possibility of user intervention and the reliance on manual user intervention.

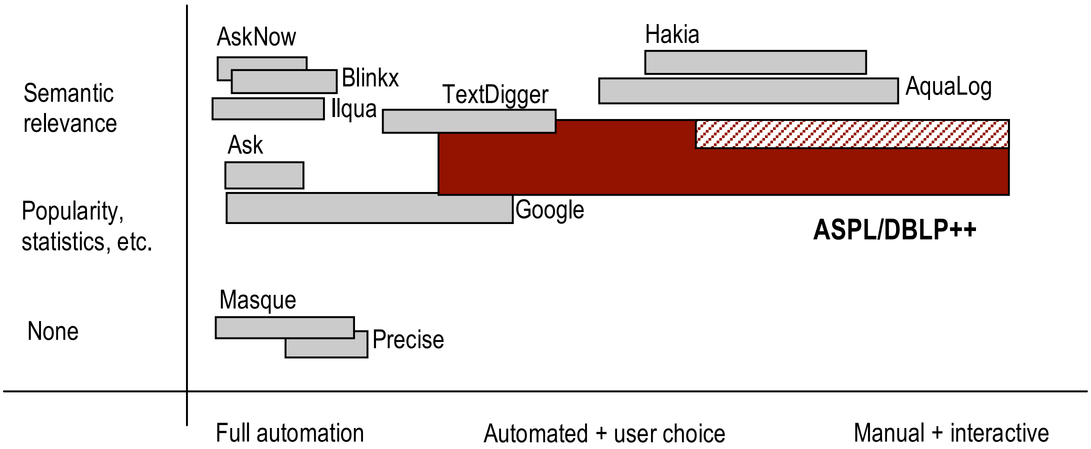


Figure 7. Positioning of tools on the result ranking capabilities

Figure 7 shows that most of the tools surveyed provide non-interactive presentation of their results or a presentation with a very limited interaction choice for the user. Here, ASPL/DBLP++ seems to functionally richer in terms of providing various options for result reorganization. For instance, its lightweight ASPL front end automatically orders the result by relevance and shows the relevance criterion currently applied. In some services this default criterion has been tentatively opened for the user to change, and thus reorder the result set temporally or according to a semantic attribute.

The situation is even more flexible in the DBLP++ rich user interface, which thanks to its faceted navigation metaphor offers numerous ways for the user to present the relevance of the results, as well as apply filtering and constraining techniques. This capability seems to be one of the strong features of our ASPL/DBLP++ couple, and as such might be worthwhile to focus on in the future to extend and thus 'cement' its competitive advantage.

3.5.2 Positioning tools on query output explanation

In terms of approaches to post-processing the results of a query, the majority of tools provide a simple list of records they consider relevant. In most cases, this is presented in the form of an ordered list, where the ordering attribute usually corresponds to some statistical measure (e.g. popularity) or solely to semantic relevance. There is limited occurrence of classification techniques and even fewer tools attempt to present their classification or clustering outcomes in a visual way (i.e. not as a list of records with respective sub-lists).

We also observed some attempts to provide summaries or précis of the retrieved items as a way to enrich the result set in the post-processing stage. The implementation of this strategy ranges from showing either syntactic neighbourhood where the terms in question can be found (e.g. in the case of Google) to more semantically oriented view of the records with some key descriptive attributes. For example, ASPL provides some associated attribute values to describe the retrieved piece of information less ambiguously, but also to provide directly the information usually sought by the user in one click.

While these summaries may help the user in explaining a particular record, we also considered whether the tools make an attempt on explaining their result sets. Here we found several strategies, both in terms of initiative and in terms of what is explained. Some tools automatically expanded such items as acronyms (e.g. Ask and also ASPL to some extent); others offered additional information (e.g. the attributes in ASPL used as the ordering criterion).

In terms of initiative, we observed a mix of automatic provision and upon-request provision. In the latter case, this was realized by means of adding information such as cached version, category or date by Google, or by means of adding action-specific links/buttons to the applicable items of the result set (as e.g. in ASPL/DBLP++).

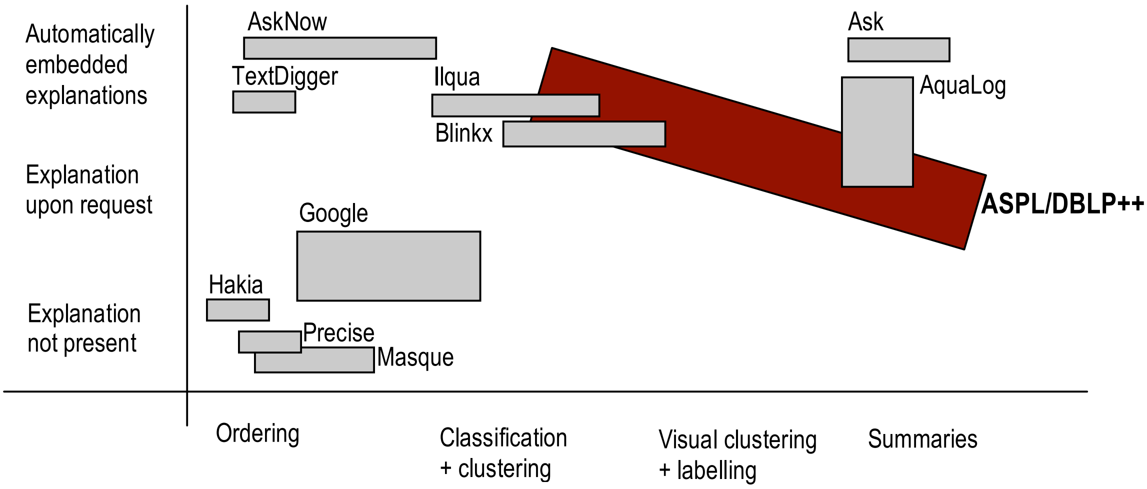


Figure 8, Distribution of tools on the explanatory power of their post-processing techniques

The distribution of tools in the space (shown in Figure 8) is similar to that discussed in the previous section – also here, we observed a fairly strong presence of our ASPL/DBLP++ in terms allowing a choice additional information that extends and, in principle, seems to have a good explanatory power for the query results.

The figure also shows a gap in the section where additional explanation may be acquired upon the user's explicit request; hence, the presence of ASPL/DBLP++ in this area is a potential strength and competitive advantage.

3.6 Positioning tools on the use of learning mechanisms

In terms of supporting some form of learning mechanisms for the purpose of improving the performance of the tool, the situation in the market is fairly clear. In principle, only one tool from our survey had an explicit learning mechanism exposed to the user (AquaLog – in order to ascertain an appropriate query disambiguation alternative). No user-centric learning was observed in other tools.

However, we may, potentially, include in this category techniques for monitoring user activity and also user profiling techniques, which are often a part of more generic engines like Ask and Google. In our opinion, these techniques did not alter the performance (precision, recall or time) in any direct way. These algorithms may have influenced other aspects of information retrieval, such as presenting to the user the notion of result popularity or hit counts – which are more covered by the query output presentation rather than learning and profiling as such.

Hence, this is an area where ASPL/DBLP++ may possibly make an attempt in the future. Since useful learning mechanisms and, more importantly, ones that would be effectively helping the user, are rare, this is an open opportunity that might be considered later.

In a nutshell, in this section we attempted to offer a generic multi-dimensional space, which can be used to compare and contrast different tools with each other based on various functional aspects – searched content, context treatment, query formulation, query analysis, query output processing, and the use of learning mechanism. Because this multi-dimensional space is generic in nature, it allows users to compare the existing tools with newly emerging ones, and, as a result, it provides a uniform basis to analyze the strengths and weaknesses of the new tool with respect to the existing ones.

However, it is important to note that this multi-dimensional space is flexible in nature; in the sense that if the users need to compare and contrast only specific functions associated with the heterogeneous tools, say, context treatment, they can safely perform such a type of evaluative scanning without having to consider other functions, which may not be relevant for their needs. Finally, this multi-dimensional positioning space also allows its users to introduce new evaluation functions, which may be beneficial for their specific purposes.

4 User-based evaluation of ASPL-v2

Having compared and contrasted performance of the different tools in the previous section, our aim here is to evaluate the performance of three of the tools mentioned before: ASPL-v2, Google (its Google Scholar sub set) and Ask by applying the methodology of task-based evaluation of the users' performance and their attitudes.

The main objectives behind conducting this study is a) to identify whether ASPL-v2 serves as a useful tool for its users to perform the functions related to the academic tasks and activities for which ASPL was designed, b) to compare and contrast the performance of ASPL-v2 with other tools from the domain, e.g. Google Scholar and Ask, and finally, c) to identify the performance scope of ASPL/DBLP in comparison with similar tools, which may provide an indicator to the users to decide when they can use tools such as ASPL and when other tools might be more suitable.

It is important to remember that comparing and contrasting tools can be difficult; particularly when some of them search in a specific domain and others are generic, and therefore can be used to find the information in any domain. We had to overcome this difficulty and therefore we designed five different tasks and activities for the users to carry out. These tasks allow us to evaluate the performance of the three tools on the same basis. In the study described in the sub-sections, the users are provided with three tested tools – ASPL-v2 (i.e. its lightweight user interface), Google Scholar, and Ask, and they are asked to perform the study tasks by using the three tools in a different and variable order.

4.1 Study setup

In total 20 users were approached to participate in the user evaluation study. They were selected in such a way that they represent different levels of skills, research background, and research expertise. Hence, we had the following categories of research staff participating: research students in early years with little experience in structuring and retrieving academic information, more senior research students with medium level of skill to deal with and analyze academic data, and research fellows with higher levels of the same skill. We believe this setup is sufficient and necessary in order to avoid biased evaluation of the tools.

This evaluation study consisted of five tasks that users needed to perform by using the following sub-set of tools reviewed and assessed in the previous section – we opted for a range including ASPL/DBLP (http://neon-project.org/aspl-v2/), Google Scholar (http://scholar.google.com/), and Ask (http://www.ask.com). We asked the users that all the tasks involved in this evaluation study were performed independently; i.e. trying not to bias their assessment of performance of second and third tool by their impression from the first one. To counter this potential confoundment of the study we swapped the order in which the tools were shown to the users.

During the study, the participants could ask for assistance from a facilitator if there were any issues or clarifications needed. The total duration of user evaluation study was 80 minutes, and a short period for familiarizing oneself with the evaluation material and tools helped by the facilitator and then performing all the five tasks involved in the evaluation study. To begin with, the facilitator addressed the main reasons for carrying out the user study by giving some examples. Then the facilitator introduced the tools used for the evaluation and demonstrated the key functional properties that we were interested in evaluating as a part of this user evaluation study.

We were focusing on gathering qualitative impressions and reactions of the participants to various aspects of the tools. We believe this is an important focus in the user-based study, especially in the domain that is as open-ended as the search for knowledge on academic connections, communities and publications. Nevertheless, we touch on this at a later point in the report.

The tasks were given a fixed duration, the participants were asked to stick to the schedule, and in case they ran out of time, they were asked to summarize the reasons they believed hindered them. There was no explicit reward for an early finish, neither was there any penalty for an unfinished task. The key requirement for each task was underlined, and this specified the material that needed to be retrieved in the task. The reason we provided this information was to avoid confounding the study by people trying to find some hidden catches in the task statements or spending time interpreting the natural languages sentences.

Next we summarize the description of the user-based evaluation study and the tasks designed for the evaluation purposes. After summarizing each task we will also give a brief rational or the reason for including this task in the study.

4.1.1 Task 1: Finding expertise in a topic

In this task, the user is required to retrieve the names of leading researchers, who are active in a specific research area. In Task 1 the users were required to use the tools provided by the facilitator in the following order: Google Scholar, ASPL/DBLP, and Ask search engine. To retrieve the information described above the users were expected to make use of the query, which is stated in the box below:

> List 5 top researchers whose research work is closely associated with the research topic Semantic Web Services. Please explain the reasons why the specific researchers are included in the list.

The main purpose for including this task in the evaluation study was to evaluate how well ASPL/DBLP and other two tools handled a domain-specific search for a very specific and well-defined query. In particular for ASPL/DBLP, this task allows us to evaluate whether the 'leading experts on topic' service, which has been introduced after the revision of ASPL in response to improving the interpretation of the results in addition to merely listing them. The 'leading experts on topic' service took in account semantically interpreted annotations of the publications in our DBLP++ store and combined these with some statistical evidence to hypothesize the leading roles of particular individuals (authors) with regard to a particular topic. The input to this task was a phrase representing the research topic in question as underlined above.

4.1.2 Task 2: From individuals to communities

In Task 1, the users identified key researchers who are known to be active and leading personalities in the research area of *Semantic Web Services*. We asked the users to assume for the purpose of this task that **Sheila McIlraith** and **Ora Lassila** were two experts featuring in the list acquired in the previous task. In Task 2, the users were asked to perform the following activity by using the tools provided by maintaining the following order: ASPL/DBLP, Google Scholar, and Ask.

> *Activity 1.*
>
> a) For both researchers identify their <u>personal areas of expertise</u> and research interest in the decade (i.e. 1990-2003) prior to the current date.
>
> b) Having identified the expertise and research interests of the two researchers, please generalize these areas of expertise so that we can describe to which <u>research communities</u> these researchers belong. Please list these 'community descriptions' and state whether (in your opinion) it is clear from the tool output if these past research activities are related to the Semantic Web Services research.

The main aim behind the development of ASPL/DBLP was to provide a semantic platform for learning that provides a means of accessing a range of services including simple glossaries and more sophisticated search services contextualized to highlighted text fragments on a web page. To this end, one important aim of ASPL is to provide assistance to its users to explore the areas of expertise not only by means of listing the individuals fitting the keywords. In addition, in the re-engineering we aimed to also support more analytic and synthetic processes whereby the user is capable of inferring the communities of interest for a particular individual, of justifying why a particular community and/or individual fits within a particular research interest, etc.

Thus, activity 1a was included to see how the three tools handle a domain-specific query to find research areas of the named researchers contextualized within a given temporal query modifier. In particular for ASPL it allowed us to evaluate the performance of the services 'person's publications & interests' and 'person's interests' that evolved from the initial simple listing of the records from DBLP in the previous version of ASPL-v1.

> *Activity 2.*
>
> Please suggest 3-5 research publications of both the researchers which would cover the top-ranked past research interests of the selected researchers.

The rationale for this task is that the user new to the research domain may try to understand the scope of a research community and its relationships to other research communities. In order to evaluate the performance of the tools to see how they support such users, Activity 1b was included that allowed us to see how these tools handle this query broadening, synthetic scenario. The context here is to take into account the research interests of the named researchers and their commonly occurring co-authors, and try to generalize this into knowledge about communities of practice. Our aim was to evaluate the performance of the following two services: 'person's community characteristics' and 'person's co-authors and community characteristics' that are also an outcome of the ASPL re-engineering activities.

And finally, Activity 2 was included to get back to the core service of the ASPL and DBLP framework – the retrieval of the actual publications for a given individual and topic. Here we looked at specific, detailed information, not merely a title, but something more like a bibliographic reference.

4.1.3 Task 3: Bibliographic lists

In Task 2, Activity 2 the users have identified top 3 publications of the researchers who were named as active in the research area of Semantic Web Services. We asked the users next to prepare a bibliography, a full list of the found publications e.g. for the purpose of a literature review. They were expected to produce detailed bibliographic information for the publications retrieved in Task 2, Activity 2. For this purpose, please perform the activity, which is given in the box below in order to achieve this task.

> *Task 3.*
>
> For all the publications that are collected in Task 2, Activity 2, please describe in detail the relevant places (e.g. name of a conference, workshop, or journal) to find the collected publications along with complete bibliographical information of these publications.

As above, the purpose of this task was to test the tool's ability to ground its search results in the additional information that can be readily reused e.g. in the literature review. This functionality was partially present in the earlier version of ASPL, so this was mainly to see whether the redesigned ASPL maintained one its key original motivation capabilities.

4.1.4 Task 4: Constraining query results

One important part of a critical literature review on a certain topic may be the need to drill into in-depth details of a research topic. As a part of literature review the users may need, for example, to compare and contrast different viewpoints that exist over a particular issue, say Ontology Alignment. For this purpose, the users are required to find only those publications about Ontology Alignment, which are technical in nature and help them to prepare the literature review. To perform Task 4 the users need to use the tools in the following order: ASPL/DBLP, Google Scholar, and Ask.

> *Task 4.*
>
> Find up to three research publications on research topic, Ontology Alignment, which provide the technical information on the research topic.

For the purpose of this task, the technical nature is seen as a publication containing such aspects as definitions, schemas, architectures, and similarly. The motivation for this activity is to see what support do the tools offer in terms of constraining the search boundaries, and to what extent are the applications knowledgeable or aware of different purposes publications and papers may serve for.

4.1.5 Task 5: Coping with incomplete inputs

In Task 5 our aim is to look at the performance of the tools to see how adequately they support the user's intentions in the cases, when the users have incomplete information and yet, they aim to retrieve a specific piece of information. Here, the main task is to retrieve a specific publication (or alternatively, its full bibliographic information) if various cases of complete reference are considered.

One objective is to observe how the users make use of the combinations of keywords. To carry out this task, the user was provided with the publication, which they were supposed to locate by using specific combinations of keywords. We want to see (and rank) which combination of keywords and tools works best to correctly locate the publication. In the box below we provide the complete bibliographic information of the publication in question.

Task 5 (objective)

Activity Dieter Fensel and Enrico Motta and V. Richard Benjamins and Monica Crubezy and Stefan Decker and Mauro Gaspari and Rix Groenboom and William Grosso and Frank van Harmelen and Mark Musen and Enric Plaza and Guus Schreiber and Rudi Studer and Bob Wielinga. *The Unified Problem-solving Method Development Language UPML*. Knowledge and Information Systems, 5 (1), pp. 83-131, 2003.

Having provided the publication details we gave users suggestions of different combination of keywords that can be used to retrieve the aforementioned publication. For instance, if the users have decided to use the keyword combination of 'Year of publication AND one of the authors from the list', they would have to make their own choice about how to submit these modifiers to particular tools.

Task 5 (Keyword combinations).

1) Journal AND Author
2) Keyword from the publication (e.g. PSMs) AND one of the authors
3) Year of publication AND one of the authors from the list
4) Author AND Author AND Keyword from Publication

The tools.

1) ASPL/DBLP
2) Google Scholar
3) Ask

Here, we complete our description of the tasks used in the study to evaluate the performance of the tools on a variety of tasks that are broadly based on the common theme of preparing materials for a literature review using publicly available resources.

4.2 Data analysis and qualitative feedback from participants

Having completed the user evaluation study, we compiled and analyzed the comments that were raised by the users reflecting their impressions and experiences from interacting with all three tools in the tasks that involved different types of searches (as described briefly in the previous section.

These comments are presented in Appendix 1, and they allow us to point toward the useful features, strengths, and also to missing features, weaknesses of the three tools included in the study. Moreover, it also allows us to determine the scope of the tools to specify that certain types of tools can be seen as useful resources only for performing certain types of search tasks. As a result, if that tool needs to be evaluated then a due care must be taken into account by evaluating this tool only with the tools with similar functions.

In Appendix 1, the comments are classified into the positive comments that provide an indication about effectiveness of the tools for performing a specific task described in the user based evaluation study, the negative comments provide an indication about how a specific tool fail to handle a task or whether such a type of task falls outside the scope of tool's functionality, and finally suggestions can be used to improve the performance of the tools.

The main aims of our analysis are to see – a) if the users successfully managed to perform the tasks and the activities associated with it in a given time, b) then we also compare and contrast the performance of all the three tools with each other in order to evaluate which a most appropriate tool in order to perform a specific type of search, c) we also analyze if any specific tool performed better as compared with other tools in general, and finally d) based on this analysis we determine the scope of ASPL/DBLP. In other words, we will provide an indication about ASPL/DBLP, which will help its users to determine the types of searching services that can be successfully handled by ASPL/DBLP when compared with more generic search tools.

4.2.1 Task 1: Finding expertise in a topic

As described earlier, in Task 1 the users used the three tools to retrieve the names of the top researchers in the domain of 'Semantic Web Services'. To this end, our main aim was to evaluate how these tools handled domain-specific search for a given query. Moreover, this task also allowed us to evaluate the performance of the service 'leading expert on topic' embedded in both ASPL services and the ASPL/DBLP web-based front end.

The total duration for performing this task by using all three tools was 10 minutes. All the 20 participants successfully managed to complete the Task 1 within allocated time without having to extend the time duration. Moreover, all the users were satisfied with the description of the task provided to them in the user evaluation material and therefore no further assistance was required to be provided to them.

We found that 90% of the users found ASPL/DBLP useful. ASPL/DBLP successfully managed to retrieve the names of the top 5 researchers in the research domain of Semantic Web Services. In some cases the users decided to cross check the information retrieved by ASPL/DBLP and therefore they changed the time interval given in the task to retrieve the names of the researchers within new time period. In the task, we asked users to find the names of the top researchers during the time interval of 2002-2006.

The users changed the time interval to 1980-2006 to get the new set of results. They were particularly happy with the results they received after changing the time interval and confirmed that ASPL/DBLP retrieved the same set of results in both time intervals. One of the main reasons why ASPL/DBLP successfully handled the change in the time duration because the retrieval function embedded in ASPL/DBLP was robust enough to handle such scenarios. This can be particularly useful, as we will show later when the temporal context becomes a key part of a query.

When compared with the performance of ASPL/DBLP, only 10% of the users stated that the Google Scholar was a useful tool for this task, and that it was simple and straightforward to carry out such type of retrieval using this tool. One important observation was made where the users have mentioned that Google Scholar did not take into account the publishing dates associated with the publications, and therefore they had to look somewhere else (e.g. Google itself or the content of the link) in order to make a decision about which researchers were more active in a specific period.

As a result, they had to spent more time to look for the correct information by performing *several* searches. The reason why the users put more stress on taking into account the publication dates as one of the indicators because if a certain researcher published higher number of publications during a specific time period then he/she can be considered to be a more active researcher during that time period. In some other cases, the users also indicated that in contrast with ASPL/DBLP, the ranking mechanism used by Google Scholar was not satisfactory because some important publications authored by the key researchers were placed lower in a list of results.

Finally, when compared with ASPL/DBLP and Google Scholar, the performance of Ask was not satisfactory at all. Typically users stated that they had to look at multiple places in order to realize who might be the top researchers in the Semantic Web Service domain. When compared specifically with Google Scholar, almost all users stated that the search with Ask was less intuitive – mainly because the search with Ask failed to find the publications and researchers, while even simple Google Scholar search was more accurate and quick.

Moreover, Ask failed to provide any indication about which were top researchers involved in a specific research area as Ask did not offer many results for academic publications and instead pointed users to Citeseer (http://citeseer.ist.psu.edu/) and it was difficult for the users to judge based on Citeseer entries whether certain researchers were more influence as compared with others in a certain time interval.

The service ‘leading experts on topic’ embedded under the tab ‘search in topic domain’ in ASPL was a key factor why this tool outperformed the other two; i.e. Google Scholar and Ask. This service certainly helped the users not only to retrieve the required results, but it also saved their time because they did not have to look elsewhere. Moreover, because the users could set a specific time period to retrieve the key researchers associated with a specific research area, it helped them to identify the leading researchers in Semantic Web Services without having to process the time interval part of a query. Figure 9 shows a sample screen, which the users encountered in this particular task. Figure 10 shows the output produced by ASPL/DBLP for the service ‘leading experts on topic’.

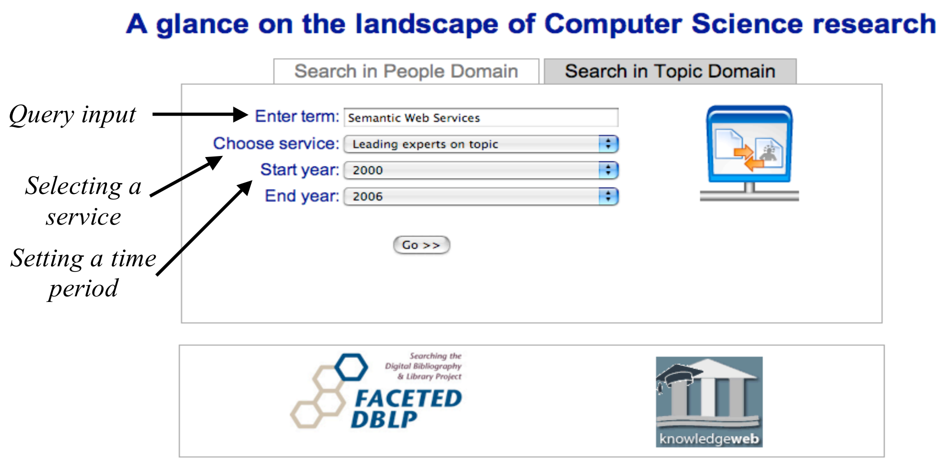


Figure 9. The interface of ASPL/DBLP for Task 1

Leading authors publishing on '*Semantic Web Services*'

The list below contains leading researchers with highest numbers of published work on a given topic. For each person we list the period in which the publications occurred, and the number authored by a given person split into 'on topic' and 'overall' output. The 'relative share' is thus the percentage of the person's published work on a given topic out of his or her total output.

Person's Name	Number of Publications			In Period
	On Topic	Overall Output	Relative share	
Dieter Fensel	17	80	21%	2003 – 2006
Katia P. Sycara	13	93	14%	2003 – 2006
Christoph Bussler	13	60	22%	2003 – 2006
Amit P. Sheth	12	74	16%	2004 – 2006
Massimo Paolucci	11	43	26%	2003 – 2006
John Domingue	11	40	28%	2003 – 2006

Figure 10. The output of ASPL/DBLP for Task 1

About 10% of the users did not prefer to make use of the existing services associated with ASPL/DBLP. In one case, the user stated that ASPL/DBLP failed to retrieve the researchers, who can be considered to be ‘gurus’ in the domain of Semantic Web Services. In our viewpoint, this cannot be considered as a limitation of ASPL, because this is a subjective opinion to determine who can be considered to be ‘gurus’ in a certain domain without using any explicit basis for selection. More interestingly, one user expected ASPL to consider the relevance of the conferences for determining the importance of the publications, and therefore the authors who authored such publications.

In some other cases, the users stated that the existing version ASPL failed to take into account the *number of citations* made about a specific publication because the authors of such a publication can be considered to be key researchers in a research area. Finally, some of the users suggested that it would be useful if ASPL/DBLP allowed its users to set multiple type of searching facility, such as leading experts based on number of publications, leading experts based on impact factor of the publications authored by the researchers. In our viewpoint, in the future version of ASPL/DBLP it will be useful to add such type of services that would enable the users to handle these personalized extensions of the primary retrieval service and thus allow custom interpretations of the 'is expert in' assertion made by ASPL by re-using DBLP.

Figure 11 shows two pie charts: The pie chart on the left shows the percentage of users who preferred to use the three tools for performing Task 1. The pie chart on the left shows percentages where users preferred not to make use of a certain tool in the future, in order to perform searches similar to the one performed in Task 1.

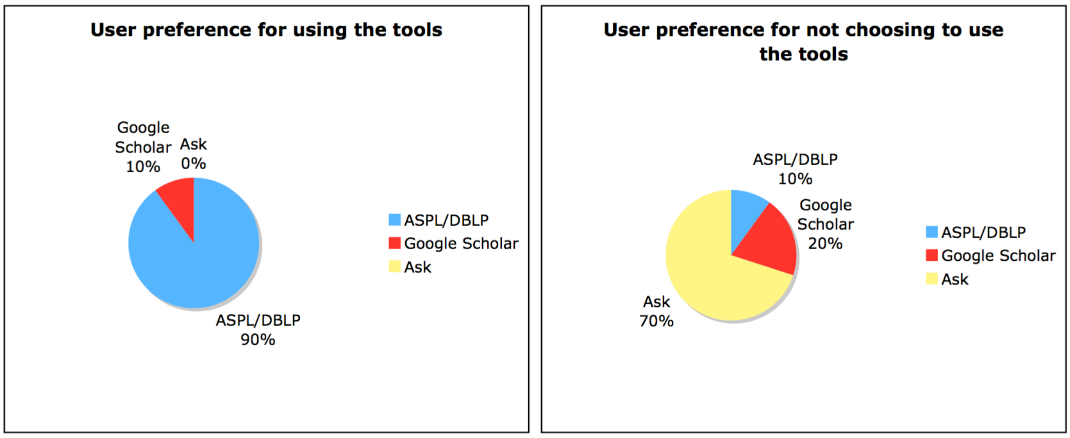


Figure 11. Representation of user preferences for using the tools

In a nutshell, based on the data gathered and observations made by the users, it can be said that for the type of search described in Task 1, ASPL/DBLP can be considered to be the best tool in comparison with the generic search engines such as Google Scholar and Ask. The users preferred not to use generic search tools because it was difficult for them to look for the certain information specific to the time period. They had to perform multiple searches and then manually process the information, which was time consuming. Moreover the users also stated that tools like ASPL/DBLP use uniform criteria to determine whether certain researchers were considered in a given time period. In this case, it was highest number of publications authored by the researchers were considered to be the key researchers. In contrast with this, the generic search engines used different criteria to determine the expertise of the researchers and it was difficult for the users to compare and contrast the results quickly.

4.2.2 Task 2: From individuals to communities

One of the main aims behind the development of ASPL is to provide a semantic platform for learning by means of accessing a range of services including simple glossaries and more sophisticated, contextualized search services associated with annotated text fragments on a web page. In order to evaluate whether the three tools provide such a support, we have designed Task 2, which consisted of two activities, Activity 1 and Activity 2.

In the activity 1a, we asked users to identify the 'personal areas of expertise' and the 'research interest' of two researchers, i.e. Sheila McIlraith and Ora Lasilla over the period of 1990-2003. Here, our aim was to evaluate the performance of all tools but particular in ASPL/DBLP we evaluated the performance of the services 'person's publications & interests' and 'person's interests'. In the activity 1b, we asked users to generalize the expertise and/or research interests of the two researchers in order to identify to which research communities they belong. By using this information the users were then asked to state if these past research activities related to the research in Semantic Web Services. The activity 2 was straightforward; we asked the users to find 5 research publications of the researchers that cover the addition research expertise, which they have identified in Activity 1a.

The total time allocated for performing Task 2 was 20 minutes. We observed that 80% users took more time to complete this task. On an average the users took 5 minutes of extra time to complete the task. The main reason why the users found it difficult to fit their search for information within time limits was due to Activity 1b – the search performed by generic search engines such as Google Scholar and Ask took more time to retrieve meaningful results, and the users were forced to re-formulate their criteria several times. We will discuss this point later in the following paragraphs.

In particular, 65% users felt that ASPL was an appropriate tool for this task, which provided them with some assistance to find out about the personal interest and research expertise of the named individuals. When looking for the past research interest of the researchers, the users voted ASPL as an easy to use tool, because they could set the time interval to find the necessary information explicitly. As a result, they didn't have to worry about processing time interval as a part of the query results.

Moreover, the two services implemented in ASPL – 'person's publications & interest' and 'person's interests' helped the users to look for the required information straightforwardly. With the help of these two services the users only had to provide as an input the name of a researcher and set the time period to get the necessary information quickly. Moreover, ASPL not only retrieved the past research interests of the researchers in a given period, but it also sorted these past research interests in the order of most recent ones to the oldest ones. Figure 12 shows how the users can set the input for the service 'person's publications & interest'.

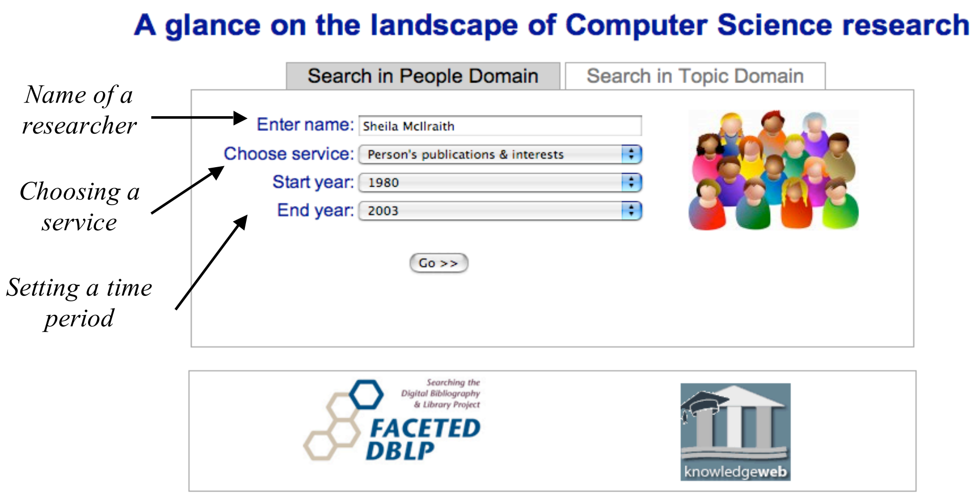
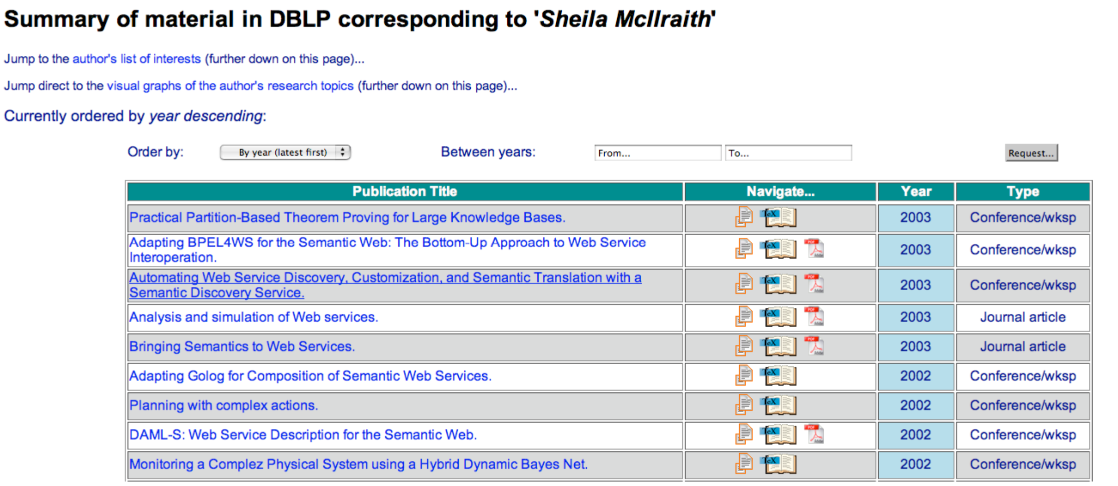


Figure 12. The interface of ASPL/DBLP showing how the users can set inputs for person's publications & interest

Figure 13 shows a partial view of the output produced by ASPL for the service 'person's publications & interests'.



Summary of material in DBLP corresponding to '*Sheila McIlraith*'

Jump to the author's list of interests (further down on this page)...

Jump direct to the visual graphs of the author's research topics (further down on this page)...

Currently ordered by *year descending*:

Order by: By year (latest first) Between years: From... To... Request...

Publication Title	Navigate...	Year	Type
Practical Partition-Based Theorem Proving for Large Knowledge Bases.		2003	Conference/wksp
Adapting BPEL4WS for the Semantic Web: The Bottom-Up Approach to Web Service Interoperation.		2003	Conference/wksp
Automating Web Service Discovery, Customization, and Semantic Translation with a Semantic Discovery Service.		2003	Conference/wksp
Analysis and simulation of Web services.		2003	Journal article
Bringing Semantics to Web Services.		2003	Journal article
Adapting Golog for Composition of Semantic Web Services.		2002	Conference/wksp
Planning with complex actions.		2002	Conference/wksp
DAML-S: Web Service Description for the Semantic Web.		2002	Conference/wksp
Monitoring a Complez Physical System using a Hybrid Dynamic Bayes Net.		2002	Conference/wksp

Figure 13. The output produced by ASPL/DBLP for the service 'person's publications & interest'

In comparison with ASPL, the performance of Google Scholar turned out to be very poor. And all the users stated that Google Scholar was not the desired tool when they were looking for the information about the personal interest and research expertise of researchers when the time duration was a crucial part of a query. In contrast with ASPL, Google Scholar did not allow users to set a specific time interval to filter out the information. As a result, the users had to use several search query combinations to get the results about personal areas of interests and research expertise of the researchers. Having received this information, it was very difficult for them to map this information over the time interval within which the researchers were active in a certain research area. In other words, they had to manually perform the mapping activity of the information retrieved by the tools over their time interval. However, one particular user stated that although it was difficult to find the necessary information by using Google Scholar but the 'related authors' function was quite useful. By using this function the users could provide as an input the name of a researcher and then look for other researchers that may be related with the given one.

In comparison with ASPL and Google Scholar, 35% users considered Ask as an appropriate tool for such a type of task. The main reason behind this was that when searching for the information, Ask led users straight to the people's web pages, where research interest were listed explicitly. However, in some other cases the users preferred not to use Ask to look for the information given in this task because similarly with Google Scholar they had to look at different places to find the relevant information and then process the information manually. Once again the two services embedded in ASPL/DBLP person's publication & interest' and 'person's interest' under the tab 'search in people domain' helped ASPL to outperform Ask.

In Activity 2, the users were required to find out the research publications of the researchers in such a way that these publications cover the different research interests of the researchers. To this end, when we analyzed the collected data to evaluate the performance of ASPL. It turned out that 90% users thought ASPL helped them to perform this part of Task 2 quickly and efficiently. Generally speaking, the users preferred ASPL because they only submitted the name of a researcher and then used the service called 'person's publications and interest' to get the necessary information. However, in some cases the users reported that in contrast with Google Scholar the numbers of publications retrieved by ASPL/DBLP were limited in numbers.

More importantly, all the users stated that ASPL failed to retrieve any information when the names of the researchers were submitted in an incomplete manner, e.g. 'McIlraith' and 'Lasilla'. Moreover, when the order of name and surnames of the researchers was reversed, in such a case ASPL failed to find any publications, too. In contrast with this both Google Scholar and Ask provided the necessary information when the queries submitted by the users were incomplete. We see this as one of major weakness of existing version of ASPL. In order to overcome this drawback our aim is to implement the name disambiguation service within ASPL framework, which would allow users to submit the names in any fashion that is suitable for them to retrieve the necessary information.

In comparison with ASPL, the users found it difficult to perform this part of Task 2 by using Google Scholar. Similarly as with the Activity 1, by using Google Scholar, they found it difficult to establish a relation between research interests of the researchers and their publication that could cover the different interests. While using Google Scholar the users used the following type of query 'name of a researcher + research publications' to retrieve the information. It was observed that Google Scholar simply retrieved the list of publication in which a researcher was an author. However no support was provided Google Scholar to determine whether these research publications covered the past research interest of a researchers. As a result, having retrieved the publications, the users had to dig out for the appropriate results from all the information.

As compared with ASPL and Google Scholar, about 10% of users voted Ask as an appropriate tool to perform activity 2. Typically the users used the query 'name of a researcher + research publications' to retrieve the information and Ask led the users to the home page of a researchers, where the users could access all the information quickly. However, for the users it was difficult to relate the research publications that would cover top-ranked past research interests of a researcher. The users preferred Ask in comparison with Google Scholar because for the same query, Ask led its users to the home page of a researcher where they accessed different types of information, whereas Google Scholar failed to retrieve the home page of a researcher.

Figure 14 shows the pie charts where we represent the comparative performance of all the three tools when used to perform Activity and Activity 2 involved in this task. In a nutshell, based on the data collected from the user evaluation study, we can say that in the context of search where the users are looking for establishing a relationship between the submitted keywords, e.g. "Sheila McIlraith" + "Research Interest" + "1990-2003", in order to retrieve more meaningful results because the information may be represented in different context, and therefore, it may be distributed at different places, ASPL/DBLP can be seen as the best choice.

Similar type of search by using generic search tools required lot of manual processing of the results, which was difficult for the users while time consuming at the same time. Moreover, we can say that ASPL not only successfully realized the context of search, but it also successfully established query term correlation between the keywords appear in a query. This helped the users in avoiding to do any manual post-processing of the search results to understand how the results can be related with each other. Finally we can say that the services that are included in ASPL, such as ‘person’s interests’, ‘person’s publications & interests’, ‘person’s community characteristics’ and ‘person’s co-authors and community characteristics’ certainly helped the users in retrieving the appropriate information quickly and with no need for any substantial manual intervention.

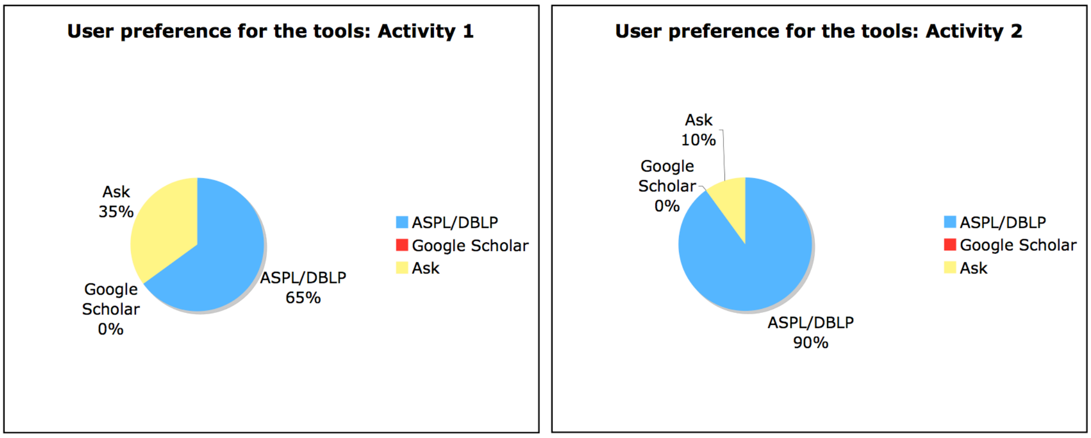


Figure 14. User preferences for using the tools during Activity 1 and Activity 2

4.2.3 Task 3: Bibliographic lists

Task 3 consists of only one activity. Here we asked the users to retrieve the complete bibliographic information of all the publications that were collected by them in Task 2, Activity 2. The total duration for completing this task was 15 minutes. Because this task was a continuation of the previous one we expected it to be quite straightforward for the users to finish it in time, but it turned out that about 15% users failed to complete the task in a given time. The main reason behind this was the users found it difficult to retrieve complete bibliographic information of the publications by using Google Scholar and Ask.

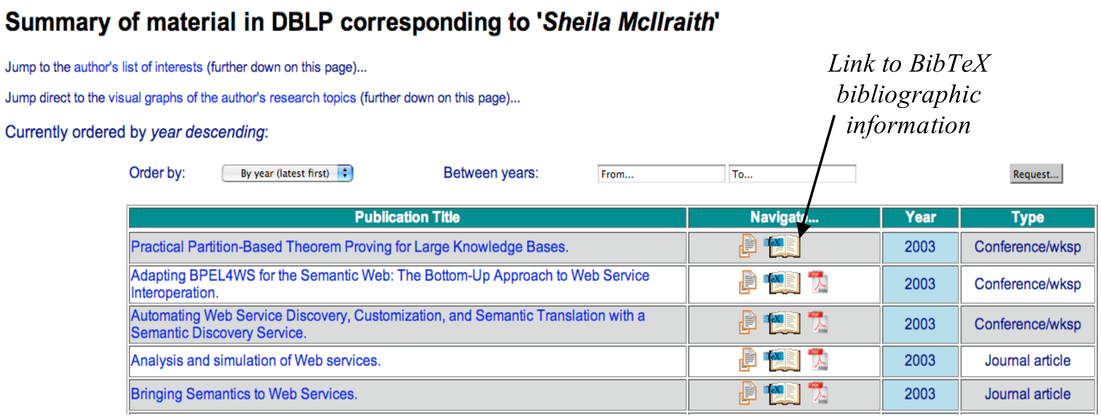


Figure 15. Navigation provided by ASPL/DBLP to access complete bibliographic information

The users were asked to use ASPL as the tool to perform this task. They typically relied on the service ‘person’s publications and interests’ under the tab ‘search in people domain’. Having retrieved the publications as shown in Figure 15 for Sheila

McIlraith, ASPL allowed users to navigate the complete bibliographic information of the publication simply by clicking on an icon with TeX written on it under the column 'Navigate'.

Generally speaking, 70% users preferred ASPL to perform this type of search for a very detailed information. The users typically considered ASPL as a useful tool, because it not only provided with complete bibliographic information of a publication but also the bibliographic information was retrieved straight from DBLP the users had a higher trust factor to the result. As a result the users did not have to cross check authenticity of bibliographic information elsewhere. This saved a crucial time while performing this task.

More importantly, the users also like the way ASPL presented the results to them in a nice and consistent format: names of the authors, title of a publication, source of publication (i.e. name of conference, workshop or journal), and the publication year. The users preferred this type of *post-processing* service performed by ASPL, because it was easier for them to interpret the results given by the tool. When the users used the other two tools they stated explicitly that the format of the retrieved results changed from one publication to the next one, which made it difficult for them to interpret the results.

When compared with ASPL, about 20% users preferred to use Google Scholar to perform such a task. Generally speaking, the users liked Google Scholar as a tool, because they were already familiarized with the interface of it and the way it worked (thanks to its Google parent). In contrast with ASPL, in Google Scholar no facility was provided to retrieve research publications straight from DBLP. As a result the users decided to cross check the results of Google Scholar with DBLP, which contributed in failing to complete the task in a given time. In contrast with ASPL, which presented the results in a consistent format as described earlier, no such post-processing service was implemented in Google Scholar. That made the overall interpretation of the results difficult for the users and in some cases the users failed to recognize the retrieved result as the correct one.

Finally, as compared with Google Scholar, the performance of Ask was not satisfactory and only 10% of the total users preferred to use Ask for performing such a task. The users stated that Ask managed to retrieve the publications authored by both the authors but retrieving the bibliographic information of these publications was not straightforward. Similarly as Google Scholar, no post processing of the results was performed by Ask and therefore the users were not satisfied as typically the format of the results changed from one publication to the next one. The main reason behind this was that in contrast with ASPL that there was no service implemented in Ask, which would check whether the bibliographic information was DBLP compliant.

In a nutshell, once again ASPL turned out to be performing reasonably well where the academic related information was required by its users quickly and also in a consistent format. In contrast with generic search engines, which performed the search over wider documents to retrieve the results ASPL had a scope only within academic domain, and some may see as a weakness its tight link to the DBLP as the sole data provider. However, tools like ASPL can be seen as a acceptable resources for the users that are interested in finding the information from the academic domain –

mainly in balancing the complexity of queries with results precision. As a result, after completing evaluation of the first three tasks, we realized that ASPL can be seen as an appropriate tool effectively supporting narrow needs within an academic domain.

4.2.4 Task 4: Constraining query results

In Task 4, our aim was to evaluate the performance of all the three tools to see how these tools handle the queries that consisted of multiple keywords to get the necessary information. We asked users to find out the technical publications on a research topic of 'ontology alignment'. In order to perform this task the users were asked first to make use of ASPL, then Google Scholar, and finally Ask. The total duration allocated to perform this task was 15 minutes, and all the users successfully managed to finish the task in a give time.

Task 4 acted as an acid test to see whether such a type of searching falls within the scope of ASPL. Based on the data gathered from the users it was realized that only about 10% users found ASPL to be useful tool in order to perform such a context specific search. 90% of the users preferred to use the conventional generic search tools instead. Generally speaking, by using ASPL the users managed to retrieve the publications, which consisted of the keyword 'ontology alignment' by using the service 'main publishing outlets' embedded under 'search in topic domain' tab. However, no assistance was provided by ASPL/DBLP that would allow its users to make a decision about whether the retrieved publications were good candidates for providing the technical information about 'ontology alignment'.

Moreover, the users had to pass through various outlets, such as finding the relevant conferences on a topic, then they had to manually look whether that conference hold a session or a paper on ontology alignment, and finally decide whether the publications were technical in nature. More importantly, the users found the existing version of ASPL/DBLP less flexible as compared with generic search tools. The main reason behind this was in contrast with generic search tools in which the users typically used additional keywords such as "ontology alignment + conference paper", "ontology alignment + technical papers" to look for the necessary information the users could not add additional keywords in ASPL. Moreover, the generic search tools provided a quick summary of a publication and also highlighted the important keywords in the retrieved publication, which helped the users to make a decision about the nature of a publication, while no such support was provided by ASPL.

In contrast with ASPL, most of the users preferred Google Scholar as a tool to perform such a type of search. One of the main reasons why the users preferred Google Scholar as compared to ASPL was that it allowed them to add multiple new keywords to perform the search. As a result the users received more accurate output fairly efficiently. As described earlier, typically Google Scholar provided users with an extract about the paper and it also highlighted the keywords from the paper similar to the submitted query. Of course, it allowed users to make their decision about the nature of the publications without having to go through the publication.

Finally, when compared with ASPL and in particular with Google Scholar, the performance of Ask was satisfactory and the users found Ask as an effective tool. The users have pointed out that Ask provided them with a URL, which had most relevant handpicked publications with their most relevant reviews. In contrast with Ask, no such service implemented in ASPL and Google Scholar. Similarly with Google Scholar and in contrast with ASPL, Ask allowed its users to add new keywords to retrieve more relevant results. Moreover, similarly with Google Scholar a short overview of a publication was provided by Ask that helped the users to realize the nature of a publication.

In comparison with Google Scholar, the users preferred to use Ask because, it suggested users the results from other search tools, e.g. Excite and Lycos, which may be relevant to their query. In some other cases, the users preformed the search by using the following combination of keywords to find the relevant information stated in this task – 'ontology alignment + technical papers' by using both Google Scholar and Ask. It was observed that the users preferred to use Google Scholar over Ask because the Google scholar provided users with the service called 'Recent Articles'. This service retrieved the most recent publications about the research topic while because no such service was implemented in Ask it failed to retrieve the latest publications. For instance, when the users analysed the first ten publications retrieved by both the tools, Google Scholar retrieved 3 publications from 2007 while Ask did not retrieve a single publication from the same year.

Introduction to the **ontology alignment** evaluation 2005 - all 5 versions »
J Euzenat, H Stuckenschmidt, M Yatskevich - Proceedings of Integrating Ontologies workshop at K-CAP, 2005 - ceur-ws.org
... Conference (I3CON) and the EON **Ontology Alignment** Contest [4 **...** as well as addressing other **technical** comments). **...** to the results provided in the fol- lowing **papers**. **...**
Cited by 23 - Related Articles - View as HTML - Web Search

A String Metric for **Ontology Alignment** - all 7 versions »
G Stoilos, G Stamou, S Kollias - International Semantic Web Conference, 2005 - xobjects.seu.edu.cn
... Electrical and Computer Engineering, National **Technical** University of **...** One important method for **ontology alignment** is the **...** In the current **paper** we present a new **...**
Cited by 22 - Related Articles - View as HTML - Web Search - BL Direct

Alignment of Ontologies: WordNet and Goi-Taikei - all 4 versions »
N Asanoma - Workshop on WordNet and Other Lexical Resources: ..., 2001 - seas.smu.edu
... 400,000 words in all: common nouns, proper nouns, **technical** terms, and **...** to align WNsub and the common noun part of GT's **Ontology**. 2.1 Automatic **alignment** **...**
Cited by 10 - Related Articles - View as HTML - Web Search

Ontologies and electronic commerce - all 11 versions »
D Fensel... - IEEE Intelligent Systems, 2001 - doi.ieeecomputersociety.org
... in the class hierarchy, and a few other **technical** issues **...** and M. Musen , "PROMPT: An Algorithm and Tool for Automated **Ontology** Merging and **Alignment**,"Proc **...**
Cited by 27 - Related Articles - Web Search

Figure 16. Results obtained by using Google Scholar

4.2.5 Task 5: Coping with incomplete inputs

The main aim of Task 5 was to evaluate the performance of the tools to see how these tools handle user queries with incomplete information. It can be imagined that in many cases the users only remember few keywords from a publication, say the surname of a researcher, or a place where the publication may be published. The tools must handle such a situations and provide users with appropriate publications. In this task, the main aim of the users was to find the following publication and they were provided with a specific set of keywords to look for a publication. The total duration allocated to perform this task was 15 minutes. All the users managed to complete this task successfully in a given time slot.

The users were asked to use ASPL in order to perform this task. Generally speaking, only 5% of the users have preferred to use ASPL in order to perform this type of search. The users have stated that ASPL gives them a high precision mainly when the task falls within the predefined pathways of ASPL. In other words, in ASPL, the users were successful in retrieving the publication by using the keyword combination 'Year of publication AND the name of one of the authors from the publication' because they could make use of the service 'person's publication and interest' embedded under the tab 'search in people domain'.

In other cases, the users failed to retrieve the intended publication with ASPL because ASPL did not allow them to set different types of keywords. The users found ASPL quite restrictive in nature about the way it allowed submission of keywords, especially if compared with Google Scholar and Ask. The generic search tools allowed their users to use heterogeneous combination of keywords to retrieve data. This task demonstrated to us that ASPL lacked the required flexibility to handle the queries that may be submitted by the users in different forms.

On a serious note, it was observed that failing to handle different types of queries submitted by the users led to failing to retrieve the publication even when that was available in the database. Most of the users after performing this task with ASPL stated that they would like to see a new service implemented in ASPL that would allow them to perform search based on keywords more than just a name of a researcher or topic. For example, data like conference venues or different research groups / institutions already exists in DBLP databases, is semantically annotated. So, to certain extent, this is a matter of joining the dots and adding a simple facility to the user interface of ASPL that would allow slightly more complex queries. To keep the user interface simple, this could be an optional, on-demand form of query re-formulation.

The users were then asked to use Google Scholar to perform the same task. After using ASPL to perform the task, the users found Google Scholar more effective to retrieve the intended publication. Of course, the main reason behind the success of Google Scholar when compared with ASPL was that Google Scholar allowed its users to use different combination of keywords. As a result, 80% users preferred to use this tool as they looked for the intended publication because this type of search was more conventional in nature for Google. However, one important observation was made when the users said that if ASPL had allowed them to submit multiple keywords they would have preferred ASPL to Google Scholar, because of the high precision and recall of ASPL as compared with other tools. At this point, it is fair to mention that this was a subjective view of the user. Also, the precision/recall of ASPL is to a great extent a function of the underlying DBLP being more topic-focused than the Google Scholar…

Finally, the users were asked to use Ask to perform the same task. And it was observed that only 15% of users preferred to use Ask when compared with Google Scholar. The users preferred Google Scholar because of its better hit rate and precision. For instance, in one case, the user used the following combination of keywords given to look for the publication by using Google Scholar and Ask – 'Dieter Fensel UPML 2003'. The user stated that Google Scholar precisely returned the required publication, which essentially appeared in the top 5 publications in the list of retrieved results, whereas Ask returned the old version of the same publication (i.e. 1999 version), which appeared as the 15^{th} publication in the retrieved results.

Moreover, in some other cases, the user used 'Dieter Fensel and UPML' keyword combination to look for the publication by using Google Scholar and Ask. In Google Scholar, the required publication was appeared as the first hit in the retrieved results, while Ask once again returned the old version as the 12th publication in the list of retrieved publications. Moreover, in some other cases, we observed that the users found it difficult to interpret the results retrieved by Ask because according to the users it was hard for them to distinguish between the publications and the web pages.

The following pie chart shows the preference expressed by the users for using the three tools in order to perform the search given in Task 5.

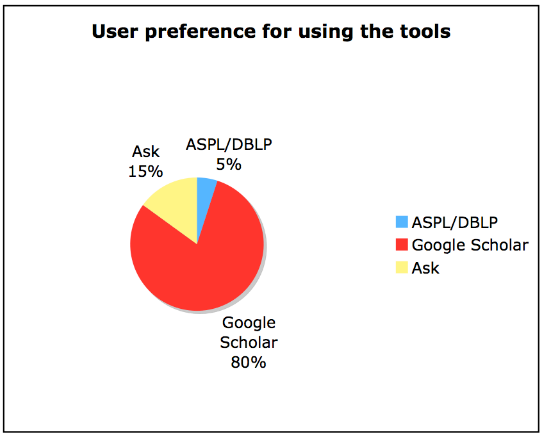


Figure 17. User preference for using the three tools to perform the search in Task 5

4.3 Data analysis summary

Here we conclude the data analysis of the data gathered by us during the user evaluation study performed to evaluate the performance of the three tools. Generally speaking, it was observed that ASPL turned out to be an appropriate tool when the users could make use of its specialized services implemented in its re-engineered version. In a nutshell, the users preferred to use ASPL to perform the search as mentioned in Task 1, Task 2 and Task 3 of the user evaluation study.

However the users preferred to use the generic search tools, i.e. Google Scholar and Ask, when the search was open ended and as a result the users had to make use of additional keywords to look for the information. In such cases ASPL failed to retrieve the necessary information because the existing version of ASPL does not allow its users to add new keywords to look for the information. Generally speaking, the users found ASPL to be useful work and they stated to make use of the tool for their search if it is within academic domain. In some other cases, the users stated that ASPL can be considered to be the best tool available for the researchers in order to find experts on a certain topic, research publications about a topic or researchers that are active in a research topic, or to find out appropriate places for retrieving the publications on a certain topic, and also to get an intuition about which journals, conferences, etc. are relevant to a specific research topic. Finally, the users stated that they were more than happy to participate in the evaluation because they not only liked the way evaluation study was designed but ASPL, Google Scholar and Ask, turned out to be quite interesting tools for performing different tasks.

Nonetheless, ASPL was observed to perform better on tasks that required a degree of synthetic or analytic thinking. It outperformed two generic tools on those tasks it has been designed for, which can be considered a successful outcome.

It can also be said that the two tasks, where the ASPL performed less well, suggest good and meaningful means for further semantic improvement of the ASPL platform, which could, in principle, give it even more credit over generic tools.

Hence, if the capabilities of ASPL to handle incompleteness by perusing the existing and inferred semantic relationships are improved, the tool is likely to gain a good support within a narrow domain of users who need to search for and interpret academic literature as well as research community relationships. Thus, we believe that ASPL can be considered a successful, albeit still open to further improvements, realization of the initial objective we set for this educational area research.

5 Summary and conclusions

In this report we evaluated how the re-designed ASPL framework addressed the needs that came out of the previous assessment. To summarize the design decisions made after the 2005 evaluation, we argued that for the purposes of web-based and semantically supported learning, the interactions of a user/learner with the content are more than mere annotation of web pages, retrieval and subsequent browsing of semantic metadata. In order to apply semantic knowledge, the re-designed version of ASPL supported a more exploratory approach to interacting with distributed learning resources, focusing on creating interpretative pathways rather than merely retrieving simple data.

Specifically, in the re-designed ASPL 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). Together, the two helped us to introduce support for analytic and synthetic learning tasks, and the value of our approach has been partially corroborated in a user-based study presented in this report (see e.g. sections 4.2.1, 4.2.2 or 4.2.3) – majority of users liked the way ASPL-v2 helped them to navigate through the problem space in a structured way, which they could mimic and thus develop a skill in analyzing academic data.

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.

The previous report (D3.3.6) concluded with a theoretical justification of our decision to re-engineer and almost completely revise the suite of learning services, so that more interaction is offered to the user, alongside with novel, semantically driven inferences. One of the outstanding tasks in the previous report was to assess whether the re-engineering actually worked. In other words, in the past period we focused more resources on two aspects:

(i) finalizing the design and implementation of learning services for the revised ASPL-v2 framework, and
(ii) carrying out a comparative assessment of ASPL-v2 vis-à-vis other tools that have a similar scope and may be commonly used by the users

In terms of the former point, L3S has put a substantial effort into their DBLP++ tool. This tool essentially draws upon the well-known database of publications in computer science, and in its interactive form, it offers the user a rich, faceted interface to access the content of DBLP. Nonetheless, this rich user interface may be somewhat daunting, so it was decided to pursue also a parallel approach promoting a lightweight user interface via the ASPL platform.

The ASPL platform is essentially about associating web services with the concepts and instances from a particular ontology, which is of interest to the user. Thus, OU and L3S developed a suite of web service end points for the DBLP++ data set, and these were later complemented with a user-friendly front end – simple, Google-style user interface for querying the content of DBLP and also for making knowledge-level inferences and connection interpretations. In particular, the following web services were exposed from the DBLP data set:

- *Person's publications and interests* … a combination of a simple data retrieval (of publications) with an interpretative inference based on the publication keywords, Semantic Web Topic Hierarchy matches, etc.
- *Person's interests* … an interpretative inference based on the occurrence of keywords and phrases (also from Semantic Web Topic Hierarchy)
- *Person's community characteristics* … an interpretative inference based on the co-occurrence of co-authors, keywords and themes allowing generalizations from the individual nodes (researchers) to their collections (communities)
- *Person's co-authors and communities* … a combination of the retrieval function with an interpretative function as described above
- *Leading experts on topic* … an interpretative inference based on the occurrence of certain phrases within individuals' profiles combined with a statistics
- *Main publication outlets for topic* … an interpretative inference allowing the user to generalize from single nodes (publications and authors) to their collections (journals, conferences, etc.)

In terms of the latter area of focus, OU has developed a two-pronged methodology to assess and position tools performing a certain class of tasks – in our case, search and data retrieval tools. The outcomes of this research activity were reported in the previous sections of the report – first, positioning the ASPL/DBLP++ combination with respect to its nearer or more distant competitors, and second, assessing the user-centred effectiveness of the ASPL-v2 re-engineering, or better the effectiveness, usefulness, pluses and shortcomings of the re-engineered learning services.

Work reported in this deliverable focused mainly on the learner's interaction with resources on the Semantic Web; in particular with the semi-structured data that can be exposed to the user via domain-specific inference templates. We assessed this capability of the service-based ASPL-v2 framework in terms of assisting users with interpreting connections in the academic domain; for example, filtering leading scientists, recognizing communities of practice, or associating research topics and issues with particular publication outlets. The outcomes of a user-based study were reported, and the ASPL-v2 was found to outperform other tools – including the generic search engine aggregator Ask and semi-specialized Google Scholar.

However, it has to be noted that ASPL only performed better as long as the input from the user exhibited a fair degree of completeness. In the case of not providing full details to the framework (e.g. a part of a person's name or partial label of a research topic), ASPL-v2 has failed to translate the incomplete query to access DBLP++ collection. Here, the more generic, keyword- rather than phrase-based search tools seemed to have had an advantage allowing users to be more flexible in their query formulations.

However, in our opinion, this shortcoming of ASPL can be addressed in the remaining duration of the project

It was highly beneficial to carry out the positioning analysis for ASPL-based application, as this has alerted us to a range of dimensions a user may take in account when decided which tool they want to deploy in a particular situation. The positioning analysis enabled 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, its 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.

In fact, 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.

Therefore, we opted to take a slightly more abstract view on the task in question, and essentially position the tools a user may come across when trying to find, discover or otherwise locate a particular piece of data, information or knowledge. Another added benefit of carrying out two-pronged analysis – i.e. first positioning and then assessing user effectiveness – is that the positioning analysis gives the latter study more credibility. This is particularly visible in our earlier argument to compare 'like with like' – the positioning analysis shows us to what extent the two tools assessed on their effectiveness share the same problem space, interest scope and strategies.

Generally speaking, it was observed that ASPL turned out to be an appropriate tool when the users could make use of its specialized services implemented in its re-engineered version. In a nutshell, the users preferred to use ASPL to perform the search as mentioned in Task 1, Task 2 and Task 3 of the user evaluation study. ASPL was observed to perform better than Google Scholar on (three out of five) tasks that required a degree of synthetic or analytic thinking. Hence, if its capabilities to handle the incompleteness are improved, the tool is likely to gain a good support within a narrow domain of academic literature and interpretation of community relationships.

The outstanding issues we intend to address in the remaining life of the project are related to the areas where ASPL-v2 was observed as underperforming. In particular,

- Development of the capability to handle incompletely and vaguely formulated queries by means of tapping into tools that came into existence in the meantime (e.g. SemSearch engine may help to disambiguate an incomplete query into a set of alternative full names)

- Explore the ways of allowing users to specify additional keywords and query modifiers extending the core query (about the people or topics)

6 References

[1] Park, J. and Kim, J. Effects of contextual navigation aids on browsing diverse web systems. In Proceedings of CHI 2000, 257-264.

[2] Jimmy Lin, Dennis Quan, Vineet Sinha, Karun Bakshi, David Huynh, Boris Katz, David R. Karger. The role of context in question answering systems conference on Human Factors in Computing Systems, CHI '03 extended abstracts on Human factors in computing systems, t. Lauderdale, Florida, US, Pages: 1006 – 1007, 2003, ACM Press, New York, NY, USA.

[3] B. J. Jansen, A. Spink, and T. Saracevic. Real life, real users, and real needs: a study and analysis of user queries on the web. Information Processing and Management, 36(2):207–227, 2000.

[4] Nenad Stojanovic: On the Role of Query Refinement in Searching for Information: The Librarian Agent Query Refinement Process. 4th International Conference on Web Information Systems Engineering (WISE 2003), 10-12 December 2003, Rome, Italy. IEEE Computer Society 2003.

[5] Lotfi A. Zadeh. From Search Engines to Question-Answering Systems - The Need for New Tools. Advances in Web Intelligence: First International Atlantic Web Intelligence Conference (AWIC 2003), Volume 2663/2003. Lecture Notes in Computer Science. Madrid, Spain, May 5-6, 2003.

[7] Gilad Mishne and Maarten de Rijke. Query Formulation for Answer Projection. Advances in Information Retrieval, Volume 3408/2005. Lecture Notes in Computer Science, Pages 523-526. Springer Berlin / Heidelberg. 2005.

[8] Efthimis N. Efthimiadis. Query Expansion. Annual Review of Information Systems and Technology (ARIST), Williams, Martha E., (ed.) vol. 31, pp. 121-187, 1996.

[9] Michael K. Buckland and Fredric Gey. The relationship between Recall and Precision. Journal of the American Society for Information Science vol. 45, no. 1, pp. 12-19, 1994.

[10] David J. C. MacKay. Information Theory, Inference & Learning Algorithms. Cambridge University Press, 1st edition, June 15, 2002. ISBN-10: 0521642981.

[11] Lopez, V., Uren, V., Motta, E. and Pasin, M. AquaLog: An ontology-driven question answering system for organizational semantic intranets. Journal of Web Semantics, Vol.5, 2, (72-105), (2007). Elsevier.

[12] Popescu, A., M., Etzioni, O., Kautz, H., A.: Towards a theory of natural language interfaces to databases. In Proceedings of the International Conference on Intelligent User Interfaces, Miami, FL, USA, Jan. 12-15 (2003): 149-157.

[13] I. Androutsopoulos and G. Ritchie and P. Thanisch. Masque/sql: An Efficient and Portable Natural Language Query Interface for Relational Databases. Proc. of the 6th International Conference on Industrial & Engineering Applications of Artificial Intelligence and Expert Systems, 1993.

[14] Min Wu, Mingyuan Duan, Samira Shaikh, Sharon Small and Tomek Strzalkowski. ILQUA--An IE-driven question answering system. In *Proceedings of TREC 2005,* Washington, D.C, 2005.

[15] Lei, Y., Uren, V., and Motta, E. SemSearch: A Search Engine for the Semantic Web. Managing Knowledge in a World of Networks. In Proceedings of 15th EKAW Conference, Podebrady, Czech Republic, October 2006.

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

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

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

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

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

[21] 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.

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

6.1 Related deliverables

Dzbor, M. and Stutt, A. (2005). D3.3.3: Prototype of advanced learning platform (ASPL-v1), Deliverable report for work package WP3.3, Knowledge Web Network of Excellence, delivered in July 2005.

Stutt, A., Dzbor, M., Maynard, D., Ronchetti, M. and Motta, E. (2005). D3.3.5: Evaluation of prototype (ASPL-1), Deliverable report for work package WP3.3, Knowledge Web Network of Excellence, delivered in January 2006.

Dzbor, M., Diederich, J., Rajpathak, D. and Dehors, S. (2006). D3.3.6v1: Report on the current status of ASPL, Deliverable report for work package WP3.3, Knowledge Web Network of Excellence, delivered in January 2007.

Dzbor, M. and Diederich, J. (2007). D3.3.8: Final demonstrator of ASPL and its components, Deliverable software demonstrator concluding work package WP3.3, Knowledge Web Network of Excellence, expected end December 2007.

6.2 Related publications

At the time of releasing this report there is no explicit publication covering the work discussed in it; i.e. evaluation of ASPL. However, the authors are shortening the report and adapting it, in order to make a submission to a journal in early 2008. Publications related to ASPL generally include the following:

Journal articles

- Dzbor, M., Motta, E. and Domingue, J. (2007). Magpie: experiences in supporting semantic web browsing. Journal of web semantics: Science, services and agents on the World Wide Web, 5 (3). pp. 204-222. ISSN 1570-8268

- Dzbor, M., Stutt, A. and Motta, E. and Collins, T. (2007). Representations for semantic learning webs: Semantic Web technology in learning support. Journal of Computer Assisted Learning, 23 (1). pp. 69-82. ISSN 0266-4909
- Dzbor, M., Motta, E. and Stutt, A. (2005). Achieving Higher-level Learning Through Adaptable Semantic Web Applications. International Journal of Knowledge and Learning, 1 (1/2). pp. 25-44. ISSN 1741-1009

Conference and workshop papers

- Dzbor, M. and Motta, E. (2007). Semantic web technology to support learning about the semantic web. In: Artificial Intelligence in education - building technology rich learning contexts that work. Frontiers in Artificial Intelligence and Applications, 158. IOS Press, pp. 25-32. ISBN 1-58603-764-1 & 978-1-58603-764-2, pp. 25-32.
- Dzbor, M. and Motta, E. (2006). Study on Integrating Semantic Applications with Magpie, 15th Conf. on AI Methodology, Systems & Applications (AIMSA), Varna, Bulgaria, eds. Jerome Euzenat and John Domingue, pp. 66-76, Springer Verlag.
- Pasin, M. and Dzbor, M. (2006). A Task Based Approach to Support Situating Learning for the Semantic Web, Workshop: Workshop on Applications of Semantic Web Technologies for E-Learning, SW-EL-06 at Adaptive Hypermedia 2006, Dublin, Ireland.

Book chapters and keynotes

- Dzbor, M. (2006). Evolution of the idea of the Semantic Web and its implications on practice. Invited keynote at Znalosti 2006 - Annual conference of Czech and Slovak Societies for AI in knowledge capture, etc., Hradec Kralove, Czech Republic, February 2006.
- Dzbor, M., Motta, E. and Stutt, A. (2005). From Knowledge Repository to Knowledge Space. Book chapter in: Intelligent Learning Infrastructure for Knowledge Intensive Organizations: A Semantic Web Perspective, eds. Lytras, M.D. and Naeve, A., Hershey PA: Information Science Publishing.