D3.3.5 Evaluation of prototype (ASPL-1)

Coordinator: Arthur Stutt (OU)
With contributions from:
Victoria Uren (OU), Diana Maynard (USFD), Marco Ronchetti (UniTn), Dan Cristea (Universitatea "Al. I. Cuza"), Martin Dzbor (OU), Enrico Motta (OU)

Abstract.
EU-IST Network of Excellence (NoE) IST-2004-507482 KWEB
Deliverable D3.3.5 (WP3.3)

This report will describe the evaluation of ASPL-v1, summarize the results and highlight possible changes to be incorporated into version 2 of ASPL (due in M36).

Document Identifier:	KWEB/2005/D3.3.5/1.3
Class Deliverable:	KWEB EU-IST-2004-507482
Version:	Final
Date:	February 9, 2006
State:	Final
Distribution:	Public

Knowledge Web Consortium

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

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
L697ZF Liverpool
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
S14DP Sheffield
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:

USFD
UniTn
Universitatea "Al. I. Cuza" Romania
L3S
INRIA

Changes

Version	Date	Author	Changes
1.0	14-11-05	Arthur Stutt	Draft version
1.1	9.1.06	Arthur Stutt	Draft 2
1.3	25.1.06	Arthur Stutt	Final

Executive Summary

This report will describe the evaluation we conducted of ASPL-v1, summarize the results and highlight possible changes to be incorporated into version 2 of ASPL (due in M36).

Since we needed to evaluate the utility of ASPL as a learning system in addition to other factors such as usability, the strategy we developed was two-pronged:

- Questionnaires/feedback from peers.
- User study.

It is clear from the questionnaires that users generally enjoy using ASPL and see it as a valuable tool. There were few critical comments and most of these were about problems which are relatively easy to fix and which we have already dealt with or will deal with in future. On the other hand there were many positive remarks which were very encouraging. While ASPL is not yet a finished system, it certainly seems to be going in the right direction. We have also received a number of interesting suggestions for future development which we will be considering for ASPL-v2.

While in general the tool is working as expected and is obviously well appreciated by users who seem to enjoy using it, the results of the user study indicate that ASPL does not as yet fully support even simple learning tasks such as building a citation list. In fact the results indicated that ASPL users performed slightly less well than non-ASPL users although the results were not statistically significant.

It is clear from this that much more needs to be done in terms of clearly thinking about what learning tasks involve and how we can support them with ASPL. One direction indicated by both the questionnaires and the user study is the development of services which filter lists of search results from sources such as REASE using criteria derived from individual or community profiles.

Contents

Introduction

According to the revised KnowledgeWeb Technical Annex, a report on the evaluation of the first version of ASPL (ASPL-1) is required at the end of the second year of the Network's existence (M24). Accordingly, this report will describe the evaluation we conducted, summarize the results and highlight possible changes to be incorporated into version 2 of ASPL (due in M36).

The prototype version of ASPL (ASPL-v1) was completed in August 2005 and made available at http://kmi.open.ac.uk/projects/kweb/ along with a 'Read me' document and the deliverable report (D3.3.3) which includes the ASPL manual.

It is important to note at the outset that we do not feel that ASPL is mature or complete enough for really meaningful results from any comparison with other such tools or with the situation where no tool is used. Our main objective in conducting these evaluative studies is thus to gain an understanding of what users perceive to be the main problems with ASPL as well as what features users think should be added.

Evaluation method

We developed a strategy for evaluating ASPL based on the pioneering work of Maynard and Uren on the evaluation of semantic systems (Maynard, 2005; Uren et al., 2005). Fortunately, both of these studies look at the role of Magpie (the semantic technology underlying ASPL) though they both look at its role as an annotation tool rather than, as in the present report, as a system to assist learners.

Since we need to evaluate the utility of ASPL as a learning system in addition to other factors such as usability, the strategy we have developed is two-pronged:

- Questionnaires;
- User study.

The questionnaires were developed in the light of usability and other criteria developed by Maynard (2005) with additional questions on the pedagogic impact of ASPL derived from the work of Bloom (1956) on higher level cognitive tasks (see KnowledgeWeb deliverable D3.1.4, section 7, Pedagogical Requirements). The user study was largely based on the experimental design developed by Uren (Uren et al., 2005) for her study of the effect of different Magpie lexica on the identification of important concepts in news stories.

While we are unconvinced that ASPL is mature enough for a complete formal evaluation we felt that the questionnaires (though adequate for informal feedback) were unlikely to indicate to what extent ASPL was able to support a particular task. We thus created the formal user study which recorded the results of a control group and a group using ASPL in completing a common academic task – the creation of a citation list.

The Questionnaire

The questionnaire (Appendix 1) had four main goals:

1. *Downloading, installation and basic operation.* To establish whether the current instructions for downloading ASPL and associated documentation are enough for users to get the system running and operating as intended.

2. *Usability.* To get some impression of the usability of the system.

3. *Use of Semantic technologies.* To get some idea of the adequacy of ASPL's use of semantic technologies (especially ontologies) in the learning context.

4. *Pedagogy.* To get some feedback on the pedagogic aspects of the system. While the user study is intended to produce results which show how well ASPL operates as a learning tool, we added some questions to the questionnaires which are intended to garner the opinion of peers such as lecturers and senior researchers rather than learners. In effect we asked them how well they thought ASPL *might* satisfy certain criteria.

The questionnaire was developed over a period of two months in August-September 2005. The initial design for the questionnaire was based on the analytical criteria developed by Diana Maynard at the University of Sheffield. It was circulated internally for comments and revised in the light of these. We received valuable feedback from Sheffield and L3S. The final version was posted on a variety of general (kweb-edu, REWERSE, ARIADNE, PROLEARN) and institutional lists (Sheffield, Open University, L3S) in November 2005.

Statistics and Analysis

We received 34 completed questionnaires, mostly from users who had completed the user study. We originally intended that the questionnaires would be completed by fairly casual users. However since we received very few responses to our canvassing efforts and since we came to the view that the participants in the user study would have greater familiarity with the tool and hence have more apposite comments, we made no further canvassing attempts.

The questionnaires were recorded in a spreadsheet which was used to produce relative percentages for a variety of question types. We also solicited and recorded user comments, the most significant of which we have included below. Both were used to produce a set of recommendations for future ASPL development.

Results of Questionnaires

In the following we have included, for each section, a table for the different types of user questions (i.e., those soliciting measures of satisfaction, agreement, adequacy or importance) followed by the comments made in response to more open-ended questions. We conclude each section with a list of important points from these results and possible actions for the ASPL development team. The most important of these will be included in the overall Conclusions section. As Figure Q-1 indicates, most of those questioned gave positive or very positive responses to these measurement questions. Figure Q-2 shows overall results for yes/no type questions.

Since the measurement questions have a variety of formats (e.g., Unhappy, Happy, Very happy; Yes, No, Don't know) we indicate which type it is using abbreviations given in the following table:

Type	*Alternatives*
u-h-v	Unhappy, Happy, Very happy
y-n-d	Yes, No, Don't know
n-r-v	Not very, Reasonably, Very
i-a-e	Inadequate, Adequate, Excellent
n-a-m	Not sufficient, About right, More than needed
tf-a-tm	Too few, Adequate, Too many
u-i-v	Unimportant, Important, Very important

Table Q-0: Types of questions

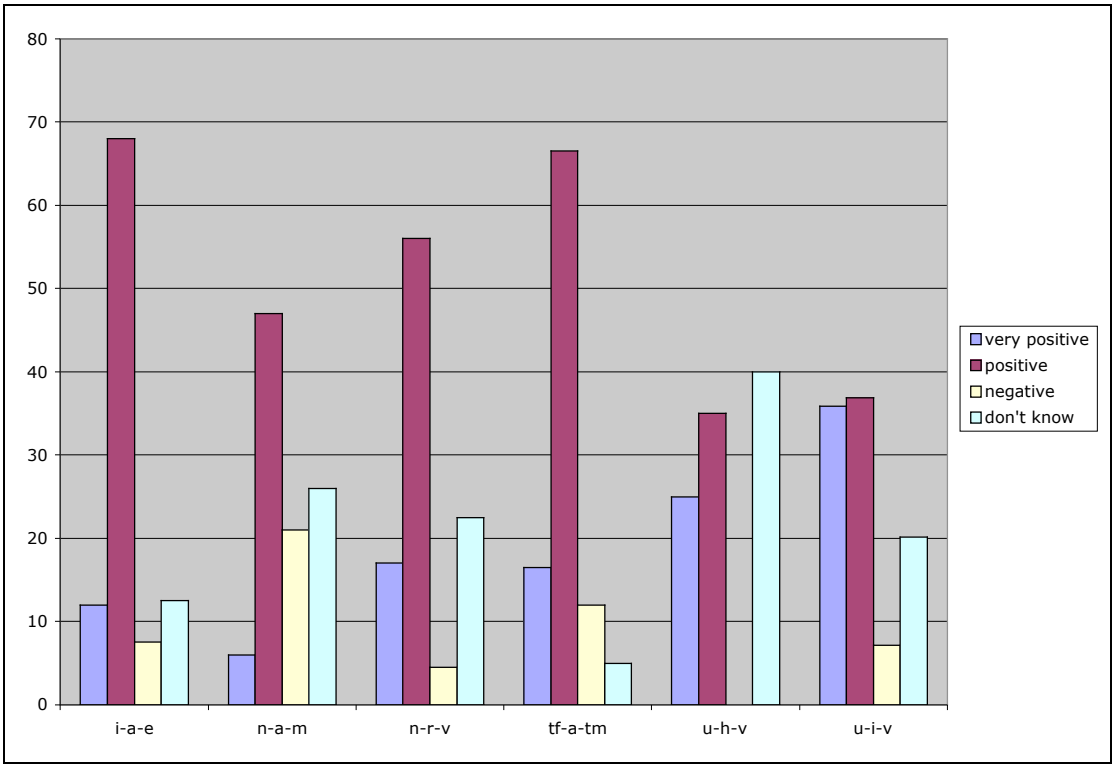


Figure Q-1: Summary of responses to above question types as percentages (Yes/No responses shown in Figure Q-2)

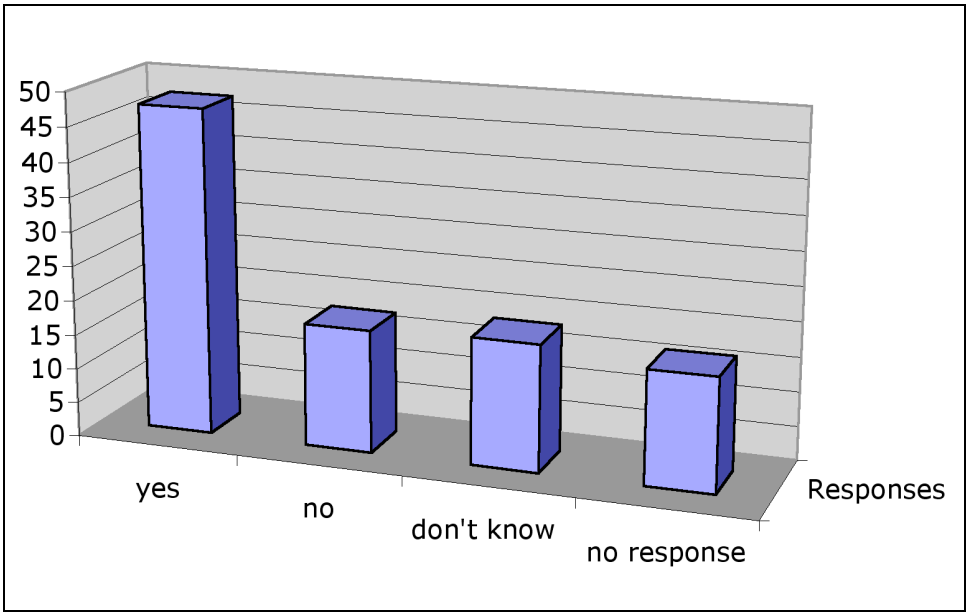


Figure Q-2: Overall Yes/No type responses shown as percentages

Downloading, installation and basic operation and usability

Qu	*Text*	*Type*	*%*	*%*	*%*	*No response*
A-1a	How happy were you with the download site?	u-h-v	0	38	18	44
A-2a	How happy were you with the installation and set-up instructions?	u-h-v	0	32	32	36
A-5	Would the addition of an uninstall option be useful?	y-n-d	41	12	12	35
A-7	Please indicate how useful you found the documentation/manual?	n-r-v	0	59	9	32
A-8	Would the addition of a help facility be useful?	y-n-d	56	9	9	26
A-9	Was the overall performance of ASPL	i-a-e	3	68	12	17
A-10	Please indicate how well designed you felt the system interface was?	n-r-v	6	65	15	14
A-11	Did you find the ASPL icons clear and legible?	n-r-v	6	38	38	18
A-12	Were the ASPL tooltips for the ontology categories sufficient?	n-a-m	21	47	6	26
A-13	Would you like to be able to modify fonts, colours, etc. more than is currently available?	y-n-d	21	41	24	14
A-14a	Are you satisfied with the accessibility of ASPL?	y-n-d	50	12	24	14
A-15a	How useful was the ability to customize the Magpie plug-in (via User preferences)?	n-r-v	6	62	6	26

Table Q-1

A-1b. Have you any suggestions as to how the download site might be improved?	Separate download area – eventually a link from main navigation should point to download area Is good but the link's font should be greater for a best view. The font should be bigger and easier to read. The download link had a small font.; should be more greater. For example, I accessed the site because I wanted to download the ASPL kit and that's why I think that, even if it is red, the download link should be made with a larger font. In this way, it will be more visible everybody will see that this kit may be used on a personal computer. Or it should be written on the top of the page. There isn't a link to download the version for Mozilla. It would be useful to make more accessible the Mozilla/Linux version of the tool. I was a bit confused by the option to "download only public ontologies". It wasn't clear if and why these might be needed, for example if these are extra ontologies to the ones already given or whether they are required to run the ASPL (if you don't have your own). Also the use of the word "only" confused me – not sure if it meant only the public ones, or only the ontologies and nothing else (both interpretations would seem to be obvious anyway). A quick link to downloads (the link is a bit buried).
A-2a. How happy were you with the installation and set-up instructions?	A warning to close all windows of Internet Explorer before running setup. This should be included in installer since not everyone reads the readme file. I've noticed that during installation, at one step, if I don't select the "Create Magpie group icon in the Start Menu" option, the installation fails. It took ages to download and open the User Guide from the web (on a fast connection but slow machine). You also have to scroll through a large chunk of the documentation to find the installation and setup instructions. It would be better to have these in a separate file somewhere where they could be more easily accessed. Otherwise it seemed pretty straightforward. I wasn't sure what to do with it once installed. Some quick start instructions would be useful. When trying to overwrite a previous version of Magpie, it fails. It would be better if it was automatic overwriting.
A-3. Did you have any problems with the browser/ASPL combination?	I did not close all browser windows while running setup and at the first use of Magpie the browser crashed. Even reinstalled, the browser shut down unexpectedly while using Magpie. I did not found a pattern yet. Had problems on IE v 6.0 on Windows XP with the toolbars. It may not have been caused by Magpie, but the toolbars for Magpie and MSN were reversed so that clicking on the MSN toolbar name brought up the Magpie one and vice versa. I haven't yet replicated the problem after uninstalling and reinstalling ASPL/Magpie however. Magpie doesn't open in pop up boxes automatically – REASE uses pops quite a bit – so you have to find the ontology again if you want to keep using Magpie. Mozilla/Linux Debian Sarge, it is quite slow.
A-4. Did you have any problems loading lexica/ontologies? If so,	Internet Explorer has been restarted once. No, but I have used Magpie before. I think many novices would have

briefly list them.	difficulty with this task. For ASPL purposes the sem-web-studies ontology should be loaded by default perhaps?
A-6. Please list any problems you had when you tried to access services from the pop up menu (i.e., when accessing the plainmoor.open.ac.uk ASPL server).	Services from pop up menu on highlighted items are not working. For certain entities, when I access the service "Find in ACM library..." the search is slow and my browser blocks . I'm still not sure what "trigger service alerts" are and whether I would want to have them enabled or not! It doesn't seem to do anything. Loads of problems accessing some of the services. e.g. one day I just had connection refused to the ASPL server. Very few of the community links worked. When I used it the "search in digital libraries " link also didn't work. Slow when finding people related Encoding special characters did not work properly on targhee (Jerome Euzenat's name looks very strange…) co-citation and co-authors did not work (always empty set) the difference between 'SW activity' and 'SW technology' (and 'other SW topics) is not very clear no tool tips in mozilla version
A-14b. Do you have any suggestions for improving accessibility?	As far as I saw it you can access Magpie options only using the mouse. It would be nice if keys can be assigned to different actions from Magpie menu, especially for people that are using the keyboard. Several problems with accessibility. It is easy to click accidentally on the Magpie symbol instead of the arrow on the icon. This then leads to the Magpie page in the browser, which takes ages to load, and clicking the Back button (on IE v6) didn't work so you then have to manually retype the address of the page you were on into the browser. When a highlighted item is too close to a link, right clicking on it does nothing unless you are very exact with the mouse. When an annotation type has been selected, and then a link is followed to another page, when you return to the original page you have to reselect the annotation type to be highlighted again. It seems strange (and is a bit annoying) that this isn't remembered. The interface was quite crowded. As a novice it was hard to know where to go first. The button labels were not very self explanatory & I couldn't deduce what the "classes" were from what was being highlighted by the tool. As mentioned before the reliance on pop-ups conflicts with how Magpie works. I'm not sure what accessibility means in this context – if you mean e.g. visually impaired users I think they might have trouble but you would need to have a proper assessment done by experts It might be better if the toggle buttons from the toolbar were kept pressed when the browser is loading another page (following a link). Some of the options must be better explained to understand what we can do with a tooltip or an option. Implement some shortcuts for various tags.
A-15b. What other forms of Magpie plug-in	An option to keep history of the latest ontologies loaded and the no of items in history.

customization would you like to see?	option to load ontology directly from history. option to highlight automatically the terms for all categories for An ontology without clicking on a category Customization personalized in relation with the knowledge of the user and the subject he is looking for. Save colours if you change them for specific annotation types. Enable/disable option.

Table Q-2

Important Points

- An uninstall option is clearly needed and in fact already implemented as a result of early responses to questionnaire.
- There is no great desire to modify the interface.
- We are aware of and have fixed the installation bug to do with overwriting – the new installation instructions will detail this fix.
- We are working on the pop-up window/need to renew categories problem. The most recent version of ASPL does not require reloading of ontologies. However it doesn't automatically highlight classes in new windows.
- The issue of category labelling will be dealt with below as will personalization.

Actions

- ✓ Documentation/manual could be improved. We could consider providing a separate manual – i.e., not as part of a KnowledgeWeb deliverable. We already have a separate document with download instructions.
- ✓ We should consider providing a help facility.
- ✓ We still need more explicit installation instructions.
- ✓ We should consider a bigger font for ASPL download.
- ✓ We should have downloads for other versions/other platforms.
- ✓ We need to provide an explanation of the "download only public ontologies" option.
- ✓ The possibility of default ontologies pre-loaded will be considered.
- ✓ We may need to consider more robust servers/maintenance cycles to ensure that services are accessible at all times especially when ASPL is used for course-work.
- ✓ Other suggestions (e.g. text coding, precision needed when clicking, keyboard access) will be considered for future versions.

Discussion

It would seem that apart from the relatively minor issues listed above, most users found ASPL easy to install and use. Obviously, as with any piece of software undergoing development, we will need to ensure that future versions take account of these points. Most of all we need to ensure that the software is relatively bug free and that the many servers used are as robust as possible.

REASE usability

Qu.	Text	Type	%	%	%	No response
B-1	Does the user need to know that different services access different databases?	y-n-d	32	38	18	12
B-2	Did you find it easy to interact with REASE to access material pointed to by ASPL services?	y-n-d	62	12	15	11
B-3	Did you find it easy to set up an account on REASE?	y-n-d	38	3	32	27

Table Q-3

B-4. List any other problems you had with REASE and/or the combined ASPL/REASE system and any suggestions for improvement.	Some of the words underlined in the text were of no relevance and sometimes 2 colours covered the same word without letting the user notice that. "Community of referrers" is I think a really odd term and I still don't really know what it means! Clicking on "Find in digital libraries" I got an error message: Unknown service requested: 'KW-FIND-OTHER-LIBRARY' Rease works very well, better than other sites in the same category. One problem could be that the ranking was not always accurate and that I couldn't find all the data I needed in some cases. I didn't need an account for this task. Some of the menu names of the magpie services are no particularly clear about what they do More details about documentation and more structurate them … I've encountered some dead documents' links that prevent me from accessing the searched documents. I am not familiar with this site, so I have problems with the usability of REASE website. 1. There is not clear description about the content in REASE from the suggestion of ASPL. 2. Too many steps supposed to be done, just only take a look what kind material are they? For related keywords, REASE suggested so many overlapping items that may be not so specific for certain subject. I did not see any login/register button until REASE asked me for login. Low visibility (now I can see it in top-right corner: "User: Anonymous" and a login link), but it was not easy to find it. There's too little material on REASE, and some of material that is there is empty.

Table Q-4

Important Points

- There's a certain equivocation about whether those questioned feel the need to know about data sources. Nonetheless we will consider this for future versions.
- In general there is a positive view about REASE's usability. However the number of Don't knows and no answers indicate that a lot of participants didn't try to set up a REASE account.

Actions

ASPL:

- ✓ We need to look at overlapping highlighting.
- ✓ While we are happy with the phrase "Community of referrers" perhaps we should have more in the manual on some of the terms we use.
- ✓ We should have more in the manual on what to expect from the services which access REASE.

REASE:

The relevant comments (about ranking, dead links, imprecise classifications, clearer login procedures) have been forwarded to the REASE maintenance team.

Discussion

We feel that the most significant critical comment about REASE concerns the relative paucity of REASE resources. Since the population of REASE is an ongoing task which depends on the goodwill of the KnowledgeWeb community, this will improve over time. However, it is really up to the KnowledgeWeb community to make use of this as their favoured publication platform for educational resources.

Semantic technologies

Qu.	*Text*	*Type*	*%*	*%*	*%*	*No response*
C-1	Did you feel that the four categories of highlighted concept (SW technologies, SW activities, Other SW topics, Community) were: Too few, Adequate, Too many	tf-a-tm	18	68	12	2
C-2	Would 8 (limit of colour differentiation) be enough?	tf-a-tm	6	65	21	8
C-4a	Do you think that there should be different sets of categories for different users (e.g., learners, developers, researchers)?	y-n-d	41	32	18	9
C-5	Did you feel that user access to/awareness of the underlying ontology was adequate?	i-a-e	12	68	12	8
C-6	Did you feel that users should be able to see where highlighted instances fit in the class hierarchy?	y-n-d	68	15	6	11
C-8	Do you think that it would be useful to have class instances from more than one ontology highlighted at a time (currently only one)? For example a learner might	y-n-d	65	3	26	6

	want to have instances from classes in a pedagogic/instructional ontology highlighted as well as those from a domain ontology.					
C-9	How important do you think it is for ASPL to have a way for users to select which classes should make up categories and which instances should be highlighted – i.e., for creating a private ontology/view?	u-i-v	6	47	29	18
C-10	Currently ASPL ontologies are written in OCML. How important do you think it is that ASPL can make use of other formats such as XML/RDF/OWL?	u-i-v	3	26	50	21
C-11	How important do you think it is that ASPL should be able to import publicly available lexica/ontologies in XML/RDF/OWL?	u-i-v	3	29	53	15
C-12	How important do you feel that it is for ASPL to have a web-service architecture with the capability of using any arbitrary web service?	u-i-v	15	41	24	20
C-13	How important do you think it is for ASPL to have a set of easy to use tools for ontology/lexicon creation?	u-i-v	9	35	32	24
C-14	How important do you think it is that ASPL should provide a tool which can produce a Magpie lexicon from any ontology submitted to it?	u-i-v	9	44	26	21
C-15	How important do you think it is for ASPL to have a set of easy to use tools for creating the services accessed from highlighted items?	u-i-v	9	41	29	21
C-16	How important is it for ASPL to have a means for automatically maintaining lexica/ontologies – e.g., populating ontologies with new instances and relations?	u-i-v	3	32	44	21

Table Q-5

C-3. Can you think of better top-level categories for Semantic Web Studies?	SW applications I think that the existing categories are ok, but it would be useful if every category button had a "description" option, defining the entities it will highlight. The ones covered by magpie were indeed relevant at least for the semantic

	web studies. I don't like the term Community as it's not really clear what it means. Why not just say "people" if that's what it means? Activities is also rather unclear. Given that I got thinks like "semantic web", "knowledge" and "links" highlighted I am not really sure what this category is about. Community is fairly clear (but because the lexicon is small it doesn't find some obvious people, like Tim Berners Lee, and I didn't guess it was meant to do organizations as well) The other 3 I couldn't sort out which covered what – they might as well have been one button – all had the same services attached I didn't think about it, but when I first looked I didn't know what was all about those buttons and the classes specified. Actually, I do not know but, Semantic web integration with video audio streaming services. Persons SW Languages SW projects Should be improved though I can't come up with a better category
C-4b. If so, can you briefly list them (different categories for different t users)?	Students, developers, researchers and custom users It is almost always a bad idea to have main links on a website that take you to different subsections of the site depending who you are (e.g. researcher, academic, student), because there is too much overlap and because you cannot predict what a particular type of person wants. It is generally better to direct people in terms of their goals ("are you looking for X, Y or Z?" rather than "are you X Y or Z?". The same applies in this case. So you might have different sets of categories depending on the task you might want to do, e.g. someone creating a bibliography might have different needs from someone who wants a better understanding of semantic web terms or wants to know who else is working on a similar topic to themselves. This could be still making use of the same ontology. Learners probably need fewer (like me). Perhaps experts need more but I can't say what. students teachers researchers Developers for technicalities Researcher for a combination Learner for basics (as a learner or researcher in initial phase) I would prefer to have some identification of the type of the material (link to material), like introductory lecture, research paper, guidelines, etc… For instance can put cites for researchers, like citation index. 1. Newbie 2. Students 3. Researcher 4. Developer

	5. Industry Elementary, familiar, advanced… No. I really think I am not prepared for that. There should be done a study of the way each user searches information and the type of information that is necessary for each of them. I can't really enumerate a list, but I suppose for learners more general categories can be defined, while for more expertise or specialised users, more specific categories can be defined. Researchers might want to distinguish between names of concepts and names of concrete instances
C-7. How would you show more of the ontology structure?	This should be an option available only for interested users – be able to check it in preference area. The user should be able to see it as a graphical collapsed tree and having the option to expand it. I think is good to use an index structure or something like this. Using different colours. Relations with hierarchically lower and higher situated terms. Link to a popup window with the ontology, probably. The roll over help could show something more useful – even a list of high level classes might help In my opinion, the ontology structure should be in a tree-view form. By using a tree control in a right panel of the browser window. In an enclosed window, like the history option. Like a tree Predefined library of similar topics Might be nice also to have relevance scale to every category on the left (generally useful mode to choose what to read) Graph Just printing information about the results and the sub-categories to which they belong (where they are in the class hierarchy). Maybe another option to show the hierarchy. Perhaps open (on-demand) a new web page with a graphviz picture of the ontology?

Table Q-6

Important Points

- It is interesting to note that those questioned were more engaged with this set of questions as indicated by fewer no responses. Perhaps this is to be expected given the users' background.
- With regard to categories (C-1 and C-2) it seems that while most are happy with four categories, some would like more than four but less than eight.

- The analysis indicates that there is a slight majority for different sets of categories for different types of user (personalization). This is one of the facilities we intend to implement.
- C-5 and C-6 indicate that users would like to see where classes fit without undue details of the ontology.
- C-8 responses indicate a desire for multiple ontologies. Again this is one of the facilities we intend to implement.
- C-9 responses indicate a desire for private views of ontologies. We will consider this for future versions.
- C-10 and C-11 indicate that users want interoperability with other formats and ontologies available on the Web. Again we will consider this for future versions.
- C-12 indicates a desire for a full web-service architecture for ASPL. This is already planned for future versions
- C-13 to C-16 indicate that developer tools should be available. This will be considered for future versions of ASPL. The clearest score here (44%: Very Important) is for automatic lexica/ontology population and maintenance.

Actions

- ✓ We certainly need to think about the names of the categories and how to reveal what they mean and what they contain. We have already implemented a function which allows the user to search for a class in the categories. However, in the next phase of development we will be looking at the ontologies used in ASPL. This will feed into changes here.
- ✓ We also need to think more about how to personalize ASPL for different sorts of users. Should we, for example, have different sets of categories or should we have a learning service which offers personalization or both? A discussion of which classes to include can be considered as part of the next - ontology development – work package task.
- ✓ We definitely need to show more of the ontology structure. Our favoured solution at present is a pop up window with class details but other options will be considered. We could combine this with some tree structure visualization.
- ✓ We need to ensure that ASPL is fully interoperable with ontologies in the standard formats and that some form of automated population and maintenance is implemented.

Discussion

This section was intended to check that our own perceptions of what was required for a satisfactory ontology based tool were correct. In general here, our predictions about responses were confirmed. However we were slightly surprised that users were fairly equivocal about the need for different categories for different types of user (yes: 41%; no:32%). We were also struck by the number who thought that it was important or very important to have automatic population and maintenance (76%).

Pedagogy and general comments

Qu.	*Text*	*Mode*	*Mean*
D1	Knowledge Recall	4	3.8
	Comprehension	4	3.3
	Application	3	3.3
	Analysis	3	3.3
	Synthesis	2	3
	Judgment/evaluation	3	3.4

Table Q-7: Please indicate how well you feel that ASPL assisted with/could assist with the following cognitive skills? (1 = not very; 5 = extremely well)

Qu.	*Text*	*Mode*	*Mean*
D2	Semantic knowledge	5	4.2
	Factual knowledge	3	3.3
	Procedural knowledge	3	3.1
	Metacognitive knowledge	3	3.1

Table Q-8: Please indicate how well you feel that ASPL assisted with/could assist with the following kinds of knowledge? (1 = not very; 5 = extremely well)

Qu.	*Text*	*Type*	*%*	*%*	*%*	*No response*
D-3	Do you feel that ASPL makes learning more efficient – e.g., Could ASPL help learners to complete a learning task such as creating a citation list more quickly?	y-n-d	68	0	15	17
D-5	Would you like a Learning Service (i.e., a service which can be associated with highlighted instances) which gathered information on persons from the World Wide Web and, if members of KnowledgeWeb, from the KnowledgeWeb portal?	y-n-d	59	6	21	14
D-6	Would you like a Learning Service which ranked results of searches?	y-n-d	69	9	12	10
D-7	Would you like a Learning Service which knows about the learner and can recommend new topics/papers/people etc? I.e., a form of personalization.	y-n-d	71	15	0	14

Table Q-10

D-3	I suspect that initially it will slow things up (learning a new task always does). A nice thing would be if there were some simple way to collect the citations (pass them to a file) without doing lots of cutting & pasting if ASPL supports the functional tasks well the user will be able to concentrate on higher level tasks (e.g. remembering what that thing they just saw was & making a connection.) The selection stage is the hard part so clever user guided rankings might also facilitate things.
D-4. What other external sources would you like ASPL to have access to (e.g., DBLP)?	Google or other search engine English Dictionary CiteSeer, DBLP, Google ACM Portal & CiteSeer are what I use most. The citation support is the critical feature in both cases

	http://scholar.google.com IEEE Explorer DBLP Citation Index I want ASPL have access to other documents like .doc .pdf etc. for better search and understanding of the document. The ASPL should have access to other external sources like pdf document Access to other digital libraries and learning resource such as: Ariadne, Mellot… Google special query: "define: semantic web" to gain several definitions about that terms
D-8. What other learning services would you like to see included in future versions of ASPL?	Is better, I think, in the contextual menu use search over the internet, like google, altavista etc. It would be interesting to have a service that samples some practical applications (examples), where it is possible. Ranking results and personalization would be very helpful. Show related concepts (i.e. things close in the ontology) In contextual menu I like to see a way to modify preferences a aspl Animated characters. For the "Explain Concept" functionality, ASPL should be able to explain by terminology rather than by word. Example, it can find a definition of "Semantic Web" rather than the word Semantic and the word Web. Can change the term: now ASPL put "semantic" and "web" as two terms. And in this trial I want to put them as one term. Because it'll effect to search result
D-9. What other functionality (apart from Magpie's semantic browsing and associated learning services) would you like to see included in future versions of ASPL? For instance we have considered adding a tool which allows learners to develop a graphical representation of their learning tasks.	I think is useful adding a tool which allows creating a summary of the items searched. Open/Save in other format like XML. Compatibility with Linux. A graphical representation would be good and also a tool that would allow learners to find a word they choose in texts with a percentage related to the number of apparitions of the word in these texts and graphics related to this. Yes, it is a good idea. Also, a "memory" of the learning task, as well as a forecasting of interesting links/articles considering the learning task would be agreeable. Something for building a personal bibliography which can be output as BibTex or for EndNote – to fit ASPL with core tools I consider very useful adding a tool that make a kind a "glossary" of user's learning tasks That's a great idea. It would be good if there were a history of the web pages visited. Animated characters.

	Web services APIs
D-10. Add any critical comments or positive suggestions as to how the system might be improved.	An option to highlight terms from all categories at once can be useful. A service to see the definition of a highlighted term – term's exact definition from a dictionary The plug-in is a useful tool, but is necessary adding a few tools, accessibility option like I suggested thought the document. I saw that the system highlights not only entire words, but also portions of words, and I thing that's not good; for example, from the word "Research", "search" is highlighted as a SW activity, or from "United Kingdom", "dom" is highlighted as Other SW topics. In magpie…the colours should be chosen in such a way that we could be able to see if one word is covered by 2 colours. It is a great and useful idea and I would like to see it developed. More data! Better lexicon – the "community" part particularly seemed a bit parochial – some obvious semantic web people missing. The categories were hard to interpret too. Magpie needs to start automatically in pop ups Critical comments: The magpie plugin contain a shortcut to site: www.kmi.open.ac.uk/project/magpie that i consider not necessary Browsing history with page ranking from the most important hierarchical page. Why using IE? Instead of Firefox, that has already proven in several OS Related keywords search may lead to quite similar list of resource, maybe it's advantage, but rather constraint and it's not easy to find other interesting resource unless you point 'next page' several times. Maybe we can turn on and off Magpie anytime we want so that we can start a 'normal syntactic' search with ease. Another point is that semantic is subjective and local, thus to provide semantic search service should meet the need of end users so I think creating and importing ontology as the database of Magpie would be very important. I think the essential factor is improving the efficiency (it runs quite slow).
Any other comments or suggestions	I think ASPL is a great and very useful tool especially for semantic web and I hope to see a new version on the market very soon. In some pages opened by ASPL, I observed that, when clicking two times on one of the four highlighting labels, it closes the pages and all the other pages opened by ASPL. The system has a lot of good parts but it at an early stage of development. I'm not sure you know what kind of user you want. The task I tested is for students but the interface is complex and might be more suitable for experts.

	I wait for the next version. Is a great tool, so could be nice extend it to other browser and use it sources like Google Scholar.

Table Q-11

Points

- Question D-1 had fairly inconclusive responses. Although users seemed to have felt that ASPL could help with these cognitive tasks they didn't distinguish among them. We could say (as we might have predicted) that users didn't think ASPL was much help with *synthesis* and more help with *recall* and *comprehension.*
- D-2 is clearer. Here, as we predicted, there is a definite view that ASPL helps with semantic knowledge (as might be expected from an ontology based system with a glossary service).
- The view that ASPL can help with learning is supported by responses to D-3 which were overwhelmingly positive in agreeing that ASPL could support learning tasks.
- D-5 to D-7 support us in the new services we would like to add. In particular, D-7 indicates support for some kind of personalization although certain users disagree: in one instance with the comment "I find that my interests change to frequently for personalisation to be helpful".

Actions

- ✓ We could possibly add a translation dictionary/dictionaries.
- ✓ We are already looking at other data sources such as CiteSeer and DBLP.
- ✓ We will consider all the suggestions for new learning services. "Show related concepts (i.e. things close in the ontology)" seems like a fruitful suggestion
- ✓ We will consider all the suggestions for new application functionality. The idea of a summarization tool seems useful as does the citation output tool.
- ✓ We will consider all other comments. In particular we will look at the mis-highlighting of parts of words (e.g. 'search' in 'research') and the possibility of more complex highlighting systems.

Discussion

We were pleased to see that users felt that ASPL was likely to be a positive aid to learning. We are well aware that the system currently only supports parts of a number of learning tasks. However, it seems that we are on the right track in treating these tasks as decomposable into a variety of sub-tasks and designing ASPL to support these sub-tasks (see Figure US-1 in the User Study section for a sub-tasks example). Future versions of ASPL will continue to support a variety of learning tasks and sub-tasks.

We were also pleased to see a large degree of agreement with our proposed new services, as well as lots of suggestions for new data sources, learning services and ASPL functionality.

Conclusions from questionnaire analysis

It is clear from the questionnaires that users generally enjoy using ASPL and see it as

a valuable tool. In general there were few critical comments and most of these were about specific issues which are easily dealt with. On the other hand there were many positive remarks which were very encouraging. While ASPL is not yet a finished system our infrastructure + service approach certainly seems to be going in the right direction. We have also received a number of interesting suggestions for future development which we will be considering for ASPL-v2.

> I think ASPL is a great and very useful tool especially for semantic web and I hope to see a new version … very soon.

The overall level of satisfaction is gratifying. However, in order to see if ASPL actually helps with learning we needed to conduct a more formal user study with groups of real learners. The results of this (in general terms, more equivocal than the questionnaires) are reported in the next section.

The User Study

The more formal evaluation took the form of a user study conducted by researchers from KMi/Open University, University of Sheffield, University of Trento and Universitatea "Al. I. Cuza" in Romania. The goal was to test the efficacy of ASPL-1 as a tool for assisting in a particular learning task both in terms of the *quality of the outcome* and the *efficiency of its production*.

The user study was developed in September and October 2005 in KMi. We were fortunate to be able to make extensive use of Victoria Uren's expertise in carrying out a similar evaluation of the different sources for lexica used in the Magpie framework underlying ASPL (see above).

The user study was extensively revised as a result of her comments and then tested on two students in KMi. This led to further refinements before the first of the main studies took place. Sheffield carried out the study in the week starting 14th November 2005, "Al. I. Cuza" in the week starting 21st November and Trento in the week starting 28th November.

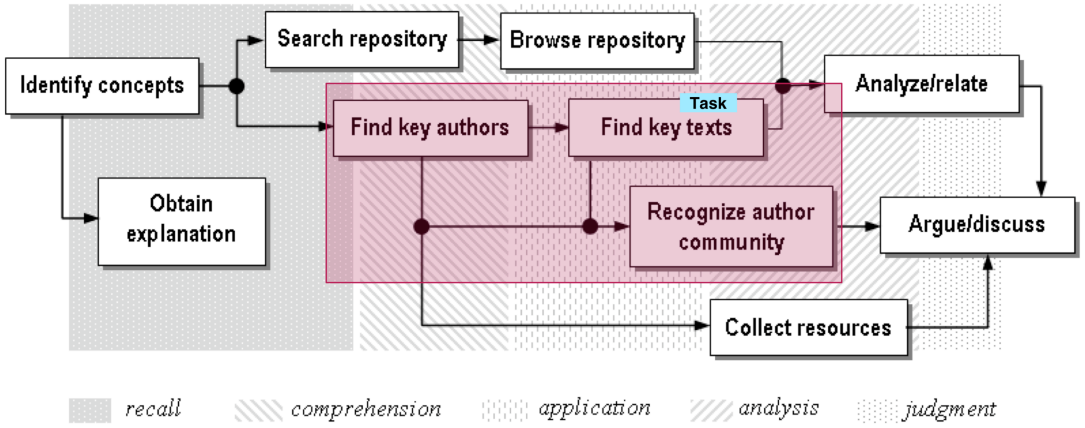


Figure US-1. Decomposition of a complex learning task — preparing a critical literature review

In particular, our aim was to see whether or not ASPL-1 makes it easier to carry out part of the task of writing a literature review: the building of a list of resources relevant to a particular topic (a citation list). See Figure US-1 for other related tasks. The performances of two groups of participants were compared on the learning task. Group A (baseline), used a particular set of resources (REASE, ACM Digital Library) but not ASPL, and Group B used ASPL as well as REASE and ACM Digital Library.

The task. Participants were asked to compile a list of key resources on the topic 'RDF and the Semantic Web'. It was a requirement that some of these should be from the EASE repository (REASE) - http://rease.semanticweb.org. This task was designed to test retrieval efficiency as well as the quality of the resources retrieved.

Facilitators. Where possible there were one or two facilitators for each study, making sure that the task was understood and completed and making notes about the study and the differences between the groups.

Group members. The participants were a mix of undergraduates, Masters and PhD students from the three universities. We asked that in choosing group members whose native language is not English, study facilitators should, as far as possible, make sure that groups were mixed – i.e. not all good speakers of English are in one group. We also asked that, where there was a range of prior knowledge about RDF and the Semantic Web, not all knowledgeable students should be included in one or the other group if possible.

Recording results. Participants were asked to record their results in a text file along with:

Their Group (A or B)
The time taken
The level of their prior knowledge about the Semantic Web
The level of their prior knowledge about RDF

For each item participants were asked to record:

Author(s)
Title
Date
Where published
Data source (e.g., ACM Digital via ASPL menu)
Process

By 'process' we meant a brief outline of the process participants underwent in producing the list (E.g., for Group A, "Found in first page of ACM Digital Library results reached using ASPL menu".)

Criteria for list items. Participants were asked to:
• Use the services from ASPL's pop-up menus as much as possible (rather than conventional search mechanisms etc.) if in Group B;
• Provide a list of TEN important items related to the subject matter with at least TWO items from REASE and at least TWO items from ACM Digital Library (reached directly or via ASPL) with the remaining SIX coming from either.

Recording interactions. As well as producing text files of results, facilitators were asked to use Camtasia to record user interactions.

Theoretical problems with the Study

There is the possibility that the sorts of results achieved depend to a large extent on the prior knowledge of the participant (though results don't really bear this out – see below). For example, someone who is doing a thesis on RDF should have no problems in producing a list of 10 citations without recourse to ASPL or any database. We attempted to overcome this by asking for student's prior knowledge level. Other factors which we didn't control for included: English language abilities, prior knowledge of the bibliographical sources (including the use of citation-based searches), of REASE, of Magpie, and of web search more generally.

Practical problems with the Study

We had two main practical problems with the study:

1. Camtasia tended to slow down PCs and the files collected were very large. Camtasia slowness *may* be a factor in the differences between Romania (where Camtasia was not used and where only 12.5% failed to get 10 citations) and Trento (where Camtasia was used and where 70% failed).
2. Due to an outage in the KMi servers handling the reasoning used by Magpie services the Trento study had to be postponed for a week.

Analysis of the Study

Metric. Since we were testing efficiency as well as the quality of the results, the metric we used was composed of two elements: a measure of the quality of the results plus a measure of the efficiency of the results (using time taken as a proxy).

Overall score = Score for quality + Score for efficiency

Quality. Results were scored in terms of the number of positives (i.e., viable resources from REASE and other sources) minus the negatives (resources which are irrelevant).

More specifically, we used the following scheme to score the results:

1. Good sources had a score of 2. This included sources which mentioned both RDF and the Semantic Web in the title or which dealt with the basics and/or theory of RDF. We also included a few citations here which did not strictly meet either of the above in their title but which were obviously relevant (e.g., 'Enabling knowledge representation on the Web by extending RDF Schema' by J. Broekstra et al., International Conference on World Wide Web, 2001).
2. Sources which mentioned either RDF or the Semantic Web in their title had a score of 1.
3. Neutral items scored 0 (i.e., not completely irrelevant but not really worth including either, such as papers on ontologies or description logics).
4. Completely irrelevant items had a score of –1.

Score for quality=sum of scores for each item

A single member of KMi scored the results. A percentage (20%) of the results were also double-scored by another member of KMi. While this scheme may be inadequate in many respects, it is at least consistent in its application to the results of both groups.

Efficiency. We used the following simple formula for scoring time taken:

Score for efficiency=(number of items/time taken) * maximum time

Thus if 10 items were found in the 30 minutes the score was 10; 10 in 25 = 12; 5 in 30 = 5; and so on.

In addition we added a score of 1 each for having the required 10 results, having 2 from REASE and 2 from ACM Digital.

Example. A full set of results with 2 from REASE and 2 from ACM Digital, all reasonable and none irrelevant taking the full 30 minutes.

Overall score = score for quality + score for efficiency
Score for quality = 10 + 3 = 13
Score for efficiency = (10/30)*30=10
Overall score = 13 + 10 = 23

Results of User Study

There were 51 user study participants in total with 27 in Group A (control) and 24 in Group B. We scored the results using the procedure given above. Since less than half the participants included their time (43%), we decided to analyze the scores for the quality of the citations and for quality+speed separately. In fact the results do not differ greatly; both show that GroupAs did slightly better than GroupBs although the results are not statistically significant in either case (see Table US-1).

The average score (including quality and time) for Group A was 19.11, for Group B, 18.19. A simple one-tailed, unequal variance t-test indicated that this difference was not statistically significant (0.285, with significance at 5%) — i.e., there's a 2 in 7 chance that the differences are due to chance.

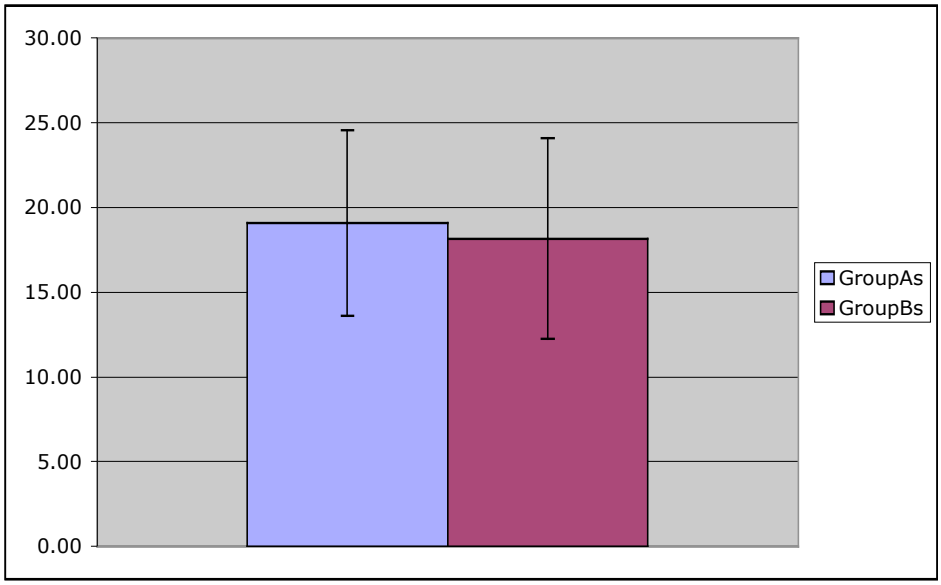


Figure US-2 Means and standard deviations for overall score

The average score (quality only) for Group A was 10.85, for Group B, 9.79. A simple t-test indicated that this difference was not statistically significant (0.130, $p<0.05$).

However when we split the groups into 4 according to institution, certain differences emerge. Note that all of these groups represent very small samples (sub-groups range from 4 to 9 members) so the statistics are likely to be of little value.

	mean quality As	mean quality Bs	t-test quality	mean As	mean Bs	t-test overall
Romania	13.67	10.29	0.01715	24.11	20.68	0.06526177
Sheffield	12.00	8.00	0.03982	20.25	13.82	0.06365032
Trento 1	8.57	12.50	0.02291	15.57	22.26	0.00469364
Trento 2	8.86	8.00	0.28111	15.57	14.71	0.39071656
Overall	10.85	9.79	0.13019	19.11	18.19	0.28510036

Table US-1

Interpretation

Overall we can conclude that while ASPL does not make the task any better or any quicker, by the same token it does not make it any harder. The difference (which is not statistically significant) can be explained by the learning curve necessary to come to grips with the system. This suggests that any future evaluations should take account of the time dimension, perhaps by staging the study over a longer period, after a more extensive period of familiarization or in multiple stages.

Prior experience

The more detailed results breakdown can be interpreted in a number of ways. While the Trento results indicate that more experienced PhD students (Trento-1) got better results with ASPL than their less experienced MSc counterparts (Trento-2), the Sheffield students (PhD/post docs) fared better without ASPL, as did the Romanian undergraduate level students.

We tested the possibility that more experience of the domain leads to higher scores by trying to find correlations between prior knowledge and scores. Note that since not all groups recorded their prior knowledge this can only be taken as indicative. Nonetheless, our results show that while the correlation is weak to moderate, we can conclude that prior knowledge has some slight effect on scores. However, the correlation between score and prior knowledge for the group intuitively most likely to benefit – PhD students – is only $r=0.16$ showing that a healthy scepticism is needed here.

Prior knowledge	*Correlated against*	*r*	r^2
Semantic Web	Quality score	0.395	0.2
	Overall	0.388	0.2
RDF	Quality score	0.418	0.2
	Overall	0.410	0.2

Table US-2

First language
While this may have been a factor in the differences, only Sheffield recorded this. In general, most of the sample were non-native English speakers.

Camtasia analysis

Of the 51 test subjects, 35 returned Camtasia files (68%). Our view before the study was carried out was that these might prove to be most useful as a means of investigating the processes which students underwent in building citation lists. This certainly has proved to be the case and we will use the data as the basis of future work on ASPL services.

The results were also helpful in highlighting certain failings in ASPL. In this deliverable, we focus our report on the results of our analysis of the Group B files from the Trento group (13 files, 25% of participants). We also looked at several of the Group A files but did not attempt any close analysis of these or detailed comparison with Group B files for the present evaluation.

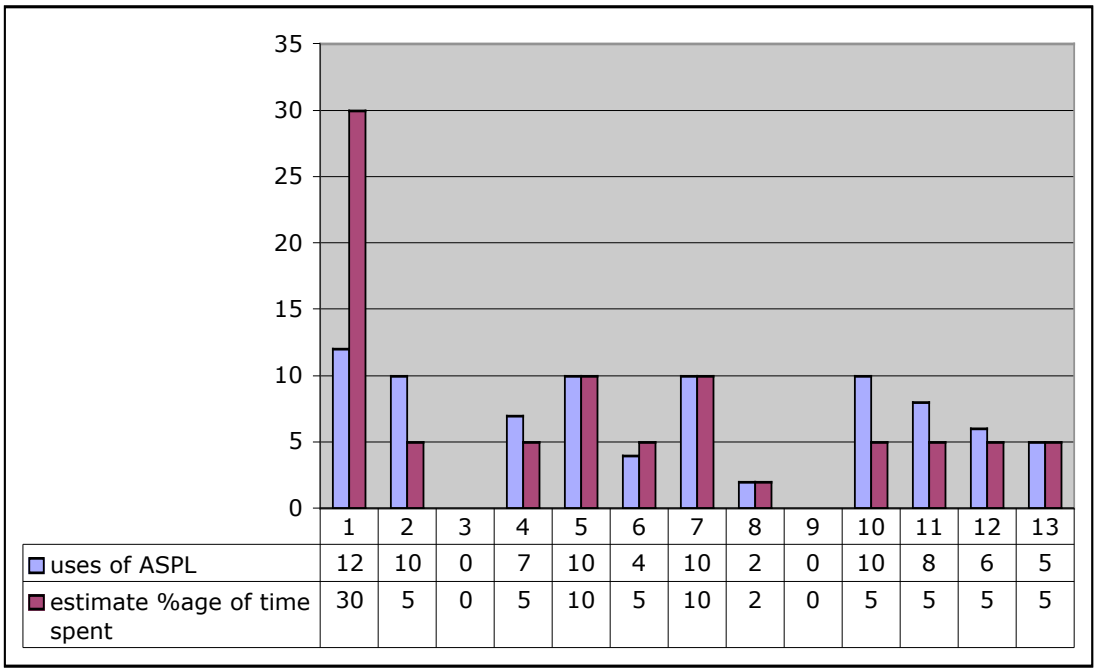
Comparison Group A and B
From our (qualitative) survey of these files, it is obvious that users spend very little time using ASPL, looking at ASPL generated windows or at other material reached via ASPL. In fact most of the time is spent using the in-built search mechanisms of the REASE and ACM Digital sites. In brief, Group A and Group B perform essentially the same tasks for most of the time.

Group B – relative use of ASPL
While we did not carry out a rigorous study of relative timings, our estimates indicate that a ratio of 5:95 is not untypical — i.e., users spend most of their time NOT using ASPL. Our counts of the numbers of times users accessed ASPL services (see Figure US-3) concur with the qualitative estimates. For example, while the 'best' user makes use of ASPL services 12 times in the session with 31 instances of highlighting and 4 returns to ASPL generated pages, most of this user's time (approximately 70%) is spent on non ASPL search, reading documents, or formatting the results file. A more typical user will access ASPL services 7 or 8 times with 4 or 5 instances of highlighting. (Note that only service accesses are recorded in Figure US-3 and Table US-3.)

This is not unexpected since (a) ASPL-v1 only has a few services relevant to this task mostly aiding focused web navigation and (b) the traditional search mechanisms for both sites are perfectly functional. Thus the typical ASPL user interleaves use of ASPL with longer periods of traditional web search and navigation. In fact most uses of ASPL in this process are very brief. For example: ASPL directs the user to a page in ACM Digital; this is accessed and highlighted using ASPL; no further access to ASPL is made although its highlighting is used to select items.

Group B – repeated use of ASPL
A more interesting question is "Do users use ASPL repeatedly? – i.e., do they find that the results are interesting enough to re-highlight, or to return to an ASPL service or ASPL generated page?" While our best and typical users above do make some repeated use of ASPL, our analysis indicates that many users only make use of ASPL

as an means of accessing a REASE or ACM Digital Library page after which they use more conventional search mechanisms. Figure US-3 shows the number of ASPL accesses for 13 users (with mean accesses of 6.5).



	1	2	3	4	5	6	7	8	9	10	11	12	13
uses of ASPL	12	10	0	7	10	4	10	2	0	10	8	6	5
estimate %age of time spent	30	5	0	5	10	5	10	2	0	5	5	5	5

Figure US-3: Per user ASPL accesses and estimates of percentage of time spent using ASPL

Group B – relative use of services

The Camtasia files also showed which services the users are more likely to access. Note that the (low) numbers given in the rightmost column in Table US-3 are for the total accesses for 13 participants (i.e., a total of only 84 service accesses in over 6.5 hours of use). The same information is given in Figure US-4.

Service	*Code*	*Usage %*	*Actual Accesses*
Explain Concept	NC1	10	8
Find related learning resource	NC2	32	27
Find in ACM Library	NC3	33	28
People active in…	NC4	13	11
Find co-citing community	C1	0	0
Find co-author	C2	1	1
Shares institution with..	C3	0	0
Authors learning material	C4	7	6
Find in digital libraries	C5	4	3

Table US-3 Percentages for service use and number of times services accessed (NC=non community; C=community oriented service)

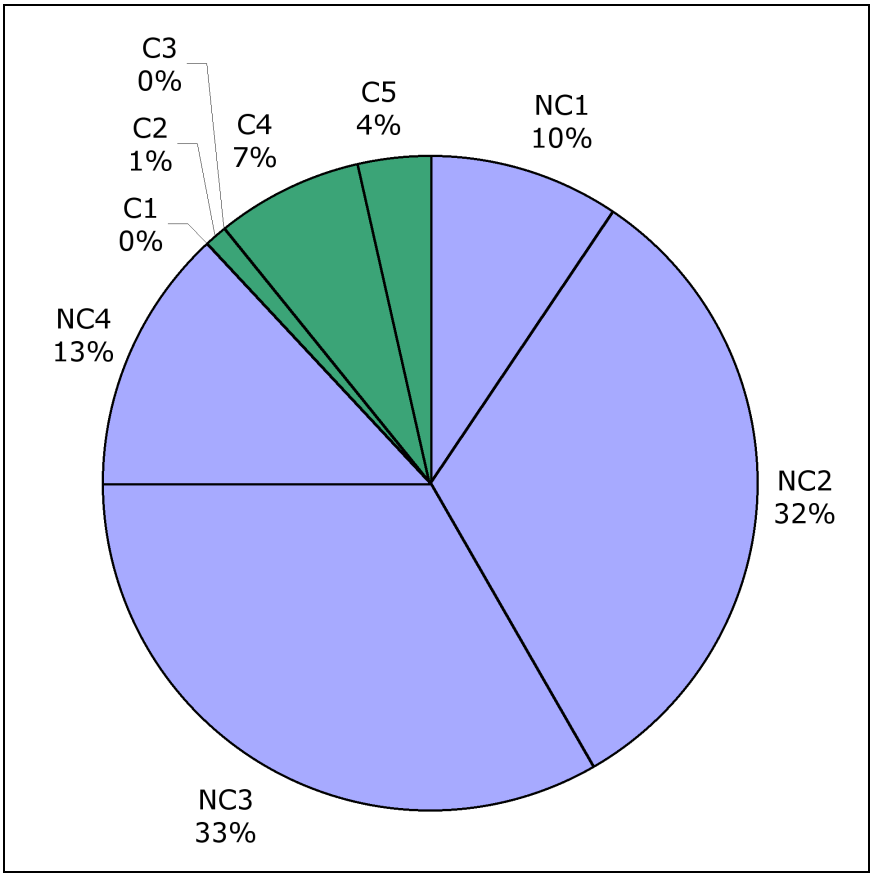


Figure US-4: Use of different learning services

Perhaps as might be expected, little use was made of any services apart from those which searched for resources in REASE or the ACM Digital Library (NC2, NC3) which represent 65% of service use. However, we were surprised that so few actually used the community oriented services. One possibility here is that users were most familiar with keyword search. If they had been familiar with citation-based search they might well have made more use of the community oriented services.

Group B – Correlations between scores and use of ASPL

We tested the hypothesis that there is a correlation between the heaviness of ASPL use (as given by the service accesses from Table US-3 and an estimate of the time spent using ASPL, Figure US-3) and scores for quality and quality+time. The results are given in Table US-4.

Correlation	*r*	r^2
quality, use of ASPL	0.258	0.066
quality, %age time	0.290	0.084
quality+time, use of ASPL	0.089	0.008
quality+time, %age time	0.152	0.023

Table US-4: Correlations of scores and use of ASPL

From this table (and Figure US-5) we can see that there is only a weak or very weak correlation between either the quality or quality plus time scores and the use of ASPL (whether given quantitatively as number of separate uses of the tool or, qualitatively, as estimated percentages of time spent using the tool).

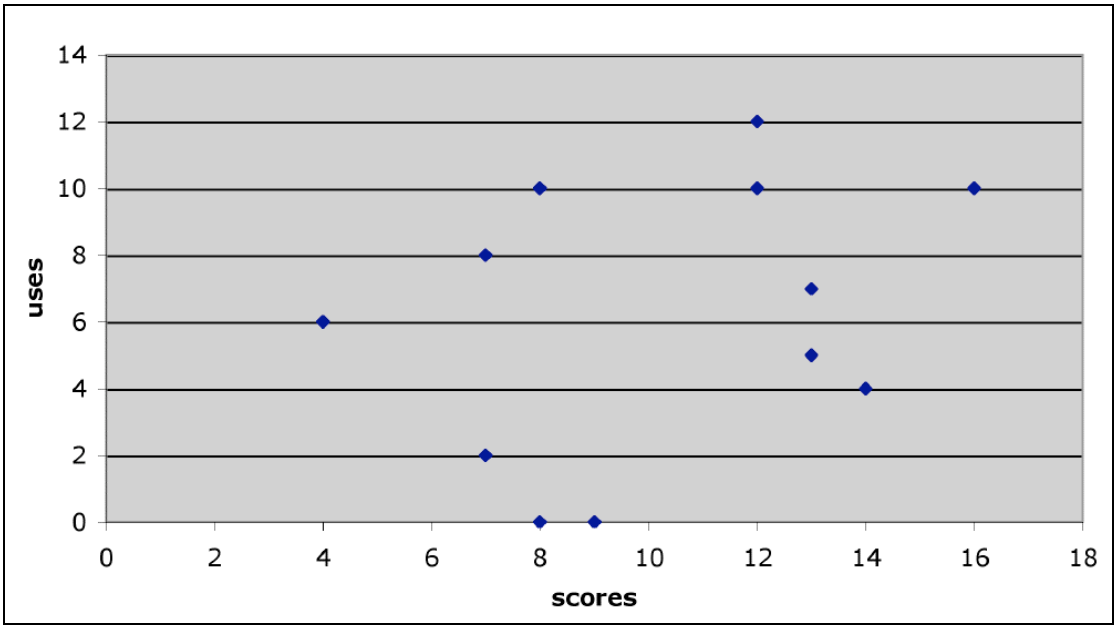


Figure US-5: An example scatter diagram of quality scores vs. ASPL uses showing little or no trend in the data

Conclusions from Camtasia analysis

From our analysis of the Camtasia files a number of points emerge.

- The main use of ASPL is as a highlighter and visual cue. Participants often switch on the highlighting to see which item in a list is worth reading. For example, some users searched using ACM's search mechanism and then used ASPL highlighting to focus on resources which had all three keywords ('semantic', 'web', 'rdf').

- However some users only use ASPL to reach a page they are interested in. For example, once a user reaches ACM Digital using ASPL, the tendency is for no further use of ASPL to be made although ASPL could at least be used to highlight items on the ACM pages. The relative paucity of use is perhaps indicative of a lack of practice in using ASPL (suggesting the need for user studies over a longer period of time). Since users are more likely to abandon ASPL and make use of the search mechanisms of both REASE and ACM Digital Library rather than ASPL services the result is that there is little difference between study groups either in the strategies followed or in the results achieved.

- There is little use of community oriented services. In fact all of the Group B services in our analysis were uses by a single participant! We thought that at least some users would use the services for finding co-authors and then access material using these. We also thought that there might be some evidence that participants tried to assess the relative importance of an author. The lack of use is probably because (a) the services available are not very powerful and (b) even with traditional search mechanisms, few users bother trying to home in on key authors as part of their search strategies. Both this and the above suggest the need for better documentation and/or training in ASPL use. It also suggests that ASPL community services might be better used by those learners who already make use of citation based searches when building bibliographies. Bearing this in mind, if we use the building of a social network as the task for a future evaluation of ASPL, we will need to ascertain users' familiarity with

different types of search. We should also consider using ASPL as a means of teaching users how to go about building a list of citations (using key authors etc.).

- While ASPL can lead to various lists of resources, the user must still choose among them and may choose badly. Frequently users seem to simply choose the first resource found. Users who actually look at resources before making their choices are few. Services such as summarization might aid in the selection task. The correlation between search behaviour and scores could also be the subject for future analysis since it is possible to have a low score but exhibit 'good' search behaviour

- With regards to the task it seems as if users spend most of their time interleaving ASPL and non ASPL tasks (e.g., start an ASPL service and look at the results of a PDF download) and, within ASPL, interleaving a variety of sub-tasks within the main task (e.g., start an ASPL service and continue to look at another ASPL highlighted window while waiting). We suspect that this is common web search behaviour although it may be due in part to the slowness of REASE/ACM.

- Much more time than we expected was spent looking at resources and completing results files.

Conclusions from User Study

Task	*Group A - control*	*Group B*	*Group C*
People	13.2	15.3	13.7
Technologies	19.2	23.4	26.7

Table US-5

While Uren (Uren et al., 2005) found that Magpie[1] users performed better in her task of identifying important concepts in KMi news stories without Magpie (Group A), with Magpie and a handcrafted lexicon (Group B), or with a lexicon built using information extraction techniques (group C) — see Table US-5 — we found that there was little difference between users in the control or experimental group. There are two important reasons for this.

- Firstly, the task is different. Uren's task was to find important concepts in news stories. This task was intuitively likely to be aided by a tool which highlights portions of texts using an ontology which represents these concepts. Indeed our findings are in agreement in that, as reported above, we often saw ASPL users use the tool as a means of focusing on relevant areas of texts.
- Secondly, in its current incarnation, ASPL does little to support the interpretational or analytic aspects of the citation gathering task.

Any seeming disadvantage of using ASPL can be attributed to inadequate prior familiarization times, or the distorting effects of one or two 'good' or 'bad' users.

Our main conclusion is that there is little difference between Group A and group B users apart from the relatively small time spent by the latter using ASPL. Both groups spend much of their time searching by entering keywords into conventional search boxes. Some also made use of the REASE topic hierarchy and the advanced search

[1] Magpie is the underlying semantic technology on which ASPL is based.

based on metadata. Interestingly, at least one of the Group A participants reconstructed some of the ASPL functionality by using the Google toolbar to highlight keywords on a page even though this page was reached using ACM/REASE search. One point worth making here is that there is bound to be less relative use of ASPL versus traditional tools in tasks which are focused on search as in the user study. Perhaps non-search based tasks will see greater relative use of ASPL (as with the Uren study).

With regard to future versions of ASPL, the analysis indicates that we need to add tools which will support more of the sub-tasks in learning tasks such as 'building a citation list' (and support them more thoroughly). Currently ASPL is simply used as one among many means of accessing materials. As it stands, ASPL could not be used on its own to build a citation list. We thus need:

- More direct search of sources such as the ACM Digital library (and filtering and personalization of results) rather than a service which simply brings up the ACM Digital search page.
- Other services for analyzing material reached – e.g., summarization. Both of these would support lower level sub-tasks such as 'homing-in' and 'selection'.
- Some means for users to access task structures (as in Figure US-1) or to build their own task structures. This would add functionality to ASPL as a tool for learning about learning tasks as well as the Semantic Web Studies domain.
- More training in ASPL use.
- More material in REASE.

Summary Conclusions from User Study

A: About the study

- ✓ We need scoring by an external expert.
- ✓ A result which can easily be scored would be preferable. Even in the case of something simple like a list of resources, there is a great deal of scope for disagreement.
- ✓ We need clearer criteria for participant selection.
- ✓ More variables should be controlled (e.g., language, machine setups etc.).
- ✓ Stricter control over results is needed to ensure that all details are correctly entered.
- ✓ We need some measure of the differences between results of the two groups – i.e., did they have totally different sets of citations?
- ✓ An alternative to Camtasia is needed. We need tools which allow us to track (and automatically analyze) how much time is spent in using ASPL as opposed to time spent using other tools. We also need to track which results are directly or indirectly reached using ASPL.
- ✓ Future studies should take account of the necessary familiarization time since it is obvious that 10 or so minutes may not be sufficient. It is probable that ASPL should really be evaluated over a longer time-frame.

B: About ASPL

- ✓ Overall, ASPL (in its present form) does not help with the citation list task but does not hinder either. The slightly higher score of the control group is not statistically significant.

- ✓ There is some slight evidence that more experienced PhD students with some prior knowledge of the topic benefit more from the use of ASPL. This is welcome since these are the main target of the system at the moment.
- ✓ Future versions of ASPL will need to support more aspects of (learning) tasks such as building citation lists. One way of doing this would be to provide summaries of material listed in ASPL generated pages. Another way would be to provide recommender or personalization services which filter the results of searches in some meaningful manner. As noted above, 71% of those who completed the questionnaire would like to see some form of personalization.
- ✓ We need some means of visualizing or building learning task structures. This was suggested in the questionnaire and seemed to be welcomed.
- ✓ Better documentation of community oriented services and examples of their use are needed. We have already decided to provide this as part of a new ASPL manual. We should also consider providing some form of training in ASPL use.

Overall Conclusions

While in general the tool is working as expected and is obviously well appreciated by users who seem to enjoy using it, there is evidence from the user study that (not unexpectedly) ASPL does not as yet fully support even simple learning tasks such as building a citation list. What this support should be and whether users will use it is problematic.

As we pointed out above, the difference between Group A and B results is not statistically significant nor does it seem from an analysis of Camtasia files that ASPL slows users down, if only because they spend relatively little of their time actually using ASPL while completing the learning task. However it is clear from the scores and from the Camtasia files that much more needs to be done in terms of clearly thinking out what such tasks involve and how we can support them with ASPL.

Our overall conclusions from the evaluation are:

1. ASPL as a service based architecture is on the right track.
2. More work on the nature of learning tasks and how to support these is needed.
3. As a first instance we should consider personalization, recommendation and summarization tools (to ensure more prolonged and consistent use of ASPL).
4. We should consider new documentation and/or training in ASPL use (to ensure use of all aspects of ASPL including community oriented services).

Future work

After developing/extending/amending the necessary ontologies (M30) and producing new services/changing current services in the light of the results given above a second version of ASPL will be produced (M36). A second evaluation will then take place by Month 42.

References

Bloom, B.S. (1956) A Taxonomy of Educational Objectives. Handbook 1: Cognitive Domain, Addison Wesley, Boston, MA.

Maynard. D. (2005) Benchmarking ontology-based annotation tools for the Semantic Web. UK e-Science Programme All Hands Meeting (AHM2005) Workshop "Text Mining, e-Research and Grid-enabled Language Technology", Nottingham, UK, 2005

Uren, V., Motta, E. Dzbor, M., Cimiano P. (2005) Browsing for Information by Highlighting Automatically Generated Annotations: A User Study and Evaluation. KCAP 2005 , Third International Conference on Knowledge Capture. October 2-5, 2005. Banff, Canada.

Appendix – the Questionnaire

In the following questionnaire we have four main goals:

1. To establish whether the current instructions for downloading ASPL and associated documentation are enough for users to get the system running and operating as intended.

2. To get some impression of the usability of the system.

3. To get some idea of the adequacy of ASPL's use of semantic technologies (especially ontologies) in the learning context.

4. To get some feedback on the pedagogic aspects of the system. We need this to inform the design of a more rigorous learner oriented experiment later in the year. At this point we are canvassing opinions from users (who are mainly researchers and other academics rather than learners) as to how well they think ASPL might satisfy certain criteria.

Please send your completed questionnaires via email to a.stutt@open.ac.uk

Thanks for your time.

ASPL development team

KMi

The Open University

Milton Keynes

MK7 6AA

UK

A. Usability

Installation and Setup

A-1a. How happy were you with the download site?

Unhappy	*Happy*	*Very Happy*

A-1b. Have you any suggestions as to how it might be improved?

A-2a. How happy were you with the installation and set-up instructions?

Unhappy	*Happy*	*Very Happy*

A-2b. Have you any suggestions as to how they might be improved?

A-3. Did you have any problems with the browser/ASPL combination? If so, can you briefly indicate what they were (and say which platform and browser you used). (E.g., problems with conflicting key combinations between browser and ASPL.)

A-4. Did you have any problems loading lexica/ontologies? If so, briefly list them.

A-5. Would the addition of an uninstall option be useful?

Yes	*No*	*Don't know*

General usage

A-6. Please list any problems you had when you tried to access services from the pop up menu (i.e., when accessing the plainmoor.open.ac.uk ASPL server).

A-7. Please indicate how useful you found the documentation/manual?

Not very	*Reasonably*	*Very*

A-8. Would the addition of a help facility be useful?

Yes	*No*	*Don't know*

A-9.Was the overall performance of ASPL:

Inadequate	*Adequate*	*Excellent*

Interface design/accessibility

A-10. Please indicate how well designed you felt the system interface was?

Not very	*Reasonably*	*Very*

A-11. Did you find the ASPL icons clear and legible?

Not very	*Reasonably*	*Very*

A-12. Were the ASPL tooltips for the ontology categories sufficient?

Not sufficient	*About right*	*More than needed*

A-13. Would you like to be able to modify fonts, colours, etc. more than is currently available?

Yes	*No*	*Don't know*

A-14a. Are you satisfied with the accessibility of ASPL?

Yes	*No*	*Don't know*

A-14b. Do you have any suggestions for improving accessibility?

A-15a. How useful was the ability to customize the Magpie plug-in (via User preferences)?

Not very	*Reasonably*	*Very*

A-15b. What other forms of Magpie plug-in customization would you like to see?

B. ASPL/REASE

B-1. Does the user need to know that different services access different databases?

Yes	*No*	*Don't know*

B-2. Did you find it easy to interact with REASE to access material pointed to by ASPL services?

Yes	*No*	*No opinion*

B-3. Did you find it easy to set up an account on REASE?

Yes	*No*	*No opinion*

B-4. List any other problems you had with REASE and/or the combined ASPL/REASE system and any suggestions for improvement.

C. Semantic technology

Visualization and customization

C-1. Did you feel that the four categories of highlighted concept (*SW technologies, SW activities, Other SW topics, Community*) were:

Too few	*Adequate*	*Too many*

C-2. Would 8 (limit of colour differentiation) be enough?

Too few	*Adequate*	*Too many*

C-3. Can you think of better top-level categories for Semantic Web Studies?

C-4a. Do you think that there should be different sets of categories for different users (e.g., learners, developers, researchers)?

Yes	*No*	*Don't know*

C-4b. If so, can you briefly list them?

C-5. Did you feel that user access to/awareness of the underlying ontology was adequate?

Inadequate	*Adequate*	*Excellent*

C-6. Did you feel that users should be able to see where highlighted instances fit in the class hierarchy?

Yes	*No*	*Don't know*

C-7. How would you show more of the ontology structure?

C-8. Do you think that it would be useful to have class instances from more than one ontology highlighted at a time (currently only one)? For example a learner might want to have instances from classes in a pedagogic/instructional ontology highlighted as well as those from a domain ontology.

Yes	*No*	*Don't know*

C-9. How important do you think it is for ASPL to have a way for users to select which classes should make up categories and which instances should be highlighted – i.e., for creating a private ontology/view?

Unimportant	*Important*	*Very Important*

Interoperability

C-10. Currently ASPL ontologies are written in OCML. How important do you think it is that ASPL can make use of other formats such as XML/RDF/OWL?

Unimportant	*Important*	*Very Important*

C-11. How important do you think it is that ASPL should be able to import publicly available lexica/ontologies in XML/RDF/OWL?

Unimportant	*Important*	*Very Important*

C-12 How important do you feel that it is for ASPL to have a web-service architecture with the capability of using any arbitrary web service?

Unimportant	*Important*	*Very Important*

Developer Tools/Issues

C-13. How important do you think it is for ASPL to have a set of easy to use tools for ontology/lexicon creation?

Unimportant	*Important*	*Very Important*

C-14. How important do you think it is that ASPL should provide a tool which can produce a Magpie lexicon from any ontology submitted to it?

Unimportant	*Important*	*Very Important*

C-15. How important do you think it is for ASPL to have a set of easy to use tools for creating the services accessed from highlighted items?

Unimportant	*Important*	*Very Important*

C-16. How important is it for ASPL to have a means for automatically maintaining lexica/ ontologies – e.g., populating ontologies with new instances and relations?

Unimportant	*Important*	*Very Important*

D. Pedagogy

Theory

D-1. Please indicate how well you feel that ASPL assisted with/could assist with the following cognitive skills?
(1 = not very; 5 = extremely well)?

Knowledge Recall		
Comprehension	*– i.e. understanding recalled knowledge*	
Application	*– i.e. ability to use knowledge in novel contexts*	
Analysis	*– i.e. ability to see how knowledge is structured*	
Synthesis	*– i.e. creation of new knowledge*	
Judgment/evaluation	*– i.e. ability to recognise value of knowledge*	

D-2. Please indicate how well you feel that ASPL assisted with/could assist with the following kinds of knowledge?
(1 = not very; 5 = extremely well)?

Semantic knowledge	*– i.e. definitions, relations etc.*	
Factual knowledge	*– i.e. statements about the world*	
Procedural knowledge	*– i.e. how to do something*	
Metacognitive knowledge	*– i.e. knowing how to know/learn*	

D-3. Do you feel that ASPL makes learning more efficient – e.g., Could ASPL help learners to complete a learning task such as creating a citation list more quickly?

Yes	*No*	*Don't know*

Practical matters

D-4. What other external sources would you like ASPL to have access to (e.g., DBLP)?

D-5. Would you like a Learning Service (i.e., a service which can be associated with highlighted instances) which gathered information on persons from the World Wide Web and, if members of KnowledgeWeb, from the KnowledgeWeb portal?

Yes	*No*	*Don't know*

D-6. Would you like a Learning Service which ranked results of searches?

Yes	*No*	*Don't know*

D-7. Would you like a Learning Service which knows about the learner and can recommend new topics/papers/people etc? I.e., a form of personalization.

Yes	*No*	*Don't know*

D-8. What other learning services would you like to see included in future versions of ASPL?

D-9. What other functionality (apart from Magpie's semantic browsing and associated learning services) would you like to see included in future versions of ASPL? For instance we have considered adding a tool which allows learners to develop a graphical representation of their learning tasks.

D-10. Add any critical comments or positive suggestions as to how the system might be improved.

Any other comments or suggestions

Finally, could you add any comments, criticisms or suggestions about any aspect of the system not covered in the above questions. Thanks for your cooperation in this; your answers will be used to improve future versions of ASPL