D3.1.5v3 Published learning resources and evaluation of REASE

Coordinator: Jörg Diederich (L3S)

Diana Maynard (USFD), Frank van Harmelen (VU), York Sure (UKARL), Sylvain Dehors (INRIA), Mustafa Jarrar (VUB), Martin Dzbor (OU), Dnyanesh Rajpathak (OU), Marco Ronchetti (UniTn), Steven Willmott (UPC), Heiner Stuckenschmidt (UniMannheim), Lyndon Nixon (FUBerlin)

Abstract.
EU-IST Network of Excellence (NoE) IST-2004-507482 KWEB
Deliverable D3.1.5v3 (WP3.1)

Abstract
This deliverable summarizes the activities related to populating REASE, the repository of EASE for learning units about Semantic Web topics, with learning resources, including the creation and evaluation of the REASE catalogue, a description of the quality management process, and an evaluation of the usage of the published learning units, based on user's experience as gathered from a formal evaluation of REASE as published in a journal paper. It also summarizes our efforts to ensure the sustainability of our results beyond the end of network funding.

Document Identifier:	KWEB/2007/D3.1.5v3/v3.0
Class Deliverable:	KWEB EU-IST-2004-507482
Version:	v3.0
Date:	05/11/2007
State:	Final
Distribution:	Public

Knowledge Web Consortium

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

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

Forschungszentrum L3S
Appelstraße 9a
30167 Hannover
Germany
Contact person: Wolfgang Nejdl
E-mail address: nejdl@l3s.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 by writing parts of this document:

CERTH
EPFL
FUB
FUBerlin
INRIA
L3S
NUIG
OU
UKARL
UniLiv
UniMannheim
UniTn
UPC
UPM
USFD
VU
VUB
VUM

REWERSE:
Stony Brook University New York
University of Göttingen
University Nova of Lisboa
LMU Munich
Linköping University
University of Turin

AgentLink:
University of Southampton

Changes

Version	Date	Author	Changes
0.1	28-10-2005	Jörg Diederich	Initial version
0.2	5-12-2005	Jörg Diederich	Completed first version without statistics
0.3	15-12-2005	Diana Maynard	Made a few small edits
0.4	16-12-2005	Jörg Diederich	Included comments from others, added published learning units and the access statistics
0.5	20-12-2005	Jörg Diederich	Included further comments from the WP
0.6	22-12-2005	Jörg Diederich	Minor fix from Luis+Sylvain
1.0	13-01-2006	Jörg Diederich	Comments from Quality control
1.01	20-01-2006	Jörg Diederich	Comments from Quality assurance
2.0pre1	4-12-2006	Jörg Diederich	Updated data, included section about evaluation of topic hierarchy and evaluation of the questionnaires
2.0pre2	15-12-2006	Diana Maynard	Included evaluation of questionnaire
2.0pre3	16-12-2006	Mustafa Jarrar, Jörg Diederich	Included evaluation of topic hierarchy
2.0pre6	22-01-2007	Jörg Diederich	Included comments from quality control
2.0final	25-01-2007	Jörg Diederich	Included further last-minute comments + final polishing
3.0pre1	28-9-2007	Jörg Diederich	Updates for 2007
3.0pre2	15-10-2007	Jörg Diederich	Further improvements
3.0pre3	05-11-2007	Jörg Diederich	Finally including October data

Executive Summary

This deliverable summarizes the activities related to populating REASE, the repository of EASE for learning units about Semantic Web topics, with learning resources. The number of learning resources published by KnowledgeWeb members or resulting from events organized by KnowledgeWeb increased from about 30 at the end of 2004 and about 50 at the end of 2005 to 81 at the end of 2006 and finally 119 in September 2007. Even though we concentrated on the creation of material for industrial education already in 2005 and 2006, our focus for 2007 still was on such industrial materials, but also on publishing material different from 'slides only' resources. Specifically, we continued to publish a series of one hour lecture recordings from the KnowledgeWeb summer school and also recordings from the ISWC industry track 2006 (provided by videolectures.net).

This deliverable also presents the evaluation of the REASE catalogue version 1.1. The underlying Semantic Web Topic Hierarchy was developed in 2005 and refined in 2006 in a more general discussion of a Semantic Web Topic Hierarchy among KnowledgeWeb and REWERSE participants. Though the evaluation of taxonomies is in general a difficult task, we have found some evidence that our taxonomy is reasonable, based on both manual inspection and on statistics taken from publications in Semantic Web topics. The topic hierarchy was refined and extended again in 2007 to reflect recent changes in the domain as well as add missing features and the outcome of the evaluation from 2006.

To control the quality of the published learning units, we have set up a list of quality guidelines, which have to be followed when publishing learning units. This is complemented by a quality management process which determines how the guidelines are actually enforced. In 2006 we implemented that quality management process involving the editorial board. However, the quality of the published learning resources is very high in general and did not require major actions neither in 2006 nor in 2007 to enforce the quality guidelines.

Finally this deliverable comprises also an evaluation of the usage of REASE and the published learning units using a log file analysis, the results of a questionnaire, sent to all 230 users of REASE (as of September 2006) and the results of a user study, conducted with more than 30 people at three different locations in November 2006. This analysis was reworked in 2007 and has been thoroughly rewritten as a journal paper about REASE, which has been accepted for publication and will appear soon. Though the users in general were quite satisfied with REASE, they made a large number of suggestions to improve REASE, which have guided our efforts related to REASE in 2007.

To summarize, the main contributions in 2007 are:

- Additional 38 learning resources by KnowledgeWeb partners
- Version 2 of the Semantic Web Topic Hierarchy
- Update of REASE: further refinement of the user interface
- Improved evaluation of REASE

To ensure the sustainability of the results of WP3.1, we have published the Semantic Topic Hierarchy on the ontoworld wiki so that its development can be continued by

anyone after the end of the KnowledgeWeb funding. In the same spirit, REASE will be taken over by EASE in 2008 to ensure that it will be continued.

Contents

1 Introduction

This deliverable is intended to document the work in the education area related to publishing educational material on REASE, the Repository of EASE for learning units[1]. It is an extension of D3.1.5v2, published a year ago, which reported about the following issues:

- Publishing more learning resources, especially ones for industry
- Extend and evaluate the REASE catalogue to allow for a more effective search
- Creating guidelines and procedures for quality management
- Evaluate formally the REASE platform, focusing on user interface issues

Besides updating the statistics about the usage of REASE, we have focused on the following issues in the past 10 months:

- Contact our industry board to find out about requirements with regard to learning resources
- Extend the REASE catalogue and the underlying Semantic Web topic hierarchy to version 2
- Continue publishing learning resources, especially ones for industrial education and those based on video recordings
- Refine the evaluation of the REASE platform by adding a log file analysis to the originally two orthogonal strategies:
    - Send a questionnaire to actual REASE users
    - Conduct a user study in a controlled classroom environment

These activities will be reported in more detail in the following sections. We start with a description of the industry questionnaire, which was developed to get feedback about what resources are expected to be found on REASE.

2 The Industry Questionnaire

One particularly important issue to achieve a high impact in the community of industrial users of REASE is to get to know what learning resources are important for this community. Hence, we prepared a questionnaire (cf. http://www.l3s.de/kweb/REASE-industry) to get feedback about the relevance of selected Semantic Web topics (either now or in the near / far future) for the business of professionals being interested in Semantic Web. We basically used the list of Semantic Web "core" topics and "special topics" available from the Semantic Web Topic Hierarchy version 1.1 plus those which we already had on our list to be included in version 2 of the topic hierarchy.

We distributed the questionnaire by email to our industrial partners and also in paper to the participants of the ESTC conference 2007 in Vienna. We received feedback from 15 persons with 7 coming from 'Aerospace' industries, 4 from 'Technology and Solution Provider', 2 from 'Telecommunications', and one from 'Banking and finance' and 'Automobile Industry' each. As a result, the following topics were mentioned most often:

[1] http://rease.semanticweb.org

- Semantic Knowledge Management
- Semantic IR / search
- XML
- Knowledge Representation
- Semantic visualization / browsing tools / plug-ins
- Ontology / knowledge repositories
- Tools for engineering semantics (editors, …)
- OWL languages
- Rules / rule languages
- Ontology integration

Even though the results of the questionnaire are not representative due to the low number of responses, at least having "Semantic Knowledge Management" as the most relevant topic coincides with the fact that the most often booked resource on REASE for industrial audiences in fact is about Knowledge Management. Even though we did not have "Semantic IR" as topic in the catalogue when the questionnaire was prepared we already had quite a number of resources dealing with this topics, which was found during the manual reclassification step of the resources after having updated the REASE catalogue to version 2 of the topic hierarchy. For "XML", we do not have many resources simply because it was not our focus (though it is an important building brick, it is not a core component of the Semantic Web) and there are other good sources for learning material (W3C XML tutorials). To suit the needs of Semantic Web users, we have added links to these tutorials on REASE.

For "knowledge representation", "tools", and "OWL" we also already have a reasonable number of resources in REASE, for "visualization/ browsing" and "repositories" there are currently very few resources yet (we are in the process to fill this gap). The following table lists the available resources for each topic:

Topic	Resources for any audience	Resources for industrial audience	In topic hierarchy v1.1
Knowledge Management	11	7	Yes
Semantic IR / Search	12	10	No
XML	6	1	Yes
Knowledge Representation	24	6	Yes
Visualization / browsing	2	2	No
Repositories	2	1	No
Tools	26	18	Yes
OWL	23	14	Yes
Rules	19	5	Yes
Ontology integration	10	10	Yes

So instead of identifying topics which are not yet covered by REASE, the questionnaire rather helped us to improve the topic hierarchy. At least, we were urgently lacking three from the 10 most often mentioned topics, most notably the category 'semantic IR /

search' which has been added in version 2 of the REASE catalogue as application are for Semantic Web technology.

3 The REASE Catalogue

As described in D3.3.2v2, the REASE catalogue initially comprised only five categories. However, when more and more learning resources were added, it became clear that this initial classification was no longer sufficient, so we initiated a general discussion about a Semantic Web curriculum, the so-called ***Semantic Web Topic Hierarchy***[2] together with the NoE REWERSE, in order to align the REASE catalogue with already started curriculum activities in REWERSE. In 2006, we moved the Semantic Web curriculum to the OntoWorld wiki, a Wiki system which is itself semantically enhanced. Specifically, we have included the Semantic Web topic hierarchy into the Wiki categorization scheme. This way, other users of the OntoWorld wiki, for example, who are using it to publish workshop or conference announcements, can use the topic hierarchy for classification of any wiki page. Furthermore, keeping the topic hierarchy and the discussion about it s development in a public wiki ensures a high degree of sustainability. As an outcome of last year's evaluation, we have published version 2 of the topic hierarchy (cf. the following section), which extends the hierarchy from 58 categories to 70 categories, adding new ones as well as deleting some.

3.1 The Semantic Web Topic Hierarchy, version 2

The Semantic Web Topic Hierarchy was developed jointly with REWERSE starting from the initial curriculum as discussed in the REWERSE deliverable E-D5. In 2007, advances were made along four lines, mainly to ensure a high sustainability of the topic hierarchy activities:

1. Discussion version 1.1 of the topic hierarchy and update to version 2
2. Adding the development history of the topic hierarchy to the ontoworld wiki
3. Adding a flexible mechanism for emerging topics which are (not yet) included in the topic hierarchy
4. Connecting different prototypes and results of KnowledgeWeb using the wiki.

More details about the history of the topic hierarchy and older versions are available at http://ontoworld.org/wiki/History_of_the_Semantic_Web_Topic_Hierarchy. The current version of the topic hierarchy with explanations, e.g., more information about at which version a topic was added, is available at http://ontoworld.org/wiki/Semantic_Web_Topic_Hierarchy, a list of potentially emerging topics can be found at http://ontoworld.org/wiki/Emerging_Semantic_Web_Topics.

Overall Structure

The structure of the topic hierarchy is in general three-fold:

- Foundations
- Semantic Web Core Topics
- Semantic Web Special Topics

[2] http://ontoworld.org/wiki/Semantic_Web_Topic_Hierarchy

3.1.1 Foundations

The foundations top-level category comprises the following categories and sub-categories:

- Knowledge Engineering / Ontology Engineering
  - Ontology Engineering Methodologies
  - Ontology Construction / Learning / Population
  - Ontology Evolution / Maintenance / Versioning (Dynamics)
  - Ontology Mapping / Translation / Matching / Aligning (Heterogeneity)
  - Ontology Evaluation / Validation
  - Ontology Interoperability / Integration / Merging
  - Ontology Modularization and Composition
  - Ontology Engineering Tools
  - Ontology Visualization
- Knowledge Representation and Reasoning
  - Logics
    - Description Logics
    - F-Logic
    - First-Order Logic
    - Temporal Logic
    - Fuzzy Logics
  - Logic Programming
    - Horn Logic
    - Datalog
    - Prolog
    - Hilog
    - Monotonic and Nonmonotonic Negations
  - Reasoning
    - Reasoning Engines / Theorem Provers
    - Fuzzy Reasoning
- Basic Web Information Technologies
  - XML
  - Web Data Integration
  - Web Services
    - Web Service Discovery
    - Web Service Composition
  - Personalization Techniques
  - Web Data Extraction / Information Extraction
- Agents
- Natural Language Processing
- Security / Trust / Privacy
- Machine Learning
- Rules
  - Deductive Rules
  - Reactive Rules
  - Rule Visualization

- Information Access
    - Query Languages
    - Browsing / Navigational Access
    - Query Algebra
    - Query Optimization
    - Visual Querying
    - Event Queries

3.1.2 Semantic Web Core Topics

The Semantic Web Core Topic top-level category contains:

- Semantic Web Information Access
    - Semantic Web Query Languages
    - Semantic Web Browsing
- Ontologies on the Semantic Web
    - OWL
    - Resource Description Framework / RDFSchema
    - Legacy Ontology Languages (DAML, DAML+OIL)
    - Ontology Repositories
    - Ontology Instances
        - Upper-Level Ontologies / Top-Level Ontologies
        - Domain Ontologies
    - Semantic Annotation / Microformats
- Web and Semantic Web Rules
    - Rule Languages
    - Rule Markup
    - Integration of Rules and Ontologies
    - Distributed Rule Processing
- Semantic Web Application Domains
    - Knowledge Management
    - elearning
    - Bioinformatics
    - Multimedia
    - ehealth
    - ebusiness
    - eGovernment
    - Information Retrieval / Search
    - eCulture
    - Human resources
    - Blogs
    - Business Rules
    - Wikis
    - Digital Libraries
    - Data Integration / Enterprise Information Integration
- Reasoning in the Semantic Web

3.1.3 Semantic Web Special Topics

- Natural Language Processing / Human Language Technologies
- Social Impact of the Semantic Web
- Social Networks and Semantic Web
- Peer-to-Peer and Semantic Web
- Agents and Semantic Web
- Semantic Grid
- Outreach to Industry
- Benchmarking / Semantic Web Scalability
- Semantic Web Services
- Semantic Desktop

3.1.4 Reclassification

In general, we tried to ensure a backward compatibility to previous versions of the topic hierarchy as far as possible. However, this has been a matter of trade-offs in the past. Most notably, some categories being in version 1.1 turned out to be too specific compared to others, so that they had to be removed for consistency reasons even though they were used (rarely). For applications using the topic hierarchy (such as REASE), a fully automatic reclassification approach could, hence, no longer be guaranteed during the update to version 2 of the topic hierarchy.

3.1.5 The Ontoworld wiki: The Central Connector

Using the ontoworld wiki as main connector of the different outcomes of KnowledgeWeb was one of the main efforts in 2007. For this purpose, we have created the following mandatory standard structure for all topics in the topic hierarchy (which can be enhanced with more optional parts):

General information about the topic:

- Short definition of the topic
- Short statement about when the topic was introduced in the topic hierarchy plus links to the main topic hierarchy and the history of the topic hierarchy.

General links to other KnowledgeWeb prototypes:

- Search for Semantic Web publications using FacetedDBLP (the search is restricted to the main venues publishing about 'Semantic Web' including the conferences ISWC, ESWC, EKAW, WWW, RuleML, Web Intelligence, ASWC, ESWS, SWWS, PPWSR, RR and the journals IEEE Intelligent Systems, and Journal of Web Semantics).
- Search for computer science publications using FacetedDBLP
- Search for computer science experts using ASPL-2
- Search for computer science publishing venues using ASPL-2
- Search for Semantic Web tools and applications for Ontology engineering methodology
- Search for Semantic Web educational resources on REASE
- Search on the web pages of the Ontology Outreach Advisory

General links to other useful Web resources

- Search for scientific publications as phrase or bag of keywords on Google Scholar
- Search as phrase or bag of keywords on Google
- Look into Wikipedia

Articles in category:
This shows the usage of the topic within the ontoworld wiki itself. These annotations are not necessarily an outcome of KnowledgeWeb but also reflect the impact of the topic hierarchy on the ontoworld users in general.

Figure 1 shows an example topic on the ontoworld wiki for ‘ontology learning’

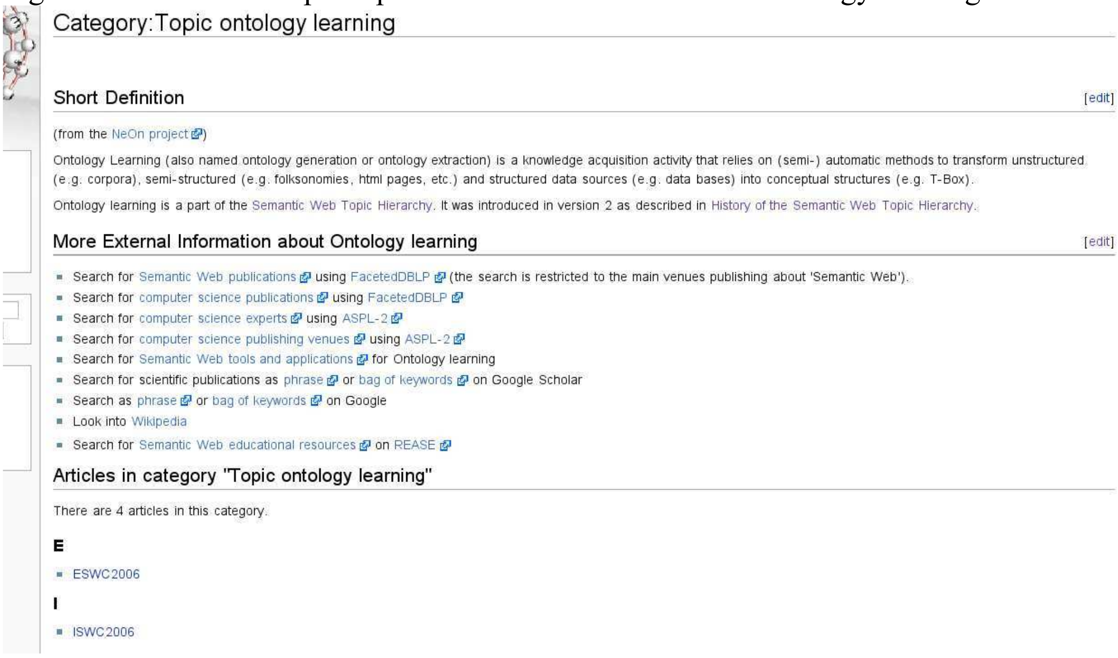


Figure 1: Ontoworld topic 'ontology learning'

Optionally, we have added further links to topic-specific resources, such as the web page of the Ontology Outreach Advisory use case studio for the topic ‘ehealth’.

All these wiki pages can be enhanced by anyone visiting the ontoworld wiki with the exception of the main topic hierarchy page, which can be changed by wiki administrators only to ensure that the version of the topic hierarchy remains unchanged. However, any ontoworld user can make a copy and open up new wiki pages with separate versions of the topic hierarchy.

3.2 Evaluation of the Semantic Web Topic Hierarchy version 1.1

Topic-based classifications are an important part of information retrieval. The topic hierarchy of REASE is intended to guide users to quickly and correctly find the learning units they seek. Although we use several metadata elements to describe a learning unit, the topic hierarchy is seen as a formal description, and thus it enables simple reasoning.

For example, if ‘ontology mapping’ is a sub-topic of ‘ontology engineering’, then all learning units that are classified under ‘ontology mapping’ are also instances of the topic ‘ontology engineering’.

In order for the topic hierarchy of REASE to be effective, first, the *learning units should be correctly classified*. To meet this requirement, we engaged human experts to review and assure that (all and only) the relevant topics appear in the description of a learning unit. Second, the *topic hierarchy itself should be expressive*. This means that the hierarchy should be descriptive enough to indicate what a learning unit is about and intuitive enough so that users can easily understand and use it. We have found that this requirement is indeed not an easy goal to achieve, because of its subjective nature. In the following we describe this difficulty and our approach to build and validate the REASE hierarchy.

Although topic classification is an old subject matter especially in the libraries world, it has always been a difficult issue. Topic classification can be viewed from a variety of perspectives ranging from the purely ad hoc and pragmatic to the purely philosophical, see [WJ99].

As the term *topic* typically means *an area of knowledge* [WJ99], the problem of classifying such knowledge depends on how mature the understanding of the area is and to what degree it is shared within a community. Some areas of knowledge change very rapidly, especially IT research topics. In addition, people with different backgrounds view topic classification differently [GMZ06]. This is due to the different expectations of what activities in an area of knowledge are.

> J. Doyle stated: “It became clear to me that the main organizing principle for indices, at least for most people, is sociological. That is, one structures the index not to reflect conceptual relations, but to reflect things like what populations of people like to work together, what do they think of as the current main topics of interest, etc.” [WJ99].
>
> C. Welty added: “as a consequence of the social nature of topic organization, the stability of a particular section of the hierarchy seems to be directly tied to the field’s maturity.”

In short, topic classifications are subjective knowledge, they are mostly influenced by personal tastes, may reflect fundamental disagreements, and change rapidly.

With this in mind, our approach to build the topic hierarchy of REASE is guided by the following two aspects.

1. The topic hierarchy should be built by active and up-to-date researchers. As such researchers are supposed to be aware of the research directions and activities in the semantic web area, we expect the topic hierarchy to be more stable and to reflect a relatively mature understanding of the subject matter.

2. It should build on existing classifications if available.

3. In order for the topic hierarchy to be intuitive and easy to understand by users, the labels in the topic hierarchy should be familiar and frequently used in the semantic web community.

Regarding the first aspect, the REASE topic hierarchy has been built by a group of active researchers in KnowledgeWeb and REWERSE. It has been discussed on the mailing lists of KnowledgeWeb and REWERSE, for over a year, and around 300 emails were exchanged about it. Furthermore, the researchers involved in this discu ssion originated from different backgrounds, such as: artificial intelligence, peer-to-peer, database, natural language processing, business informatics, etc.

The Semantic Web Topic Hierarchy is also built upon existing classifications, namely the ACM Computer Classification System as mentioned before and as explained in more detail in the REWERSE deliverable E-D7.

For the third aspect, we cross-validated our topic hierarchy with frequently used keywords from Semantic Web publications. For this purpose, we extracted the author keywords of all Semantic Web publications in the popular DBLP archive, performed a standard set of methods for cleaning them (e.g., stemming), and cr eated a list of co-occurring keywords, sorted by the number of co-occurrences, restricting ourselves to those keywords that occur at least 5 times with Semantic Web (cf. Table 1)

Table 1: Co-occurring Keywords from Semantic Web publications in DBLP

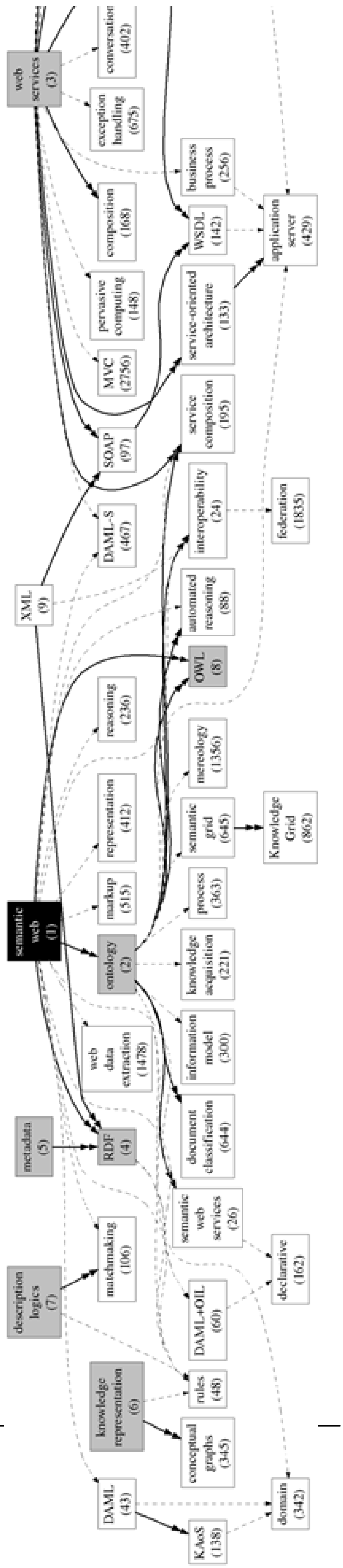
ontology	95	Matchmaking	7
web services	39	Personalization	6
RDF	34	Ontology mapping	6
OWL	19	Information integration	6
XML	18	DAML+OIL	6
metadata	16	e-Learning	6
knowledge representation	12	Semantic Web Services	6
information retrieval	11	Interoperability	6
knowledge management	11	machine learning	6
Agents	10	OWL-S	5
Description Logics	10	User interface	5
annotation	9	Web service composition	5
P2P	8	RDFS	5
Semantic annotation	8	Rules	5
Information extraction	8	Automated reasoning	5
Search	7	multimedia	5
DAML	7	e-commerce	5

The REASE topic hierarchy comprises directly contains 22 out of the 34 terms in Table 1, namely 'ontology', 'web services', 'RDF', 'OWL', 'XML', 'knowledge representation', 'knowledge management', 'agents', 'description logics', 'peer-to-peer', 'information extraction', 'personalization', 'ontology mapping', 'information integration', 'e-Learning', 'Semantic Web Services', 'interoperability', 'RDFS', 'rules', '(automated) reasoning', 'multimedia', and 'e-commerce' (in our hierarchy called eBusiness). 4 out of the remaining 12 are very general including 'metadata', 'annotation', 'semantic annotation', and 'user interface'. The two keywords about predecessor languages of OWL ('DAML' and 'DAML+OIL') can be easily classified into the category 'Ontology Representation / Ontology Languages / OWL' and are, hence, also included in the Semantic Web Topic Hierarchy. Three of the remaining topics are subtopics of 'Semantic Web Services', namely 'matchmaking', 'OWL-S', and 'Web service composition' They might be added to refine the Semantic Web Service category in the future.

As a result of this analysis, we considered including the topic 'search / information retrieval', and 'machine learning' as an application area in the Semantic Web Topic Hierarchy.

The remaining 50 terms in the Semantic Web Topic Hierarchy do not co-occur with 'Semantic Web' more than five times and are, hence, not listed in Table 1. This is mostly because they are very specific and typically not used as author keywords (like the subtopics of 'XML') or very rarely only (like 'Logic Programming' co-occurring 4 times with 'Semantic Web'). We will analyze the usage of the categories in REASE in more detail in Section 3.

We also developed the Semantic GrowBag approach [DBT06] to automatically find tag graphs, i.e. relations between topics (which are not necessarily hierarchical) from tagged object collections, for example, publication databases annotated with author keywords. We used the same DBLP dataset, enhanced with author keywords to get a tag graph for 'Semantic Web' as shown in Figure 1.
In this figure (generated using author keywords from publications in the period 2001-2005), 'Semantic Web' is depicted with a black background, and the main related concepts (i.e. 'ontology', 'web services', 'RDF', 'metadata', 'knowledge representation', 'description logics', and 'OWL'), as found from a co-occurrence analysis, are shown with a grey background. Arrows in general can be best described



to mean ‘is related to and more specific’, but quite some of them are also ‘hierarchical’ and actually mean ‘subsume’. The confidence in the automatically computed relations is shown with bold lines and two-headed arrows for strong confidence and with a dashed line for a weak confidence. More details about the GrowBag scheme can be found in [DBT06] and [DB07].

As a result, GrowBag confirms the Semantic Web Topic Hierarchy in the following aspects:

- The topics ‘web services’, ‘description logics’, and ‘knowledge representation’ are all important, but not a subtopic of ‘Semantic Web’ (in the Semantic Web Topic Hierarchy they all belong to the ‘foundation’ part)
- The topics ‘ontology’, ‘RDF’, and ‘OWL’ are all subtopics of ‘Semantic Web’ (strong confidence)
- The topics ‘rules’, ‘reasoning’, and ‘Semantic Web Services’ are subtopics of ‘Semantic Web’ (weak confidence)

Further interesting findings (those involving strong relations) are:

- ‘RDF’ is a subtopic of ‘metadata’
- ‘Ontology’ is a super-topic of ‘service composition’ as well as ‘Web service’ is a super-topic of ‘service composition’. This is the connection between the tag graph of ‘semantic web’ and ‘web service’ (“semantic web service” does not take this role because authors do neither use the pair ‘semantic web’ and ‘semantic web se rvice’ nor the pair ‘web service’ and ‘semantic web service’).

No further keywords are found in the tag graph of ‘Semantic Web’ which appeared to be missing in the Semantic Web Topic Hierarchy.

3.3 The REASE Catalogue

While the Semantic Web Topic Hierarchy reflects, of course, a compromise among the different opinions within the Semantic Web community (e.g., some consider `natural language processing’ as a foundational topic while others treat it as special topic), we had to generate an even more simplified version for technical reasons: the REASE catalogue, though customizable, can only handle up to two hierarchical levels at maximum. This has also the advantage that the number of categories is more limited, so REASE users are not `lost’ in too many catalogue categories.
As a result, we skipped the first-level hierarchy of `foundations’, `Semantic Web core topics’ as there sometimes also is no real distinction between them (there was, for example, quite some discussion during the creation of the topic hierarchy whether ontologies are foundational or belong to the core topics). Furthermore, we ignored all subcategories of a depth larger than 2 (e.g., the subcategories of `Logics’, `Logic Programming’, and `XML’), since it was not expected that learning material in REASE will deal specifically with one of the subtopics. Instead, it is expected that learning units in these topics give an overview, for example, on `Logics’ and discuss most of the sub-categories.

As a result, the REASE catalogue comprises the following topics:

- Knowledge Engineering / Ontology Engineering
    - Ontology Engineering Methodologies
    - Ontology Construction / Learning / Population
    - Ontology Evolution / Maintenance / Versioning (Dynamics)
    - Ontology Mapping / Translation / Matching / Aligning (Heterogeneity)
    - Ontology Evaluation / Validation
    - Ontology Interoperability / Integration / Merging
    - Ontology Modularization and Composition
    - Ontology Engineering Tools
    - Ontology Visualization
- Knowledge Representation and Reasoning
    - Logics
    - Logic Programming
    - Reasoning
- Basic Web Information Technologies
    - XML
    - Web Data Integration
    - Web Services
    - Personalization Techniques
    - Information Extraction / Web Data Extraction
- Information Access
    - Query Languages
    - Browsing
    - Query Algebra
    - Query Optimization
    - Visual Querying
    - Event Queries
- Ontologies on the Semantic Web
    - OWL
    - Resource Description Framework / RDFSchema
    - Legacy Languages (DAML, DAML+OIL)
    - Ontology Repositories
    - Ontology Instances
    - Semantic Annotation
- Rules
    - Rule Languages
    - Rule Markup
    - Integration of Rules and Ontologies
    - Distributed Rule Processing
    - Deductive Rules
    - Reactive Rules
    - Rule Visualization
- Security / Trust / Privacy in the Semantic Web

- Application Domains
    - Knowledge Management
    - elearning
    - Bioinformatics
    - Multimedia
    - ehealth
    - ebusiness
    - eGovernment
    - Information Retrieval / Search
    - eCulture
    - Human resources
    - Blogs
    - Business Rules
    - Wikis
    - Digital Libraries
    - Data Integration / Enterprise Information Integration
- Special Topics
    - Natural Language Processing / Human Language Technologies
    - Social Impact of the Semantic Web
    - Social Networks and Semantic Web
    - Peer-to-Peer and Semantic Web
    - Agents and Semantic Web
    - Semantic Grid
    - Outreach to Industry
    - Benchmarking Scalability
    - Semantic Web Services
    - Semantic Desktop

Of course, this catalogue will be discussed further in the future to adapt to the dynamics of the Semantic Web domain.

4 List of Published Learning Units

This section summarizes the learning units that have been published on REASE by the end of 2007. In contrast to previous versions, we do not list all deliverables separately for space reasons. All resources can be viewed directly on REASE at http://rease.semanticweb.org.

4.1 Overview and Statistics

The following Figure 2 depicts the number of learning resources available on REASE since it was put online in July 2004.

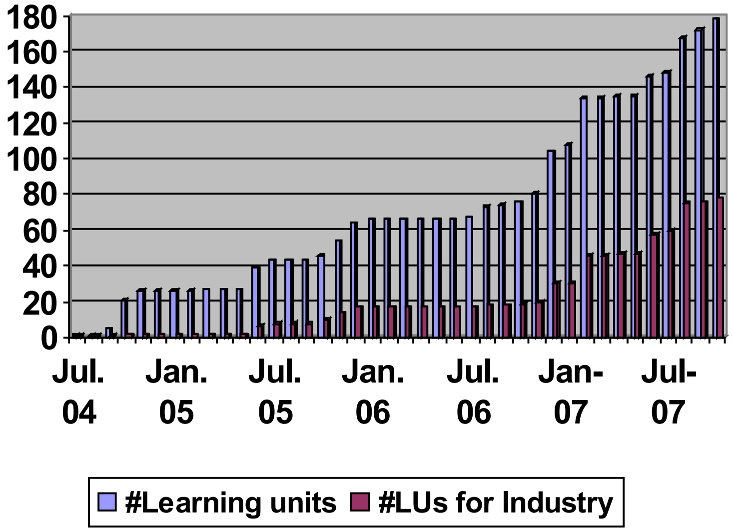


Figure 2: Published Learning Units on REASE

In total 178 learning units were published on REASE, of which 119 were published by KnowledgeWeb partners. 78 of these learning units (44%) are especially suited for industrial education. (as of 1st November 2007).

A couple of main events can be identified, In October / November 2004, an initial set of learning units was published as a result of the first public announcement of REASE in October 2004. A second significantly large set of resources was added in June / July 2005 by the tutors of the REWERSE summer school, who were required to add their resources before the start of the summer school. More resources were added step by step at the end of 2005 as a result of further educational activities in KnowledgeWeb, such as the industry-education events (reported in D3.2.9). The REWERSE summer school slides of 2006 were added during the whole year 2006 while the KnowledgeWeb summer school recordings of 2005 and the presentations of the ESWC industrial day were added in December 2006. In February 2007, we announcements for educational events provided by the Semantic Web School in Austria were added as well as the video recordings of the KnowledgeWeb summer school 2006. In June 2007, the presentations of the industry awareness day of Semantics 2006 were included in REASE. The industry track presentations of ISWC 2006 were added in August 2007 together with the keynote presentations of ESWC 2007 and the presentations from the ESTC 2007.

In summary, we focused on publishing material for industrial audiences in the past month and the percentage of courses suited for industrial education has grown from less than 10% at the beginning of 2005 and 25% at the end of 2005 to 30% at the end of 2006 and 44% in November 2007.

5 Evaluation of REASE

As discussed in the previous deliverable D3.1.4, the development of REASE has reached a stage of maturity where it requires some formal evaluation of its quality and effectiveness. This evaluation needs to be carried out on various dimensions:

- Does the *existence* of REASE fulfill a needed role?
- Is the *quality of material* in REASE of a suitable standard?
- Is there *sufficient* material in REASE to make it worthwhile?
- Are the mechanisms for *finding information* within REASE adequate?
- Are there mechanisms for *providing* material in REASE adequate?
- Is the usability of REASE acceptable?

We have therefore carried out a two-part evaluation in 2006: first, in the form of a questionnaire sent to all users of REASE, and an analysis of their responses; and second, a task-based study completed by selected volunteers. As a follow-up to the evaluation in the last version of this report, we have prepared in 2007 a journal paper about the evaluation, which also includes a transaction log file analysis. This journal paper has been accepted for publication and is available in the appendix of this report. It supersedes the original version of the evaluation report provided in the previous version of this report.

6 Quality Guidelines and Procedures

To assure a high quality of the material stored in REASE, a review process is required, especially since REASE is now moving more towards the public (we could assume a reasonable degree of quality for the material published from KnowledgeWeb / REWERSE partners up to now, but this will not necessarily be the case if people from outside both projects start uploading material). For this reason, we have set up a list of quality guidelines which are to be fulfilled before the learning material is finally accepted to be published at REASE. This is necessary to ensure that REASE can achieve a high reputation in the area of 'learning about Semantic Web'. The quality guidelines will evolve over time, so this section only describes the current state of the quality guidelines.
The quality of each learning unit is related to two major areas: technical requirements and requirements regarding the content.

6.1 Technical Requirements

The technical requirements define all issues which are not related to the content of a learning unit. Specifically, this comprises:

6.1.1 Non-Proprietary File Formats

To ensure that learning units do not depend on specific applications to be able to use them, they should not be published in proprietary file formats. As an example, the very popular file formats for Microsoft Office applications are very difficult to read for users from other operating systems.
Therefore, we require strictly that learning units must be provided at least in one non-proprietary format. However, we do want to keep the proprietary (source) formats

additionally as many people work with them and reuse them for their own purposes (if the licence allows this).

Therefore, to support providing proprietary (editable) source files together with non-proprietary (read-only) versions, we have integrated an automatic conversion tool into REASE (Linbox[3]). If the learning resource provider uploads their material in one of the Office file formats, they will be automatically converted into a PDF file and a selection between both is presented to REASE users, who want to access the learning resource.

6.1.2 Uploading Material vs. Linking

Basically, each learning material provider has the choice to either upload their material to the REASE server or to provide a URL to where the learning material is located.
Providing a link basically has the potential advantage that updates are available instantaneously and automatically. However, it carries the risk that the material will not be available at all, for example, after a re-organization of the web server or if the provider changes institution. Furthermore, it is not possible to automatically convert proprietary file formats (as mentioned in the section above). Therefore, we require that material is uploaded instead of providing a URL only, unless the material is itself in HTML.

6.1.3 Metadata

To implement a reasonable search service on REASE, it is essential that a sufficient number of metadata fields is specified for each resource. The main part of verifying this metadata is already done by the system. On the one hand, the REASE catalogue provides a classification into the most popular Semantic Web Topics, on the other hand the most important additional metadata fields apart from the classification are 'mandatory' in the sense that the system will not allow the user to complete the upload of the material until the mandatory metadata fields are specified. However, if the metadata is to be described in free text, people might fill in wrong values such that a manual post-control of the metadata field is necessary.

6.1.4 File Formats

As mentioned above, learning objects that are provided in an editable format (the source code) are highly valuable for persons who are teachers themselves. Such editable formats may also be valuable for EASE, for example, if they are only available to EASE members, generating a higher interest for EASE in this way. However, we do not force providers to upload their material in a source format as this might prevent too many people to use REASE at all to provide their learning units.

6.1.5 Modularization

The utility of a learning resource also depends on its size. Oversized resources are difficult to use for a potentially interested learner and they are difficult to classify according to the REASE catalogue. For example, if someone uploads a lecture on Semantic Web covering a 6-month-course at university, all topics can be associated with this course. To avoid this problem of too-common learning materials, we require that

[3] http://www.linbox.com/en/converter

such material is to be broken into several subunits before it is published in REASE. As a rule of thumb, material that covers more than 12 hours is considered to be too long to constitute a single learning unit in REASE, but this has to be decided on a case-by-case basis within the quality management process as described below.

6.1.6 Questionnaire

To be able to get feedback from users of learning resources, REASE allows each provider to attach a questionnaire to each learning unit. However, each provider has to individually decide whether her material is associated with a questionnaire or not. We are currently not demanding that they do this as the questionnaire support of REASE is incomplete. However, we might change this in the future to get more feedback from REASE users.

6.2 Non-Technical Requirements

The non-technical requirements are mainly related to the content of each learning resource. We basically have to verify two issues:

- Relation to the Semantic Web
- Quality of the actual content.

The first requirement is necessary to ensure that REASE keeps its focus on Semantic Web topics and the necessary basics to understand the Semantic Web. As an example, we are allowing material around the topics 'XML' (as RDF is often expressed in its XML variant), but a general tutorial about `HTML' or `computer networks' is out-of-focus.

6.3 Quality Management Procedures

Quality management in REASE is intended to ensure that all published learning units are in accordance with the above listed requirements. We can distinguish between automatically controlled requirements and those that have to be verified manually.

6.3.1 Controlling Requirements Automatically

The fulfilment of the technical requirements is as often as possible ensured automatically. For example, the most important metadata fields describing the learning units are mandatory such that REASE will not accept a new learning unit without these metadata fields being filled in. Furthermore, we implemented an automated conversion of the most popular proprietary formats (Microsoft Office) into the PDF format using the Linbox technology (http://www.linbox.com/converter).

6.3.2 Controlling Requirements Manually

This manual quality management process has to be effective and efficient. Therefore, REASE is required to support this process, which is already partly available. Each time an author publishes a new learning unit / updates an existing one, the administrator of REASE has to approve the changes. In this manner, we can avoid the publication of low-quality material, which is not related to REASE at all. This is a sustainable approach regarding the number of learning units and the expected low frequency of updates (which is different from other large-scale approaches, such as wikipedia, as REASE is only about a limited topic).

To ensure that the quality of the content of all Semantic-Web related learning units is high, we envision the following process:

1. The REASE administrator (currently Jörg Diederich, L3S) verifies the remaining technical requirements (those that cannot be validated automatically, or only with difficulty).
2. He also assesses the content of each learning unit to filter out the non-borderline cases. These include, on the one hand, learning units from KnowledgeWeb partners or cooperating NoEs, which have a very high probability of being excellent and can thus be assumed to match the content requirements. On the other hand, the administrator can also easily filter out 'spammers', who try to use the platform for exchanging material completely unrelated to Semantic Web topics.
3. For borderline cases, we have installed an editorial board that will review the remaining units for their suitability to REASE in accordance with the quality guidelines. The current members of the editorial board are:
    - Holger Wache, VU (knowledge representation and reasoning: ontologies, representation languages, reasoning techniques)
    - Diana Maynard, USFD (human language technology)
    - York Sure, UKARL (ontology engineering, ontology management, semantic web infrastructure)
    - Lyndon Nixon, FUBerlin (materials for business professionals, multimedia, Semantic Web services)
    - Sylvain Dehors, INRIA (basic web information technology, ontologies for the Semantic Web, Resource Description Framework (RDF) / RDFSchema, e-learning)
    - Enrico Franconi, FUB (logics, Semantic Web languages)
    - Martin Dzbor, OU (interoperability & integration, dynamics, tools, architecture of information systems, personalization techniques, Semantic Web infrastructure/architecture, security/privacy/trust, information management)
    - John Breslin, NUIG (Semantic Web infrastructure, social networks in the Semantic Web)
    - Yiannis Kompatsiaris, CERTH (multimedia ontologies, semantic analysis and reasoning of multimedia content, multimedia and Semantic Web)
    - Mustafa Yarrar, VUB (knowledge engineering / ontology engineering, knowledge representation and reasoning, ontologies for the Semantic Web, Semantic Web special topics)

Finally, some learning units are expected to be highlighted using some kind of 'KnowledgeWeb certificate', which can either be requested by other members of KnowledgeWeb (for example, if they have successfully used the learning unit for their own courses) or by other REASE users, who can express their opinion of the learning unit using the REASE feedback mechanism and rating scheme. This feedback mechanism is currently, however, non-public and might be extended to become public.

Depending on the different communities represented in KnowledgeWeb (Description Logics, Ontology Engineering,…), we also envision recommendations for reading, which might be different depending on the community. These recommendations might be generated automatically / semi-automatically, depending on the advanced semantic platform for learning (ASPL), which will be developed in WP3.3. In autumn 2007, the Semantic Web Education and Outreach working group (SWEO) of W3C has started to work towards a set of 'SWEO recommended' Semantic Web tutorials, which are planned to be published using REASE.

7 Usage of Learning Resources

In this section, we report about the usage of REASE and the provided resources. The presented numbers are gathered from log files of the underlying web server and from the bookings and access information of the database, on which REASE is based.

7.1 General Usage of the REASE Web Pages

The usage of the REASE web pages since it went online in July 2004 is shown in the following Figure 3 (the statistics were taken on Nov-1 2007 from the web server log file excluding accesses from popular web robots and accesses from within the hosting domain of REASE):

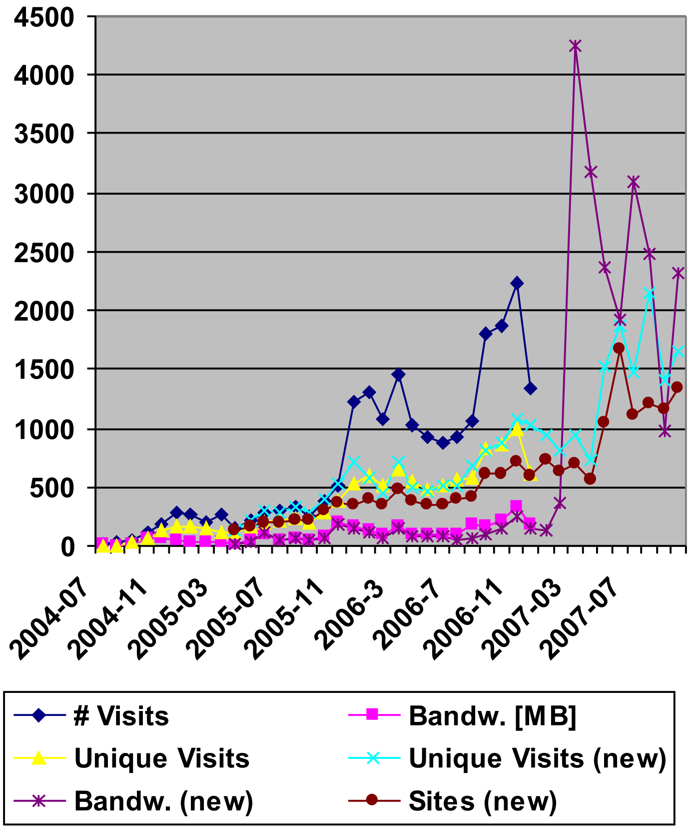


Figure 3: Accesses to the REASE web server

The first public announcement of REASE was issued in October 2004, leading to an initial increase in the access statistics, because a first set of learning resources became available in November 2004. Whereas the number of accesses remained stable in the first half of 2005, it increased again in summer 2005, mainly because of the summer school activities of KnowledgeWeb and REWERSE. Especially, the teachers of the REWERSE summer school were required to upload their material before the summer school starts so that the students could access them from REASE directly. Finally, the usage of the REASE web pages increased again starting from October 2005. As an example, the REASE web pages were visited about 500 times from about 380 unique v isitors in November 2005, downloading an approximate amount of 200 MB of data. Even though especially the increase in the number of non-unique visitors was partly caused by the evaluation activities in work package 3.3 (REASE is one service connected to ASPL-1, which was evaluated in November 2005 at USFD, OU, and Universitatea "Al. I. Cuza" Romania as reported in D3.3.5), the main increase could not be associated with a single or few events. After the addition of several learning resources in the end of 2005, the general usage of REASE increased significantly at the beginning of 2006 with the usual decrease in the summer. The peak in November 2006 is again partly caused by the evaluation of the platform.

In 2007, we had to use a different log analyzing tool (webalizer) because of several security flaws in the old analyzer and an insufficient support for filtering search engine traffic and web spammers. We have shown the results of both analysis tools for 2006 simultaneously for comparison reasons. As a result, '(unique) visits' are still increasing overall and have exceeded the 1000 visits per month since summer 2007, which is also true for the 'sites' measure, which means that REASE users are currently spread over different institutions. Furthermore, the consumed bandwidth has increased significantly because of REASE users accessing the video recordings from the KnowledgeWeb summer schools. Unfortunately, the new web log analyzing tool does only provide the notion of 'unique' visits (the light blue curve) and does not allow judging repeated visits anymore. However, a manual analysis for the evaluation of REASE in the journal paper (see appendix) has shown that users return on average about 2 times (measured for both visitors from web sessions as well as log-ins from registered users).

More details are discussed in the following sections about registered users and institutions and the actual access patterns of the learning material.

7.2 Registrations on REASE

To access most of the material on REASE, users have to register first and specify (very few) information about their hosting institution (i.e., university or company and their country). The following Figure 4 depicts the number of registered users / institutions on REASE.

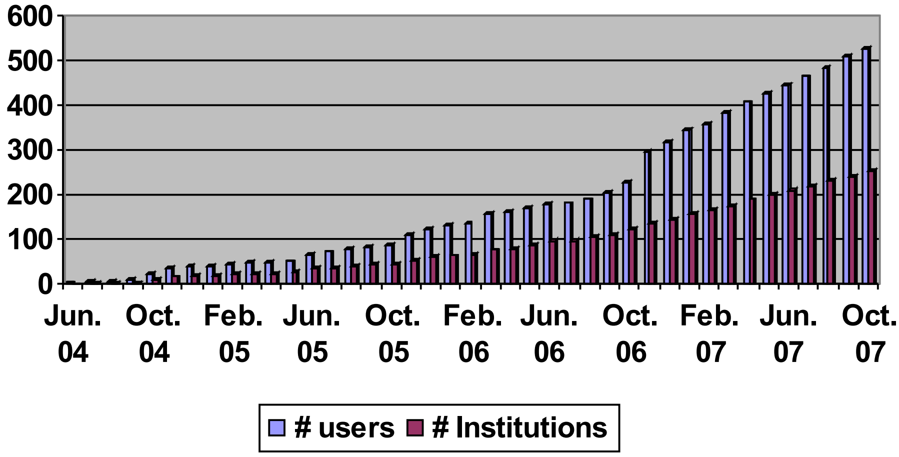


Figure 4: Registered users and institutions

The first public announcement of REASE in October / November 2004 led to the registration of users and institutions from KnowledgeWeb mainly. The second peak in June 2005 is mainly caused by the fact that REASE was used to distribute the learning material for the REWERSE summer school as mentioned above. The increase in November 2005, however, is not dominated by KnowledgeWeb or REWERSE activities, only 2 from the 12 additionally registered institutions were actually directly related to one of these NoEs. In 2006, the increase in the number of registered institutions co ntinued with a slight increase in the rate of change. The number of registered users also increased and doubled compared with the end of 2005. However, about 50 additionally registered users are due to the user study in November 2006. In 2007, the increase in registered users and institutions continued further, again with a slightly increased rated of change as compared to 2006.

7.3 Access to REASE Resources

REASE resources were accessed as shown in the following Figure 5:

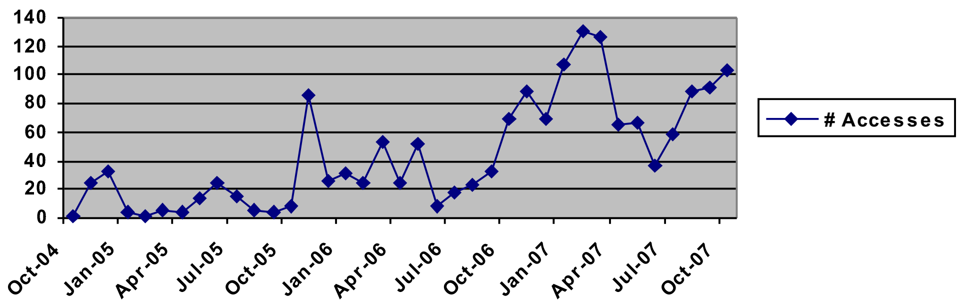


Figure 5: Access to REASE resources

The peaks in October and November 2004 were caused by a few users who accessed quite a large set of learning units, obviously playing around with the platform. This included people from KnowledgeWeb or REWERSE, but also one person from outside both NoEs. The peak in July 2005 could be because of the KnowledgeWeb and REWERSE summer schools, which took place at that time. The peak in November 2005 is partly (about 40 from the 85 accesses) caused by the evaluation activity of WP3.3. However, 39 accesses were from users all over the world (Malaysia, Germany, USA, France, Brazil, Canada, and Greece), who were definitely not involved in KnowledgeWeb or REWERSE! Accesses continued on a much higher base level in 2006 with the usual 'summer break' in June-August. As in 2005, the peak in November is partly a result of the user study. In 2007, again the overall number of accesses to REASE has increased with the (again) usual decrease in accesses during the summer.

7.4 Most Popular Resources on REASE

Based on the access pattern by REASE users, the following learning units are the 10 most popular ones on REASE (as of October 1st 2007):

1. Semantic Web Tutorial (UKARL)
2. Semantic Web Lecture – Logics (L3S)
3. Ontological Engineering (UPM)
4. OWL – Web Ontology Language (TU Wien)
5. Fundamental Research Challenges generated by the Semantic Web (VU)
6. Semantic Web Lecture – Basic Building Blocks (L3S)
7. Semantic Web Lecture - Introduction and Overview (L3S)
8. Semantic Web Use Cases (FUBerlin)
9. A short tutorial on the Semantic Web (UKARL)
10. Ontology Engineering Best Practices – Building and applying the SWRC ontology (UKARL)

Resources 1 and 8.-10. on this list are also suited for industrial audiences.

While analyzing why the "Semantic Web Tutorial" became most popular we noted the following:

- It is the only English material on REASE providing an introduction to Semantic Web for people from industry
- It is on rank 6 on Google for the search 'Semantic Web Tutorial industry' (probably because the KnowledgeWeb portal is on rank 5 for the Google query 'Semantic Web industry'). [both ranks validated on 2005-12-19]. In October 2007, it was still on rank 12 on Google.

Of course, becoming the most popular resource is only possible because the absolute number of bookings in REASE is still not very high (about 70 overall for the most popular resource).

8 Sustainability

Overall, we have undertaken several measures to ensure the sustainability of the work related to REASE:

- Foundation of EASE which will host REASE starting from 2008
- Publication of the topic hierarchy in the ontoworld wiki, allowing anyone being interested in the development of the topic hierarchy to participate; also

publication of the history of the topic hierarchy in the ontoworld wiki for a better understanding of the design decisions made during the development process.
- Publication of high-quality learning resources (such as the summer school videos) to attract users to REASE
- Connection of the different components created in KnowledgeWeb, such as between REASE and the topic hierarchy pages in ontoworld or between REASE and the ASPL-2 demonstrator.

As an incentive for author to continue publishing resources on REASE, all resources on REASE were assigned a Digitial Object Identifier (DOI) in 2007 so that they can also be counted as 'educational publication' in the same way as 'scientific publications' from journals or conferences.
Finally, the SWEO working group at W3C is currently considering using REASE for publishing a list of recommended Semantic Web tutorials in the spirit of a 'W3C' certificate.

9 Summary and Future Work

In summary, the following main contributions were made regarding the activities related to REASE in 2007:

- A new version 2 of the classification system in the REASE catalogue.
- 119 learning units from KnowledgeWeb partners available on REASE.
- Increase of the percentage of KnowledgeWeb learning units for industrial education from 30% to 44%.
- A refined evaluation of REASE based on a questionnaire sent to all REASE users, a classroom-style user study with different user groups, and a transaction log analysis.
- An evaluation of the usage of REASE which shows a promising increase in usage during 2007.

The discussion about the REASE catalogue took place between KnowledgeWeb and REWERSE partners and finally merged into a global discussion of a Semantic Web Topic Hierarchy, as described on the ontoworld wiki. It was also used for organizing Semantic Web conferences like the ESWC (session organization) and for the creation of the KnowledgeWeb Technology Roadmap. Both the topic hierarchy as well as the REASE catalogue is expected to be subject of a further evolution since the research area 'Semantic Web' will also evolve over time.

Future work regarding REASE comprises the following issues:

- Continue publishing learning units, again focused on material for industrial education, but also trying to fill those categories in the topic hierarchy, which are not covered yet by existing material. It should be noted, however, that the topic hierarchy is for general purpose usage and that for some of its categories, while important, for example, for the organization of sessions in a research conference, no educational material may be available at all.
- Continue to publicize REASE and recruit new users

- Continuous monitoring of the quality management process and application to new resources.

10 References

[WJ99] Welty, Chris and Jessica Jenkins. 1999. An Ontology for Subject. J. Data and Knowledge Engineering. 31(2)155-181. September, 1999. Elsevier.

[GMZ06] F. Giunchiglia, M. Marchese and I. Zaihrayeu, "Encoding Classifications as Lightweight Ontologies", accepted for publication in Journal of Data Semantic, volume VIII, Springer, Berlin, Germany, 2006.

[DBT06] J. Diederich, W.-T. Balke, and U. Thaden, „The Semantic GrowBag Demonstrator for Automatically Organizing Topic Facets", SIGIR06 Workshop on Faceted Search, Seattle, USA, 2006.

[DB07] J. Diederich and W.-T. Balke: The Semantic GrowBag Algorithm: Automatically Deriving categorization systems., In: 11th European Conference on Research and Advanced Technology for Digital Libraries (ECDL 2007), Budapest, Hungary, September 16-21, 2007

11 Appendix

The following paper is a pre-print of a journal paper accepted for publication in the journal "The New Review of Hypermedia and Multimedia". It is an extended version of the evaluation section of the previous version of this deliverable.

New Review of Hypermedia and Multimedia,
Vol. 00, No. 00, Month 200x, 1–19

REASE: The Repository for Learning Units about the Semantic Web

Jörg Diederich*[1] and Martin Dzbor[2] and Diana Maynard[3]

[1]L3S Research Center, Leibniz Universität Hannover, Appelstr. 9a, 30167 Hannover, Germany
[2]Knowledge Media Institute, The Open University, Walton Hall, Milton Keynes, MK7 6AA, UK
[3]Department of Computer Science, Regent Court, 211 Portobello St, University of Sheffield, UK

REASE, EASE's repository for Semantic Web learning units, is a unique repository containing a diverse set of learning resources, ranging from annotated slides to video recordings and from one-hour tutorials to fully-fledged university courses, for both academic and industrial audiences. It aims to accommodate the heterogeneous requirements of different users trying to learn about the Semantic Web. This article provides an introductory description of REASE and describes the lessons learnt while creating and designing the repository, based on a mixed-method evaluation involving actual REASE users, a clean-room evaluation, and usage statistics gained from a transaction log analysis.

Keywords: Digital libraries; REASE; Semantic Web; education; user evaluation

1 Introduction

In recent years, Semantic Web technologies (Berners-Lee *et al.* (2001)) have matured significantly and are increasingly deployed outside the academic domain in industrial settings (Angele and Gesmann (2006)). As news spreads about these successful early use cases, more and more companies are becoming interested in such technologies, moving the technology further from early adopters (Benjamins (2006)). For a successful adoption in specific business domains, well-founded knowledge is required about both Semantic Web technologies and specific characteristics of the business domain, since experts in both areas are often not readily available.

To support educating such people in Semantic Web technologies, we have set up the EASE foundation [1] (Diederich *et al.* (2006)) and an associated repository for learning units (REASE), available at http://rease.semanticweb.org, in the context of the European projects KnowledgeWeb [2] and REWERSE [3]. The main goal is to provide a collection of high-quality learning resources about topics related to the Semantic Web, which supports both academic and industrial communities in learning about the Semantic Web. Specifically, REASE has the following main characteristics:

- Its collection of learning resources is unique; many resources are exclusively available on REASE only.
- REASE serves as a central entry point about the Semantic Web, which is especially helpful for users who do not know what to search for.
- Without the need for user training, REASE supports different types of user interaction (as recommended by Fuhr *et al.* (2007), McKay *et al.* (2004) for Digital Libraries in general) to access the collection through the following mechanisms:
  - Keyword-based search,

*Corresponding author. Email: diederich@l3s.de

[1]http://ease.semanticweb.org
[2]http://knowledgeweb.semanticweb.org
[3]http://rewerse.net

New Review of Hypermedia and Multimedia
ISSN 1361-4568 print / ISSN 1740-7842 online © 200x Taylor & Francis
http://www.tandf.co.uk/journals
DOI: 10.1080/1361456YYxxxxxxxx

- Browsing the resource collection using a catalogue based on the Semantic Web Topic Hierarchy [4] or based on social navigation (Brusilovsky *et al.* (2005)) ('users who accessed this resource also accessed...') to support serendipity,
- Flexible filtering of result lists allowing users to browse the set of available resources according to several facets best suited for their context (Fuhr *et al.* (2007), Borgman (2002a)),
- Ranking of resources based on popularity in terms of downloads (such information would be scattered around in different web server log files without a central repository).

The main contribution of this article is two-fold. After a short description of related work in Section 2, Section 3 presents a brief introduction to REASE, which is based on well-known principles in the area of pedagogics and accessibility, and describes its user interface. Second, we evaluate REASE, focusing in particular on the user interface, using a mixed-method approach along three dimensions: (a) using feedback from actual users gained from a questionnaire; (b) using REASE in a clean-room user study; and finally (c) evaluating the transaction log files from our web server running REASE. The main findings are:

- REASE, i.e. the resources it provides, appears to be useful for its users in general,
- The user interface of REASE in general has a high usability,
- It is useful to have both keyword-based search and facilities for navigating the set of available resources, but ease of searching and navigation has to be achieved simultaneously,
- The value of the flexible filtering remains unclear,
- Accidental resource discovery is highly valuable,
- Ranking of resources based on popularity is often used.

2 Related Work

While most of the main features of REASE exist in other repositories already (e.g., browsing is available in all legacy Digital Libraries such as the ACM Digital Library), it is the combination of these features which makes REASE unique, especially in the domain of Semantic Web education, where there is no such central portal yet [1]. The following paragraphs present the existing repositories most closely related to REASE.

The EducaNext portal (EducaNext (2007)) is a repository for learning resources, similar to REASE, and actually building upon the same technology, the Universal Brokerage Platform (Simon (2004), Simon *et al.* (2003)). However, EducaNext collects learning resources from any domain and hence does not enable learning about a specific domain such as the 'Semantic Web' by browsing the collection of available resources. Furthermore, it has not been designed according to well-known accessibility guidelines, it does not support an easy search and navigation (as revealed by the evaluation of the first version of REASE, which was based on the EducaNext design), and it provides no support for filtering resources according to specific target audiences such as people from industry.

The videolectures.net portal (Videolectures (2007)) provides a large collection of about 2000 video recordings from courses, tutorials and interviews mainly from the domain of Knowledge Management, which includes many topics important for the Semantic Web domain. While it also contains a 'most popular resource' notion similar to REASE, it does not rely on a learning-based metadata schema like LOM (IEEE LOM (2002)) (on which REASE is based) and, moreover, the existing metadata is very incomplete; it basically contains only metadata about authors and the resource type (tutorial, course, interview,...). Hence, this portal does not support filtering for specific user groups such as people from industry, and does not allow browsing the domain along a domain-specific classification system.

Some educational resources about Semantic Web topics are already available on the web (e.g., on the homepages of university staff teaching about Semantic Web topics) and can be found using web search engines such as Google. However, especially for novices in the Semantic Web domain, it is not trivial to

[4] http://ontoworld.org/wiki/Semantic_Web_Topic_Hierarchy

[1] The SWEO group within W3C (http://www.w3.org/2001/sw/sweo/) is currently working towards such a portal; the authors are in close contact with the SWEO group.

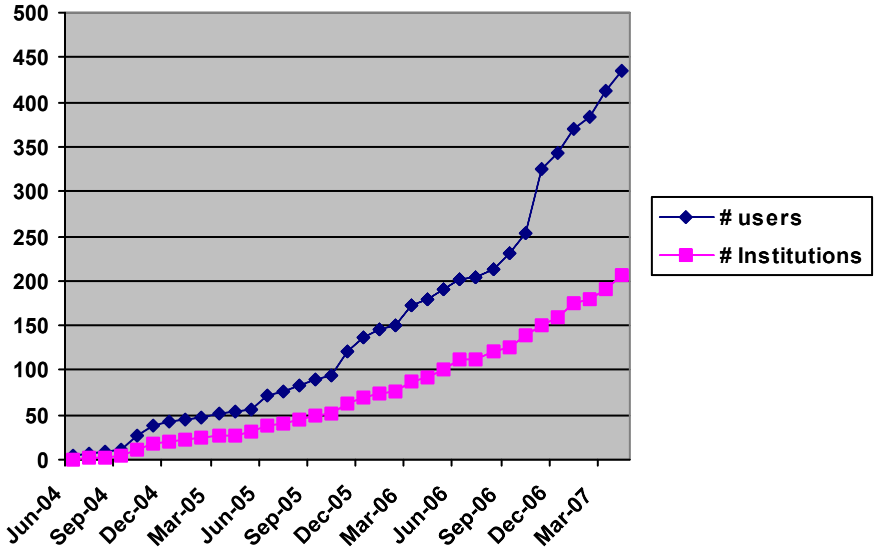


Figure 1. Registered users / institutions on REASE

get to know the relevant sub-topics, such as ontology engineering, logics, or knowledge representation, to be used as query terms. Also, unless one knows the names of relevant authors in the field, it is not possible to find interesting papers by random browsing through web pages. The REASE catalogue, which allows the user to navigate through the available set of resources, is based on the Semantic Web Topic Hierarchy, a classification system created by an expert group. This hierarchy helps in assessing the Semantic Web domain while searching for learning resources in REASE, which is not possible with existing web search engines. Furthermore, REASE contains a large set of resources which are not available elsewhere in the web. It also associates each learning resource with a specific copyright license which makes it easier to reuse in commercial scenarios, as the copyright situation remains unclear for most educational resources being published on university pages.

3 REASE: EASE's Repository for Learning Units about the Semantic Web

Though REASE is not a fully-fledged Digital Library (in particular, it lacks any organisational component behind it; it is currently run by a single person only), it can be described using the components 'content', 'users', and 'system / technology' according to the conceptual model for Digital Libraries of Fuhr *et al.* (2007).

Content. REASE contains a unique collection of currently more than 150 learning resources, such as slide sets or videos from lectures, tutorials, and keynote speeches from Semantic Web conferences. Over 60 of these have also been found useful for education of professionals, according to manual inspection by experts. In addition, REASE disseminates information about some announcements for events, where participants can learn about Semantic Web topics.

Users. Users of REASE are intended to be a member of the academic, research, or professional community interested in Semantic Web topics, which is currently a rather small, but dynamically growing community. Although access to REASE resources is open to anyone in general, it requires users to register in case they want to upload resources or access certain resources, which the providers have specified to be accessible only for registered users (about half of all resources).

Figure 1 depicts the number of registered users and institutions on REASE. As of May 2007, there are 430 users from 200 institutions, with about 150 users and 30 institutions coming from the partners being involved in setting up REASE (June'04 to June'05) and evaluating it (Nov.'05 + Oct./Nov.'06). Hence, REASE has increasingly attracted interested users during the course of 2006 and 2007, who invested the

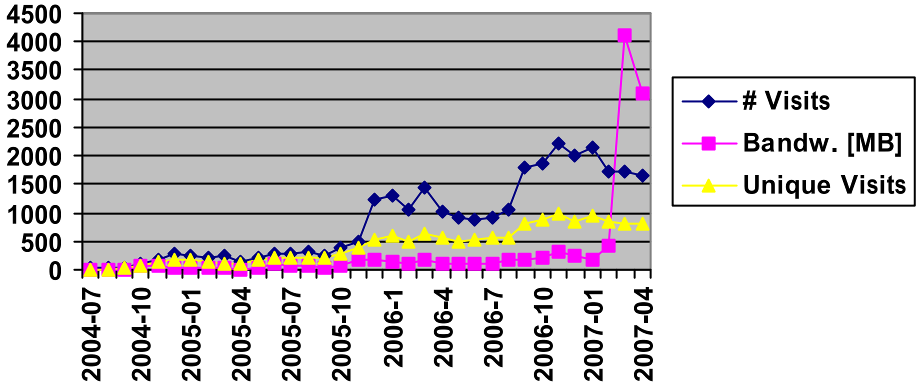


Figure 2. Web server statistics for REASE

time to register themselves and their institution to be able to access all learning resources on REASE. This is also shown in Fig. 2 where the gap between the single visits and the total number of visits has increased significantly, indicating that the same users are making many visits.

System. The web application functionality of REASE is provided by the Universal Brokerage Platform (Simon *et al.* (2003)), which stores metadata in a database, the learning resources in a distributed storage connected with web service technology, and uses Java Server Page technology to access and visualise the data on the REASE web pages.

Our main contribution to the REASE system was related to two parts:

(i) An adapted metadata scheme to better support different target audiences and to improve the search functionality;
(ii) A completely renewed *user interface design*.

Both are described in more detail in the following two sections.

3.1 *Metadata*

Regarding metadata in REASE, we required all providers of resource to specify at least metadata about:

- The resource language
- The title
- A short description of the content (important for search, as we currently cannot search inside the provided material, which would be feasible for textual documents but rather difficult for uploaded videos).
- The matching categories in the Semantic Web Topic Hierarchy (needed for browsing / navigating)
- The educational material type (tutorial, recorded lecture, presentation,. . .)
- The target audience.

Resource providers can specify additional metadata such as prerequisite knowledge in order to understand the learning resource, but many providers are not willing to spend much time in annotating their uploaded resources. Hence, we restricted ourselves to the above minimal set. All resources in REASE also have a Digital Object Identifier (DOI) automatically assigned by the system.

3.2 *User Interface Design*

Designing the user interface of REASE has been an iterative process involving several stages of evaluation, whose results directly informed the development of the next stage of design (Marchionini *et al.* (2003), Borgman (2002b)). This led to a total of three versions, which have been evaluated in two stages:

(i) Starting from the original EducaNext design, we performed only community-specific adaptations to address the intended user community of REASE (people interested in learning about the Semantic

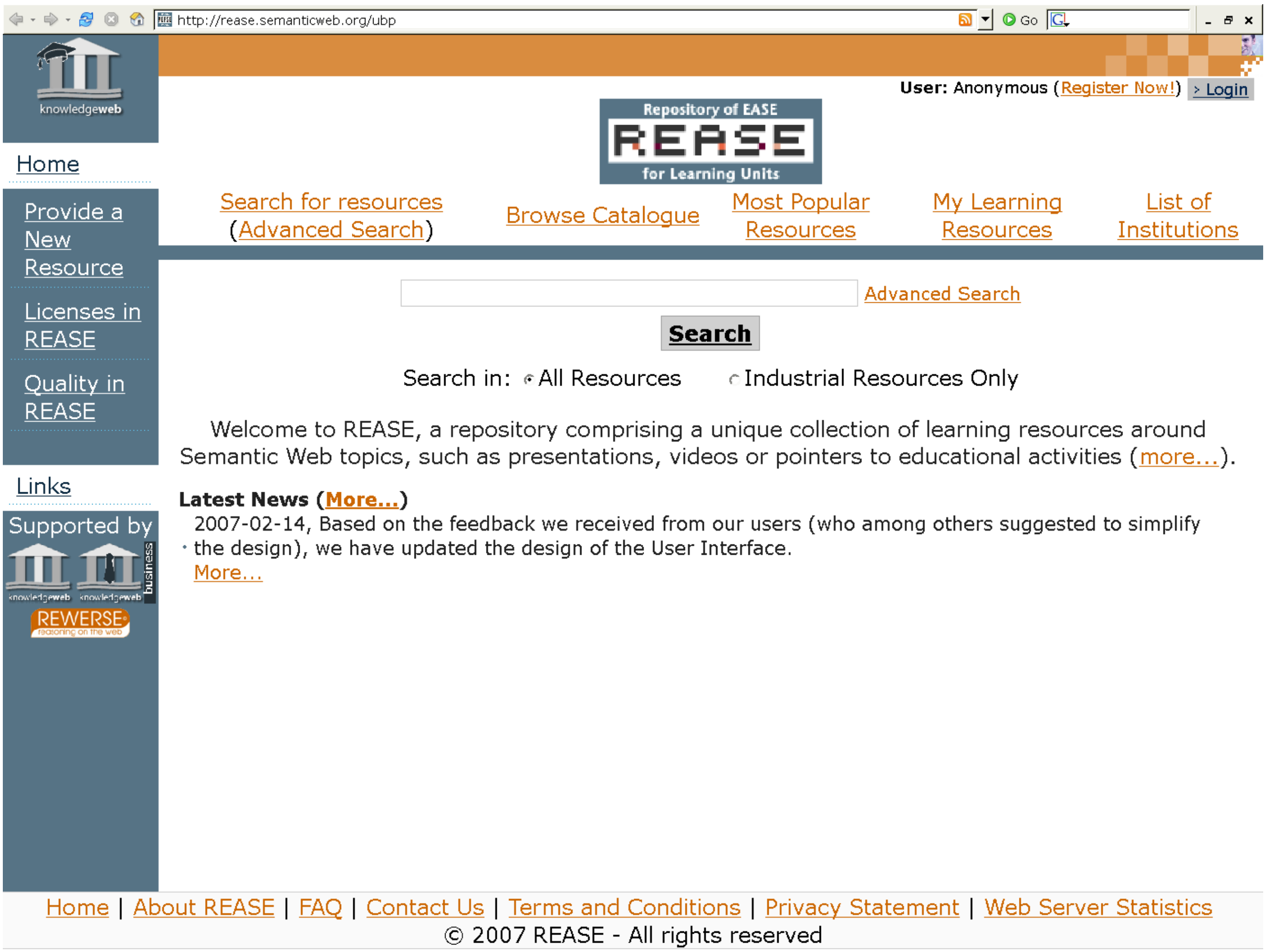


Figure 3. The main page of REASE

Web). Hence, we changed only the content of the web pages and the catalogue to the domain 'Semantic Web'. This version went online in August 2005 and formed the basis for the questionnaire-based evaluation performed in October 2006 (described in Section 4.4).

(ii) Using the outcome of the questionnaire, we developed a second prototype, essentially performing evolutionary changes to the organisation of web pages and the web design, and adding functionality such as a separate filter for industrial resources. This version was evaluated in a clean-room-style user study, conducted in November 2006 with about 35 participants (cf. Section 4.5).

(iii) Based on the feedback from the user study, we decided on a radical redesign of the main page, to give it a more search engine-like structure. This version has been online since February 2007 and is shown in Fig. 3. We will give some insights about changed user behaviour after the last version has been published using a transaction-log analysis in Section 4.6.

The REASE design exemplifies the 'middle ground' between the two extremes of a purely top-down "build it and they will come" design policy and a purely bottom-up design policy continuously involving users in the design process (Marchionini *et al.* (2003)): We have started using an existing design from a different web site, adapted it to our target user community and refined it based upon the feedback from our users.

In its current design, REASE's main web page (cf. Fig. 3) comprises a search box for a full-text search in the metadata of all learning resources in the centre of the screen. The search can be restricted to industrial resources only as a special option for users from industrial contexts. Above the search box, five further web pages are linked, leading to the main functions provided by REASE (i.e., search/advanced search, browsing the catalogue / most popular resources, accessing uploaded resources (for resource providers),

Table 1. Extracts from Bloom's taxonomy and their applicability to REASE

Category	Example role in REASE
Explanation	finding and recalling deposited learning materials based on a range of criteria
Comprehension	interpretation, clustering and classification of recalled knowledge into different sub-sets
Application	abstraction and re-use of knowledge in different context (e.g., in the web services of associated e-learning toolkit)
Analysis	formation of inference from patterns in the existing repository content (incl. popularity of resources, user feedback, etc.)
Synthesis	not applicable in the current version of REASE
Judgement/evaluation	not applicable in the current version of REASE

and a list of those institutions with registered users in REASE). More information about REASE can be found using the links in the footer section or in the left-hand menu where information about quality management and licenses is available.

The following two sections describe the underlying guidelines from pedagogics and web design regarding accessibility, as both were found to be highly important for successfully creating learning repositories.

3.3 *Pedagogical Considerations*

REASE is designed to support e-learning, and this focus is reflected in its functionality and evaluation. We evaluated REASE on general usability and user acceptance, but we also investigated its "fitness for purpose". For this purpose we based a part of the REASE evaluation on a skills-oriented pedagogic framework of Bloom. His taxonomy (Bloom (1965)), exemplified by Dzbor *et al.* (2005), is widely used in learning theory, and includes affective, psychomotoric and cognitive skills, but in REASE we concentrated only on the *cognitive level*. Skills relevant to REASE as a repository open to user exploration are summarised in Table 1 (see also Dzbor *et al.* (2005)).

The examples in Table 1 reflect the kinds of skills our learners need to acquire. In the prototype which was the subject of evaluation in this article, we focused on knowledge recall and comprehension. Given the pedagogic interest of REASE, the following questions are used to discuss its pedagogic benefits for learners:

- Does REASE help learners in performing tasks that require a particular prior knowledge, which the learner may not have acquired yet?
- Does REASE facilitate useful (i.e., not only precise but pedagogically helpful) recall of learning materials?

To answer the first question, REASE provides access and recall for a range of materials. The recall is executed efficiently and is contextualised within a small number of "context dimensions" (e.g., preferred audience or format).

With regard to the second question, REASE provides an access to the repository content; thus, it facilitates effective recall of the stored knowledge. REASE design supports recall on the 'macro level', on the level of materials, rather than smaller chunks of knowledge or indeed solutions to particular research problems. Here the main feature affecting REASE's effectiveness is the nature of learning materials. Much of the REASE content takes the form of lecture notes, presentations or handouts. These are all context-dependent resources that rely on someone putting the individual slides into a narrative. Since the narrative is not captured in REASE, the system only supports recall of materials that are semantically relevant to the learner's query. In other words, REASE cannot support recall on the pragmatic level, which leads to a situation where a resource may be semantically close to keywords, but less fitting for a very specific need the user had in mind.

REASE combines keyword-based search in metadata and search using the classification hierarchy. The combination helps learners by not only *recalling* learning material but also *comprehending* its position in the repository, its relationship to other materials. REASE does not directly support Bloom's levels from Table 1 beyond recall and basic comprehension. This is a deliberate design decision to keep the feature set of the core REASE search to an optimum size. The advanced skills, however, can be supported by optional external components, such as those mentioned for the ASPL application (Dzbor and Motta (2007)).

Thus, the REASE infrastructure contains features that contribute to supporting users in effectively carrying out (and possibly acquiring) several types of skills. As can be expected, the higher-level pedagogic skills are computationally expensive, and as such may reduce general user acceptance; e.g., by introducing time delays. This concern is addressed in REASE by splitting the core functionality of the infrastructure from the supportive technologies used by optional extensions (such as the aforementioned ASPL). REASE focuses on the keyword- and phrase-based recall and features such as the classification hierarchy have an auxiliary role – e.g., to show how the retrieved materials might be generalised or specialised, or how the community interprets a particular subject.

3.4 *Accessibility of REASE*

Accessibility is a crucial but often overlooked factor in the design of good websites, portals and software in general (Maynard (2005)). Accessibility is not about designing specifically for people with disabilities, but about making them things both easy to use and useful for everyone, regardless of the method in which they access it. For example, people accessing the Internet or web-enabled software via different modalities such as PDA or mobile phone are particularly susceptible to the problems caused by inattention to accessibility of design.

Furthermore, inaccessible websites are rarely future-proof. Accessibility covers many different issues, for example: design and appearance (fonts, colours, spacing, layout), use of hypertext, navigation ease, mystery meat navigation, appropriate use of images, sound and colour, aesthetics, formats for interactive behaviour (form filling, checkboxes, radio buttons, etc.), alternative modes of use, keyboard shortcuts and other mouse alternatives, ability to modify the setup, speed of familiarity acquisition, compliance with existing standards and so on.

Not only does a lack of accessibility limit the potential size of the user group, and actively antagonise those whom it affects, but it also has many legal implications. Aside from these, there are many other reasons why designers and content providers should take responsibility for providing accessible material. First, as already mentioned, they cannot know in advance who will be looking at their site, and via what medium. Second, good websites should be future-proof, and one of the best mechanisms for ensuring this (as much as is possible) is to conform to established accessibility guidelines and standards. For example, just because a website works in Internet Explorer does not mean it conforms to standards, because Internet Explorer itself does not. Third, a site that is discriminatory (for example, against people using "alternative browsers" or those who require large fonts), is not good publicity. Finally, an inaccessible website is simply evidence of bad web design. A truly reflection of web design skills is not about filling pages with animated graphics, Flash, Shockwave or unusual methods of navigation, it is about a website that everyone can use, but which is still aesthetic.

In the light of this, we ensured that our first version of REASE fulfilled a set of recommendations based largely on the W3C guidelines for accessibility [1], but also including some additional aspects devised by accessibility experts. Some of the main features are given below:

- Fonts and Colours:
  Colour combinations have to take into account the requirements for a high contrast. The most accessible colours to use are combinations of light/dark or dark/light for fonts and background, preferably in complementary shades (e.g., light blue and dark blue are better than light blue and dark red). For people with vision problems, no amount of increasing the font size can compensate for unreadable colour combinations.
- Images:
  Images have to be properly labelled and the information they contain should also be accessible by other means; graphics should be easy to interpret and should be resizable.
- Navigation:
  One of the standard accessibility rules is that links should be underlined wherever possible / practical

[1] http://www.w3.org/WAI/

(not just on mouseover), and that items which are not links (e.g., headings) should not be underlined. Not only does it make it more obvious what is a link and what is not, without having to mouse over an item to discover this, it also provides the user with the ability to distinguish previously visited links from unvisited links. Links which are identified solely by colour do not provide this, and can also be problematic for people with vision and colour problems.

- Layout:
  One of the most important accessibility issues that is often overlooked is the maintenance of a clear and simple layout that is consistent across the whole site. Cluttered or complicated pages, along with changes to the layout on different parts of the site make navigation difficult and cause confusion.

4 Evaluation

As REASE already found quite a number of users in autumn 2006 (about 240 registered users from about 110 institutions), we decided to carry out an evaluation to get insights about the users of REASE and also feedback from them.

We did not employ a fully-fledged formal evaluation (Reeves *et al.* (2003)) mainly because of the basically unknown user information needs (Marchionini *et al.* (2003), Fuhr *et al.* (2007)), which are hard to get for dynamically evolving communities such as the Semantic Web community. In fact, one particular objective of REASE is to help in moving Semantic Web technology from academia to industry, where it is just starting to achieve a higher attention. Hence, it was expected that only a few professionals will use REASE initially.

The evaluation reported in this section is aligned with known theoretical frameworks on Digital Library evaluation (Reeves *et al.* (2003), Marchionini (2000), Fuhr *et al.* (2007)), which cover aspects such as the evaluation objectives ("why and what to evaluate"), the evaluation methodology and methods ("how to evaluate / measure"), and the result analysis.

4.1 *Evaluation Objectives*

The purposes of evaluation in general are many-fold (Marchionini *et al.* (2003)) on a continuum ranging from evaluation research to product/system testing (Marchionini (2000)); evaluation objectives can range from very high-level, e.g., to measure the overall impact of a Digital Library (Marchionini (2000)) on its users' behaviour, to more procedural objectives, such as the usability of the user interface (McKay *et al.* (2004)).

Our evaluation was focused on testing the system aspect of REASE, especially when interacting with users (the usability of the most important user interface design elements, viz. for searching and browsing). We also tried to cover other aspects such as getting insights about users (are actual users from the intended user community?) or about the actual usefulness of the provided content (satisfaction of users).

4.2 *Evaluation Methodology and Methods*

Evaluation research in general recommends an evaluation methodology based on a mixture of qualitative and quantitative metrics gained from using several different methods of evaluation. The combination of methods in so-called *mixed-method evaluation designs* (as known from social sciences, cf. Greene *et al.* (1989)) or *multifaceted* evaluation (Marchionini *et al.* (2003)) can serve different purposes. For example, triangulation can enhance the validity of the evaluation results by comparing the results gained from different methods or even from evaluations of other (similar) systems.

According to Reeves *et al.* (2003) evaluation methods are also many-fold and can comprise observation (e.g., transaction logs), interview (including questionnaires and online forms), document analysis, and learning analysis. They should especially always integrate statistical data (Marchionini (2000)).

For the evaluation of REASE, we adopted a mixed-method approach including three different methods:

(i) As a first step, we sent a short and rather general questionnaire to all registered users of REASE in October 2006 focusing mainly on web design and the different search functionalities, but also trying to

Table 2. Overview on the evaluation

Metric	Covered by	Results
Usefulness: Users & Content		
Repeated Visits	Questionnaire, transaction logs	2.2 on average
Content	Questionnaire, (transaction logs)	At least some content valuable overall
Usability of the user interface		
... in general	questionnaire, transaction logs	High usability
... search & browse	questionnaire, user study, transaction logs	High usability
... filter	user study, transaction logs	Usability unclear
... browse most popular resources	transaction logs	Increasing use
... for accidental resource discovery	questionnaire, (user study), transaction logs	Usability high
Performance		
Retrieval	Questionnaire	In general fine, sometimes slow

find out about general user background.

(ii) After improving the user interface according to the outcome, we conducted a task-based user study in November 2006, similar to usability testing with a predetermined protocol as described by Reeves *et al.* (2003). It was completed by 35 selected volunteers at four different institutions and focused on the different information exploration functionalities of REASE (keyword search, advanced search, browsing). Its main purpose here was to receive immediate feedback from the users while they were performing a certain set of tasks, as opposed to the questionnaire, where people reported on their overall experience with REASE.

(iii) Finally, we analysed our log files to derive additional information for the evaluation, especially to compare with the first two evaluation stages for triangulation of the results. [1]

We also compared our results with evaluation results of other Digital Library systems as mentioned in the following sections. In this way, the evaluation combines aspects of an iterative evaluation and a comparative evaluation (Borgman (2002b)).

4.3 *Result Analysis*

To provide a better overview, Table 2 summarises the evaluated metrics and which evaluation method contributed to the results (if a part is mentioned in parenthesis, some contribution could be derived only). We group them according to Fuhr *et al.* (2007) along usefulness, usability and performance, though the latter was not an objective of our evaluation because of the relatively low number of simultaneous users. The questionnaire was intended to cover as many aspects as possible, while the user study focused on the usability of those parts of the user interface dealing with search and browse. The transaction log analysis finally complemented both the user study (for a restricted set of users and restricted to the period of time when the study was done) and the questionnaire (restricted to the period of time the first version was online) as well as giving some insights about the successfulness of the final adaptations of the user interface for version 3 (though this has not been validated using another evaluation method).

4.4 *Questionnaire-based Evaluation of REASE Version 1*

We conducted a user-based evaluation of REASE, addressing mainly the issues of usability, accessibility, as well as the quality of the material in the repository. It consisted of a simple questionnaire sent to all registered members of REASE, asking them about their general experiences and satisfaction (we achieved a return rate of about 30%). We highlight below some of the most important aspects of the responses. Unlike the SUS questionnaire (Brooke (1996)), which uses a score of 1-5 and alternates questions expecting positive and negative answers, we opted for a non-numerical scale where the questions are always worded positively, because it is less confusing for users (especially those who are not native English speakers), and because we can still calculate the scores from the responses, as described below. We also allowed for free-text comments to combine both qualitative and quantitative feedback.

[1] Even though the independence of the different methods was not ensured strictly in our evaluation for a theoretically sound triangulation, the results of the evaluation provided some insights and helped to improve REASE.

Table 3. Answers by Section

Score	5	4	3	2	1	Total
Usability	85	**147**	124	28	0	384
Information Finding	53	55	**57**	19	3	187
Information Providing	14	29	**31**	11	1	86

Table 3 shows the counts for answers to each section, where a score of 5 is the highest answer (e.g., "very easy") and is always the first answer out of the selection provided, ranging down to a score of 1 for the lowest answer (e.g., "very difficult") which is always the last answer out of the selection provided.

4.4.1 ***General Background.*** We were interested to see what type of people were using REASE, partly so that we could better analyse the results, and partly because we were just interested to see which kind of person REASE appeals to or may be useful for. So we asked some general questions about the users. Most people had visited it only a few times, but as time goes on, we expect the number of repeat visits to increase: it is encouraging at least that very few people had only visited once.

The majority of people visited REASE to find specific material, while the next highest reason was for general browsing. Others visited to upload material. Reasons for visiting ranged from searching for specific material in preparation for a lecture, through general curiosity, to increasing their knowledge of the domain.

The majority of REASE users first heard about it through Knowledge Web or REWERSE. This means that the results gained from the questionnaire can include a certain bias (i.e., overly positive feedback) as the initiators of the questionnaire were all well-known within both projects. Other users came across it via a search engine or via a colleague, while some students had been recommended it by their professor.

4.4.2 ***Usability and Accessibility Issues.*** As discussed earlier, usability and accessibility issues are extremely important for websites, portals and software. Our questionnaire focused on asking the users for feedback on issues such as those discussed in Section 3.4. All these questions scored reasonable marks, with aesthetics, layout, general usability and speed rated mainly as "quite good", while ease of navigation was rated slightly lower. In fact, some of the highest overall scores in the questionnaire were for this section, which is encouraging.

The system performance was also primarily rated as "quite good" although slow speeds were sometimes associated with downloading material, particularly at busy times of the day. One suggestion was to try to speed up the step between finding and previewing the material, which actually turned out to be a user interface problem.

Regarding content, we asked the users directly if the format of the material was suitable for their needs, to which in most cases the answer was "yes".

4.4.3 ***Finding Information.*** An important aspect of REASE is how easy it is to actually find the information one is looking for, since the quality of the information, and in fact the existence itself of REASE is irrelevant if the information cannot be found easily. This issue can be divided into two parts: first, layout and navigation issues, and second, the design of the catalogue topic hierarchy. The search mechanism itself is reasonably efficient regarding performance since the amount of metadata for 150 resources can easily be handled by standard server technology. However, it was established that the original topic hierarchy was inadequate and to some extent misleading.

Most people found it quite easy to find what they were looking for, although there were some problems associated with having to login to discover that there are some resources only available to registered users. This has now been modified. The search mechanism, on the other hand, was clearly more of a problem for many people. For example, some people got errors when performing a search (though it is not clear whether this was through user error or a problem with the system); other people found the hierarchy very unintuitive and difficult to use. It was also suggested that a more semantic search was incorporated, such as making suggestions based on a user's past history.

While the responses to these questions are not negative overall, they still show that people cannot find things as easily as they might, and that the search mechanisms could definitely be improved (as we did for version 2).

We also asked the users about "accidental information discovery" as we were interested to know about the browsing potential of the site, where people find things of interest that they were not originally searching for. An example of this is Amazon's "other people who bought X also bought Y" mechanism (and in fact this was explicitly mentioned in one comment). REASE has a similar mechanism "Users interested in this learning resource were also interested in the following learning resources:" which appears when a user looks at the page containing specific information about a certain resource. We found that a high proportion of users did indeed benefit from this mechanism.

4.4.4 ***Information Quality.*** As mentioned earlier, we wanted to know how users perceived the quality of the information they found on REASE. The majority of answers to this question were positive, and there were many comments about this topic, ranging from very complimentary to suggestions for improvement. It is clear, as one person said, that there is quite a wide range of quality in the resources, as there is only minimal quality control on the material uploaded. Some people liked the fact that there is material provided by PhD students as well as by more established academics, reflecting a wide coverage of topics. Another request was for more multimedia resources such as videos.

4.4.5 ***Providing Resources.*** Another important aspect of REASE is specific to the provision of resources. If the system is not conducive to users providing material with minimal effort and time, then they simply will not do so since there is little benefit to themselves apart from the wider dissemination of their work (and they will therefore find other methods of dissemination). Much of the questionnaire relates to both uptakers and providers of material, but we devote a small part specifically to the issue of information provision. It is important to find a balance between ensuring that there is sufficient information about the material provided (so that the uptakers are suitably informed about what is available and do not waste time looking at material which is irrelevant to them or not suitable for their needs), and on the other hand ensuring that for the providers the process of uploading their material is as easy and streamlined an experience as possible.

The responses showed that while information providers liked the system in general, they found the process of uploading information quite tedious. The main comments were that the process took too long and there was too much information to be filled in. Clearly there is a difficult tradeoff here, with information providers wanting to provide less information about their resources, but information seekers wanting more information about resources to be available. In fact, six users are responsible for more than 50% of the uploads, so we clearly need more active users in this respect for a proper evaluation.

4.4.6 ***Summary of the Questionnaire-based Evaluation.*** Due to the slightly lower scores from the questionnaire for usability and the free-text comments, the main result of the questionnaire was a desired improvement of the user interface with respect to 'ease of navigation'. We tried to address this in the second version of the user interface where we especially focused on achieving a less 'busy-page' layout (McKay *et al.* (2004)) and improved the browsing to guide users more quickly to their preferred pages. For example, we modified the presentation of the catalogue in several ways, e.g., adding an expand/collapse mechanism and avoiding that the currently selected topic is always shown on top of the catalogue.

Regarding content, we focused on providing more resources including video recordings; currently there are 34 resources with video on REASE.

4.5 *Evaluation of REASE Version 2 based on the User Study*

The user study had several intentions: first, to verify the findings from the questionnaire, which could only give a rather superficial feedback as such. Second, we wanted to check if version 2 of the user interface had an improved usability especially regarding 'ease of navigation', and third, to gain more insights about

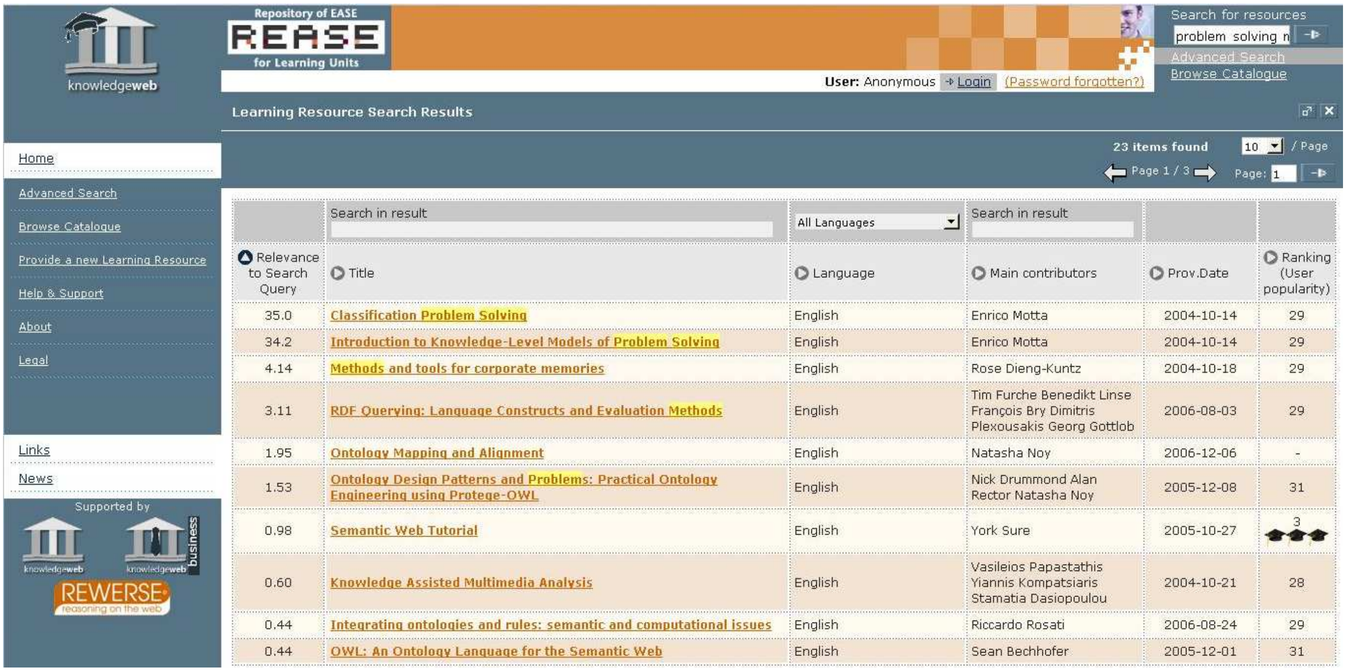


Figure 4. Result of the query 'problem solving methods'

particular features such as the filters. The users had to conduct a series of four short and fairly simple tasks (after a short time to become familiar with REASE):

(i) Finding material about the topic 'Problem solving methods' using simple keyword search,
(ii) Using the catalogue to find material about the topic 'Human language technologies',
(iii) Using the advanced search facilities to find resources about 'Description Logics',
(iv) Finding material about 'OWL' with any (favourite) search strategy.

Users were recruited from KnowledgeWeb partners but were not directly involved in the project. The results of the study are presented in more detail in the upcoming sections using data from transaction log files to enhance the understanding and assist in the interpretation of the results.

4.5.1 *Task 1: Find material about 'problem solving methods (PSMs)'.* For this task, people were basically intended to type in the term 'problem solving methods' (or any other term they considered useful) into the main keyword search box. We considered the first two results of the query 'problem solving methods' to be highly relevant as also indicated by the relevance score of > 34 from the underlying Lucene search engine, (cf. Figure 4), because they match on the phrase 'problem solving methods', not only on the single words in the phrase.

While the success rate in total and, hence, the effectiveness of search was very high (91% found these two resources), the users had many different strategies to get to the final results (which helped us to better understand how users search). Since we provided space for up to four result resources in the web-based form, the participants tried several alternative queries most likely to get more than the two required result resources. Table 4 summarises the issued queries, their percentage compared with all 82 queries, and how many of the two relevant resources could be found with the query.

The users posed on average 2.4 unique queries, an average also reported by other studies (e.g. 2.43 from Jones *et al.* (2000)) ranging from one query per user to eight at maximum (some even used the advanced search feature, which they should have used in task 3 only). 25 users returned the two result resources only, while two users from the 35 in total finally decided to put three result resources, eight users provided four resources even though the search engine indicated a low relevance of the resource with respect to the query.

Table 4. Query terms for task 1

Query term	Percentage	Hits
'problem solving method[s]'	27%	2 / 2
'psm[s]'	22%	0 / 2
'problem solving methods psm	9%	2 / 2
'problem solving methods' (advanced search)	7%	2 / 2
'"problem solving method[s]"'	6%	1 / 2
'problem solving'	5%	2 / 2

The main outcome of this task is that users try different queries to be able to find additional resources as they were not sure whether their initial query had a sufficiently high recall. Two main strategies were using acronyms and trying to relax the query manually (which does not change recall as the underlying search engine returns also results matching only single terms in the query). This could be due to lack of explanations on the web page (which we added later), though manuals have been found to be rarely used anyway (Jones *et al.* (2000)).

4.5.2 ***Task 2: Find materials suited for an industrial audience, discussing the role of Human Language Technologies (HLT) in the Semantic Web tools and applications.*** In this task, the participants were asked to use the catalogue to find material about Human language technologies and tools/applications. We considered one resource to be very relevant here (being in both catalogue categories 'human language technologies' and 'Semantic Web applications' and being suited for industrial education and being tagged as 'tutorial').

As a result, 33 of the 34 users (one had technical problems and dropped out from the study) found this resource and returned it as first result and all users again provided further resources as results (again perhaps due to the fact that we provided them with room for additional answers).

However, a closer log file analysis revealed that, while users were able to filter for industrial resources, only four users (12%) actually went directly to the two catalogue categories 'human language technology' or 'Semantic Web applications'. Instead, seven participants (21%) used the keyword search to finally find the result resources, the remaining ones just used the 'browse catalogue' functionality where they could also search in the title of all resources. Actually, this way of searching REASE was never intended to be a major search option.

4.5.3 ***Task 3: Find materials containing definitions of reasoning techniques; in particular deduction, in the materials about Description Logic (DL).*** In this task, the participants were intended to use the advanced search facilities. From the 91 answers, 65 (71%) were modules from the available Description Logics resources, that score highest on the query 'description logics'. Another 11 answers (12%) described the resulting top two resources of the query 'reasoning' / 'reasoning deduction', while 5 pointed to a resource that was returned only if the category 'knowledge representation and reasoning' was selected in the advanced search dialogue. Hence, the results were not as uniform because of the more complex task description that left room for several different queries.

To get to these results, the participants posed in total 111 queries about 'reasoning', 'deduction', 'description', 'logics', and/or 'DL', of which 103 (93%) were using the advanced search facility. Hence, users were more willing to follow our 'suggestion' to use advanced keyword search compared to task 2, where many did not use the catalogue as suggested. The users did not have a preference for any of the terms, so the frequency distribution of the used query keywords was rather uniform.

82 (80%) of the advanced queries were using the catalogue category filter to find appropriate results. From the 267 category filters used (3.3 per query), 137 (51%) were for the catalogue entry 'Knowledge Representation and Reasoning' and its three subentries ("logics", "logic programming", and "reasoning"). A further 39 (15%) were for 'Semantic Web Rules and Logics' including its sub-category "reasoning languages". Hence, most of the participants actually found those categories relevant for the given task. We conclude that the advanced search functionality is well-received by the users (at least when asked to use it) and could be used successfully to fulfil this task.

4.5.4 ***Task 4: Find material and the categories associated with the material in the catalogue containing description of species or layers in Web Ontology Language (OWL).*** This task was intended to find the favourite search mechanisms users preferred after having tried all of them in the previous tasks. However, most users did not really understand the formulation of the task and were looking for 'species' or 'layers' (for which there is nothing available in REASE). We could still find that 86 queries (61%) were posed using the advanced search dialogue and 47 (33%) of the advanced queries actually used the filter for catalogue classification, though this extensive usage might be caused by the fact that no resources were found at all and users gave up searching at some point in time.

Furthermore, only 16 users in total clicked directly on the category 'Ontology representation / Ontology Languages / OWL' in the catalogue, where these resources are listed as mentioned above, but 10 of them got to the category by clicking on the category mentioned in a summary page of another resource (hence, they were not browsing the catalogue as we expected but used search first and then navigated using the connecting links).

4.5.5 ***Summary.*** While the user study indicated that users have little difficulty in finding specific resources, they seem to have a preference for keyword-based search functionality and do not use the catalogue much, even if they are specifically requested to do so. In our view, this can have two reasons:

(i) In spite of the improvements we made for version 2, it is still too difficult to use the catalogue in general (which made users try other search facilities)
(ii) Users prefer keyword search technology in this task because the task description already contained specific keywords for a keyword search.

Another outcome of the user study was a set of very useful free-text comments , which was also reported by Marchionini (2000), including "Predefined filters such as 'learning resource type', 'target audience', or educational material type eased the search a lot." There were also comments about bugs that had occurred or how to improve the user interface, for example, regarding the comprehensiveness of some button labels. There were also still comments about the main page being too busy (similar to the feedback from the questionnaire). Hence, we finally did a more radical redesign of the main page towards a more search-engine like structure with a prominent keyword search box in the centre (cf. Fig. 3) which should have improved the usability of the latest version of REASE.

4.6 *Statistics and Usage*

In this section, we try to underline and enhance the findings from the questionnaire and the user study with statistics from our transaction log files based on all REASE users (not only the study participants as in the previous section). The presented numbers are gathered from the transaction log files of the underlying web server (some of them being available at http://rease.semanticweb.org/stats) and from the bookings and access information for REASE resources. In the web logs we have excluded accesses from the main search engines crawlers and access from within L3S Research Center, which is currently hosting REASE. We also had a significant problem with referrer-log spamming, which had to be handled semi-manually to get to a clean set of data for the analysis.

4.6.1 ***Overall Usage of the REASE User Interface.*** In general, a badly designed user interface can still lead to a high usage of any web platform if there is sufficient information for search engines to do a reasonable indexing of the offered content which then help users to find material using search engines. Hence, Table 5 shows the origin of users for REASE (as determined from referrer logging on the web server). As a result, the user interface functionality is in general highly used, most of the accesses to REASE pages have come from internal pages ($> 80\%$). For a further $13 - 14\%$, no referrer is known, meaning that REASE is accessed either directly (e.g., as a bookmark) or referrer logging is disallowed in the browser. Finally, $1 - 2\%$ of all accesses came from one of Google's search sites, the highest occurring external referring site.

Table 5. Origins of users

Origin	**Version 1** (04/2006-10/2006)	**Version 2** (11/2006-01/2007)	**Version 3** (02/2007-04/2007)	**Overall** (04/2006-04/2007)
REASE internal	58706/70890 = 83%	79231/92795 = 85%	49986/60839 = 82%	195345/233920 = 84%
Direct access / unknown	10336/70890 = 15%	11851/92795 = 13%	8769/60839 = 14%	32518/233920 = 14%
Google	928/70890 = 1%	1087/92795 = 1%	1107/60839 = 2%	2833/233920 = 1%

Table 6. Usage of search functionality

Search functionality	**Version 1** (07-10/2006)	**Version 2** (11/2006-01/2007)	**Version 3** (02/2007-04/2007)
Keyword search	221 (10%)	895 (17%)	448 (22%)
Advanced search	78 (4%)	631 (12%)	80 (4%)
. . . of which using catalogue filter	11 (14%)	336 (53%)	21 (26%)
Browse	1931 (87%)	3682 (71%)	1500 (74%)
. . . of which ‘Browse most popular resources’	0	43 (2%)	295 (20%)
Browse with filters	60 (3%)	364 (10%)	111 (7%)
. . . of which using title filter	31 (52%)	317 (87%)	55 (50%)
. . . of which using industrial filter	0	0	66 (60%)
Total	2230	5208	2028

Table 7. Going to the resource details page from. . .

Search functionality	**Version 1** (07-10/2006)	**Version 2** (11/2006-01/2007)	**Version 3** (02/2007-04/2007)
. . . direct access	312 (18%)	698 (20%)	420 (16%)
. . . the catalogue	562 (33%)	908 (26%)	554 (22%)
. . . of which browsing the most popular resources	0	30 (3%)	134 (24%)
. . . keyword search on the main page	194 (11%)	940 (27%)	509 (20%)
. . . another details page (recommendation)	33 (1%)	46 (1%)	107 (4%)
. . . from Google	426 (25%)	782 (23%)	811 (32%)
Overall	1695	3470	2564

Furthermore, about 65% of all visitors started from the main page (decreasing from 74% for version 1 to 49% for version 3) while 18% of all visitors started directly on one of the pages describing the details of a learning resource. This has increased from 14% for version 1 to 18% for version 2 and 26% for version 3, which is an indication that users are increasingly arriving from a search engine result page, from bookmarks or from other pages linking permanently to specific resources on REASE.

4.6.2 ***Searching and Browsing REASE.*** Table 6 shows which of the search functionalities (i.e., keyword search, advanced search, and browsing) is used how often, in relation to the overall number of browse/search events for each version. It can be seen that REASE users do use the browse functionality (about 75% of all search functionality) even though the keyword search functionality has gained more attention in the latest version (from 10% to 20%). The filtering options, however, are used rather rarely (less than 7%, not considering version 2 here, which included the user study where we explicitly requested people to use the filter), the filter to show resources for an industrial audience only was used by less than 4% of all users. For those users who actually do use this special filter to support professional users, however, it seems to be as valuable as the title filter, which was previously used most often. This is in line with a report from Jones *et al.* (2000) about users only rarely changing default values for search options.

Furthermore, about 20% of the keyword queries were posed using advanced search (leaving aside again the data from version 2). From the advanced queries, 14% used the catalogue filter in the first version and 26% in the third version.

Table 7 shows how many users actually looked at the details of a resource, arriving from different pages such as direct access, keyword search, browsing, or social navigation, to give an idea about the actual effectiveness (‘success’) of different search mechanisms. In general, the results from version 2 are highly biased by the user study and will not be considered further (they, however underline that participants in the user study seemed to have a particular preference for keyword search). The following analysis will focus on the evolution from version 1 to version 3.

In version 1 most users (33%) found interesting results to look at more closely from the catalogue (33%), followed by Google (25%), direct access (18%), and keyword search (11%). In version 3, the ‘success’ of Google increased to 32%, which in general is good news since it actually means that the popularity of

REASE in terms of connectivity in the web has increased. More important for the user interface, however, is the increased percentage of keyword search (20%) now being almost at the same level as the one of the catalogue (22%). This actually shows the effect of the more prominent position of the search interface on the main page. One open issue, however, is that the catalogue is based on the Semantic Web Topic Hierarchy, which was created during the past two years as a community effort to establish a classification system for the Semantic Web domain. This hierarchy is currently not as well-known and accepted compared to well-established classification systems in other domains, which might also effect the usability of browsing REASE and might make users prefer using keyword search.

Compared to the total number of browse/search events as shown in Table 6, the 'success' of browsing has actually increased from version 1 to version 3: in version 1 only 562/1931 = 29% of the browse event actually led to the user looking into the details of a resource, which increased to 554/1500 = 37%. Hence, the overall usability of the catalogue seems to have improved from version 1 to version 3 even though the overall usage of the catalogue has decreased due to the more prominent position of the keyword search box on the main page. Regarding keyword search, the success rate has increased slightly from 194/221 = 0.9 for version 1 to = 1.1 resource detail views per query for version 3.

Finally, both browsing the 'most popular resources' and the social navigation feature, where users clicked on the 'recommended related learning resource' links (based on actual access behaviour of users), are used increasingly in version 3, accumulating 5% and 4% of all accesses to resource detail pages, respectively. In particular, browsing the 'most popular resources' accounts for 20% of all browse events in version 3.

4.6.3 ***Most Popular Resources on REASE.*** Looking at the most popular resources in REASE provides some useful indicators about preferences of users with regard to the content and the format of the provided resources. Based on the resource downloads, the following resources are the ten most popular on REASE as of 16th May 2007:

(i) Semantic Web Lecture - Logics
(ii) Semantic Web Tutorial
(iii) Ontological Engineering
(iv) Semantic Web Lecture - Basic Building Blocks
(v) Fundamental Research Challenges generated by the Semantic Web (lecture recording)
(vi) OWL - Web Ontology Language
(vii) Semantic Web Lecture - Introduction and Overview
(viii) Semantic Web Use Cases
(ix) Web and Semantic Web Query Languages: A Survey
(x) Ontology Engineering Best Practices - Building and Applying the SWRC Ontology

In general, there is a lot of introductory material in this list (7 out of 10), which indicates that many REASE users are new to the topic. There also seem to be some professional users, to which the second, eighth and tenth topic were addressed. For the second topic, 'Semantic Web Tutorial', we note that it is the only English introductory resource being suited for professional education (provided in October 2005). It was highly ranked on Google for the search 'Semantic Web Tutorial industry' at the end of 2005 (in fact, 44 users found this resource coming from Google). However, a high visibility in Google does not necessarily translate into a high download rate on REASE: the resource 'Rules and Ontologies in F-logic' achieved most referrals from Google (120) but is placed on rank 59 when ranking resources according to user downloads.

4.7 *Comparison and Discussion*

This section compares and discusses the results of the different evaluation parts to provide some triangulation of the findings across the methods. These are grouped according to finding about users and content on the one hand, and usability as the main target of the evaluation on the other hand.

Table 8. Revisits / re-logins per month

Method	**Version 1** (04/2006-10/2006)	**Version 2** (11/2006-01/2007)	**Version 3** (02/2007-04/2007)	**Overall** (04/2006-04/2007)
Visits per visitor	2.0	2.3	2.0	2.1
Logins per user	2.6	1.8	2.5	2.2

4.7.1 ***Users & Content.*** From the questionnaire, we discovered that users were generally revisiting REASE a few times. This is supported by an analysis based on our web server logs (cf. Table 8) resulting in an average of 2.1 visits per visitor with non-significant deviation between the versions (defining a visit as 'new' when a visitor (e.g., its IP address) has not been active for more than 30 minutes). With regard to logging into the platform, registered people logged in about 2.2 times within each month with small variation between the versions (from 273 persons performing a login during 07/2006-05/2007, 122 (45%) were logging in more than once).

Under the assumption that users only come back when there is some value on the platform, we can also deduce that some information on REASE seems to haven proven useful to the users. While REASE can of course also be used to inform users in general about trends in Semantic Web education (McKay *et al.* (2004)), we assume that the resources are the main reason for users to visit the REASE web pages (which is at least the main objective of users, according to the questionnaire). This seems to be valid at least for those resources which are most popular overall (as mentioned in Section 4.6.2). This is, therefore, an indicator that at least part of the resources, i.e. the content, are useful for the REASE users.

4.7.2 ***Usability of the User Interface in General.*** From the referrer logging analysis, we found that most accesses to REASE are from within REASE itself. Furthermore, most users actually start from the main page (e.g., when referred to by someone as indicated in the questionnaire), and jump directly to the resource pages only in less than 10% of the cases (e.g., when pointed by a search engine). Together with the fact that users actually do return to REASE, we can infer that there is at least some general utility in the user interface.

4.7.3 ***Usability of Search & Browse.*** In the questionnaire, users specifically remarked that the usability of browsing the catalogue can be improved, which is why we focused on finding out more details about this part of the user interface in the user study. While the user study resulted in a pretty low usage of the catalogue (12%), the general usage of the catalogue in version 2 and the increase in its 'success' (in terms of a lower number of browse events to finally get to a resource details page) indicate that the improvements to the catalogue were well-received. We presume that the low usage of browsing in the user study is rather an artefact of the user study design (where users simply have used the keyword search in a previous task) and is not due to a low usability of the catalogue.

A contradictory result is related to the filters and the advanced search features; while the user study showed a high involvement of filters in task 3, this is not the case when looking at the log files in general. This could again be an artefact of the user study: the task was designed in a way that using advanced search and the filters would help to find the right results. In reality, users might be able to find relevant results already without these advanced features, perhaps because there are still too few resources to let this filter become useful in most cases, or simply because the usability of these filters in not high enough. The low usage of the special filter for industrial audiences, in particular, could also be caused by too few users having a professional background (which is somewhat underlined by the fact that the vast majority of the registered institutions in REASE have an academic background). This remains, however, an open issue overall.

In general, the log file analysis has revealed that browsing is used more often than searching, which is in line with results from other studies (McKay *et al.* (2004), Brusilovsky *et al.* (2005)). However, the 'success' of searching and browsing has become almost balanced in version 3 of the user interface. While we assume for the moment that this is due to the more prominent position of the search interface, this is subject to triangulation with other data and, hence, for further studies.

Note that the result of the questionnaire having 'searching for specific material' as a more important objective for visiting REASE than 'general browsing' cannot be compared readily to the actual usage of keyword search and browsing; users can just arrive with the intention of finding something specific and then end up browsing, for example, for additional interesting material.

Accidental discovery of resources by means of the social navigation feature of the resource was found to be beneficial for users in the questionnaire and also in the user study, where 30% of the users used it to search for material in task 4. The transaction logs also show some usage here, which is, however, small compared to the 'success' of browsing or searching. One reason for this might be that such social navigation is only available for a subset of the resources, another could be the less prominent position of this feature at the bottom of a resource description page.

5 Summary and Conclusions

This article has introduced the REASE repository, which comprises a unique collection of educational material about Semantic Web topics, providing different means to access them including keyword search, browsing, filtering, and social navigation.

Overall, the evaluation revealed that an accessible web design with a high usability is important for users of REASE. Furthermore, from the different search facilities on REASE, both keyword search and browsing the catalogue are used extensively and can lead users to relevant resources. There are indicators that the popularity of the different search facilities strongly depend on how prominent they are on the main page, but this is subject of further studies. In particular, because users use the catalogue for finding material, the efforts of the resource providers to classify each resource according to the categories in the Semantic Web Topic Hierarchy can be said to be worthwhile. When faced with a specific problem, users prefer using advanced search over using the catalogue (as shown in the user study) and use filters for catalogue topics. Other metadata, however, such as whether a resource is suitable for an industrial audience or not, is used rather rarely for searching. Finally, the evaluation has also demonstrated that browsing along the most popular resources is used: in the latest version about 20% of all browse actions were looking at the most popular resources.

We have also learnt that users tend to change their queries in systematic ways using query relaxation and acronym expansion if they cannot find a sufficiently high number of results they expected. To support this process in the future, we are planning to add an automatic query relaxation, which is possible due to the existence of the Semantic Web Topic Hierarchy. We will also continue running REASE and collecting further relevant resources. To ensure this will be done even after the end of the initiating projects KnowledgeWeb and REWERSE at the beginning of 2008, we have founded the EASE association which will take over the REASE repository at that point in time. This will also give the chance of doing a longitudinal evaluation as suggested by Marchionini (2000).

6 Acknowledgements

This research has been funded partly by the European Networks of Excellence KnowledgeWeb and REWERSE.

References

Angele, J. and Gesmann, M., Data Integration using Semantic Technology: A use case., in *Proceedings of the RuleML*, pp. 58–66, 2006.

Benjamins, R., The Semantic Web: Challenges and Applications, Talk held at the 4th Summer School on Ontological Engineering and the Semantic Web (SSSW'06) 2006.

Berners-Lee, T., Hendler, J. and Lassila, O., The Semantic Web. *Scientific American*, 2001.

Bloom, B., *A Taxonomy of Educational Objectives Handbook 1: Cognitive Domain*, 2nd 1965.

Borgman, C.L., Challenges in Building Digital Libraries for the 21st Century, in *Proceedings of the ICADL*, pp. 1–13, 2002a.

Borgman, C., Fourth DELOS Workshop. Evaluation of Digital Libraries: Testbeds, Measurements, and Metrics. 2002b, Technical report, Final report to NSF computer and information science directorate., Budapest, Hungary.

Brooke, J., SUS: a "quick and dirty" usability scale. In *Usability Evaluation in Industry*, pp. 189–194, 1996.

Brusilovsky, P., Farzan, R. and wook Ahn, J., Comprehensive personalized information access in an educational digital library, in *Proceedings of the JCDL*, pp. 9–18, 2005.

Diederich, J., Tolksdorf, R. and Nejdl, W., EASE: The European Association for Semantic Web Education, in *Proceedings of the Semantic Web Education and Training Workshop (SWET'06), co-located with the First Asian Semantic Web Conference (ASWC 2006)*, Sep., 2006, Beijing, China.

Dzbor, M. and Motta, E., Semantic Web Technology to Support Learning about the Semantic Web, in *Proceedings of the Proc. of the Intl. Conf. on Artificial Intelligence in Education (AIED)*, California, US, Jul., p. in press, 2007.

Dzbor, M., Motta, E. and Stutt, A., Achieving higher-level learning through adaptable Semantic Web applications. *Intl. J. of Knowledge and Learning*, 2005, **1**, 25–43.

EducaNext, The EducaNext home page: http://www.educanext.org, 2007.

Fuhr, N., Tsakonas, G., Aalberg, T., Agosti, M., Hansen, P., Kapidakis, S., Klas, C.P., Kovacs, L., Landoni, M., Micsik, A., Papatheodorou, C., Peters, C. and Solvberg, I., Evaluation of digital libraries. *Int J Digi Libr*, 2007.

Greene, J., Caracelli, V. and Graham, W., Toward a Conceptual Framework for Mixed-Method Evaluation Designs. *Educational Evaluation and Policy Analysis*, 1989, **11**(3), 255–274.

IEEE LOM, IEEE Standard for Learning Objects Metadata (LOM), IEEE Standard 1484.12.1-2002 2002.

Jones, S., Cunningham, S.J., McNab, R.J. and Boddie, S.J., A transaction log analysis of a digital library. *Int. J. on Digital Libraries*, 2000, **3**(2), 152–169.

Marchionini, G., Plaisant, C. and Komlodi, A., 6. In *The People in Digital Libraries: Multifaceted Approaches to Assessing Needs and Impact*, 2003 (MIT press).

Marchionini, G., Evaluating Digital Libraries: A Longitudinal and Multifaceted View. *Library Trends*, 2000, **49**(2).

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

McKay, D., Shukla, P., Hunt, R. and Cunningham, S.J., Enhanced browsing in digital libraries: three new approaches to browsing in Greenstone. *Int. J. on Digital Libraries*, 2004, **4**(4), 283–297.

Reeves, T., Apedoe, X. and Woo, Y., Evaluating Digital Libraries: A User-Friendly Guide. 2003, Technical report, University Corporation for Atmospheric Research.

Simon, B., Enabling Faculty to Share Knowledge over the Web: The EducaNext Approach., in *Proceedings of the World Conference on Educational Multimedia, Hypermedia and Telecommunications 2004*, pp. 2276–2280, 2004.

Simon, B., Retalis, S. and Brantner, S., Building Interoperability among Learning Content Management Systems., in *Proceedings of the WWW (Posters)*, 2003.

Videolectures, The videolectures.net portal: http://videolectures.net, 2007.