EU-IST Project IST-2003-506826 SEKT

SEKT: Semantically Enabled Knowledge Technologies

D11.2.1 BT Digital Library Scenarios

David Alsmeyer[1], Allyson Cheung[1], Michael Engler[2], Nick Kings[1], Ian Thurlow[1], Paul Warren[1].

[1]British Telecommunications plc.
[2]University of Applied Sciences (FH) Furtwangen.

Abstract

This document describes a set of example use cases and user scenarios for a suite of semantically enabled applications that are proposed for use in the BT digital library case study. The use cases and scenarios described in this document are targeted primarily at the end user. Use cases and scenarios for the domain expert and the system administrator roles are also considered, although to a lesser extent. The set of use cases and scenarios are not necessarily exhaustive; other use cases are likely to emerge as a result of further research and development within the technology-based work packages and as a result of feedback received from user validation of the applications. More coverage is given to the applications that are expected to be prototyped first, e.g. the search and browse, and knowledge sharing applications.

The expected benefits of using the applications are explained for each use case. The required technologies and the SEKT tasks and deliverables required to support the applications are identified. Mock-up screenshots are used, where appropriate, to help clarify the use cases. The expected benefits of using the applications are explained for each use case.

Keyword list: BT digital library, requirements, use cases, user scenarios.

WP11 Case study – retrieving and sharing knowledge in a digital library

Report	PU
Contractual date of delivery 31/12/04	Actual date of delivery 10/02/05

SEKT Consortium

This document is part of a research project partially funded by the IST Programme of the Commission of the European Communities as project number IST-2003-506826.

British Telecommunications plc.
Orion 5/12, Adastral Park
Ipswich IP5 3RE
UK
Tel: +44 1473 609583, Fax: +44 1473 609832
Contact person: John Davies
E-mail: john.nj.davies@bt.com

Empolis GmbH
Europaallee 10
67657 Kaiserslautern
Germany
Tel: +49 631 303 5540
Fax: +49 631 303 5507
Contact person: Ralph Traphöner
E-mail: ralph.traphoener@empolis.com

Jozef Stefan Institute
Jamova 39
1000 Ljubljana
Slovenia
Tel: +386 1 4773 778, Fax: +386 1 4251 038
Contact person: Marko Grobelnik
E-mail: marko.grobelnik@ijs.si

University of Karlsruhe, Institute AIFB
Englerstr. 28
D-76128 Karlsruhe
Germany
Tel: +49 721 608 6592
Fax: +49 721 608 6580
Contact person: York Sure
E-mail: sure@aifb.uni-karlsruhe.de

University of Sheffield
Department of Computer Science
Regent Court, 211 Portobello St.
Sheffield S1 4DP
UK
Tel: +44 114 222 1891
Fax: +44 114 222 1810
Contact person: Hamish Cunningham
E-mail: hamish@dcs.shef.ac.uk

University of Innsbruck
Institute of Computer Science
Technikerstraße 13
6020 Innsbruck
Austria
Tel: +43 512 507 6475
Fax: +43 512 507 9872
Contact person: Jos de Bruijn
E-mail: jos.de-bruijn@deri.ie

Intelligent Software Components S.A.
Pedro de Valdivia, 10
28006
Madrid
Spain
Tel: +34 913 349 797
Fax: +49 34 913 349 799
Contact person: Richard Benjamins
E-mail: rbenjamins@isoco.com

Kea-pro GmbH
Tal
6464 Springen
Switzerland
Tel: +41 41 879 00
Fax: 41 41 879 00 13
Contact person: Tom Bösser
E-mail: tb@keapro.net

Ontoprise GmbH
Amalienbadstr. 36
76227 Karlsruhe
Germany
Tel: +49 721 50980912
Fax: +49 721 50980911
Contact person: Hans-Peter Schnurr
E-mail: schnurr@ontoprise.de

Sirma AI EAD, Ontotext Lab
135 Tsarigradsko Shose
Sofia 1784
Bulgaria
Tel: +359 2 9768 303, Fax: +359 2 9768 311
Contact person: Atanas Kiryakov
E-mail: naso@sirma.bg

Vrije Universiteit Amsterdam (VUA)
Department of Computer Sciences
De Boelelaan 1081a
1081 HV Amsterdam
The Netherlands
Tel: +31 20 444 7731, Fax: +31 84 221 4294
Contact person: Frank van Harmelen
E-mail: frank.van.harmelen@cs.vu.nl

Universitat Autonoma de Barcelona
Edifici B, Campus de la UAB
08193 Bellaterra (Cerdanyola del Vall` es)
Barcelona
Spain
Tel: +34 93 581 22 35, Fax: +34 93 581 29 88
Contact person: Pompeu Casanovas Romeu
E-mail: pompeu.casanovas@uab.es

Executive Summary

This document describes a set of example use cases and user scenarios for a suite of semantically enabled applications that are proposed for use in the BT digital library case study. The use cases and scenarios are targeted primarily at the end user. Use cases and scenarios for the domain expert and system administrator roles are considered, although to a lesser extent.

The use cases are explained with reference to user scenarios that explain the functionality of the applications. The expected benefits of using the applications are explained for each use case. Potential issues are highlighted. The required technologies and the SEKT[1] tasks and deliverables required to support the applications are identified. Mock-up screenshots are provided, where appropriate, to help explain the functionality and the user interaction.

The applications developed from these user scenarios will be supported by technology researched and developed in the technical work packages of the SEKT project. The technical work packages will provide an architecture for the semantic annotation, indexing and retrieval of documents with regard to extensive semantic repositories. All applications will make use a common domain-specific ontology-based on the PROTON ontology [1].

A number of prototype applications are expected to be developed from these use cases and scenarios. Other use cases are likely to emerge as a result of further research and development within the technology-based work packages and as a result of feedback received from user validation of the prototype applications. Moreover, user validation of the use cases and prototypes will establish the functionality that will be integrated into the final set of tools and applications.

[1] http://www.sekt-project.com

Contents

1 Introduction

This document describes a set of example use cases and user scenarios for a suite of semantically enabled applications that are proposed for use in the BT digital library case study. The use cases and scenarios described in this document are targeted primarily at the end user. Use cases and scenarios for the domain expert and the system administrator roles are also considered, although to a lesser extent. The set of use cases and scenarios are not necessarily exhaustive; other use cases are likely to emerge as a result of further research and development within the technology-based work packages and as a result of feedback received from user validation of the applications. More coverage is given to the applications that are expected to be prototyped first, e.g. the search and browse, and knowledge sharing applications. The expected benefits of using the applications are explained for each use case. Potential issues are highlighted. The required technologies and the SEKT tasks and deliverables required to support the applications are identified. Mock-up screenshots are used, where appropriate, to help clarify the use cases. The expected benefits of using the applications are explained for each use case.

Section 2 gives an overview of the BT digital library as it is today, and provides an outline description of a semantically-enabled digital library that could be developed from the use cases and scenarios described in this document. Use cases targeted at the end user are described in section 3. Section 4 provides descriptions of the use cases that support the domain expert and the system administrator. A description of a preliminary BT digital library domain ontology[2] is given in section 5. Finally, Appendix A gives an outline design of a knowledge sharing application derived from the use cases. This design will be modified and elaborated in the second year of the project as the case study prototype is developed.

[2] Work on the BT digital library ontology is ongoing; the ontology will develop further as the tools and applications are researched and developed.

2 The BT digital library

2.1 The BT digital library today

The BT digital library offers users the capability to search and browse an extensive on-line collection of technical and business journals, conference proceedings, and electronic books. The Digital library provides access to over 4 million records from the Inspec[3] and ABI/INFORM[4] databases. BT subscribes to approximately 1000 on-line publications, giving users access to the full-text of over 900,000 scientific and business articles and papers.

The digital library is used by a wide-range of BT employees[5], many with a strong technical or research background. Increasingly people interested in publications with a greater commercial focus are using the library. Users can search the content of the digital library through a simple keyword-based search engine. Alternatively, users can browse the contents and abstracts of the library's journals. A limited selection of advanced search options, e.g. search by author's name and search by title, are provided for the specialist/expert user.

In addition to the basic search and browse functions, users can browse and subscribe to *public information spaces*; special domains of the digital library purposely created to gather information relating to topics known to be of interest to people. These information spaces bring together, in a single point of access, relevant content from the library's Inspec and ABI databases, and details of new books related to that topic area. Users subscribing to an information space receive alerts when new articles and recently published books become available. Users can also create their own *private information spaces*, enabling them to gather information on articles and books that are of relevance to their own topics of interest.

An alerting function enables users to register to receive updates on the contents of new issues of selected journals. A limited degree of personalisation enables information space membership, journal update alerts, and a list of recently read articles to be presented to the user.

2.2 A semantically-enabled digital library

Based on an earlier requirements capture exercise [3], and known limitations of some elements of the digital library, the following applications are proposed for a semantically-enabled digital library:

2.2.1 End user applications

1. A semantic search and browse application that will help improve the relevancy of search results by establishing the user’s search context.
2. Semantically enabled public and private information spaces that enable users to receive updates on more fine-grained topics of interest from the digital library, without loosing the 'sense of community' that exists with the existing information spaces.

[3] http://www.iee.org/Publish/INSPEC/

[4] http://www.il.proquest.com/products/pt-product-ABI.shtml

[5] The BT digital library has approximately 9500 registered users.

3. A knowledge sharing application that enables users to annotate and share digital library documents, WWW or Intranet pages with information from the digital library topic ontology, making it easier to recall those pages at a later date.
4. An expertise location function that enables people to identify experts based on other people's interaction with the digital library and their level of readership in a particular topic area.
5. A personalised semantic search agent that will collect relevant content from the digital library and WWW on behalf of a user. The user will configure the search agent's search profile.
6. A personal information-based content delivery agent that identifies and delivers relevant content based upon items in a user's calendar, tasks list, etc.

Many of the user applications will be supported by a user profile that is derived from a user's interaction with the digital library, and the content of visited WWW or Intranet pages. A common notification component will enable the user to schedule the delivery of notifications on a per application basis[6].

2.2.2 Domain expert and system administrator tools and applications

1. A focused crawler that extends the content of the digital library (these crawlers can be associated with a public information space).
2. A configuration tool that enables a public information space to be tailored to the needs of a particular topic or a domain of interest (a domain of interest can cover a number of topics).
3. A tool that enables the digital library topic ontology to be populated with information from the ABI and Inspec[7] databases.
4. An ontology extension and merging tool those enables new topics to be identified and added to the digital library ontology.
5. A tool that enables additions to the domain ontology to be merged with new topics specified in new releases of the ABI and Inspec[8] thesauruses.

All applications developed from these user scenarios will be supported by technology researched and developed in the technical work packages of the SEKT project. Access to applications will be given through a digital library portal. Users will access the

[6] Notifications can be scheduled separately for each application.

[7] Produced by the Institution of Electrical Engineers (IEE), Inspec is one of the world's leading bibliographic databases, providing comprehensive coverage of scientific and technical literature in the fields of physics, electrical engineering, electronics and computer science.

[8] Each Inspec record is indexed using controlled terms chosen from the Inspec Thesaurus. Furthermore, the Thesaurus provides cross-reference terms, the relationships between terms (e.g. broader terms, narrower terms), classification codes, the dates on which terms were added, and the terms in use before these dates. Recent thesaurus updates have been released in 1999, 2003 (CD only), and 2004. As each new edition of the Inspec thesaurus is released, information regarding new terms and relationships between existing terms will be added to the databases. Subsequent Inspec records will be classified against these new terms. Updates to the digital library topic ontology are likely to occur in the period between thesaurus updates; the topic ontology is likely to evolve significantly. The new thesaurus may or may not contain the topics/terms that have been added to the ontology. As a consequence there is a need to merge the evolving digital library topic ontology with new topics/terms in the latest thesaurus.

portal via a range of devices, e.g. PC, laptop, pocket PC, etc. Content will be purposed to suit the capabilities of the device, access network bandwidth restrictions, and any preferences set by the user.

All applications will make use of a common domain-specific ontology based on the PROTON ontology [1]. The initial BT digital library topic ontology will be derived from the thesauri of the Inspec and ABI/INFORM databases. The topic ontology will evolve as new versions of the ABI and Inspec thesauruses are released, and as new topics are added to the ontology by the domain experts.

3 End user use cases

3.1 Requirements

The requirements for the end user use cases and scenarios have been derived from:

1. The overall aims and objectives of the SEKT project [2].
2. A requirements capture activity for the BT digital library, comprising an analysis of the responses given to a questionnaire on the usage of the digital library, and a focus group study [3]. The questions and detailed results of the questionnaire are provided as an Annex to [3].
3. The state of the art report on the use of semantic and collaboration tools in digital libraries [4].

Other requirements have been derived from known limitations of the digital library's search application and information spaces.

3.2 End user scenarios and use cases

Scenarios and use cases relating to the following end user tools and applications are described:

1. End user use case 1: Search and browse (p. 9);
2. End user use case 2: Information spaces (p. 26);
3. End user use case 3: Knowledge sharing (p. 35);
4. End user use case 4: Expertise location (p. 45);
5. End user use case 5: Personal search agent (p. 49);
6. End user use case 6: Personal information-based content delivery (p. 53);
7. End user use case 7: Profile management (p. 58); and
8. End user use case 8: Notification (p. 61).

3.3 End user use case 1: Search and browse

3.3.1 Description

The generally accepted view is that search engines that rely on keyword and phrase matching techniques between the query and index alone tend to offer high recall and low precision[9]. As a consequence, the users of these search engines are often faced with the prospect of sifting through a large number of results, many of which are often irrelevant to their search goals.

The current search interface to the digital library is based upon a conventional Information Retrieval (IR) search engine. Searches initiated on the digital library tend to offer high recall and low precision, particularly for searches entered as simple one

[9] Recall is defined as the ratio of the number of relevant documents retrieved to the total number of relevant documents in the whole collection. Precision is defined as the number of relevant documents retrieved to the total number of documents retrieved.

to two term queries[10]. Although advanced search options are available, they require the user to add non-intuitive modifiers to their query, e.g. to search for articles authored by Davies, the user must enter the query: `au=davies` into the search engine's query interface. Furthermore, the digital library's search engine treats a user's search in isolation, i.e. the results for a given query are identical, being independent of the user submitting the query and the context in which the user made the request.

An additional issue is concerned with the low number of results that a user is likely to view; the majority of digital library users tend to only view the first couple of pages of results returned from the search engine. As a consequence many useful results could be overlooked.

The State-of-the-Art report on Knowledge Access [5] recommends an approach of providing a semantically enabled search engine, based upon general and domain specific ontologies, that employs automated systems to facilitate the annotation, indexing and retrieval of documents with respect to *named-entities*, e.g. people, companies, organisations, locations, etc. The report also recommends an approach of personalising search based-upon a user's context. Furthermore, an approach to allow structured querying in a digital library scenario [6] finds that better results could be achieved with structured queries vs. unstructured queries.

The search and browse application proposed for the BT digital library will establish the users search context based on their interaction with the digital library and other information sources, and from the composition and preferences set in their user profile. The topics and named entity associations, i.e. named entities that are defined in the user profile[11], can be used to enrich a search query which, when used in conjunction with the digital library topic ontology, will help identify semantically close topic areas, i.e. topics in the digital library ontology that are most relevant to the user's current search context (see section 3.3.2.3 for an example). Furthermore, the search and browse application will identify and highlight named entities within the search results, and offer supplementary information[12] about those entities (e.g. for an entity of type *company*, the supplementary information could include the location of its headquarters, its web address, sector, chairman, and CEO). This information will enable the user to quickly identify the context of the result, and filter further results based on the entities of interest.

An attribute-based search interface that enables users to base their search on specific attributes of selected resources is also proposed. This search interface will, for example, enable users to search for publications authored by a particular person, for publications of a specified publication type (a book, an article, a journal), or a publication aimed at a specific level of readership in a particular topic area. The

[10] An analysis of the queries submitted to BT's Intranet search engine over a 4-month period between January 2004 and May 2004 showed that 99% of the submitted queries only contained a single phrase and that, on average, each phrase contained 1.8 keywords. An analysis of the queries submitted the BT digital library search engine over the same period also revealed a similar usage of short length queries (on average, 1.5 keywords per query phrase), although a larger proportion of the queries contained more than one phrase per query (approximately 24%).

[11] The topics defined in the user profile are derived from the users interaction with the content of the digital library and other information sources. Key features extracted from visited WWW and Intranet pages are mapped to topics in the BT digital library topic ontology.

[12] The supplementary information is contained in a knowledge base.

attribute-based search is expected to help the user get to the information they seek more quickly. Furthermore, implicit statements can be inferred from the inherent transitivity of some properties in the digital library ontology, e.g. if a resource is of type *Article*, where *Article* is a subclass of *PublishedMaterial*, then that resource is also of type *Document* because *PublishedMaterial* is a subclass *Document*. A user instigating a search for resources of type *Document* will also be searching its subclasses, e.g. resources of type *Article* and type *Book*. The attribute-based search therefore has the capability to offer a very fine grained search, e.g. when the user searches against a specific resource type such as an entity of type *Book*, or alternatively, it can offer a wider scope search.

The BT digital library topic ontology will be extensive. Users could experience difficulty browsing the ontology using a simple link-based interface. Different users have different information requirements, and very few users will need to access all topic areas of the library. Furthermore, much of the content, and therefore the parts of the ontology that give access to that content, will not be of interest to users. It is anticipated that the majority of users will only search and browse for information in those segments of the digital library that are relevant to their work and their interests, i.e. the nature of their work is likely to restrict their browsing activity. Users will be able to browse a more limited view of the ontology based on their particular interests (as defined in their profile).

The following application scenarios describe how a user will use various functions to search and browse for information resources in an ontology-based digital library.

Scenario 1: a general search for a specific resource (section 3.3.2.1).

Scenario 2: searching for unspecific resources on a specific topic area (section 3.3.2.2).

Scenario 3: named entity search (section 3.3.2.3).

Scenario 4: attribute-based search (section 3.3.2.4).

3.3.2 Application scenarios

3.3.2.1 Scenario 1: a general search for a specific resource

Intention of scenario/setting: This scenario describes a search where a user can remember minimal information about the content of an article, but cannot identify the article using attributes such as the author's name or the title of the article. The worker believes that the paper was published in the last six months, and expects details of the article, but not necessarily its full-text, to exist in the digital library.

User expectations: The user expects to find the article reasonably quickly, but appreciates that initial searches using simple, short-length, keyword-based queries can produce a large number of results. The user does, however, expect to be able to narrow down the results quickly based upon their actions and the interaction of their user profile with the digital library topic ontology.

Pre-conditions: The user has opened a web browser and navigated to the digital library home page (see section 3.3.9.1 for an example screenshot). The user has a specific target resource in mind. Selected topics of interest are active in the user's

profile[13]. Named entity recognition functions and named entity associations are disabled.

1. The user instigates a search using the simple search interface, basing their search on keywords that they believe to best describe the content of the article. The query is modified with topic information from the user's profile.
2. The search engine matches the query against the digital library index, and returns a set of results matching the query.
3. A summary of each result is displayed in the web browser, e.g. the title of the article, its author, an extract from its abstract, and publication date. Furthermore, the user is presented with a hierarchy of semantically similar topics from the digital library topic ontology.
4. The user selects one or more topics from the topic ontology. This action refines their search to include only those papers that have been classified against the selected topics. The user also refines their search to only include articles that have been published within the last 6 months. The modified query is submitted to the search engine.
5. The search engine matches the query against the index, and returns a new set of results (ideally, this would include the result for the target document).
6. The user views the full text of the article, and decides to print it.

Post-conditions: The system collects information about the user's interaction with the digital library, e.g. when the user selects to view the description of an article. Depending on previous search behaviour and content viewed, the system may suggest modifications to the user's profile (the user can accept or reject the system's recommendations). Furthermore, in this example scenario, the user chose to print the document. Consistently similar behaviour, e.g. viewing and printing similar articles, could signify an interest in the associated topic areas.

Alternative scenarios: Rather than modifying the query automatically, the system could suggest possible topics for query expansion to the user. The user could then choose to expand their query using one or more of the system recommended topics.

The user could also use the search history function to locate a document that has been searched/viewed previously.

3.3.2.2 Scenario 2: searching for unspecific resources on a specific topic area

Intention of scenario/setting: This scenario describes how a user tries to find information related to a topic area known to exist in the digital library. In contrast to scenario 1, the user does not have a specific target resource in mind. The user is likely to modify their search with topics selected from the topic ontology. The termination criterion is reached when the user is satisfied that sufficient information, relevant to their area of interest, has been found. In this example scenario, a BT worker would like to find information about networks for Automated Teller Machines (ATMs). Furthermore, they would like to find information on companies that manufacture

[13] The user profile will also contain details of the current topic of interest, i.e. the topic that is most relevant to the user's recent browsing activity (e.g. on the WWW), or interaction with the digital library. The search engine will place more significance on the user's current topic of interest (if active in the profile). Also see section 3.3.5 for issues regarding the user's search context.

ATMs, and be notified whenever new information about those companies is added to the digital library. The user is not familiar with the topic area.

User expectations: The user expects to find a number of relevant resources in the digital library, and is willing to spend more time searching and browsing for information. Their search sessions are likely to be spread over a period of a few days. The user expects the application to provide supplementary information about any named entities that are identified in the results summaries, e.g. companies.

Alternative scenarios: A search for information on a specific company is described in more detail in the named entity search scenario (section 3.3.2.3).

Pre-conditions: The user has opened a web browser and navigated to the digital library home page. Although some topics of interest are active, there are no topics of interest related to ATM networks (either active or inactive) as this is a new area of interest for the user. For the initial part of the search, named entity recognition functions and named entity associations are turned off.

1. The user enters the keywords ‘ATM’ and ‘networks’ in the textbox of the simple search interface, and clicks the search button to initiate the search.
2. As the user’s profile does not contain information relating to ATM networks, the query is submitted to the search engine unmodified.
3. The search engine matches the query against the index and returns a set of results matching the query (see section 3.3.9.2 for an example screenshot).
4. A summary of each result is displayed in the web browser. Furthermore, occurrences of named entities are highlighted in the summaries.
5. The user scans the summaries of the results, clicks on some links to view the full-text, and stores references to some of the articles (see section 3.3.9.3 for an example screenshot).
6. The user modifies their search by clicking on one of the related topics that are displayed in the browser. In this example, the user clicks on the topic ‘Point of sale systems’.
7. The search engine matches the modified query against the index and returns a different set of results.
8. The user continues to search in this way, selecting related topics, and storing relevant articles for future reference.
9. The user repeats this process until sufficient articles about ATMs are found, and then stops their search.
10. The user resumes their search the following day, viewing details of their previous search as a reminder, e.g. previously used topics and queries. After finding further relevant results, the user decides to query for information about companies working with ATMs.
11. The user enables company-based named entity searching, and enters the name of a company in the simple search text box. In this example scenario the company name ‘NCR’ is entered.
12. The search engine matches the text ‘NCR’ against the index and returns results for articles mentioning NCR. The text ‘NCR’ is highlighted in the results summaries.

The names of other companies occurring in the summaries are also highlighted as the user has not specified any further named entity filtering options.

13. The user clicks the highlighted text ‘NCR’ in one of the summaries. Supplementary information about that company, in the form of a company summary, is retrieved and displayed. The user views the supplementary information, and indicates to the system that this is the company in which they have an interest.
14. The user accesses the full text of articles that look interesting.
15. The user creates a private information space for the query (a scenario for creating a private information space is described in section 3.4.2.4).
16. The user completes similar searches on other companies, e.g. those companies also mentioned in the results summaries (or the full-text of the articles).

Post-conditions: The user’s profile gets modified as they interact with the digital library, e.g. topics of interest and named entity associations are updated in response to system suggestions. After the user has been using the digital library for some time, the system will have learned some of their preferences. Generally, it will have learned their topics of interest, regularly viewed publications, and possibly their level of readership in a particular topic area (section 3.6.2.1).

3.3.2.3 Scenario 3: named entity search

Intention of scenario/setting: This scenario describes in more detail how named entity recognition can be used in the digital library to help disambiguate a user's search. In this scenario the user searches for information in relation to a particular company.

User expectations: The user expects the search engine to provide additional functions and information that will help to disambiguate their query, thereby allowing them to get to the information they seek much quicker.

Alternative scenarios: The user makes use of the attribute-based search interface to the digital library (described in section 3.3.2.4).

Pre-conditions: The user has opened a web browser and navigated to the digital library home page. Although the user’s profile is enabled, they have indicated to the application that the search should not use previously established topics (topics of interest that are represented in the user’s profile), current interests or previous named entity associations. Named entity recognition functions are turned on. In this example, the user wants to search for information on the (fictitious) telecommunications company ‘Whitecross’.

1. The user initiates a basic search using the keyword 'Whitecross'.
2. The search engine matches the keyword ‘Whitecross’ against the index and returns a set of results (such a search is likely to produce a large number of results, as many articles contain the text 'Whitecross').
3. A summary of each result is displayed in the web browser. Occurrences of the term 'Whitecross' are highlighted in the summaries.
4. A number of the results do not relate to the telecommunications company Whitecross, but to other entities such as products, places, people's names, etc.

References to companies of the same name operating in different industries, e.g. the UK Dental chain, are also highlighted in the results.

5. The user clicks on a highlighted occurrence of the text string 'Whitecross'. The application returns and displays supplementary information about that entity in the form of a summary. This enables users to find results that reference the particular instance of 'Whitecross' in which they have an interest (a named entity of type 'company' would have associated properties such as HQ location, web address, sector, chairman, and CEO). Alternatively the system could indicate that results were available for 'Whitecross (Telecommunications)' and 'Whitecross (Dental), and prompt the user to refine their search according to their particular interest.
6. The user refines their search to reduce the set of results to only those that mention the particular instance of the entity 'Whitecross' in which they are interested.
7. The modified query is passed to the search engine.
8. The search engine matches the modified query against the index, and returns a set of results that reference the selected 'Whitecross', i.e. the company 'Whitecross'. Note: had the user anticipated that they would have an issue with multiple uses of the name 'Whitecross', they could have adopted an alternative search strategy (please refer to setting 2).
9. The user selects to view the full text of one of the results. All occurrences of the term 'Whitecross' are highlighted.

Post-conditions: After the user has been using the digital library for some time, the system will have learned some of their preferences. For this example, the system would have learned that the user is specifically interested in companies in a particular sector, e.g. telecommunications. As a consequence, when the user searches for any particular company name, the system will bias the results to those that contain information about a company in the telecommunications sector. The system will also inform the user if there are companies of the same name in another sector, and give them the opportunity to view information about those companies. More specifically, when searching for the text string 'Whitecross' in future, the system will use the association of 'Whitecross' with a company in the telecoms sector and will invoke the search accordingly.

Setting 2: Instead of simply starting with a text search, the user will be able to specify the type of entity for which they wish to search (in this example scenario, entities of type company[14]). This action will result in the search engine returning an initial set of results that contain references to companies called 'Whitecross', and not to other entities, such as products, places, or people, having the name 'Whitecross'. The list of results would still include documents containing references to companies with the same name as the one in which the user is interested, although these results, too, could be filtered-out through careful selection of the search parameters associated with that type of entity, e.g. the industry sector. By specifying an industry sector, the user could limit their search to documents containing references to a company called 'Whitecross' operating in the telecommunications sector.

[14] Depending on the granularity of the topic hierarchy, the user could also have searched for 'Whitecross' as an instance of a 'Telecommunications company'.

Setting 3: There are times when the user's search preferences or named entity associations, e.g. the 'Whitecross' company in the telecommunications sector, could inconvenience a user. After months of being interested in ‘Whitecross’ the telecoms company, the user might become interested in ‘Whitecross’ the advertising agency, or 'Whitecross' the author. For the user who feels constrained by their learned profile, the system provides the function to disable selected parts of the profile, and then enable them again as required, e.g. following the initial part of a new search. The user has manual control over their profile, enabling them to remove topics of interest and named entity associations for example. Furthermore, some users will habitually search for the same type of entity, e.g. somebody working in competitor intelligence may habitually search for information about companies. The system learns this from the user’s behaviour and, in the first instance, interprets all search strings as companies. For a particular search, the system will inform the user that there are other interpretations of the string (e.g. names, places) so that these entities can be investigated if required. Again, this facility can be turned off by the user by disabling that particular feature in their user profile.

3.3.2.4 Scenario 4: attribute-based search

Intention of scenario/setting: This scenario describes an attribute-based search of the digital library. The attributes presented to the user are dependent on the type of resource being searched for. More attributes will be presented to the user as the resource type becomes more specialised, e.g. a more fine-grained attribute search will presented to a user searching for a journal article, when compared to a publication, which is a more general resource that covers books, magazines, journals, etc. The application also presents information on the user’s previous attribute-based searches to save them from having to select the resource type from the ontology each time, e.g. previous type of resource searched for (some users may tend to search for just a few types of resource).

Expectations: The user expects the interface to present only those attributes that are associated with the type of resource being searched for. The user expects to be able to recall details of resource types that have been searched for in the past.

Pre-conditions: The user has opened a web browser, has navigated to the digital library home page, and from there has selected the attribute-based search function. Information about previous attribute-based searches is displayed (through the user profile). An ontology of resource types is also displayed (resource types that have been searched for previously are highlighted).

1. The user navigates the ontology and selects the type of resource for the attribute-based search (a brief description of the resource is displayed when the user positions the pointer over the resource type). In this example the user is searching for a journal article (an example screenshot is shown in section 3.3.9.4).
2. A search interface is presented to the user with search attributes specified by the resource type representing a journal article (had the user wanted a more general search, e.g. a search for any type of publication, the application would present attributes that are common to all publication types).
3. The user enters appropriate text against some of the attributes, e.g. surname of author, and selects other criteria from drop down lists, e.g. a published since date. The user instigates the search.

4. The search engine matches the query against the index, and returns a set of results.
5. The user is given the options to browse other content based upon the attributes returned with each result, e.g. the user could select other related articles published in the same journal, or select other publications authored by the same author.

Post-conditions: The user's profile is modified as they interact with the digital library. In this scenario the profile would also be updated with details of the type of resource that the user searches for (over a period of time the system will learn a user's search preferences).

Setting 2: The BT digital library ontology will be extensive. Users could experience difficulty browsing the ontology using a simple link-based interface. Furthermore, different users have different information requirements, and very few will need to access all topic areas of the library. A focused view of the ontology, based upon their interaction with the BT digital library and the content of WWW pages that they have accessed recently, can be presented to the user. The segments of the ontology presented to the user include the topics that are currently active in the user's profile, and those topics that are semantically close to those topics. The challenge for the application is to reveal as much of the ontology as necessary to be for the benefit to the user, but to reduce the complexity as much as possible by hiding any unimportant parts (unimportant to a particular user).

As well as the search and link-based navigation of the digital library's topic ontology, the user will also be able to search and browse the ontology through a graphical interface. The user interface consists of two parts; a hierarchical view of the topics and sub-topics on the left-hand side of the browser, and a graphical representation of the data in the form of a cluster map in the main part of the browser (information spaces and topics are represented by clusters of dots, where each dot represents a resource within a topic). It is anticipated that this type of interface should encourage users to browse the ontology. Users will navigate the ontology by clicking on nodes on the cluster map. As the user browses, the map will reflect their current position in the ontology and their topics of interest. Users will have the added benefit of being able to see related topics displayed as adjacent nodes on the map, thereby enabling them to visualise relationships between topics that they might not have previously considered.

3.3.3 Other scenarios/functions to consider

The following functions should also be considered when designing the search and browse application for the BT digital library:

1. Enable users to gather a collection of documents in a work space, cluster those documents (e.g. based on their content), and search for related items based on those clusters. A sub-topic ontology could be generated for each of those clusters, enabling the user to instigate a very specific search[15].
2. The full text of an article will not always be available to users (BT subscribes to approximately 1000 electronic publications). In some cases pre-prints of articles are available on the WWW. The user should be able to invoke a function that searches for pre-prints of articles on the WWW, based upon known information

[15] Further research is required to establish whether the geneation of the sub-topic ontologies would be a semi-automatic or an automatic procedure.

about that article, e.g. its title and the authors. Such a tool should present the user with a list of candidate articles. Users should then be given the opportunity to select, and add a reference to that pre-print in the digital library. A link to the pre-print would be provided (a suitable icon would denote that the full-text was taken from the WWW, and was not the actual article).

3. Link and display the citations of articles in the digital library, and provide links to articles which reference that particular document.
4. A 'find similar' function based on documents returned in a search (using both the content and the classification in the ontology).
5. Provide an export function that enables the metadata from selected results to be exported.
6. Learn a specialised domain ontology for each information space (see section 3.4). Such an ontology would be more fine-grained than the global topic hierarchy (or a subset of it); the discovery of new topics could be integrated into the global topic hierarchy.
7. Search history functions (enables a user to search on previously searched/viewed resources).

3.3.4 Expected benefits

A search for information must be fast, particularly when the user has a clear idea of what they are looking for. The results obtained through an attribute-based search of a repository of semantic annotations are expected to be more relevant than results returned from a more conventional search engine, such as the one currently used in the digital library. In its most simple form, the attribute-based search will enable the user to search for explicit statements about recognised entities and relations, e.g. search for a person holding a particular position within an organisation, or search for an organisation located in a specific location. The attribute-based search is expected to help the user get to the information they seek much quicker. Furthermore, implicit statements can be inferred from the inherent transitivity of some properties in the digital library ontology (see for section 3.3.1 for an example).

Despite the advantages to be gained by using the attribute-based search, some inexperienced users may be overwhelmed by the relative complexity of the interface when compared to a more conventional keyword-based search engine like that used in the digital library. The simple search interface proposed for the home page, a textbox into which the user types some keywords, will hide the topic ontology from the user initially, and will avoid the need to list a large number of entry points to the ontology. As users get more familiar with the ontology, they can reduce the scope of their search by navigating to a more specific topic area before beginning their search. Furthermore, the system will learn a user's preferences and specific topics of interest; simple searches will be augmented with topics selected from the ontology, enabling the search engine to give more importance to resources classified into those topic areas. A user should find that searches for resources are more aligned to their information needs.

The use of named entity recognition, when coupled with the provision of supplementary information about the recognised entities, will enable the user to quickly establish the context of a search result where previously the potential for ambiguity would have existed, e.g. when searching for articles authored by somebody

with a common surname, supplementary information such as the author's affiliation and a list of their published articles will help the user find the particular author they are interested in. The user could then apply a filter to restrict the results to articles authored by that person, or widen the scope of their search to find any *PublishedMaterial* connected with that author.

Although it could be argued that the advanced search functions that are currently available in the digital library provide a similar function to the attribute-based search, e.g. include the string `au=kings` in the search interface to restrict the results to where *Kings* is the author, its use is not particularly intuitive for the inexperienced user. Furthermore the current search engine does not provide a simple browse facility. The search and browse application proposed for the digital library offers the user a simple means to navigate the ontology based upon the results returned from a simple search.

Users are expected to browse the digital library via the topic ontology. An analysis of the usage logs will enable highly used topics to be identified and presented to the user (topics that are not used could be hidden form the user).

3.3.5 Potential issues

The main function of the user profile is to capture the concepts (topics of interest) defined by the digital library ontology that best reflect the interests of the searcher. The system will learn user preferences as people interact with the digital library and other information sources. Topics of interest are used to augment the user's query. Some topics will, however, be of more significance to the user than others. This needs to be reflected in the user's profile. In some cases the users most recent search behaviour will be of more relevance. In other cases the user's past behaviour could be of more significance. Some topics may be very significant to a user's search, whilst others are less significant (but not irrelevant). The system therefore needs to be able to establish the user's current search context very quickly in order to prevent a user's queries from being augmented with inappropriate topic information (which is likely to result in poor search precision, which may in turn stop users from using their profile to their best advantage; they may then choose to deactivate all topics in their profile). Even if the search application can detect changes in the user's search context, the continual prompting of the user to alter their profile may simply become too annoying.

From a usability perspective, users that are familiar with the current digital library search engine (and other similar keyword-based search engines) may find some of the more advanced search functions too complex, and may continue to use the simple search interface. Suitable user training will need to be put in place to smooth the transition from the old search to the new, and to make sure that users get maximum benefit from the search and browse application.

Suitable topic-to-topic and keyword-to-topic algorithms need to be investigated further. Further research is required to find the best way to represent a user's interests in a profile, especially in the area of when more or less emphasis needs to be placed against certain topics of interest. The need to model a user's changing interests over time and detect user context switches so as not to augment a query with irrelevant terms must not be underestimated. Privacy issues associated with monitoring user activity may also be a concern for some users [3]. Further research is required to

determine the best way to structure the topic ontology to enable natural language text descriptions to be generated.

If a specialised domain-ontology is extracted from the domain of an information space (see 3.4), it will need to be updated each time there is an update to the ABI or Inspec databases.

The BT digital library ontology, comprising: a) the domain-specific classes and properties that model the digital library's topics and resources, b) application-specific classes and properties that model the data required by the applications, and c) the instance data. The initial set of topics, and the relationships between those topics, will be derived from the ABI and Inspec thesauri. The initial set of resources will be derived from the ABI and Inspec databases. The topic ontology will need to be designed to facilitate the automatic generation of descriptions for each topic. Furthermore, the topic ontology will evolve as new topics are added to the ABI and Inspec thesauruses, and as new topics are discovered, e.g. through the focused crawler (section 4.3). It is anticipated that ontology evolution mechanisms will need to be applied in order to assure the consistency of the topic hierarchy when new topics are added to the thesauruses.

The results acquired through an attribute-based search of a repository of semantic annotations are expected to be more relevant than results returned from a more conventional search engine. Further work is required to design a suitable set of tests that will enable the results of the two types of search to be compared. Likewise, suitable tests will need to be designed that will enable the use of the graphical interface to be compared with the more traditional link-based navigation.

3.3.6 Supporting technology and functions

The following technologies are required to support the search engine.

1. A user profile-enhanced search and browse application/search engine that enables users to search for digital library resources (e.g. articles, papers, books).
2. Ontology-based query expansion based upon topics of interest active in the user profile, and user interaction with the topic ontology.
3. An attribute-based search interface (the attributes will be based upon the selected resource type).
4. Components to monitor and capture the user's interaction with the digital library, determine their interests from content accessed during WWW browsing activity, and map those to topics in the digital library ontology. Furthermore, the user's interests will change over time, and they are likely to switch context frequently; this activity must be captured and represented in the user's profile.
5. Functions to manage a user profile (elements of the user profile will be common to a number of tools, e.g. search and browse, and knowledge sharing) that gives the user final control over its active content.
6. Named entity recognition, e.g. identify companies, people, places, etc.
7. Algorithms to calculate the semantic similarity between topics (e.g. topics in the user profile and topics in the digital library ontology), and between keyword-based queries and topics in the ontology.

8. Alternative visualisations of the search interface and results, e.g. present the user with interfaces appropriate to the device being used and any specified user preferences.
9. Natural Language Generation software to generate descriptions of the topics in the topic ontology.

3.3.7 Supporting data

The search and browse tool will be supported by the following data objects:

1. The BT digital library ontology, comprising: a) the domain-specific classes and properties that model the digital library's topics and resources, b) application-specific classes and properties that model the data required by the applications, and c) the instance data.
2. User profiles for holding users' interests (topics from the digital library topic ontology), user preferences (e.g. an indication on whether a topic is active or inactive in the profile), named-entity associations, search and browse history. The profile will also contain data of a more static nature, e.g. contact details.
3. A knowledge base[16] containing supplementary data about named entities, e.g. people, companies, and places.
4. A visualisation ontology that will transform the machine oriented data of the digital library into a form suitable for presentation to the user.

[16] The knowledge base of named-entities and supplementary data (e.g. company information) will need to be updated and maintained (this task is likely to be undertaken by the domain experts and the system administrator). Use cases and scenarios will needs to be developed for this function.

3.3.8 Relevant SEKT tasks

The following SEKT tasks will support the search and browse application:

WP/Task number	System/Component/Function
T1.1, T1.2	Algorithms/functions to calculate the semantic similarity between user profiles and resources classified by the digital library topic ontology.
T2.6, T6.3, T6.4.	An API to the underlying semantic annotation platform and knowledge base or an API to the SEKT integrated platform. Named entity recognition functions will be integrated into the annotation platform.
T1.8, T11.2	Proton ontology and BT digital library domain-specific ontology.
T3.1	Components for managing the evolution of the ontology.
T3.2	Components to discover popular (and unpopular) topic areas, and suggest changes to the ontology to reflect user behaviour (data driven change discovery).
T3.3	Components to extract a fine-grain ontology from the domain of an information space.
T5.2, T11.3	Semantically enabled search engine combining free-text search with the capability to exploit semantic metadata, e.g. ontology-based query expansion, named-entity recognition. Attribute-based search interface.
T5.4, T11.3	Visualisation component that takes the BT digital library ontology and transforms it into a form that is suitable for presentation to the user.
T5.5, T11.3	Components (or functions) to derive, update, store and manage a user profile, i.e. model a user's interests, roles and context, based on their interaction with the digital library and content viewed whilst browsing the WWW. The profile will be common to all end-user applications.
T5.6, T11.3	Automatic generation of descriptions of topics in the BT digital library ontology. These descriptions could be edited manually by a domain expert if required.
T5.7, T11.3	A content purposing component to transform user interfaces and results into a format suitable for display on alternative devices, e.g. PDAs.

Table 3.1. SEKT tasks that support the search and browse application.

3.3.9 Example screenshots

This section of the document provides some example screenshots of the search and browse application.

3.3.9.1 Digital library home page

It is assumed that many users will usually start with a simple keyword-based search. The simple search interface on the BT digital library home page is shown in figure 3.1.

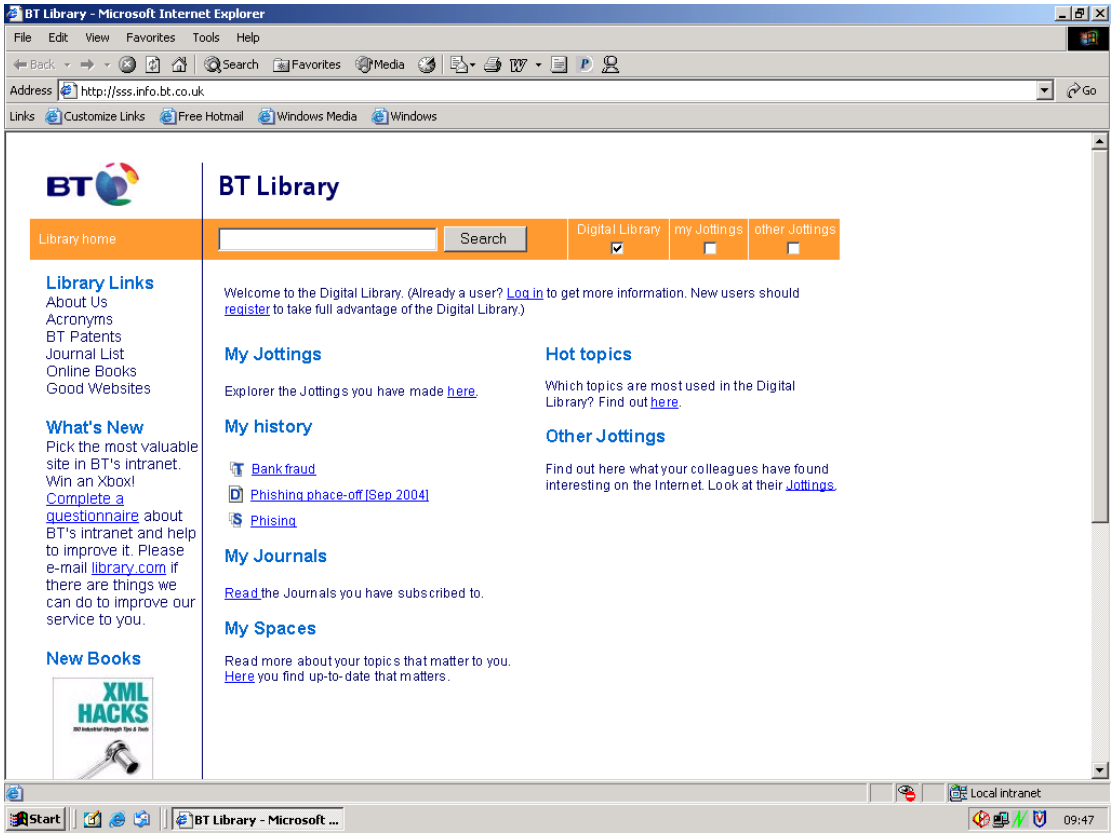

Figure 3.1. BT digital library home page.

This simple interface does not confront the user with abstract topics which may not, at first, be meaningful. The additional search options shown include a web search and a news search.

The user can browse the ontology from the ‘Hot topics’ link, which when clicked displays a page containing topics that have been used heavily recently. The category ‘My Topics’ (not in screenshot), which consists of the topics that best match the users interests, can also be used as an additional starting point for browsing the topic hierarchy.

3.3.9.2 Search results page

A typical results page returned from the search engine is shown in figure 3.2.

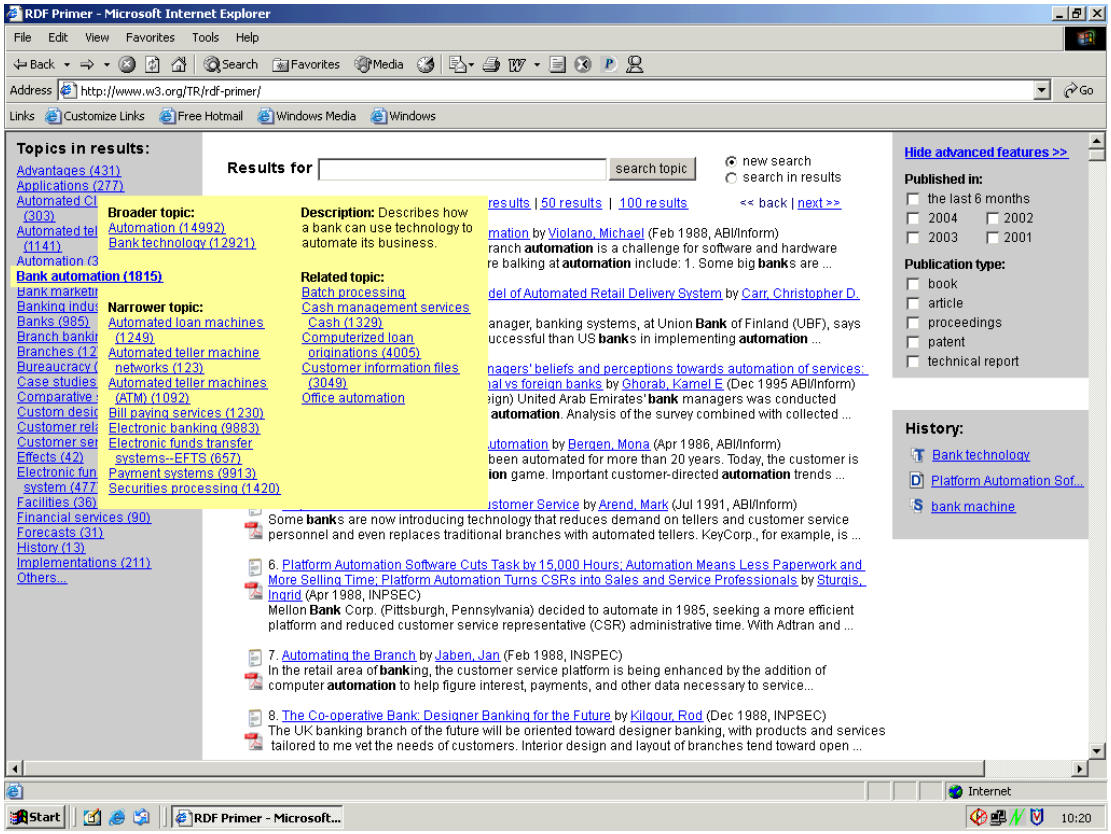

Figure 3.2. Search and browse results.

The left hand column lists the topics that annotate the documents in the result set (these topics can also be ranked by their relevance). The highlighted yellow box appears when a user positions the mouse pointer over a topic. A user can further navigate the ontology or choose to make a new query using related topics, or topics of a broader or narrower scope. A description of the topic is given to the user when a user places the mouse pointer over the topic description icon (not shown in diagram).

Typically, search results are displayed with the title, author, publication date, source (data base), and an extract from the abstract. Where appropriate, an indication is given that the full-text of the document is available, e.g. a PDF document or a text file. Keywords, topics used in the search, and any identified named entities are displayed in bold font.

Features that enable the user to refine the search are shown on the right-hand side of the screenshot, e.g. select a publication date, or a publication type (a field for search by author is missing from the screenshot).

The search history displays all keywords, topics and documents that were used or visited. The history is ordered so that the latest search is presented at the top.

3.3.9.3 Summary information for a single document/article

A view of the summary information that is presented to the user is shown in figure 3.3.

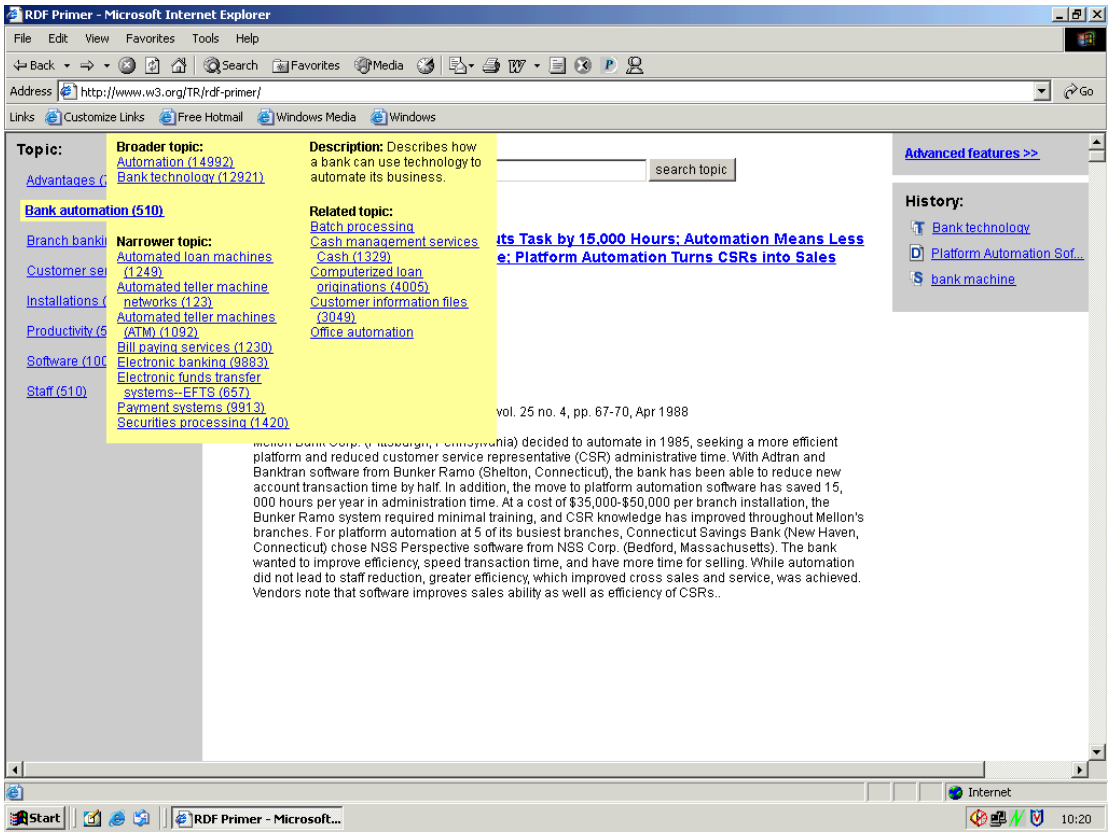

Figure 3.3 Summary of information for an article.

The left hand side presents the topics, which annotate the document. All bibliographic data is shown, and in addition an abstract.

3.3.9.4 Attribute-based search interface

An example of the attribute-based search interface is shown in figure 3.4.

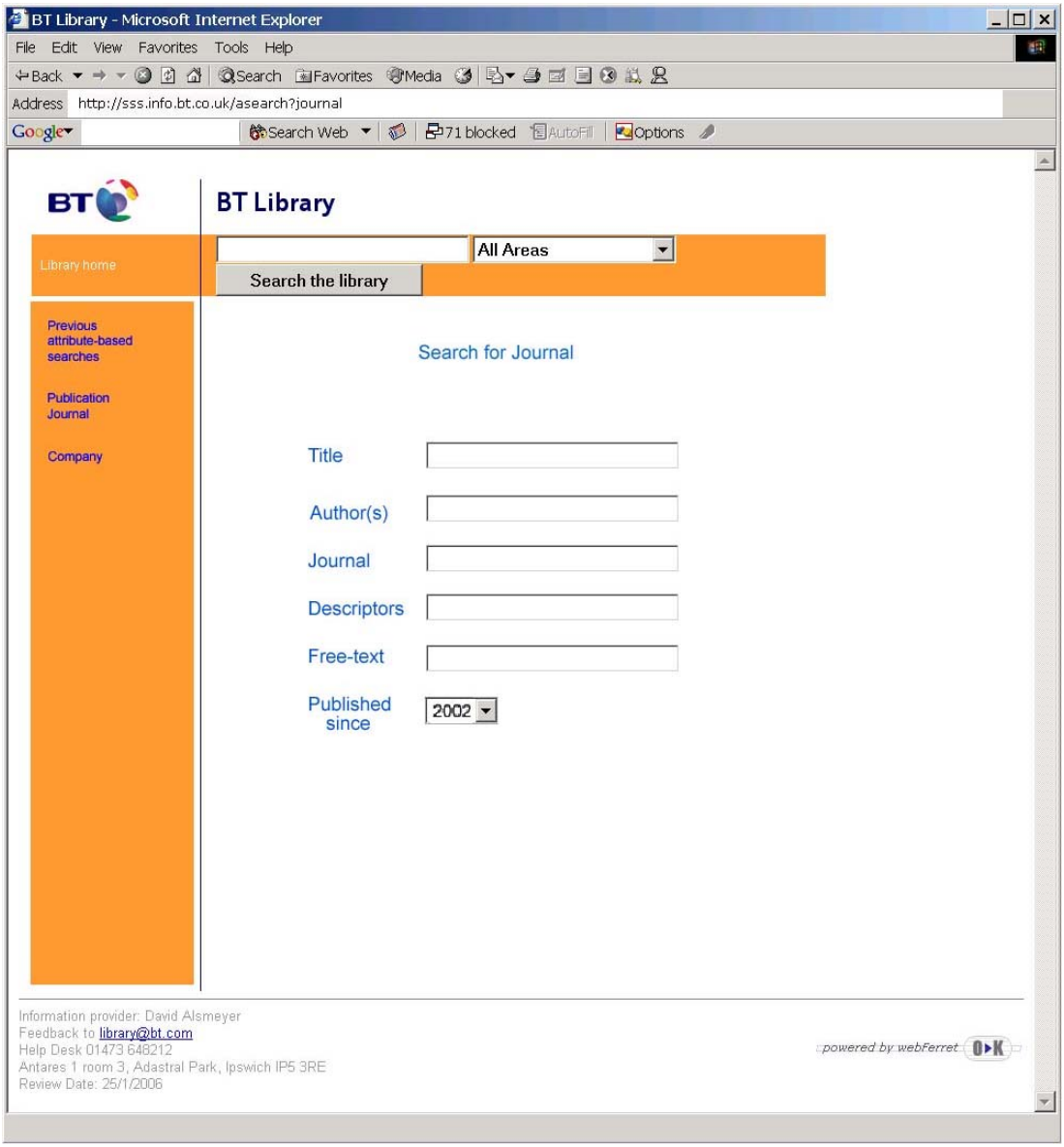

Figure 3.4. Attribute-based search interface.

3.4 End user use case 2: Information spaces

3.4.1 Description

Existing public information spaces tend to be very broad in scope, and often retrieve information on a wide range of subjects related to an area of interest. A good example is the 'knowledge management' information space, which brings together a very broad spectrum of topics connected with knowledge management. Although this has the advantage of grouping all information relating to a general topic in one place, the topics covered are considered too broad for some users; a more fine-grained information space is required. In order to retrieve information on a wide-range of related topics, the public information spaces are usually configured with very simple queries and, just like the digital library search engine, can suffer from high recall (some may argue that this is acceptable, and it's up to the user to filter the results according to their needs). As a consequence users are usually presented with too many results that may not be directly relevant to the branch of the topic area in which they

are interested. Although users can get around this problem by configuring a private information space with a more complex query, the benefits of using a public information space to serve the needs of a small group of users, i.e. the beginnings of a community of interest, is lost.

This use case describes a more fine-grained semantic information space that uses queries based on topics selected from the BT digital library ontology (the domain expert can also add free-text keywords to the information space query – refer to section 4.4). Users are given the functions to subscribe to selected sub-topics within the domain of a public information space, e.g. for ‘knowledge management’ this could include the sub-topics ‘communities of practice’, ‘organisational learning’, ‘tacit knowledge’, and 'intellectual capital'. Moreover, the digital library will suggest subscriptions to relevant public information spaces based on the user's search behaviour, documents they have read, information they have shared, and their given responses to digital library notifications. Descriptions of the information spaces and topics from the digital library ontology will be generated automatically, giving users a better idea of the scope of a space. The domain expert will be able to edit these descriptions.

The notification service associated with the information spaces will be configurable, enabling users to specify additional filtering criteria, e.g. an article’s perceived level of difficulty. The user will also be able to select the time, and the form of delivery, of information space notifications (section 3.10).

Similar functionality will also be made available to users configuring private information spaces. Users setting up a private information space will be able to invite other users to join their space, thereby providing a means to share information between small groups of users. Alternatively a user could encourage other people to join their space by making their private information spaces visible to other users of the digital library. Users could also ask the domain expert to change the class of their information space from private to public.

3.4.2 Application scenarios

3.4.2.1 Scenario 1: a user subscribes to public information spaces

Intention of scenario/setting: This scenario describes how a user subscribes to a public information space based upon their navigation and browsing of the information space segment of the digital library ontology. In this example scenario, a BT worker wishes to find some further information on the subject of 'knowledge management', specifically in the area of the integration of knowledge management systems into business processes. As this is a long-term interest, the user wants to be notified of any updates to the digital library. The user decides to join some public information spaces. In this example the user would like to receive notifications of relevant updates via email at the point when the digital library database gets updated.

Expectations: Users expect the description of the information spaces to be sufficiently clear. Users do not expect to have to read articles associated with the information space for them to judge whether that public space is likely to be of use to them.

Pre-conditions: The user has opened a web browser and navigated to a page displaying the public information space segment of the digital library ontology. The user’s profile is enabled.

1. The user browses the ontology of public information spaces, and positions the mouse pointer over the 'description' link of the 'knowledge management' space.
2. A brief description of the information space, and summary information relating to any sub-spaces or related information spaces are displayed (a brief description for each related space is displayed when the user positions the mouse pointer over the name of the space).
3. The user clicks on the ‘knowledge management’ information space link.
4. The ‘knowledge management’ information space gets displayed in the browser (an example screenshot is shown in section 3.4.9)
5. The user clicks on the 'description' link of the 'Management information systems' sub-space.
6. The system returns the full description for that space.
7. The user selects the ‘Management information systems’ space and clicks the 'subscribe' link to subscribe to that space.
8. A list of people already subscribed to that space is displayed. Contact details of each person are displayed when the user positions the mouse pointer over their name (a user can indicate to the system that they wish to remain anonymous, in which case their membership to the information space is not visible to other users).
9. The user is given the options to filter the results of the information space rather than just receiving all updates, e.g. the user can specify the type of resources that are of interest to them, the type of document, its level of readership (general reader, professional, expert), and the source of the information (ABI, Inspec, or WWW).
10. The system prompts the user to give details of how they would like to receive notifications for updates to the space (please refer to section 3.10). The system also prompts the user to indicate whether they would prefer their details to be visible to other users of the space, or remain private.
11. The user indicates that they would like to be notified about new information by email whenever the ‘business processes’ information space is updated with specific types of publication. The user selects the ‘anonymous’ option as they would prefer their details to be kept private.

Post-conditions: The public information space is updated with details of the user. The user’s profile is also updated with the subscription and notification details for the information space. The user’s details are not visible to other digital library users viewing or subscribing to that information space (the user opted to keep their details private). The information space queries are run whenever the digital library ontology is updated with new information from the Inspec or ABI databases.

3.4.2.2 Scenario 2: digital library recommends relevant public information spaces

Intention of scenario/setting: This scenario describes how a user subscribes to a public information space based on recommendations from the digital library. The recommendation is based on the user’s interaction with the digital library and the topics of interest active in their profile. This example scenario considers a relatively

new user who, having spent some time browsing the content of digital library, decides that they would like information of relevance to certain parts of their job to be sent to them. The user would like the digital library to recommend some relevant information spaces.

Expectations: The user expects the digital library to recommend relevant public information spaces based upon their interests, their browsing activity, and their use of the digital library and knowledge sharing tools (section 3.5). The user would also like some information space recommendations based upon a description of their current project.

Pre-conditions: The user has opened a web browser and navigated to the public information space section of the digital library ontology. Topics of interest in user's profile are active.

1. Having browsed the ontology of public information spaces, the user does not have a clear idea of which spaces to subscribe to.
2. The user requests the system to suggest some public information spaces based upon articles they have accessed in the digital library over the last 4 weeks.
3. The system suggests some public information spaces based upon the user's criteria. Details of any private information spaces that the owning users have opted to make visible to other users are also displayed. The reasons why the system recommended each space are shown to the user.
4. The user views the descriptions of the recommended spaces, and decides to subscribe to some of them. The user specifies their notification requirements for each information space. The user also indicates that they would like their subscriptions to be visible to other digital library users.
5. The user selects a system function that will suggest information spaces based upon a submitted text description. The user submits a text description of their project.
6. The system extracts key features from the text and suggests some subscriptions to relevant information spaces.

Post-conditions: The subscribed information spaces are updated with details of the user. The user's profile is also updated with the subscription and notification details for each of the information spaces they decided to subscribe to. The user's details are visible to other digital library users. The information space queries are run whenever the digital library ontology is updated with new information from the Inspec or ABI databases.

3.4.2.3 Scenario 3: presenting updates to the information space

Intention of scenario/setting: Some information spaces will be configured to retrieve a broad range of subject material from multiple sources, whilst others will be configured to retrieve a very narrow range of material, possibly from just a single source. The display of information for an information space covering a broad range of topics, e.g. an information space gathering information on the subject of 'knowledge management', will be presented to the user in the form of a navigable ontology of the topics associated with the information space. This scenario describes how updates to a public information space are presented to the user.

Expectations: For an information space covering a broad range of topic areas, the user expects to get to the information that is relevant to their needs reasonably quickly. The user expects the application to make use of their personal profile to highlight information that matches their interests.

Pre-conditions: The user has opened a web browser and navigated to the 'latest updates' page of a public information space. Selected topics of interest in the user's profile are active.

1. The results from the information space are displayed. Each result is in the form of a summary that includes the title of article, its author, a taster, publication, and publication date. Results are biased towards the topics contained in the user's profile. In this example scenario, the 'knowledge management' information space covers a wide range of topic areas. The topic ontology associated with that information space is also presented to the user (topics are only presented if they have associated results). An indication of the number of results classified into each topic area is displayed.
2. The user positions the mouse pointer over some of the topics presented in the information space ontology. A description for each topic is presented.
3. The user selects the 'knowledge engineering' topic area.
4. The system displays the information space updates that have been classified into that topic area.
5. The user views selected results from the 'knowledge engineering' sub-space of the 'knowledge management' information space.
6. The user saves some articles to their personal work space.
7. The user repeats this process for other topic areas associated with the information space.
8. The user decides to update their information space notifications, so they only receive email updates on selected sub-topics of the broader topic 'knowledge management'. The user makes private subscriptions to some topics areas, and public subscriptions to other topic areas.

Post-conditions: The information spaces are updated with details of the user's subscriptions. The user's profile is also updated with the subscription and notification details for each of the information spaces they decided to subscribe to. The user's details are made visible to other digital library users. The information space queries are run whenever the digital library ontology is updated with new information from the Inspec or ABI databases.

Setting 2: The digital library monitors the user's reactions to the information space notifications, e.g. whether the user views new items following an alert or whether they choose to ignore the notifications. The library continually analyses the user's search and viewing habits and updates their user profile accordingly, e.g. with topics of interest. The system will notify the user when changes to their profile give a good match to a previously unsubscribed information space (the system will also inform the user when a new information space is created that matches interests in the user's profile). When the user consistently fails to respond to notifications from a particular information space, the system will suggest that they may wish to unsubscribe from that space. The user is given the opportunity to inform the system that they will be

away from the office until specified date; the system will not send unsubscribe messages under these conditions.

3.4.2.4 Scenario 4: the user creates a private information space

Intention of scenario/setting: This scenario describes the scenario where a user configures a private information space. A BT worker has just been informed that their current project is scheduled to finish within the next 2-months, and that they will be moving to work on a joint project concerning the secure storage of patient records. The person is interested in collecting introductory level articles concerning the adoption of information technology within the health service.

Expectations: The user expects to be able to create a private information space based upon one or more search queries. The user expects to set additional filtering criteria based upon the types of resource that they wish to be notified of. The user expects to be able to schedule the forwarding of notification updates for times that are convenient for them.

Pre-conditions: The user opens a web browser, navigates to the digital library home page, and from there to a list of public information spaces. Having browsed the list of public information spaces (and their descriptions), the user decides that the scope of the information spaces do not suit their needs. The topics of interest and named entity associations in the user's profile are switched off (so as not to bias the results of the initial search).

1. The user selects the option to create a private information space.
2. The user is prompted to enter some search criteria for the information space, e.g. source of information, some keywords that best explain their information requirements, a publication date, and a level of readership.
3. The user searches for information in the digital library, e.g. using a search strategy similar to that described in search and browse scenario 3 (section 3.3.2.2). Alternatively, the user can select previously used queries from their search history.
4. The search engine matches the query against the digital library's index, and returns a set of results.
5. The user refines their search until the results returned from the search engine match their needs, e.g. the user enables named entity recognition functions to retrieve only articles that mention the NHS[17] (in a similar way to that described in search and browse scenario 3 - section 3.3.2.3).
6. The user stores the query in their private information space. The system prompts the user to give the information space a friendly name.
7. The user instigates other queries, and stores selected queries in the private information space.
8. The user specifies their private information space notification options.

Post-conditions: A number of user-specified queries are stored against the private information space. The information space queries are run whenever the digital library

[17] National Health Service.

ontology is updated with new information from the Inspec or ABI databases. The user is alerted according to their alerting preferences.

3.4.3 Other scenarios/functions to consider

Further consideration should be given to the following scenarios and functions when designing the public information spaces:

1. Extend the scope of the information space to include other sources of information, e.g. articles retrieved by automatic WWW crawlers, or users' annotations of other content. Users may wish to receive updates on new content retrieved by the crawlers. Alternatively users may prefer to view the results of latest updates on accessing the information space.
2. Consider learning a domain ontology for each information space[18]. Such a learned ontology would be much more fine-grained than the global topic hierarchy (or a subset of it). Such an ontology could be used for: i) describing the contents of the information space, i.e. the concepts, instances and relations it includes, ii) browsing the information space (each concept, instance or relation could be linked to the documents or the text fragments it has been extracted from), iii) identifying sub-topics within the information space, and iv) discovering new topics which could be integrated into the global topic hierarchy.
3. Give the user a personalised view of the information space ontology, i.e. biased by the topics of interest in their profile.

3.4.4 Expected benefits

The more fine-grained information spaces will enable users to receive updates that are targeted at more specific topics areas rather than the general topic areas covered by the current information spaces. Results from the information spaces can be biased towards the user's profile, giving further relevancy. Users will be able to schedule notifications according to their preferences, rather than at the time of the database update. Scroll over descriptions will provide the user with further information about the scope of the information space. The 'sense of community' that exists with the current information spaces is expected to be retained.

3.4.5 Potential issues

The title of an information space is usually sufficient for giving the user an appreciation of its scope, e.g. for the general topics knowledge management or semantic web. A finer-grained information space that covers a narrower topic of a more general information space may require further descriptive information to help the user understand the scope of that space (users could view the resources already identified by the finer-grained information space to gain an appreciation of its scope, but this would not be satisfactory if there are a significant number of spaces to browse). Hence there is a requirement to provide descriptions of the spaces based upon the topic ontology and the attributes of the resources associated with that space. The quality of the automatically generated descriptions will rely on the structure and

[18] When the content of an information space changes it would be quite expensive to start learning a new ontology from the scratch. With data-driven change discovery the existing information space ontology will just be adapted according to the changes within the information space.

population of the segments of the ontology associated with the information spaces and the applicable parts of the topic ontology. Further research is required to determine the best way to structure the topic and information space parts of the ontology to enable reasonable natural language text descriptions to be generated. Suitable 'information space-to-user profile', and 'information space-to-new content' matching algorithms need to be investigated further.

3.4.6 Supporting technology and functions

The following technologies are required to support public and private information spaces:

1. A user profile-enhanced search and browse application/search engine (for searching for information prior to creating a private information space).
2. An attribute-based search interface (the resulting query can be stored in the information space).
3. Components to monitor and capture the user's interaction with the digital library, e.g. their information space subscriptions and their responses to notifications.
4. Functions to manage the user profile (e.g. information space subscriptions).
5. Ontology-based query expansion based upon the topics and relationships between topics in the digital library ontology.
6. Named entity recognition (enables companies, people, places, etc to be specified as part of the information space query).
7. Algorithms/functions to calculating the semantic similarity between: a) the topics defined in the information space and the topic classifications of the resources (match information space to content), and b) between the topics defined in the information space and the topics of interest in the user profiles (match user to information space).
8. Components for producing alternative visualisations of the information spaces, e.g. present the user with interfaces appropriate to the device being used and any specified user preferences.
9. A component to generate natural language descriptions of the topics covered by the information space.
10. Components to learn individual information space ontologies.

3.4.7 Supporting data

Public and private information spaces will be supported by the following data objects:

1. The BT digital library ontology. Information space specific information, e.g. its title, description, queries, and membership, will be stored in the ontology. The ontology will be designed to facilitate the automatic generation of descriptions for the information spaces.
2. A user profile holding personal details, information space memberships, topics of interest, notification preferences, etc.
3. A visualisation ontology that will transform the machine oriented data of the digital library into a form suitable for presentation to the user.

3.4.8 Relevant SEKT tasks

The following SEKT tasks will support the digital library information spaces:

WP/Task number	System/Component/Function
T1.1, T1.2	Algorithms/functions to calculate the semantic similarity between: i) information space queries and content updates, and ii) the information space topics and topics of interest in the user's profiles.
T2.6, T6.3, T6.4.	An API to the underlying semantic annotation platform and knowledge base or an API to the SEKT integrated platform. Named entity recognition functions will be integrated into the annotation platform.
T1.8, T11.2	Proton ontology and BT digital library domain-specific ontology.
T3.3	Components to learn individual information space ontologies using 'data-driven change discovery'.
T5.2, T11.3	Semantically enabled search engine combining free-text search with capability to exploit semantic metadata, e.g. ontology-based query expansion, named-entity recognition. Attribute-based search interface (which is used to structure a query for a private information space).
T5.4, T11.3	Visualisation component that takes the information space section of the digital library ontology and transforms them into a form suitable for presentation to the user.
T5.5, T11.3	Component to update the user profile with information space subscriptions.
T5.6, T11.3	Automatic generation of descriptions of information spaces based on topics in the BT digital library ontology.
T5.7, T11.3	Content purposing component to transform user interfaces into format suitable for display on alternative devices, e.g. PDAs.

Table 3.2. SEKT tasks that support the digital library information spaces.

3.4.9 Example screenshots

An example screenshot for the 'knowledge management' public information space is shown in figure 3.5.

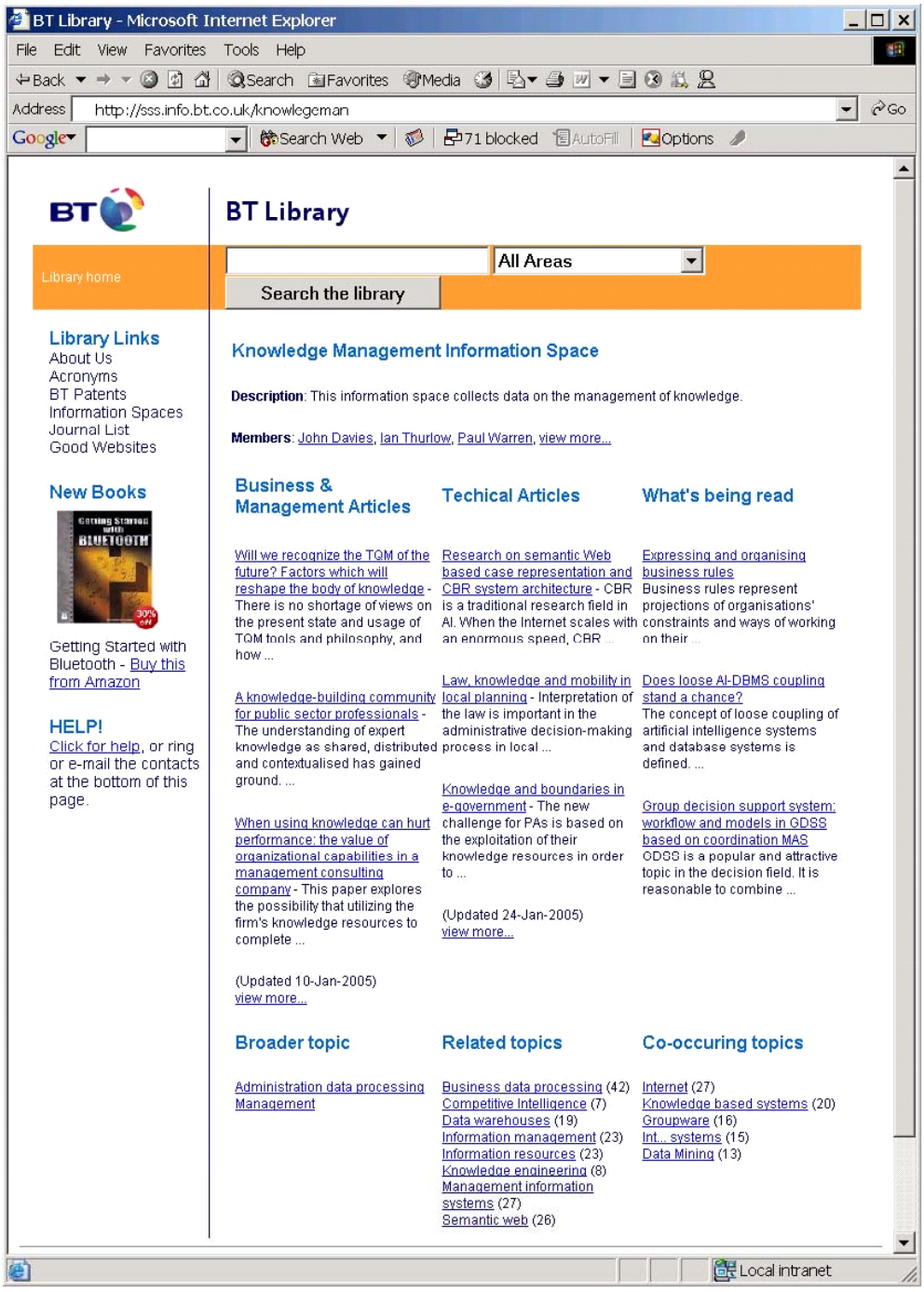

Figure 3.5 Knowledge management information space.

3.5 End user use case 3: Knowledge sharing

3.5.1 Description

Adding new information to the library is one way of contributing to the store of knowledge. Another is to add a comment on existing information. This use case describes how a user can organise URLs in a way that is comparable to ‘bookmarks’ or ‘favourites’ in a web browser. Sharing this information with other users occurs as a side-effect of the user annotating the URL for their own purposes. The proposed annotation and sharing application provides a semi-automatic way to assign

meaningful topics to URLs, thereby improving the ability of a user to find a certain URL again.

The combination of the URL with a user defined title, a description, a list of topics assigned from the ontology, and additional user comments is called a *jotting* (or a *jot*). The sharing of the jotting occurs automatically; with the system matching the jotting against the profiles of other users (the user is also given the option to specify other users that they would like to share the jotting). Users can search through the jottings other users have made. This approach to sharing knowledge is expected to get around one of the main problems of knowledge sharing systems; what makes somebody share information if it is not of direct benefit to them personally? With this approach, a user creating a jotting has self-interest to do so, in that they believe that they may want to find a certain web page again at a later date. As the knowledge sharing is simply a side-effect of creating the jotting it does not require any altruistic action by the user; the user creates the jotting primarily for their benefit (the exception to this occurs when a user decides to keep their jottings private or restricted to a specified set of users).

Other users will be able to comment on public jottings, thereby encouraging discussion within the digital library community. The originator of the jotting will be notified of the comment, and of the user who made the comment. It is anticipated that the jotting application will promote discussion, and lead to the sharing of knowledge amongst digital library users. In addition to creating a jotting, users can either edit or delete an existing jotting. The jotting also identifies starting points for search and browse activities within the digital library, e.g. users can navigate to other web pages that are classified against the topics used to classify the jotting.

The application will enable a user to extend the digital library ontology with their own, private concepts in accordance with the local adaptation step of the DILIGENT process [7]. A user will be able to introduce new topics, consisting of a name and a description. Furthermore, a user will be able to edit and delete the topics they have created themselves (users will not be able to introduce new relations between topics of the digital library ontology).

The primary application functions will be provided via the toolbar of the web browser, thereby integrating the application into the daily task of searching and browsing.

3.5.2 Application Scenarios

3.5.2.1 Scenario 1: add a jotting

Intention of scenario/setting: A typical interaction with the jot application would begin with a user finding a web page that they may want to recall at a later date. The user decides to create a new jotting for this page, and clicks on the ‘Create jotting’ button in the browser. The application extracts the URL and title of the web page and presents this information to the user in a pop-up window. The user is able to edit the title of the web page. Additionally the user can enter a short description for the jotting that describes the web page. The application selects some topics from the digital library topic ontology that are most relevant to the content of the web page, and suggests that these topics should be assigned to the jotting. Other less relevant topics are also displayed, which can be selected by the user. To give users a better understanding of the topics, every topic has a short description. A user can assign

their own topics to their jottings. In cases where a user creates a new topic, they can also enter a short description for that topic.

This scenario describes how a user will add a new jotting about a web page to the digital library. A BT worker is using a web browser to browse the Internet, and finds an interesting web page. The worker anticipates that they may want to recall the web page at a later date. Furthermore, the worker wants to share this information and discuss it with some work colleagues.

Expectations: The user expects the application to derive relevant information from the content of target web page, including its URL and title, and suggest a number of relevant topics from the digital library topic ontology that can be used to classify that web page. Furthermore, the user expects the system to inform them if another user of the digital library had previously jotted that WWW page. The user expects the application to structure their jottings in a way that makes it easy for them to retrieve at a later date.

Pre-conditions: The web page to be 'jotted' is displayed in the user's web browser. The web page has not been jotted by another user previously. Topics of interest are disabled in the user's profile. Named-entity recognition functions and named entity associations are disabled.

1. The user clicks on the 'Create jotting' button in their web browser's toolbar (an example screenshot is shown in section 3.5.9).
2. The system checks that the URL has not been jotted previously.
3. A pop-up window opens in the user's browser, giving details of the URL of the page, its title, suggestions for relevant topics to assign to the web page, and a blank text field for providing a personal description (an example screenshot is shown in section 3.5.9). All information fields can be edited by the user if desired.
4. The user is satisfied with all of the pre-populated information, adds their own description of the web page into a free-text field (extracts from the target web page can be 'pasted' into the text field), and clicks the next button to submit the jotting to the system.
5. The user is given the opportunity to inform other users of the jotting (an example screenshot is shown in section 3.5.9). The user enters the names of the people they want to notify. The application accesses the user's address book and prompts the user to confirm the name of the person in cases where there is ambiguity (this is quite common in a large organisation). The user provides a reason for why they chose to share the jotting.
6. The application generates a jotting notification and alerts the people the user wants to share the jotting with. Furthermore the system identifies other users with profiles that match the jotting (e.g. the topics of interest in others user's profiles provide a good match to the topics that have been used to classify the jotting). The jotting notification is delivered to the recipients according to their profile preferences, e.g. instant messaging, email, or by showing a message when the user next logs on to the digital library (users can opt to not receive jotting notifications, or opt to restrict the delivery of notifications to a specified set of users. The jotting will show up on the personal topic hierarchies of those selected by the user (with a note indicating why the user creating the initial jotting thought this would be a useful item for them).

Setting 2: A user has invoked the application, but notices that the topics suggested by the application do not match their particular needs for describing its content. The user deselects some of the topics suggested by the system. The user is given the option to browse the digital library ontology for more suitable topics, or search for topics based on a keyword search. Alternatively the user can create topics of their own choice to describe the document, e.g. the user wants a project-based view of the information, and would therefore like to classify a particular web page against an internal project name, rather than simply use suggested topics from the digital library ontology. If the user chooses to create their own topic, they will be prompted to enter a title and a short description (the user is extending the official topic hierarchy with their own private topics). The user can request the application to compare the title and description of the new topic with topics in the digital library ontology and suggest any relevant topics (the user can choose to ignore any suggestions).

3.5.2.2 Scenario 2: use a jotting

Intention of scenario/setting: This scenario describes how a jotting can be used to recall previously jotted web pages.

A BT worker wants to revisit a web page about a competitor's business that they Jotted 6 months ago. The user does not remember the correct name or the precise topics that were used to describe the web page.

Expectations: Users expect to find previously jotted WWW pages faster and more easily.

Pre-conditions: The user has opened a web browser. Topics of interest in the user's profile are disabled.

1. The user enters some keywords that they believe will occur in the target web page into the search field of the web browser's jottings toolbar. The user selects the option to search only in their jottings.
2. The application processes the request, matching the query against the index.
3. The results are shown in the browser. Related topics are also shown, giving the user the opportunity to browse through the ontology to find the jotting they are looking for (topics that have not been used to classify jottings will not be shown).
4. The user analyses the results, but does not find the search target they seek.
5. The user clicks on a link for one of the related topics. The application retrieves the user's jottings that were classified using that topic.
6. The user finds the jotting they were looking for.
7. The user clicks on a link for the target web page. The correct web page is displayed.

Post-conditions: The jotting is added to the user's 'most recent jottings' list (which is associated with the user's profile).

Setting 2: Had the user jotted the target web page only a few days before, they would have checked the 'Most recent jottings' function. The system would respond by returning a time ordered set of jottings.

3.5.2.3 Scenario 3: edit a jotting

Intentions of scenario/setting: This scenario describes how a user can edit the jotting after becoming aware of new information. In this example the URL of the original web page had changed.

Expectations: The user expects to be able to change any jotting information, e.g. its title, the URL (say the URL to an original jotting had changed), the user's comment, or the topic classification.

Pre-conditions: The user has opened a web browser, and located a previously submitted jotting.

1. The user clicks a link in the jotting that references the original web page.
2. The browser displays a 404 error message (which includes a referrer).
3. The user views the web page at its current URL. The application asks the user to confirm that the current URL is valid. The application also identifies and displays any new topics (this could be due to either the content of the target web page changing or the topic ontology changing).
4. The user views the target web page at the new URL, and confirms to the application that the URL is correct.
5. The user decides to change one of the jotting's original topic classifications by deselecting it. The user requests a list of further related topics to be displayed.
6. The application suggests some further topics. The user selects appropriate topics to classify the jotting, and communicates to the application that they have finished editing the jotting.

Note: the application will access the web pages of all jottings from time to time and if a 404 error message is received it will mark those jottings to have a broken link.

Post-conditions: The jotting is added to the user's 'Most recent jottings' list (which is associated with the user's profile). The jotting is marked as being reclassified by the original user. Details of the reclassification are given.

Setting 2: The user could also classify the jotting using a new topic of their choosing, rather than use a topic from the ontology (please refer to scenario 1, setting 2, section 3.5.2.1)

3.5.2.4 Scenario 4: remove a jotting

Intention of scenario/setting: This scenario describes how a user can remove metadata and the URL from their jotting perspective[19]. Other users will still be able to see the jotting (although it will be amended to indicate that the person who submitted the original jotting had removed it from their personal jottings).

Expectations: The user expects to be able to delete their jottings easily. Other users expect to be able to view all jottings submitted to the digital library, even if other users have deleted some jottings from their personal list.

[19] A jotting perspective enables people to have a different view of the jotting. By allowing different views on the same jotting, it is possible for users to assign their own topics to a shared (public) jotting.

Pre-conditions: A user has invoked the jot application and identified a jotting that is no longer perceived to be of use. The user decides to delete the jotting.

1. The user clicks on the 'Delete jotting' button.
2. The jotting is removed from the user's list of jottings.
3. The jotting will be marked accordingly so that other users are made aware that the originator of the jotting has deleted it.

3.5.3 Other scenarios/functions to consider

The Jot application gives an example how the local adaptation step in the DILIGENT process could be used in a software application. It shows how an ontology can be extended by users in an unobtrusive way to provide their own private view of content classified by that ontology. The DILIGENT process facilitates the review of private changes made by users (with the aim of identifying similarities between users' topics). Topic similarities are indicators for new topics that might be introduced into the shared ontology. New topics introduced into the ontology could be used for trend spotting, and could be presented to the users of the digital library whose personal profile contains topics that are related to that topic. Trend spotting could be supported by machine learning/ontology learning techniques.

3.5.4 Expected benefits

The jot application enables users to share information within the classification framework of the digital library, enabling them to retrieve information more easily than, for example, using bookmarks in their browser (and allow access to jottings from multiple devices). Indeed, given the natural human reluctance to spend time on knowledge and information sharing activities, it is thought that the primary motivation for 'jotting' will be easy retrieval by the user; sharing will be a by-product of this activity. By making sure that the sharing is not an altruistic act, the sharing should be more efficient. People will use the tool because they see an instant benefit for themselves. The sharing of jottings is expected to promote discussion and help build communities of practice.

The jot application enables a user to extend to the digital library ontology with their personal topics. This is expected to facilitate easier recall at a later date, and enables a user to classify using topics that not represented in the digital library ontology.

3.5.5 Potential issues

The processing of a WWW page and the matching of its key features against the index of digital library resources the digital library may require considerable processing time. The performance of the application from the perspective of the user must be reasonably fast.

The creation of new jottings must be integrated into the normal work process as tightly as possible as most users will not be willing to spend a lot of time completing the jotting process. The effort a user has to invest in completing all of the fields for a jotting may be too great for some.

This example scenario has considered a user accessing jot from a Web browser on Intranet connected PC. Requirements for users accessing the Jot application from a mobile device also need to be considered.

3.5.6 Supporting technology and functions

The following technologies are required to support the jotting application.

1. A user profile-enhanced search and browse application/search engine that enables a user to search for jottings. The keywords and topics specified by the user will be augmented with active topics of interest from their profile, and compared to the index of submitted jottings.
2. Algorithms and functions that extract the content of a target web page, compare it to the features representing topics in the topic ontology index, and rank the topics in order of relevance.
3. Components that generate alternative visualisations of the jottings user interface to support users accessing information form mobile devices such as PDAs, and pocket PCs.
4. A component is required that maps between the user's ontology and the digital library ontology, thereby allowing other users of the digital library to query and access other peoples' jottings. This enables users to use their own topic ontologies to group documents/articles (these ontologies will differ between users).
5. Named entity recognition components, e.g. to enable an article or news item to be classified against a company name, a person, a place, etc.
6. Functions to manage a user profile (the user may wish to activate/deactivate topics of interest as they search for jottings).

3.5.7 Supporting data

The following data items are required by the Jot application:

1. The BT digital library ontology (topics, resource ontology, and jottings ontology), classes, attributes and relations that hold details of the jotted web pages, e.g. URL, title, extracted keyword/phrases, user supplied description, digital library topic classification, personal classification, date/time of jotting, and private/public flag.
2. User profiles for holding users' interests (topics from the digital library topic ontology), user preferences, named-entity associations, search and browse history. The profile will also contain data of a more static nature, e.g. the user's contact details.
3. A visualisation ontology that will be used to translate the machine oriented data of the digital library into a user-friendly visualisation.

3.5.8 Relevant SEKT tasks

The following SEKT tasks will support the Jot application:

WP/Task number	System/Component/Function
T1.1, T1.2	Algorithms/functions to calculate the semantic similarity between information extracted from WWW pages and topics of interest in a user profile.
T2.6, T6.3, T6.4.	An API to the underlying semantic annotation platform and knowledge base or an API to the SEKT integrated platform. Named entity recognition functions will be integrated into the annotation platform.
T1.8, T11.2	Proton ontology and BT digital library domain-specific ontology.
T5.2	Semantically enabled search engine combining free-text search with capability to exploit semantic metadata, e.g. ontology-based query expansion, named-entity recognition. Attribute-based search interface. Used to search for jottings
T5.4	Visualisation component that takes the jottings section of the digital library ontology and transforms it to present the user interface and output of Jot (via a visualisation ontology).
T5.5	Component to update the user profile with information related to the user's personal topics.
T5.6	Automatic generation of descriptions of topics in the BT digital library ontology.
T5.7	Content purposing component to transform user interfaces and output into format suitable for display on alternative devices, e.g. PDAs.

Table 3.3. SEKT tasks that support the Jot application.

3.5.9 Example screenshots

This section shows some example screenshots of what the jot application may look like. A browser window with the installed jotting toolbar is shown in figure 3.6.

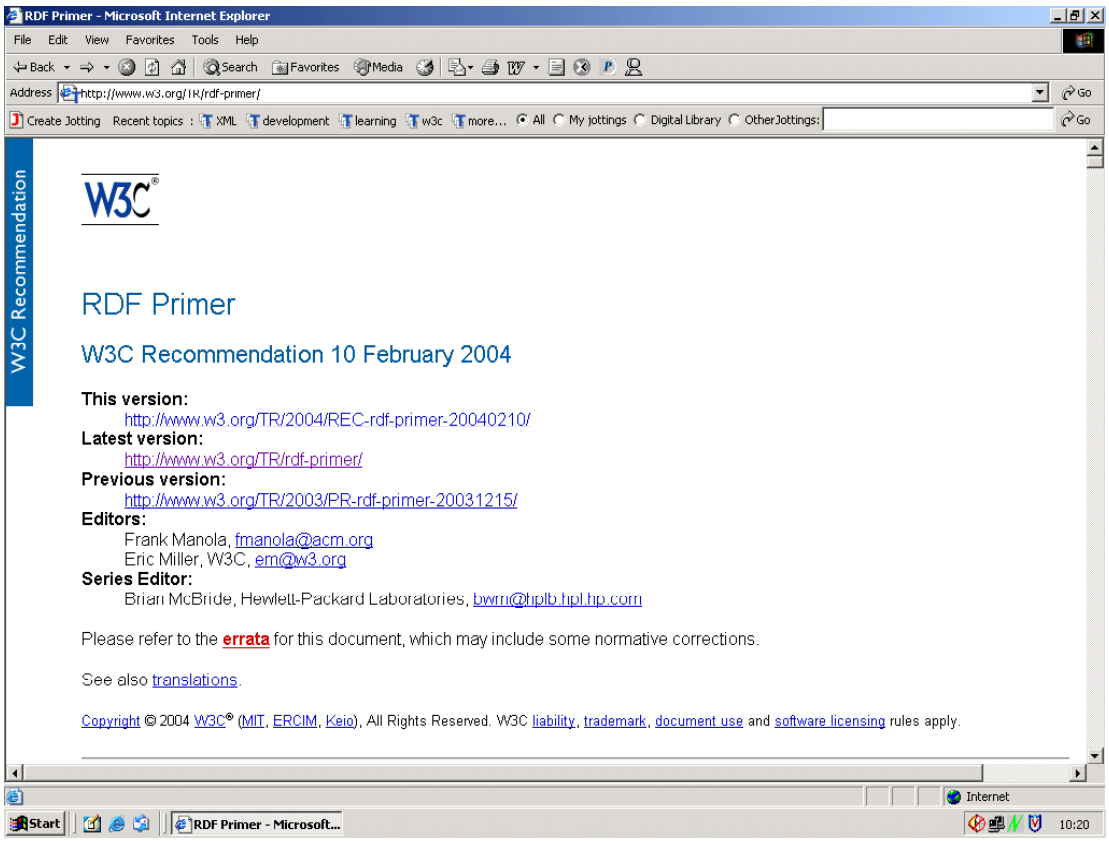

Figure 3.6 Browser showing jotting toolbar.

Figure 3.7 shows the pop-up window that will be presented when a user creates a new jotting. The yellow rectangle contains a description for the topic, and will be displayed on the event of a “mouse over”. Topics in the left-hand column are suggested by the application for assigning to the jotting. Topics in the centre column do not match the content of the web page to the same level of relevancy (of minor importance). The right hand column shows topics that have been recently used to create new jottings.

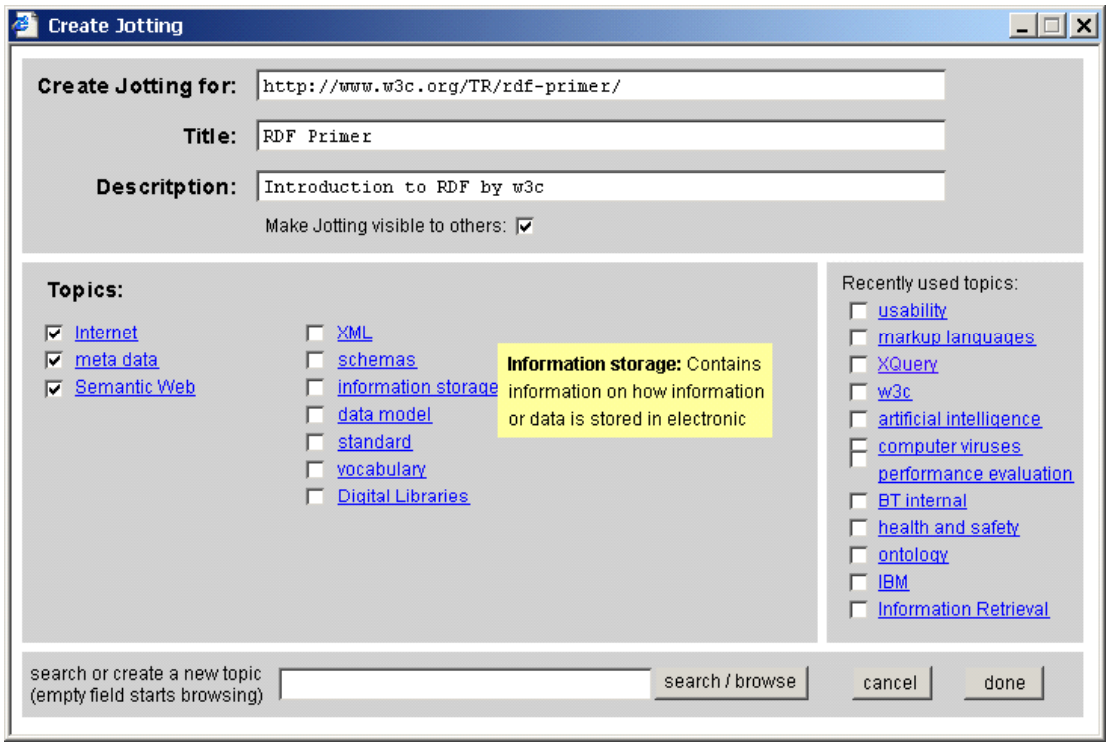

Figure 3.7 jotting pop-up window.

Figure 3.8 shows the topic ontology. In this example screenshot, the active topic is "data structures". Relations to other topics, such as broader topics, narrower topics and related topics, are shown. The user is able to search for existing topics and create a new topic.

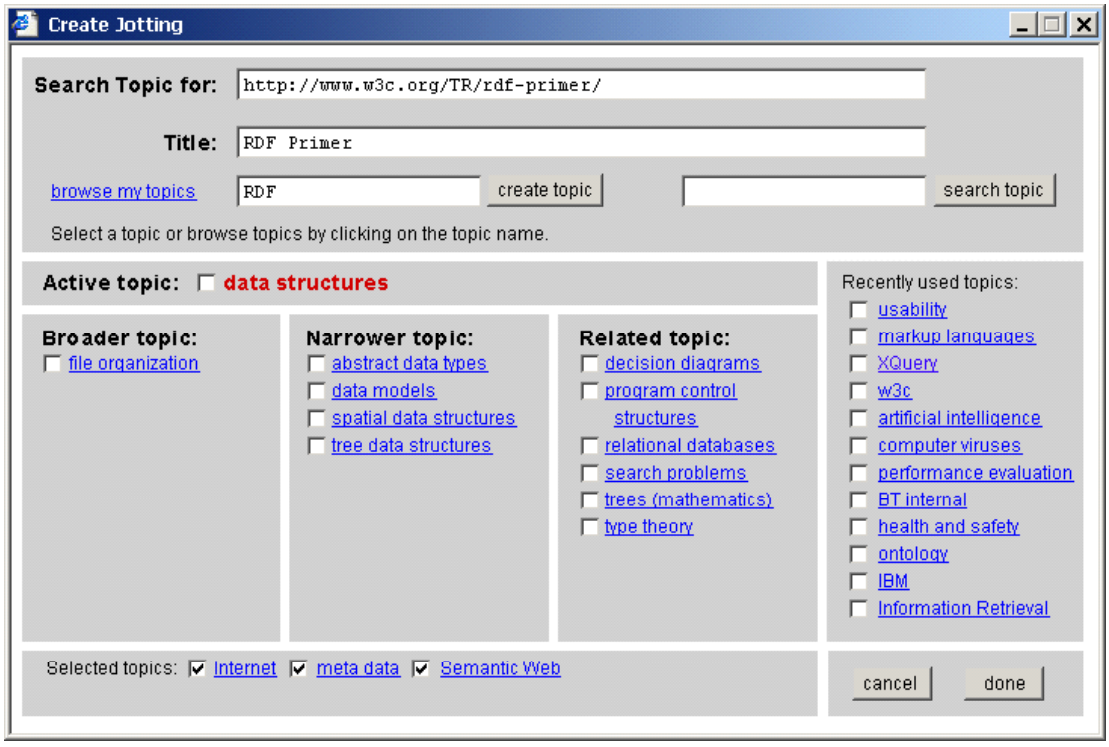

Figure 3.8. Topic ontology.

The final screen that is presented to the user when creating a jotting is shown in figure 3.9. The user can choose to actively share the jotting with other people (the application will also push the jotting to users where the jotting matches their user profile.).

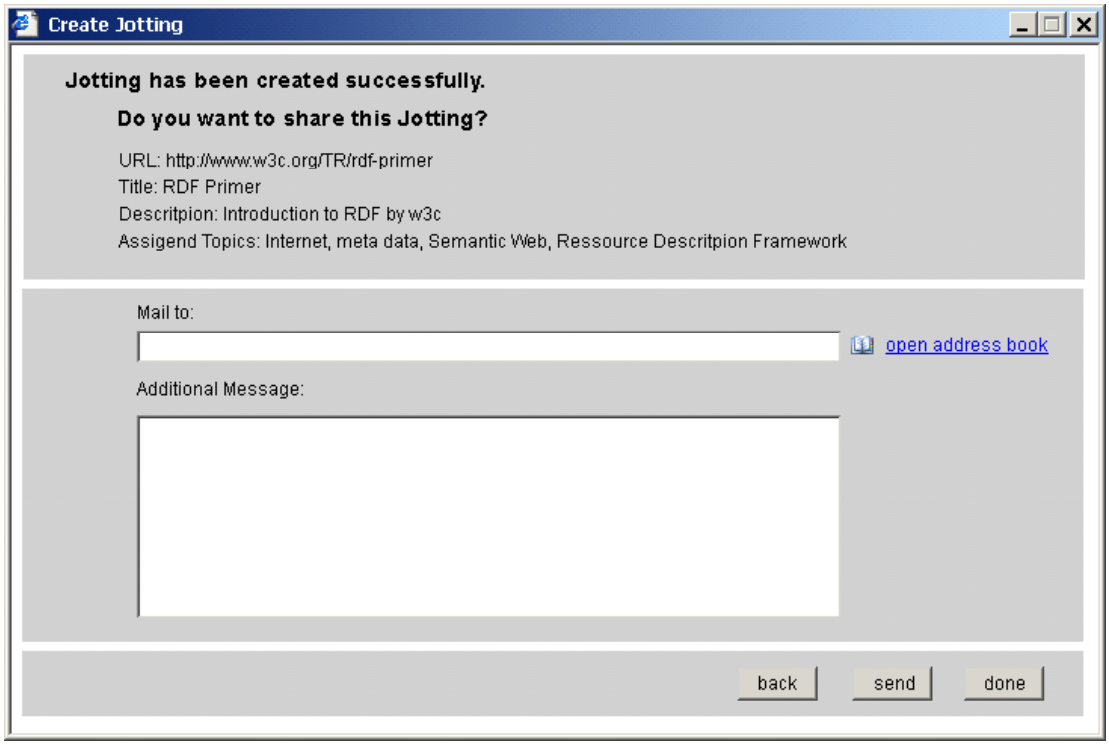

Figure 3.9. Sharing a jotting.

3.6 End user use case 4: Expertise location

3.6.1 Description

Sometimes written information by itself is not enough; a person faced with a particular problem to solve needs to speak to an expert. The identification of experts can be aided by monitoring the level of difficulty of what people are reading in specific topic areas. It is a reasonable assumption that an individual who habitually reads 'expert level' documents on a particular topic area is indeed expert themselves in that area. A user wishing to speak to an expert in a particular topic area could use the digital library as a means to identify such a person. Publications in the digital library will therefore need to be classified according to a level of reader difficulty, e.g. articles could be classified as being aimed at the general reader, the professional, or the expert in a particular topic area. Furthermore, an expert could also be identified through the classification of the content of their emails into a similar level of expertise. On other occasions an individual may be not so much interested in identifying an expert in a particular area as in finding other like-minded people interested in a particular topic area, perhaps at the same level of understanding.

Whether based on readership of the digital library or on an analysis of emails, expertise location is clearly subject to concerns about privacy and confidentiality [3]. Experts, for example, may not wish to be identified, if only because of a fear of being

overwhelmed with requests for help. One solution to this might be for such experts to remain anonymous, with all queries to be channelled through the digital library itself.

This use case describes how users can identify experts and establish communities of interest through the digital library.

3.6.2 Application scenarios

3.6.2.1 Scenario 1: finding an expert

Intention of scenario/setting: This scenario describes how a user uses the digital library to find an expert in a particular topic area. A worker in Human Resources (HR) needs help with regard to a case involving bullying and harassment of somebody registered with a disability. The HR worker has found relevant background papers and information about pertinent legislation in the digital library. However, before deciding on how to handle the case, the HR worker wants to speak to somebody within the company who has a deeper understanding of the key issues, and how it may be best to handle them.

Expectations: The user expects to be able to locate an expert as a result of collecting relevant information about a particular area of interest and requesting the library to identify experts in that area.

Pre-conditions: The user has identified a number of relevant papers from the library and stored these in a personal work space. The user wants to identify the people who have been reading papers on related topics, e.g. bullying and disability.

1. The user invokes the expertise location tool from their personal work space.
2. The tool returns a list of topics that best match the documents stored in their work space.
3. The user browses the list of topics. In this example, the topics 'harassment' and 'disabilities' appear to be the most useful.
4. The user expands the topic 'harassment' to identify its sub-topics and any related topics. The sub-topic ‘minority harassment’ appears to be appropriate.
5. The user requests details of people who have been reading articles in the topic of 'minority harassment', and related topics, at the ‘expert’ level.
6. The system returns a set of results matching the request. Each result gives a summary of the user details, e.g. contact details and links to publications they have read recently.
7. The user selects a number of results and indicates that they wish to store references to those users against their work space.
8. The system prompts the user to provide some text explaining their reasons for storing the references.
9. The user repeats the process of requesting information on users at the expert level.

Post-conditions: The user's interaction with the digital library, i.e. with the personal work space and expertise location functions, is recorded throughout their session. The user's profile is updated with information such as topics of interest, and useful contacts within those areas (including levels of expertise).

3.6.2.2 Scenario 2: finding a community of interest

Intention of scenario: This scenario describes how expertise location can also be used to establish the beginnings of a community of interest.

Expectations: The user expects to be able to identify other people who are interested in a particular subject or topic area.

Setting: A worker within the HR department is studying for their professional exams, and would like to make contact with similar people at the same level or, more preferably, at a slightly more advanced level than them. The user would like to find people classified at the ‘general reader’ or ‘professional’ level in the topic areas covered by the examination, preferably with a work location in close proximity.

Pre-conditions: The user has opened a web browser and navigated to their personal work spaces.

1. The user selects the personal work space which, in this example, contains information and articles relevant to the topics covered in the exam.
2. The tool returns a list of topics that best match the information stored in the work space.
3. The user browses the list of related topics. In this example, the topic 'human resource management' appear to be the most useful. The user expands the topic to identify its sub-topics and related topics, e.g. 'employment policies', 'recruitment', and 'relocating personnel'. The sub-topic ‘relocating personnel’ is most appropriate.
4. The user makes a requests to find details of all people who have been reading articles in the topic of 'relocating personnel' (and related topics), at the ‘general reader’ and 'professional level', and who have primary work locations close to the user's primary work address.
5. The system returns a set of results matching the request. Each result gives a summary of the user, e.g. contact details and links to publications that they have read recently.
6. The user selects a number of results and indicates that they wish to store references to selected users against their work space.
7. The system prompts the user to provide some text explaining why they stored the references to the selected users.
8. The user writes an email to selected users, asking if they can meet. The user also asks if they would be interested in forming an HR community.

3.6.3 Other scenarios/functions to consider

Privacy was one of the key concerns of the focus group study [3]. The functionality, and user interaction, that would give users anonymity should be considered when designing the expertise location tool.

Consider setting up community portals, e.g. with own information spaces, domain experts, focused crawlers.

3.6.4 Expected benefits

The expertise location application meets a number of requirements identified in the analysis of the responses to questionnaires and the focus group study [3], e.g. users expressed a view that they would like to have a personal folder, which could give them options to pick up work from where they left off. Participants were also interested in a library assistant that would be able to put users in contact with people with expertise in particular areas.

3.6.5 Potential issues

The Focus group raised some concerns about privacy and anonymity when using the proposed tools and applications, e.g. knowledge sharing tools. Users would like to remain anonymous unless they choose to contact other users personally.

In applications that require a conception of a user's level of expertise, users should be able to choose the level of expertise that is assigned to them. Participants did not like the idea of people being contacted on an individual basis as they felt both the experts and the user's privacy should be kept (unless they specifically choose to get in touch). It has been suggested that any contact with experts should be brokered through the library. The user requesting the help of an expert would not be given their direct contact details, but instead would be supplied with sufficient background information about the expert. Based on this information the user could request contact through the digital library. Experts will be given complete control over the level of expertise that is assigned to them by the system; they will be given the opportunity to revoke this responsibility at any time.

Focus group participants liked the idea of having documents classified by different levels of expertise, but were worried about the implications of such a classification, e.g. participants expressed a concern that although their knowledge in a particular topic may be at a basic level, they may still wish obtain information at an expert level (or vice versa). All users must still be given access to documents at all levels of readership (the user must have ultimate control over what they can view).

3.6.6 Supporting technology and functions

The expertise location tool will be supported by the following technologies:

1. A tool/component that can assign a level of readership to a document in a particular topic area based on a set of documents already classified into levels of readership.
2. Components to monitor and capture the user’s interaction with the digital library, ascertain their interests, and map those interests to topics in the digital library ontology (the document will be assigned a level of readership/expertise).
3. Functions to manage a user profile (giving the user final control over its active content). The profile management tool must enable users to specify their levels of expertise in their topics of interest.
4. A personal work space application, including algorithms/functions to cluster documents in the work space.

3.6.7 Supporting data

The expertise location tool will be supported by the following data:

1. User profiles for holding users' topics of interest and levels of expertise in those topics.
2. PROTON and the BT digital library domain-specific ontology.
3. Application data specific to the expertise location tool and the user's work space.
4. A visualisation ontology that will transform the machine oriented data of the digital library into a form suitable for presentation to the user.

3.6.8 Relevant SEKT tasks

The following SEKT tasks will support the expertise location functions:

WP/Task number	System/Component/Function
To be agreed (T1.1, T1.2, T1.5).	Algorithms/functions to calculating the level of readership of a document.
T2.6, T6.3, T6.4.	An API to the underlying semantic annotation platform and knowledge base or an API to the SEKT integrated platform.
T1.8, T11.2	Proton ontology and BT digital library domain-specific ontology.
T5.2	Semantically enabled search engine (with API) that enables the expertise location tool to query against levels of readership).
T5.4	Visualisation component that takes the expertise location and work space data and transforms it (via a visualisation ontology) into a form suitable for presentation to the user.
T5.5	Component to update user profile with levels of expertise in topics of interest.
T5.7	Content purposing component to transform user interfaces and output into format suitable for display on alternative devices, e.g. PDAs.

Table 3.4. SEKT tasks that support the expertise location functions.

3.7 End user use case 5: Personal search agent

3.7.1 Description

Search agents can reduce the overhead of completing a manual search for information. This use case describes a semantic search agent that can be configured to query the digital library, WWW or Intranet based search engines for information on behalf of the user.

The semantic search functions will allow users to specify semantic queries based upon the BT digital library ontology, e.g. a very simple query would search for articles in the digital library that are authored by a specified person[20]. In this example, the agent would return results for publications of type article, where a named entity of type person, having a role of author, and identified with the specified name (alias) had been identified in the text. The search agent will notify the user of the results based upon their notification preferences (please refer to section 3.10). When each subsequent query is executed, the search agent checks for new results, and notifies the user according to their notification preferences.

[20] The search agent function that queries the digital library is similar to that of the private information space, with the exception that the private information space query is run whenever the digital library ontology is updated from the ABI or Inspec databases, whereas the search agent's semantic query is scheduled to run according to the preferences set by the user.

3.7.2 Application scenarios

3.7.2.1 Scenario 1: searching the digital library

Intention of scenario/setting: This scenario describes how a personal search agent queries the BT digital library for information.

Expectations: The user expects the search agent to retrieve highly relevant information. The user expects to spend minimal time configuring the agent, but is aware that some effort will be required to 'train' the agent through user feedback.

Pre-conditions: The user has opened a web browser, navigated to the BT digital library home page, and from there to a page giving details of their search agents.

1. The user clicks the 'New Agent' button to create a new search agent.
2. The system prompts the user to give the agent a friendly name and to select the information sources to search (a screenshot for the agent configuration page is shown in section 3.7.9).
3. The user provides a name for the agent and chooses to search the digital library only.
4. The system prompts the user to select the type of search function required, e.g. a) a keyword only based search, or b) a semantically enhanced search.
5. The user selects the semantically enhanced search.
6. The system prompts the user to select semantic query functions, e.g. named entity search, or attribute-based searching.
7. The user selects the attribute-based search option.
8. The system prompts the user to enter details of their search (similar to that described in the attribute-based search - section 3.3.2.4), i.e. the user selects the appropriate resource type from the ontology, enters text against selected attributes, and selects a published since date.
9. The user selects the option for previously viewed agent results to be excluded from subsequent agent searches, and for the system to accept user feedback based on the results. The alternative is for the system to infer user feedback from the user's interaction with the results that are found by the agent, e.g. a user selecting to view and print the full-text of an article found by an agent is likely to have an interest in that article.
10. The system prompts the user to schedule the date and time of the first search, select the frequency of subsequent searches, indicate the number of results to be found, and specify a notification schedule (notification on search results is selected by default).
11. The user enters the necessary details and saves the agent configuration.

Post-conditions: The agent queries the digital library index at the scheduled time, and generates a page of ranked results. A summary of each result is provided, e.g. a search for resources of type article would include, the title of article, its author, an extract from the abstract, publication, and if available a link to the full-text. Links to the properties of the resource enable users to navigate the digital library ontology further. A notification is sent to the user according to their notification preferences.

The user accesses the agent results page, and is given the option to provide feedback on the relevance of each result. In cases where the user indicates that a result is relevant, the system will suggest ways in which the user could modify the agent's search criteria so that it finds similar resources. In cases where the user indicates that a result is not relevant, the agent will suggest ways in which the search criteria could be modified so as not to find similar results on subsequent searches. The user modifies the agent search criteria and logs out of the digital library. The process of the agent searching and returning some results, and the user providing feedback is repeated until the user is satisfied with the type of results being returned meets their needs

Setting 2: The user modifies the search agent to only return results where a named entity of a particular type matches the agent's search criteria, e.g. find information about a particular company, in a specified location, that does not mention the CEO of that company. The results summaries from the agent are displayed in a list. Each result summary includes information such as the title of article, its author, publication, and publication date. Named entities that have been detected within the text are highlighted. The results are displayed in a list, with the document summary showing the main entities that have been detected within the text. The user provides feedback on the results, thereby modifying the agents query. The search agent searches for further information at its next scheduled run time.

3.7.3 Other scenarios/functions to consider

The following scenarios and functions should also be considered when designing the personal search agent:

1. The user modifies their personal search agent to query a WWW-based search engine for information.
2. Queries specified by the users' agents and the results of those queries can be used to enhance the indexing and extraction process for the digital library's focused crawlers. This will enable the crawlers to find additional information that is closely aligned with the information needs of the users of digital library.
3. Methods of providing feedback to the agents (given that users may not like to give explicit feedback [3]).

3.7.4 Expected benefits

Personal search agents that query the digital library using an attribute-based search of a repository of semantic annotations are expected to find more relevant results than a search agent using purely syntactic queries. The agents will be capable of searching for explicit statements about recognised entities and relations, e.g. search for all articles, published by a specified publisher, since a specified date. Furthermore, just like the attribute-based search, implicit statements can be inferred from the inherent transitivity of some properties in the digital library ontology (an example of this is given in section 3.3.4).

3.7.5 Potential issues

The questionnaire/focus group study analysis [3] found that users do not like to be prompted by systems to give feedback, e.g. on the relevancy of an article to the user's search requirements. Focus group participants expressed that they would be reluctant

to give prompted feedback, e.g. in a search application. Similar beliefs (and actions) could apply to providing feedback to a search agent application. Alternative, implicit methods of obtaining user feedback need to be investigated.

3.7.6 Supporting technology and functions

The following technologies are required to support the personal search agent:

1. API to a platform that provides an attribute/entity-based search of a repository of semantic annotations.
2. An attribute/named entity-based agent interface (the attributes will be based upon the selected resource type).
3. Ontology-based query expansion based upon topics selected from the digital library topic ontology.
4. Relevance feedback algorithms and functions (explicit and implicit).
5. Alternative visualisations of the search agent interface and results, e.g. present the user with interfaces appropriate to the device being used and any specified user preferences.

3.7.7 Supporting data

The personal search agent will be supported by the following data:

1. A profile for holding information about the agent, e.g. title, description, query, date of first search, frequency of search, and number of results to be found.
2. Proton and the BT digital library domain-specific ontology.
3. A knowledge base that contains supplementary data about named entities, e.g. people, companies, and places.
4. A visualisation ontology that will be used for translating the machine oriented data of the agent into a form suitable for presentation to the user.

3.7.8 Relevant SEKT tasks

The following SEKT tasks will support the personal search agent application:

WP/Task number	System/Component/Function
T1.1, T1.2	Algorithms/functions to calculate the semantic similarity between the agent queries and the feature vectors that represent the resources in the digital library.
T2.6, T6.3, T6.4.	An API to the underlying semantic annotation platform and knowledge base, or an API to the SEKT integrated platform. Named entity recognition functions will be integrated into the annotation platform.
T1.8, T11.2	Proton ontology and BT digital library domain-specific ontology.
T5.2	Semantically enabled search engine (or API) combining free-text search with capability to exploit semantic metadata, e.g. ontology-based query expansion, named-entity recognition. Attribute-based search (similar requirements to search and browse application).
T5.4	Visualisation component that takes the search agent interface and results and transforms them (via a visualisation ontology) into a form suitable for presentation to the user.
T5.7	Content purposing component to transform the agent's user interfaces and results into format suitable for display on alternative devices, e.g. PDAs.

Table 3.5. SEKT tasks that support the personal search agent application.

3.7.9 Example screenshots

An example screenshot showing the agent configuration interface is shown in figure *3.10.*

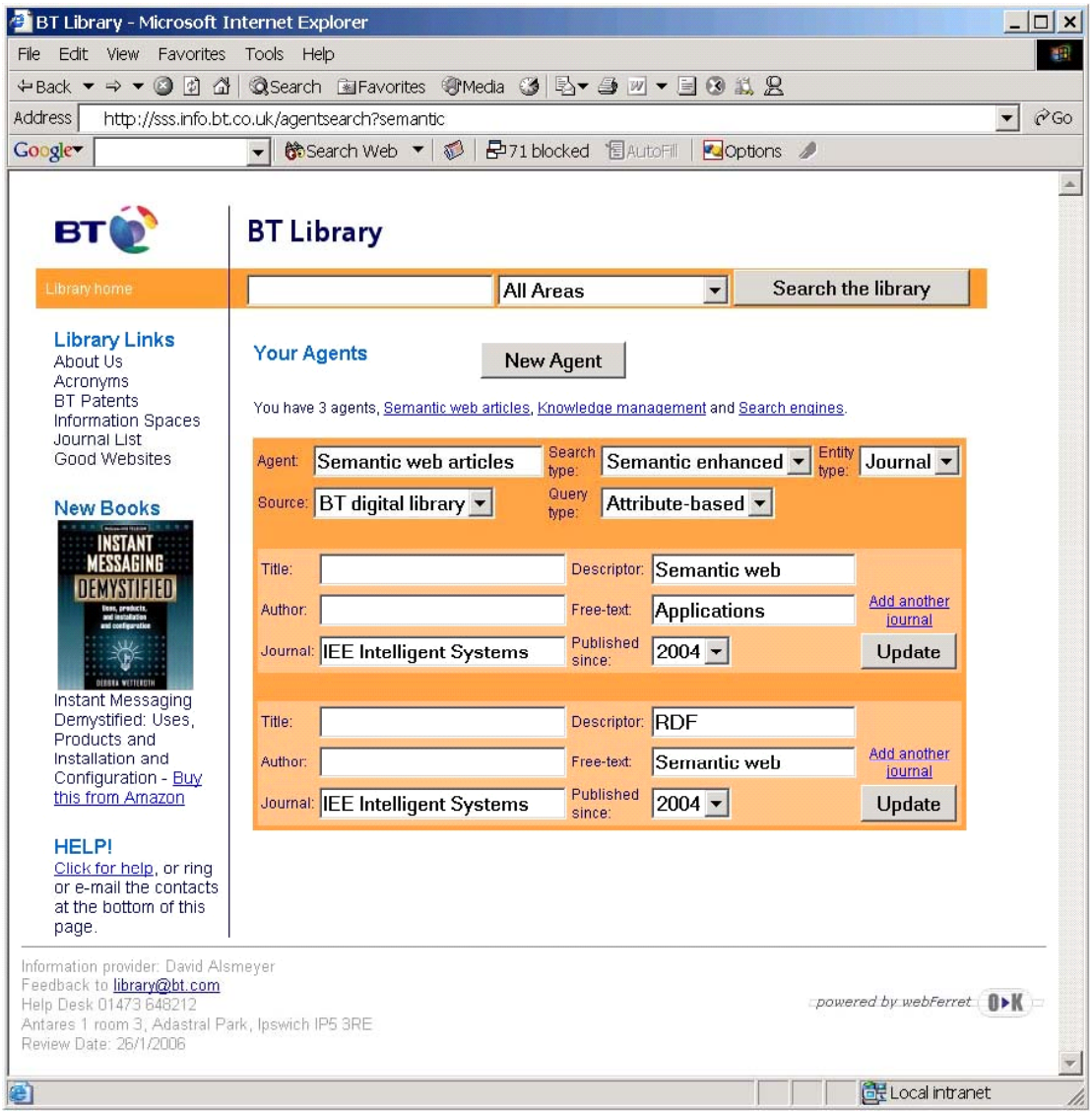

Figure 3.10. Search agent configuration interface.

3.8 End user use case 6: Personal information-based content delivery

3.8.1 Description

Technology has enabled people to become more mobile in their work; people often spend long periods of time out of their base office environment. Users need to be kept up-to-date with information that is relevant to their work. Not only do users need to be kept abreast of events within their own organisation but they may also require external information as well, such as breaking news on a client they are about to visit.

To avoid information overload, data such as emails or news items need to be prioritised and filtered by relevance to the individual's current tasks (or events in their diaries). Similarly, information in the knowledge base (e.g. digital library, newsfeeds)

should be analysed in context with a user's diary events, in particular where a *named entity* is recorded. Users should be alerted to significant information that is relevant to the *named entity*. Furthermore, people use a variety of mobile devices whilst out of the office. Information must be delivered in an appropriate format to the device that they are using and any preferences they have set.

3.8.2 Application scenarios

3.8.3 Scenario 1: personal content delivery to a mobile user

Intention of scenario/setting: This scenario describes how content relevant to personal information, such as that held in a user's Outlook Calendar, is delivered to a mobile user.

A BT worker is constantly on the move and carries a mobile device, e.g. a mobile phone or a PDA. The worker needs to be constantly updated with information that is relevant to their work, in particular with the latest information concerning their customers. Some of this information is internal to BT, e.g. business magazine articles held in the digital library, whilst other information is external to BT, e.g. information about competitors taken from external news feeds.

In this example scenario, the BT worker has scheduled a meeting for two days time with an important customer, Mr. Johnson of the (fictitious) company Rapier Pharmaceuticals. As the user will be away from their normal office environment, they wish to be updated with any breaking news items or any new business magazine articles in the digital library that mention their customers.

Expectations: The user expects information to be delivered in a form that is appropriate to the device being used, and that takes into account the bandwidth available to deliver the information. The user expects information from all sources to be prioritised, based on their preferences set in their user profile. Furthermore, information updates must be related to scheduled meetings and actions in their Microsoft Outlook Calendar and task list.

Pre-conditions: The user has previously enabled access to their Microsoft Outlook Calendar, and has set personal content delivery notifications to be visible whenever they log-in to BT's Intranet. Furthermore, the user has also specified that, whilst connected to the Intranet, they would like to be alerted about breaking news concerning companies with which they have scheduled meetings over the next 5 days. The personal information delivery application has extracted information, such as the named entities Rapier Pharmaceuticals, and Mr. Johnson, from the user's Outlook Calendar. The user is currently out of the office travelling. The user will be using a PDA/phone device to access applications on BT's Intranet at various points during the day.

1. The user uses their mobile device to log-in to BT's Intranet.
2. On completion of the login sequence, the user is taken to a web page that lists, in chronological order, their latest personal content notifications from the digital library and external news feeds (the external news feed is processed by the digital library, e.g. named entities are extracted the news items are annotated and indexed).

3. A summary of the content of each item is displayed, e.g. for a news item, this will be its title, the original news source, date/time of news item, a summary of the news item, and a link to its full text.
4. The user browses the full-text of some news items, and reads some articles that were delivered over the previous few days.
5. The user spends some time using other applications.
6. A news item breaks that is related to a named entity that had been extracted from one of the user's diary entries. In this example scenario, a statement has just announced its new CEO is published on the news feed.
7. The digital library retrieves and processes the news items from the news feed (including the statement about Rapier Pharmaceuticals); all named entities of type person and of type company are extracted, the news items are annotated, and the digital library index is updated.
8. The personal content delivery application sends notifications to all users who have expressed an interest in Rapier Pharmaceuticals or to users where Rapier Pharmaceuticals has been extracted from their Outlook Calendar items. The notification contains key information about the news item, e.g. its title, the original source of information, date/time of the news item, and a summary of the text. This information is pushed to the top of the user's personal information alerting page.

Post-conditions: The user's notifications page is updated with the latest relevant news items. The digital library index is updated, enabling users or personal search agents to search for the news item.

Setting 2: The user is interested in collecting background reading material, particularly articles published in digital library's business magazines that are related to events planned in their Outlook Calendar. In this example scenario the user is scheduled to attend a conference on Bank Security systems.

Expectations: the user expects to be provided with references to articles in the digital library that are relevant to events planned in their Outlook Calendar. The user expects the application to make references to relevant articles available in their personal work space.

Pre-conditions: The user has previously enabled the personal content application to access to their Microsoft Outlook Calendar. The user has set their personal content delivery agent to find references to introductory-level articles that have been published in specific business magazines of the digital library since a specified date, and which are of relevance to events the user is scheduled to attend in the next two months. The user has indicated that these references should be delivered to their personal work space.

1. The user makes a new entry in their Outlook Calendar giving outline details of a conference on Bank Security systems. The user includes the URL for the conference.
2. The personal content delivery application is notified of the new Calendar entry.
3. The application extracts information, including named entities, from the Calendar entry and the conference web page. The information extracted from the

conference web page is matched against topics in the digital library topic ontology.

4. The application queries the digital library index for introductory-level articles that have been published in specific business magazines (pre-selected by the user).
5. The results of the search are delivered to the user's personal work space.

Post-conditions: Relevant information is made available in the user's personal work space. This work space can be accessed using a variety of devices, the content being purposed for optimum display on that device.

Setting 3: The user is interested in gaining further information about named entities that are of importance to them, i.e. named entities that occur in their Outlook Calendar. Whenever those entities are mentioned in news items, or are present in the summaries of digital library articles, the user would like to be presented with supplementary information about that entity extracted from the digital library knowledge base.

Expectations: The user expects named entities of interest, i.e. named entities that appear in their Outlook Calendar, to be highlighted in news items and in the summaries of digital library articles.

Pre-conditions: The user has previously enabled the personal content application to access to their Microsoft Outlook Calendar. The user has set their personal content delivery agent to identify named entities in their Outlook Calendar.

1. The user makes a new entry in their Outlook Calendar, giving details of a planned meeting with a customer.
2. The personal content delivery application is notified of the new Calendar entry.
3. The application extracts details of named entities from the Calendar entry and updates the user's profile.
4. When the users accesses news items (previously processed by the digital library) or content that mention the entities held in their user profile, those entities will be highlighted and linked to supplementary information in the knowledge base.

Post-conditions: The user's profile is updated with details of named entities that occur in their Outlook Calendar.

3.8.4 Other scenarios/functions to consider

Configuration of the personal content delivery service.

3.8.5 Expected benefits

Mobile users are kept up to date with the latest information, e.g. news items that are relevant to their customers. Relevant content is delivered based on information that is contained in the tools that users tend use every day. Content is therefore delivered as a result of common work practices, e.g. schedule a meeting with a customer, rather than as a separate search activity.

3.8.6 Potential issues

Many Calendar entries are likely to be very brief, and may contain acronyms, abbreviation or errors. It may be difficult to derive anything meaningful from such a

sparse source of information if taken alone[21]. Information stored in the user's profile will be used to supplement the information contained in the user's Outlook Calendar items.

Not all news items that mention a certain named entity are likely to be interesting for the user, e.g. a named entity such as "Rapier Pharmaceuticals" i) might not be the main topic of that news item. Alternatively the context in which the named entity is mentioned, i.e. the event it is involved in, may not be of interest to the user. The user should be given the possibility to specify in greater detail the kind of news they are interested in.

3.8.7 Supporting technology and functions

The following technologies are required to support the personal information-based content delivery application:

1. An API to user profile-enhanced search and browse application/search engine to enable an attribute-based search.
2. Components to monitor and capture the user's Calendar entries.
3. Named entity recognition (enables companies, people, places, etc to be identified in news feeds).
4. A notification service that enables a user to select the preferred method of delivery for an update notification
5. Access to user's Outlook Calendar, Notes and Tasks.
6. A personal work space application.
7. Algorithms/functions to calculate the semantic similarity between information extracted from Calendar entries and digital library resources and news items.
8. Functions to manage a user profile (the user may wish to activate/deactivate named-entities that have been extracted form their Calendar).
9. Components for producing alternative visualisations of digital library content, e.g. present the user with content appropriate to the device being used and any specified user preferences.

3.8.8 Supporting data

The personal information-based content delivery application will be supported by the following data:

1. Articles/postings from appropriate news feeds (which will be processed by the digital library, e.g. to extract named entities, update the digital library index, etc.
2. Proton and the BT digital library domain-specific ontology.
3. A user's Outlook Calendar, Notes and Tasks.
4. A visualisation ontology that will be used for translating the machine oriented data of the agent into a form suitable for presentation to the user.

[21] One of this document's contributors uses the acronym PUR in their Outlook Calenadar; this entry has no connection with PUR Technology.

3.8.9 Relevant SEKT tasks

The following SEKT tasks will support the personal networking application:

WP/Task number	System/Component/Function
T2.1, T.2.6	Information extraction/classification and named entity extraction technology/algorithms for extracting meaningful content from Outlook items, and establishing a measure of semantic similarity to the indexed feature vectors that represent the resources in the digital library.
T2.6, T6.3, T6.4.	API to underlying semantic annotation platform and knowledge base, or SEKT integrated platform (named entity recognition functions will be integrated into the annotation platform).
T1.8, T11.2	The Proton ontology and BT digital library domain-specific ontology.
T5.2	Semantically enabled search engine (or API) combining free-text search with capability to exploit semantic metadata, e.g. ontology/profile-based query expansion, named-entity recognition.
T5.4	Visualisation component that transforms the configuration interface and results (via a visualisation ontology) into a form suitable for presentation to the user.
T5.7	Content purposing component to transform the application's interfaces and results into format suitable for display on alternative devices, e.g. PDAs.

Table 3.6. SEKT tasks that support the personal networking application.

3.9 End user use case 7: Profile management

3.9.1 Description

The user profile is common to a number of applications, e.g. search and browse, and the 'Jot' annotation tool. The applications will make profile update recommendations to the user as they interact with the system, e.g. suggest membership of relevant information spaces based on content accessed, or suggest that a user add a new topic of interest to their profile during a search. Ultimately, the user must have control of their profile. Users are therefore able to accept, reject or modify recommendations made by the various applications. This use case describes how a user will manage their user profile.

3.9.2 Application scenarios

3.9.2.1 Scenario 1: updating a user profile

Intention of scenario/setting: This scenario describes how a user will update their profile manually[22].

Expectations: The user expects to have complete control over their profile, e.g. leave named-entity associations active, but disable some of their topics of interest. The user expects to be able to do this on a per application basis, i.e. a topic of interest could be active for search and browse, yet be marked as inactive for the knowledge sharing application.

[22] In cases where the system recommends a profile update, the reasons for that recommendation will be given to the user.

Pre-conditions: The user has a number of topics of interest enabled in their profile. The user has opened a web browser and navigated to the digital library home page.

1. The user clicks the 'Manage profile' link to gain access to their profile.
2. The user is presented with a page that gives access to the various parts of their profile, e.g. topics of interest, public information space membership, alerting preferences, named entity associations, personal details, and device profile.
3. The user selects the 'topics of interest' element of their profile.
4. The system displays the list of digital library ontology topics that have been added to the user's profile (an example screenshot is shown in section 3.9.9). For each topic of interest, an active or inactive indication is given against each applicable application. A description of each topic is displayed when the user positions the mouse pointer over the topic name.
5. The user activates some previously used topics and disables some currently used topics. The user is also given the options to browse and add topics from the ontology, to remove topics from their profile completely, or initiate a keyword based search (in which case the system tries to match the supplied keywords against the ontology).
6. The user clicks the public information space membership part of their profile. A list of information space memberships is displayed in a similar manner to that of the topics of interest, i.e. some memberships are active, whilst other memberships are inactive (but still retained in their user profile). The user is also given options to subscribe to new public information spaces (please refer to the information spaces use case scenario- section 3.4.2.1).
7. The user adjusts other parts of their profile in a similar way.

Post-conditions: The user profile is updated and is used by the digital library applications.

3.9.3 Other scenarios/functions to consider

The applications will make profile update recommendations to the user as they interact with the system, e.g. suggest membership of relevant information spaces based on content accessed, or suggest that a user add a new topic of interest to their profile during a search. Users can accept, reject or modify the recommendations made by the various applications. Further scenarios should consider the way in which the user will interact with the system for various forms of recommendation.

3.9.4 Expected benefits

Although it was recognised that there are benefits to be gained through the system advising and helping users, the user is given overriding choice and control over their profile. The user has a single point of access from which they can modify their profile. Furthermore, common elements of the profile are shared by the digital library applications, giving consistency of use.

3.9.5 Potential issues

The user's interests will change over time as the user interacts with the digital library and browses other web-based content. Without careful management, profiles can become very general, particularly if the user has a wide range of interests and has

many interests enabled. A poorly managed profile is likely to affect the performance of the applications adversely, e.g. a user's search query augmented with a wide range of topics is likely to result in an excessive number of irrelevant search results being retrieved. Some topics will be of more significance to the user than others. This needs to be reflected in the user's profile. In some cases the user's most recent search behaviour will be of more relevance. In other cases the user's past behaviour could be of more significance. Some topics may be very significant to a user's search, whilst others are less significant (but not irrelevant). The system therefore needs to be able to establish the user's current search context very quickly in order to prevent a user's queries from being augmented with inappropriate topic information (which is likely to result in poor search precision, which may in turn stop users from using their profile to their best advantage; they may then choose to deactivate all topics in their profile). System recommendations must therefore be presented in a way that makes the user aware that their profile needs to be updated, but in doing so must not irritate the user.

3.9.6 Supporting technology and functions

The following technologies will support the profile management tool:

1. Components for producing alternative visualisations and access to the profile management tool, e.g. present the user with interfaces appropriate to the device being used and any specified user preferences.

3.9.7 Supporting data

The profile management tool will be supported by the following data:

1. The Proton ontology and the BT digital library domain-specific ontology.
2. BT digital library user profiles.
1. A visualisation ontology that will be used to translate the machine oriented data of the digital library into a form suitable for presentation to the user.

3.9.8 Relevant SEKT tasks

The following SEKT tasks will support the profile modification functions:

WP/Task number	System/Component/Function
T1.8, T11.2	Proton ontology and BT digital library domain ontology.
T5.4	Visualisation component that takes the search agent interface and results and transforms them (via a visualisation ontology) into a form suitable for presentation to the user.
T5.7	Content purposing component to transform the agent's user interfaces and results into format suitable for display on alternative devices, e.g. PDAs.

Table 3.7. SEKT tasks that support the profile modification functions.

3.9.9 Example screenshots

An example screenshot of the profile management interface is shown in figure 3.11.

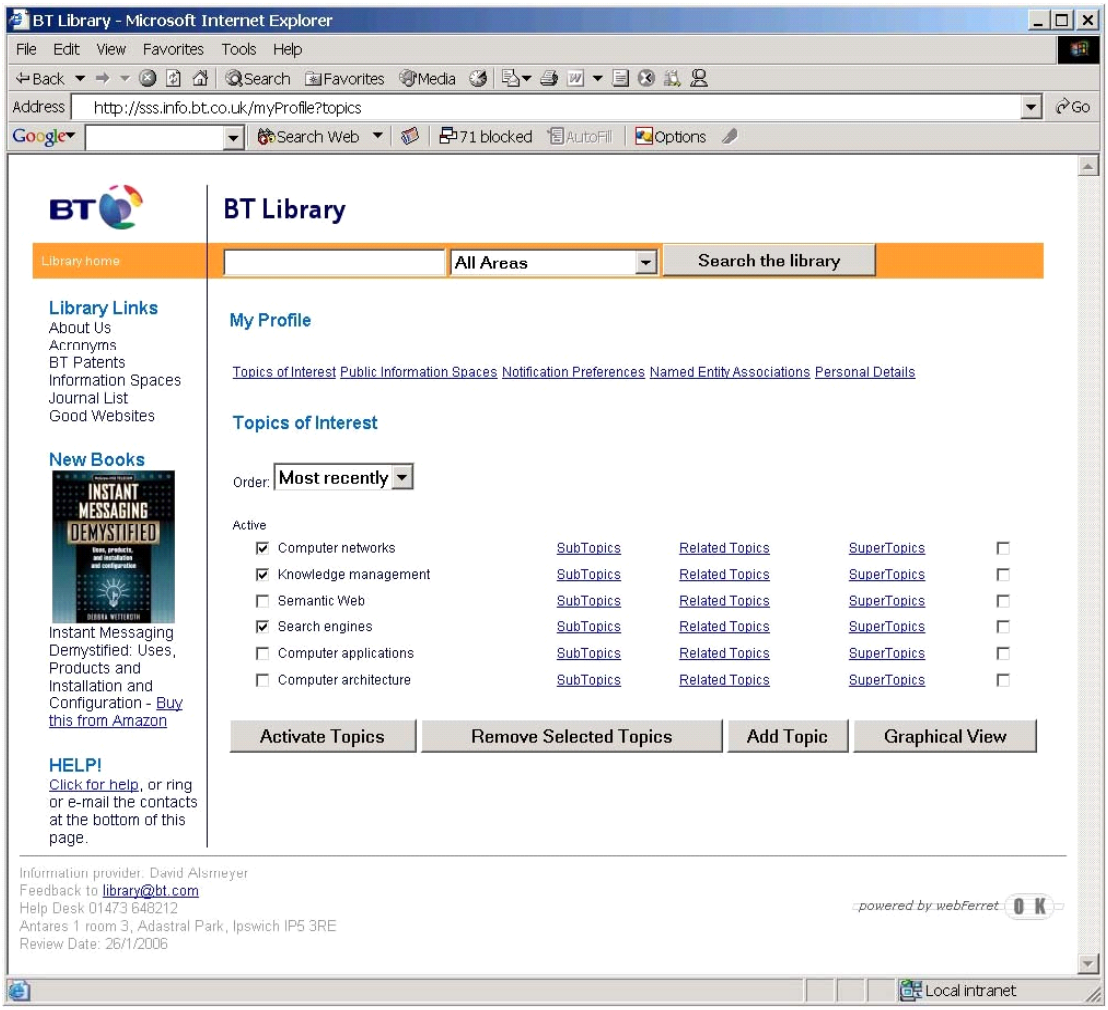

Figure 3.11. Example profile management interface.

3.10 End user use case 8: Notification

3.10.1 Description

A number of the BT digital library tools send notifications to user, e.g. the search agent application. Users have different requirements for receiving these updates; some users want to be updated immediately, whilst others prefer to schedule a single notification that aggregates alerts from multiple applications at a convenient point (consolidate notifications according to relevance and importance). This use case describes the user's interaction with the notification service.

3.10.2 Application scenarios

3.10.2.1 Scenario 1: configuring the alerting service

Intention of scenario/setting: This scenario describes how a user will set their notification preferences for the digital library applications.

Expectations: The user expects to be able to specify the way in which notifications are delivered, and be able to set preferences for the time they wish to receive those notifications.

Pre-conditions: The user has opened a web browser, navigated to the digital library home page, and from there clicked the link to access their notifications. The user has previously set a number of alerting preferences.

1. The user is given the option to edit their preferences on a per application basis. The following options are given: a) public information spaces (if subscribed to); b) private information spaces (if configured), c) personal search agents (if configured), d) expertise location, and e) personal information-based content delivery (if enabled).
2. The user selects the option to modify the way in which they will be notified of results from their personal search agents.
3. The notification delivery method and schedule is shown for each agent.
4. The user is given the options to select the form of delivery, e.g. via email or on completion of login to the digital library. The user is given the option to select the time of delivery, e.g. on completion of search, or at a specified time. The user is also given the option to a) turn-off notifications, b) group notifications from multiple agents, or c) group the results of invocations of the same agent into a single notification (an example screenshot is shown in section 3.10.9).
5. The user selects various options for each agent and updates their notification schedule.
6. Similar functions are available for other applications.

Post-conditions: Notifications are sent as scheduled by the user.

3.10.3 Other scenarios/functions to consider

Consider the batching of alerts from multiple applications.

3.10.4 Expected benefits

The notification application meets many of the requirements that were expressed by the focus group [3], e.g. it gives users control over how and when e-mail notifications are distributed, and enables notifications to be sent out in bulk, and not necessarily as and when updates to the digital library occur. The user has a single point from which they can configure all their alerting preferences.

3.10.5 Potential issues

The notification tool requires careful management. Users could be overwhelmed with alerts from the digital library if the notifications are not managed properly. If users receive an excessive number of alerts, they may decide to turn off the notifications completely.

3.10.6 Supporting technology and functions

The following technologies will support the profile management tool:

1. Components for producing alternative visualisations and giving access to the notification component, e.g. present the user with interfaces appropriate to the device being used and any specified user preferences.
2. The BT digital library applications, e.g. personal search agents, will need to provide an interface to the alerting service.

3.10.7 Supporting data

The notification service will be supported by the following data objects:

1. A notification schedule for each application. Where appropriate, each application notification schedule will be sub-divided into schedules that are associated with different invocations of the tool or application, e.g. to enable different forms of notification to be associated with individual search agents.
2. The Proton ontology and the BT digital library domain-specific ontology.
3. A visualisation ontology that will be used to translate the machine oriented data into a form suitable for presentation to the user.

3.10.8 Relevant SEKT tasks

The following SEKT deliverables/tasks will support the alerting functions:

WP/Task number	System/Component/Function
T1.8, T11.2	Proton ontology and BT digital library domain ontology.
T5.4	Visualisation component that takes the search agent interface and results and transforms them (via a visualisation ontology) into a form suitable for presentation to the user.
T5.7	Content purposing component to transform the agent's user interfaces and results into format suitable for display on alternative devices, e.g. PDAs.

Table 3.8. SEKT tasks that support the alerting functions.

3.10.9 Example screenshots

An example screenshot of the interface to the notification service is shown in figure 3.12.

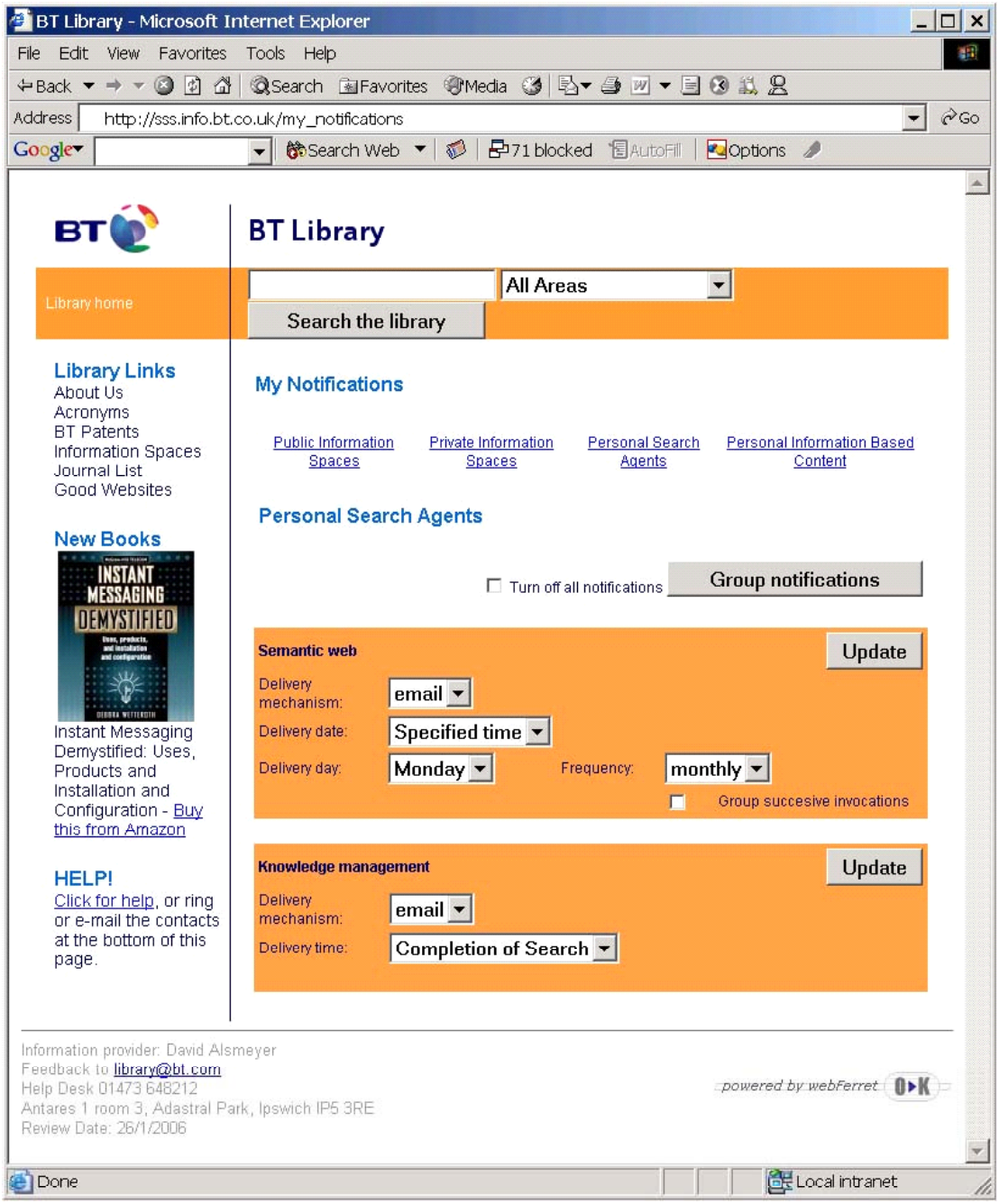

Figure 3.12. Example interface to the notification service.

4 Domain expert and system administrator use cases

4.1 Requirements

The requirements for the domain expert and system administrator use cases and scenarios have been derived from:

1. The overall aims and objectives of the SEKT project.
2. A requirements capture activity for the BT digital library, comprising an analysis of the responses given to a questionnaire on the usage of the digital library, and a focus group study [3].

4.2 Domain expert and system administrator scenarios and use cases

Scenarios and use cases relating to the following domain expert and system administrator tools and applications are described:

1. Domain expert use case 1: Focused crawling (p. 65);
2. Domain expert use case 2: Information space configuration tool (p. 68);
3. Administrator use case 1: Instance update tool (p. 70); and
4. Domain expert and system administrator use case 1: Ontology extension and merging tool (p. 73);

4.3 Domain expert use case 1: Focused crawling

4.3.1 Description

One of the key requirements for the digital library is to extend the content of the library with relevant information from the WWW, e.g. research papers, articles on business innovation, and new publications. Focused crawlers will selectively retrieve information of relevance from WWW sources, extract semantic information, e.g. named entities, annotate the content, and populate the appropriate parts of digital library ontology. Furthermore, a domain expert can associate the focused crawler with a public information space, thereby supplementing that information space with content from the WWW.

4.3.2 Application scenarios

4.3.2.1 Scenario 1: configuring a focused crawler

Intention of scenario/setting: This scenario describes how a domain expert can configure a focused crawler to retrieve information on a particular topic area and associate that crawler with an information space.

Expectations: The domain expert expects the crawler to retrieve information that is highly relevant to a topic of interest or an information space.

Pre-conditions: The domain expert has been asked to set up a focused crawler to retrieve information from the WWW on the topic ‘knowledge management’; more specifically in the area of ‘knowledge representation’. The ‘knowledge management’ information space and a number of sub-spaces (sub-topics) have already been configured, and are alerting users to Inspec and ABI database updates. The domain expert has a good knowledge of the ‘knowledge management’ domain and has used a

WWW search engine to find some WWW pages that will provide good starting points for the crawler.

The domain expert has opened a web browser, navigated to the BT digital library home page, and from there to the domain administrator functions (only domain administrators get visibility of these functions).

1. The domain expert selects a function to create a new focused crawler.
2. The system presents an interface that enables the domain expert to specify a number of crawl starting points, and the depth and breadth of the crawl from those starting points. The system also prompts the user to give a friendly name to the crawler.
3. The system prompts the domain expert to schedule the date and time for the first crawl, and the frequency of subsequent crawls for each starting point.
4. The domain expert provides the information required.
5. The system prompts the user to associate the crawler with an information space.
6. The domain expert selects the 'knowledge management' information space.
7. The system prompts the user to select a sub-topic of the 'knowledge management' information space. All sub-topics that already have associated information spaces and crawlers are identified with appropriate icons, which when clicked, take the domain expert to the crawler configuration page for that sub-topic or the information space.
8. The domain expert browses the ontology and positions the mouse pointer over the sub-topic 'knowledge engineering'. A pop-up box shows the sub-topics of the 'knowledge engineering' topic, one of which is 'knowledge representation'. The 'knowledge representation' topic already has an associated public information space (it is a sub-space of the 'knowledge management' space). The user selects this topic.
9. The system prompts the user to associate further topics from the ontology with that crawler. These topics and their associated topic descriptors will be used to determine the relevancy of the pages returned by the crawler. Pages of sufficient relevancy will be processed further, e.g. to identify and extract named entities. Web pages will be discarded if they that are considered not to be sufficiently relevant.
10. The domain expert selects relevant topics from the ontology and saves the configuration of the crawler.

Post-conditions: The crawler starts its crawl from the starting points at the scheduled times. In essence, the system operates as a typical web crawler, but with the added stage of automatically identifying and extracting meta-data, e.g. named entities, and annotating local copies of retrieved web pages. New statements are added to the digital library semantic repository as a result of the information extraction process. The crawler executes a simple algorithm, which only travels down to a configurable depth from the starting URL (the starting point); the crawler processes all links extracted from all sites, for a particular depth, before proceeding to process contents from the next level down.

4.3.3 Other scenarios/functions to consider

The crawler uses an existing search engine index, such as Google, to find the most relevant URLs that link to a specified starting point.

Augment the default stating points with a list of starting points derived from the contents of the local index, e.g. content populated by users' personal search agents (refer to section 3.7).

4.3.4 Expected benefits

The digital library will be supplemented with relevant information from the WWW, e.g. articles related to information spaces. The metadata extracted from the crawled WWW pages will add value, e.g. a user will be able to perform an attribute-based search on that data.

4.3.5 Potential issues

The digital library is a source of high quality information. People can rely on its content as most of the data is derived from databases provided by established content providers (ABI and Inspec). In contrast, a range of low and high quality information exists on the WWW. There is a danger that WWW-based automatic/semi-automatic crawlers will populate the digital library ontology with low-quality information. Furthermore, the quality of metadata extracted automatically from Web content is unlikely to be of as good quality as that provided by ABI and Inspec. Nevertheless, the extracted metadata will be of a sufficient quality to be useful. Not only must the domain expert be confident that high quality information is being found by the crawler, but they must also be confident and that the metadata extracted form WWW content is of an acceptable quality. Measurement of quality of WWW pages retrieved by the crawler, and measurement of quality of metadata extracted from such pages will be crucial to the successful operation of the crawlers. Information retrieved from the WWW will be marked appropriately, and will include information such as the domain from which the page was retrieved.

Integration of components/functions for the prototype tool, e.g. interface the focused crawler with digital library semantic annotation platform or SEKT integrated platform.

4.3.6 Supporting technology and functions

The following technologies are required to support the focused crawler functions:

1. A configurable focused crawler (with published API) that can crawl the WWW for content relevant to public information spaces and topics in the BT digital library topic ontology.
2. Similarity matching functions and algorithms that can calculate the relevancy of a retrieved WWW page compared with the metadata associated with a topic or an information space. Relevant pages are forwarded for further processing, e.g. named entity extraction. The domain expert must be able to configure any relevancy thresholds.
3. Interface to the named entity extraction software (or semantic annotation platform) for identifying the named entities and creating the necessary annotations.

4. Software/interfaces that will enable the domain expert to navigate the BT digital library ontology, e.g. topic areas and public information spaces, and associate the crawler with that topic or information space.
5. Domain experts may be mobile; in which case components for producing alternative visualisations of the information spaces will be required, e.g. present the user with interfaces appropriate to the device being used and any specified user preferences.

4.3.7 Supporting data

The focused crawler software will be supported by the following data:

1. The PROTON ontology and BT digital library domain ontology (specifically the metadata associated with the scope of an information space or topic).
2. Crawler configuration data, e.g. starting points, depth of crawl, frequency of crawl.
3. Access rights, e.g. types of users who are allowed to administer the crawler.
4. A visualisation ontology that will be used to translate the machine oriented data of the digital library into a user-friendly visualisation of the information space.

4.3.8 Relevant SEKT tasks

The following SEKT tasks will support the focused crawler functions:

WP/Task number	System/Component/Function
T1.1, T1.2, T1.3, T1.8.	Analyse content of crawled WWW content, cluster, extract features (e.g. keywords/phrases), and calculate relevance to the features that characterise the information space or topic areas in the BT digital library ontology.
T2.6, T6.3, T6.4.	An API to the underlying semantic annotation platform and knowledge base, or an API to the SEKT integrated platform
T1.8, T11.2	The Proton ontology and BT digital library domain ontology.

Table 4.1. SEKT tasks that support the focused crawler functions.

4.4 Domain expert use case 2: Information space configuration tool

4.4.1 Description

Users can subscribe to public information spaces (refer to section 3.4). The domain expert is responsible for configuring and administrating those information spaces. This use case describes how a domain expert will configure a public information space (this is very similar to the process that a user follows when creating a private information space – section 3.4.2.4).

4.4.2 Application scenarios

4.4.2.1 Scenario 1: configuring a public information space

Intention of scenario/setting: A domain expert has just been asked to configure a public information space for the topic ‘Information retrieval systems’.

Expectations: The domain expert expects to create the public information space as a sub-space of a higher level information space (in the case of ‘Information retrieval systems', this will be the ‘Computer applications’ information space).

Pre-conditions: The domain expert opens a web browser, navigates to the digital library home page, and from there to the public information spaces administration functions. The domain expert has searched (or browsed) the ontology and identified the top-level public information space, which in this example is 'Computer applications'.

1. The domain expert selects the option to create a new public information space.
2. The domain expert is prompted to enter some search criteria for the information space, e.g. source of information, some keywords that best explain the information requirements, a publication date, and the level of readership.
3. The domain expert searches for information in the digital library, e.g. using a search strategy similar to that described in search and browse scenario 3 (section 3.3.2.2).
4. The search engine matches the query against the digital library's index, and returns a set of results.
5. The domain expert refines the search until the results returned from the search engine match what they believe to be needs of the information space.
6. The domain expert stores the query against the new public information space. The system prompts the user to give the information space a friendly name. The system generates a description of the information space (the domain expert can edit the description).
7. The domain expert instigates further queries, and stores selected queries against the public information space.

Post-update conditions: A number of domain expert-specified queries are stored against the new public information space. The information space queries are run whenever the digital library ontology is updated with new information from the Inspec or ABI databases. Users subscribing to the new information space are notified of updates according to their notification preferences.

4.4.3 Other scenarios/functions to be considered

The following scenarios/functions should be given further consideration:

1. Removal of a public information space.
2. Editing a public information space.
3. Associate an information space with an alternative information space. For example, assume that the 'knowledge acquisition' information space has been configured as a sub-space of the more general topic area 'knowledge management'. The subsequent creation of the topic 'knowledge engineering' suggests that the 'knowledge acquisition' information space should now be defined as a sub-space of the 'knowledge engineering' sub-space, and not as a sub-space of the more general 'knowledge management' information space.
4. Management (create/edit/delete/extend) a more-fine grained information space ontology that offers greater levels of detail than the topic ontology alone.

4.4.4 Expected benefits

The public information spaces will be simple to configure and maintain, e.g. when compared to the current method of editing configuration files. Benefits of using the semantically enabled information spaces are given in section 3.4.4.

4.4.5 Potential issues

No major issues are anticipated with the domain expert tools. Potential issues with public information spaces are given in section 3.4.5.

4.4.6 Supporting technology and functions

The following technologies are required to support the requirements of the public information space configuration tool:

1. Software/interfaces that will enable the domain expert to navigate the BT digital library ontology, e.g. topic areas and public information spaces.
2. Search technology and interfaces that will enable the domain expert to initiate and refine a search, and store/edit queries against the information space.
3. Software/interfaces to manage information space configuration data.

4.4.7 Supporting data

The public information space configuration tool will be supported by the following data:

1. The PROTON ontology and BT digital library domain ontology (specifically the metadata associated with the scope of an information space or topic).
2. Public information space configuration data, e.g. starting points, depth of crawl, frequency of crawl.
3. Access rights, e.g. types of user who are allowed to administer the crawler.

4.4.8 Relevant SEKT tasks

The following SEKT tasks will support the information space configuration tool:

WP/Task number	System/Component/Function
T1.8, T11.2	Proton ontology and BT digital library domain ontology.
T5.2	Search and browse tool/interfaces.

Table 4.2. SEKT tasks that support the information space configuration tool.

4.4.9 Issues to be resolved

Integration of components/functions for the prototype tools, e.g. functions to navigate the ontology and search interface.

4.5 Administrator use case 1: Instance update tool

4.5.1 Description

The BT digital library is primarily built upon information contained in the ABI and Inspec databases. The initial BT digital library domain-specific ontology and instance data will be derived from these databases. The instance data will be updated with content from: a) successive weekly updates of the databases, b) from documents

discovered by automated search agents and crawlers, and c) from electronic documents added by people using the knowledge sharing tools. New topic areas are likely to be identified when articles are added by people or the automated crawlers. Topics in the topic segment of the ontology will therefore evolve as new subject matter, not covered by the initial ontology, is exposed. The topics covered by the ontology are likely to extend. Furthermore, existing topics are likely to become more fine-grained.

This use case describes some administration functions that will be undertaken by the administrator of the digital library.

4.5.2 Application scenarios

4.5.2.1 Scenario 1: updating the ABI and Inspec instance data

Intention of scenario/setting: This scenario describes how the administrator will update instance data in the digital library ontology from weekly updates of the ABI and Inspec databases.

Expectations: The administrator expects the process to be as simple as possible, requiring minimal intervention. The administrator needs to be informed of potential errors when transforming the ABI and Inspec data into a format suitable for generating new instances and, where appropriate, linking those instances to existing instances, e.g. the details of an author of a new article may already have been instantiated in the digital library ontology, so the new article needs to be linked to the existing instance of the person through an appropriate relationship.

Pre-conditions: The initial digital library ontology has been built and a number of ABI and Inspec weekly updates have been instantiated. The latest Inspec database update has been transferred onto the digital library server. The administrator opens a web browser, navigates to the digital library home page, and from there to the administrator functions (only administrators get visibility of these functions).

1. The administrator selects the prepare data function. The system processes the Inspec database update.
2. Potential errors or conflicts are reported to the administrator, e.g. where a new article is authored by a person with a relatively common surname, and where the author's affiliation can not be established, there is potential for error when trying to assign that article to the right instance of person (e.g. via the hasAuthor relationship). Moreover, data regarding the author of the article may not be instantiated, in which case a new instance of person needs to be created.
3. The system instantiates all records where the probability of error is below a set threshold.
4. The system notifies the administrator of potential conflicts or errors.
5. The administrator resolves any errors/conflicts manually, and then instantiates the ontology with those records.

Post-conditions: The digital library instance data is added to and updated from the ABI and Inspec databases. The administrator will be required to resolve any errors manually.

4.5.3 Other scenarios/functions to be considered

Update the instance data with metadata extracted from WWW pages collected by focused crawlers, and shared through the knowledge sharing tools (refer to section 4.6).

4.5.4 Expected benefits

The benefits of using the system administrator tool will be negligible (when compared to the existing process of updating the digital library with information from the Inspec and ABI databases). In actual fact, a greater workload is likely to be placed on the administrator, e.g. if they have to resolve a significant number of errors or conflicts.

4.5.5 Potential issues

Fragmenting Inspec or ABI records into an ontology introduces the potential for error. In order to reduce the work of the system administrator, these errors must be kept to a minimum. The software that fragments each record and assigns appropriate parts of that record to existing entities (or creates new entities) will need further analysis. There is potential for introducing errors, e.g. assigning an article to the incorrect author (there are no author identifiers in the Inspec records). For example, are 'Davies', 'N. J. Davies' and 'John Davies' the same author? It is important to establish this when assigning a document to a person with the role of *author*, especially when the author's name has been extracted from an external source such as a Web page, rather than from an Inspec or an ABI record.

The data associated with an instantiated object can change, e.g. change of author affiliation, where the affiliation for a new article is different from their previous affiliation. The system can not rely on author affiliation alone to disambiguate an author, where two authors share a common name.

The initial set of topics for the digital library ontology, and the relationships between those topics, will be derived from the ABI and Inspec thesauri. The initial set of resources will be derived from the ABI and Inspec databases. This approach will map data directly from the databases into the ontology. An alternative approach would wrap the ABI data in one ontology (o1), the Inspec in a further ontology (o2), with data from other sources being wrapped in further ontologies, (o3, o4, etc). A global ontology could then be created, with 'mappings' from the individual ontologies (o1, o2, etc) to/from the global ontology. These mappings can then be used to query the underlying ABI/Inspec data sources (many fields will be common, e.g. author, title, etc).

4.5.6 Supporting technology and functions

The following technology is required to support the requirements of the instance update tool:

1. Software to fragment an Inspec or ABI record, and associate record fragments with existing instances, e.g. assign a new article to an existing instance of type person via a hasAuthor type relationship, or associate the record fragments with new instances, e.g. assign a new article to a new person instance. In cases where authors have a common surname (and initial) it will be necessary to extract features form the Inspec and ABI records (and the full-text) that enable this classification (or disambiguation) to occur. The software must notify the

administrator of potential errors, and enable the administrator to manually add records once the conflict/error has been resolved.

4.5.7 Supporting data

1. The PROTON ontology and BT digital library domain-specific ontology (in particular, the metadata associated with the scope of an information space or topic).
2. A knowledge base containing supplementary information about authors and their affiliations to help resolve ambiguity.
3. Regular ABI and Inspec database updates.

4.5.8 Relevant SEKT tasks

The following SEKT tasks will support the instance update tool:

WP/Task number	System/Component/Function
T1.8, T11.2	Proton ontology and BT digital library domain ontology.
T2.1	Reference disambiguation using context.
T3.1	Components for managing the evolution of the ontology.
WP11	Fragment ABI and Inspec records and assign fragments to the correct instances (or create existing instances)

Table 4.3. SEKT tasks that support the instance update tool.

4.6 Domain expert and system administrator use case 1: Ontology extension and merging tool

4.6.1 Description

For the initial version of the digital library ontology, the topic classification is likely to be based on the metadata provided in the Inspec and ABI records, e.g. for an Inspec record the 'Preferred Terms' (as defined by the Inspec Thesaurus) will be used to classify the document into topics. Abstracts and metadata for new articles or documents acquired through the regular ABI and Inspec database updates (please refer to section 4.5) will be classified into these initial topic areas. Furthermore, metadata extracted from WWW pages, documents, news items, etc, that have been retrieved by search agent and focussed crawler software will also be classified against the digital library ontology. As metadata for new articles is extracted new subject/topic areas are expected to be identified that are not defined in either the Inspec or the ABI thesaurus. The digital library will be capable of (semi-)automatically learning an ontology (or extensions to an ontology) from an additional corpora of documents.

In addition to the evolution of new topic areas, topics may become increasingly associated with other thesaurus-based topics. As a consequence there is a need to provide functions and processes that help identify new topic areas and topic associations, and enable the digital library's topic ontology to be updated accordingly. The process of generating the ontology extensions should be interactive, e.g. domain experts will be presented with the current ontology and the software will suggest new topic areas for discussion. The domain experts should then consider whether the topic should be added to the ontology, where the topic should be added, and identify other related topics. The domain expert will liaise with the system administrator (and possibly other domain experts) to discuss whether the topic and its related terms

should be added to the ontology. When consensus is reached the administrator will modify the ontology accordingly.

As each new edition of the ABI and Inspec thesaurus are released, information regarding new terms and relationships between existing terms will be added to the library's databases. Subsequent ABI and Inspec records will be classified against these new terms. Updates to the digital library topic ontology are likely to occur in the period between thesaurus updates; the topic ontology may evolve significantly. The new thesaurus may or may not contain the topics/terms that have been added to the ontology. As a consequence there is a need to merge the evolving digital library topic ontology with new topics/terms in the latest thesaurus. Additionally, the domain experts should be given the facilities to identify equivalences between the metadata used in the various purchased databases (thesaurus) and the evolved ontology, in particular the topics in the ontology. These equivalences should be reviewed, discussed and corrected, by the domain experts.

4.6.2 Application scenarios

4.6.2.2 Scenario 1: extending the digital library

Intention of scenario/setting: This scenario describes how a domain expert and system administrator will extend the digital library ontology with new topics, e.g. as new topics are identified in content retrieved by focused crawlers.

Expectations: The domain experts expect to be notified of candidate new topic areas.

Pre-conditions: A focused crawler has extracted a number of documents from a WWW site which although relevant to a particular topic area, contain a few common terms that appear to be significant but which have not been associated with that topic area previously (these terms could be candidates for new topics or new topic associations). The system sends an alert to the domain expert responsible for that crawler. The notification gives details of the focused crawler that retrieved the information, the URLs for the pages retrieved by the crawler, and a list of candidate terms.

1. The domain expert receives the notification and logs into the digital library with domain expert privileges.
2. The domain expert browses the content of the notification and accesses the URLs of the pages retrieved by the crawler.
3. Having studied the content of the URLs, the domain expert searches for further information on the candidate topics in the digital library. Some of topics already exist in the digital library ontology, but have not been previously associated with the topic area of the focused crawl.
4. The domain expert uses an external search engine to look for more content that relates the topic of the crawler to the associated topics.
5. The domain expert decides that this is a valid association and sends a notification to the system administrator to make a change to the ontology, giving the reason for the change. Alternatively the domain expert could discuss the proposed changes with other experts in that domain.
6. The system administrator makes the proposed changes to the BT digital library ontology.

Post-conditions: The digital library ontology is updated with the new topic association (relationship).

4.6.3 Other scenarios/functions to be considered

The following scenarios/functions should be considered:

1. Scenarios for identifying completely new topics, i.e. topics not covered by the current digital library ontology.
2. Scenarios for merging the evolving digital library topic ontology with new topics/terms in the latest editions of the ABI and Inspec thesauruses.
3. Maintaining the knowledge base of named-entities and supplementary data (e.g. company information).

4.6.4 Expected benefits

The digital library topic ontology, and the relationships between topics in the ontology, will evolve rather than remain static. WWW content can therefore be classified against this evolving ontology rather than being 'forced' into a static ontology which may not represent the features of that document adequately.

The use of the DILIGENT [7] process for the controlled vocabulary should protect the quality of the topic hierarchy. Additionally DILIGENT will also allow the development of the topic hierarchy from being very closely aligned to the ABI/INFORM and INSPEC thesauruses towards a topic hierarchy that is more tailored to the needs of the BT digital library.

4.6.5 Potential issues

Domain experts may become overwhelmed with notifications to update the ontology, e.g. as candidate topics or topic relationships are identified.

4.6.6 Supporting technology and functions

The following technology is required to support the instance update tool:

1. Focused crawler software to retrieve relevant content from the WWW (refer to section 4.3).
2. A component that detects new topics (or topic associations) with an existing topic area.
3. A notification component that notifies the domain expert of new topics or topic associations.

4.6.7 Supporting data

1. The PROTON ontology and BT digital library domain-specific ontology (specifically the metadata associated with the scope of an information space or topic).
2. Crawler data, e.g. candidate topics and the URLs of WWW pages in which the candidate topics occur.

4.6.8 Relevant SEKT tasks

The following SEKT tasks will support the ontology extension and merging tool:

WP/Task number	System/Component/Function
T1.8, T11.2	Proton ontology and BT digital library domain ontology.
T1.1, T1.2, T1.3, WP11	Provide focused crawler software and candidate topic recognition software.
To be discussed and agreed	Software to identify and resolve new topics in the evolving digital library ontology with new topics in the ABI and Inspec thesauruses.

Table 4.4. SEKT tasks that support the ontology extension and merging tool.

5 BT Digital Library Domain Ontology

5.1 Background

All applications developed for use in the BT digital library case study will make use of a common domain-specific ontology. A preliminary BT digital library ontology has been designed for the purposes of testing some ideas for a semantic search and browse application (WP5). The SEKT PROTON [1] ontology was extended as necessary to give a sufficiently fine-grained representation of the digital library's data. Data from the Inspec and ABI/INFORM databases representing approximately 12000 articles was used to populate the preliminary digital library ontology. Note that this domain ontology will be subject to change as SEKT tools are deployed in the BT digital library case study context.

5.2 BT digital library domain ontology (prototype)

The following classes and properties are used in the preliminary BT digital library ontology to represent people, articles, publications, etc.

Person (PROTON, subclass of Agent) represents a person and as its properties has the following: *hasFirstName* (Literal; Person's first name), *hasSurname* (Literal; Person's surname) and *hasAffiliation* (Affiliation; Affiliation person works for).

Affiliation (DL, subclass of JobPosition) represents person's affiliation, e.g. company for which person works.

Profile (SKULO, subclass of InformationResource) is a class which holds information about a profile.

UserProfile (SKULO, subclasss of Profile) class contains information about user's profile. It has properties *whatsNew* (Literal; Yes/No option for receiving "What's New"), *libraryOnline* (Literal; Yes/No option for receiving "Library Online"), *isSubscribedTo* (InspecRecord; Link to Accession Numbers of Journals user is subscribed to), *hasLog* (UserLog; Log file contains records of last 25 viewed journals sorted by date) and *hasLog2* (UserLog2; Log2 file contains records of all viewed journals sorted by accession number).

User (SKULO, subclasss of Agent) class represents a user and it has properties *hasUserID* (Literal), *hasUserProfile* (UserProfile) and *hasJottings* (Jottings; Link to Jottings).

PublishedMaterial (PROTON, subclass of Document) is a class which represents published material. It has properties *isOnPageNo* (Literal; Page number where PublishedMaterial is, e.g. page number of a Journal where Article is), *hasArticle* (Article; Link between Journal and Article), *datePublished* (Literal*; inherited from Document*), *documentAbstract* (Literal; Original Abstract of the Document - Author's Abstract), *hasSubject* (Topic).

Topic (PROTON, subclass of Abstract) represents a topic and has property *topicName* (Literal; Name of the Topic).

General Term (PROTON, subclass of Abstract) represents a general term.

InspecRecord (DL, subclass of Dataset) is a class that contains information about record in Inspec database. InspecRecord has properties *hasAccessionNo* (Literal ; Accession Number), *hasPublication* (PublishedMaterial; Link between InspecRecord and PublishedMaterial it is describing), *hasUncontrolledIndex* (Literal; Uncontrolled Index - free text), *hasDescriptor* (Descriptor; Information about Inspec Record), *hasClassificationCodes* (Literal; Classification Codes, e.g.

C7820), *hasCoden* (Literal; Journal Codens) and *hasInspecAbstract* (Literal; Abstract in Inspec database – it can be different from Author's Abstract).

Descriptor (DL, subclass of Dataset) is a class that contains information about InspecRecord. It has properties *descriptorTerm* (Literal; Title of the Descriptor) and *descriptorBroaderTerm* (Descriptor; Description of the Descriptor).

ISDate (DL, subclass of Dataset) is a class that contains information about Information Space. It contains list of results to a query (sorted ascending by document's ID) using the property *resultID* (Article; Accession Number).

ISMembers (DL, subclass of Dataset) is a class that contains information about Information Space. It contains list of members (user ID's) subscribed to a certain Information Space – *hasSubscriber* (UserProfile; UserID).

ISList (DL, subclass of Dataset) is a class that contains information about Information Space. It contains information about Information Space ID, UserID, Information Space keyword and full name – properties *hasUserID* (UserProfile; UserID), *hasFullName* (Literal; Full name of the Information Space), *hasID* (Literal; Information Space ID) and *hasKeyword* (GeneralTerm; Information Space keyword).

ISDef (DL, subclass of Dataset) is a class that contains information about Information Space. It contains elements of a query – property *hasQuery* (Literal; Elements of a query).

UserLog (DL, subclass of Dataset) is a class that contains information about journals which user has read/looked at. It has properties *dateAndTime* (Literal; Date and time when Journal was read/looked at, e.g. "Wed Dec 15 13:13 2004"), *database* (Literal; Database where journal is located - a/ABI; i/INSPEC), *complete* (Literal; +/-; depends if the whole article has been seen or just abstract - works only with DL html articles) and *hasInspecRecord* (InspecRecord; Link to Journal Accession No).

Jotting (DL, subclass of Dataset) is a class that contains information about user's jottings. It has properties *Clippings* (Literal), *YourRef* (Literal), *Share* (Literal). It inherits properties from its super class Dataset and from those following are being used: *derivedFromSource* (InformationResource; URL), *description* (Literal; Notes), *hasDate* (Literal; Date jotting was added), *hasSubject* (Topic; Topic title), *isOwnedBy* (Agent; UserID of the user who added the jotting), *title* (Literal; Title of jotting).

6 Conclusions

A number of applications have been proposed for use in the BT digital library case study. The requirements for these applications were based on a requirements capture exercise undertaken in WP11 [3]. Furthermore, some known limitations of the digital library's search engine and information spaces were considered; these limitations provided a further source of requirements.

Example use cases and user scenarios for a suite of semantically enabled applications have been described. Although the use cases and scenarios have been targeted primarily at the end user, use cases and scenarios for the domain expert and the system administrator roles have also been considered (although to a somewhat lesser extent). More coverage has been given to the applications that are expected to be prototyped first, e.g. the search and browse, and knowledge sharing applications.

The expected benefits of using the applications have been explained for each use case. The required technologies and the SEKT tasks and deliverables required to support the applications have been identified. Mock-up screenshots have been presented, where appropriate, to help clarify the use cases. The expected benefits of using the applications have been explained for each use case.

A number of prototype applications are expected to be developed from these use cases and scenarios. Other use cases are likely to emerge as a result of further research and development within the technology-based work packages and as a result of feedback received from user validation of the prototype applications. Moreover, user validation of the use cases and prototypes will establish the functionality that will be integrated into the final set of tools and applications.

7 Appendix A: Outline design of the knowledge sharing application

7.1.1 Envisioned Application Design

This section describes some of the main ideas for the design of the Jot application, in particular the use of an ontology as underlying data structure. The design is described in a scenario, where the personas (users) Daniel and Joe are using the application to create new jottings. The scenario focuses on technical aspects of the application rather than on the user experience as presented in the use cases. Figure 7.1 shows how the topic ontology is extended by the action the users are taking.

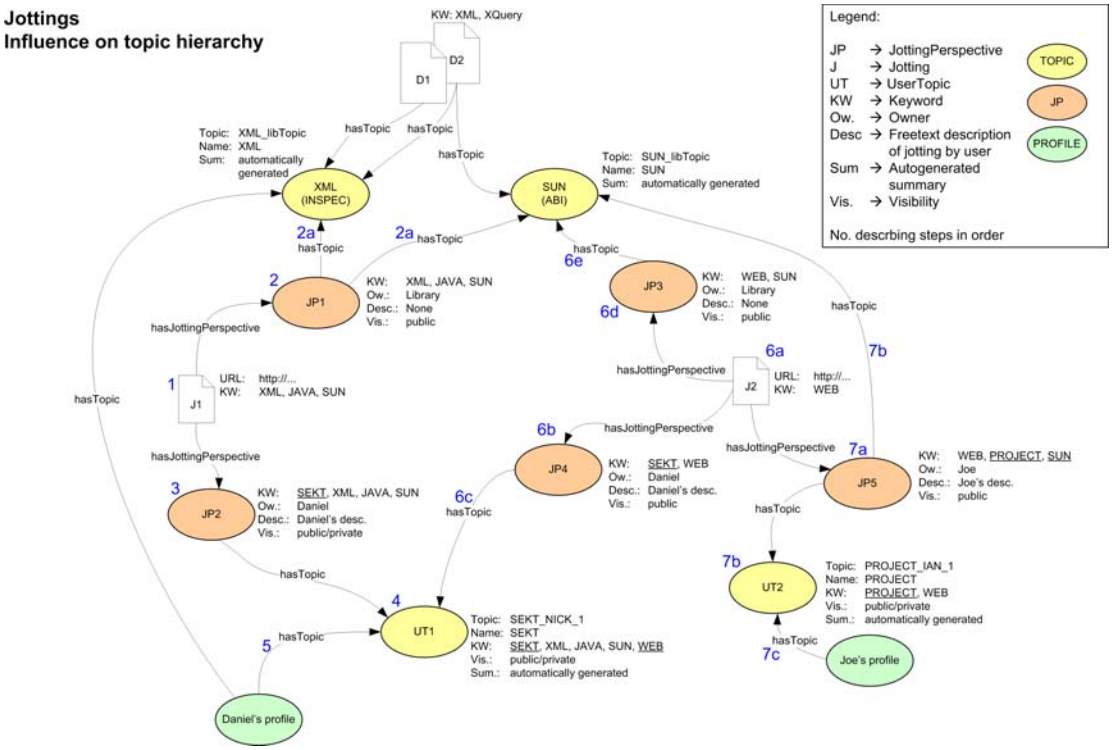

Figure 7.1. Influence of jottings on the topic hierarchy.

The user Daniel creates a jotting with a URL and enters his description. None of the topics suggested by the application fit the jotting, so he creates his own new topic "SEKT". He does not choose to keep the jotting private.

1) The application creates the Jotting (J1) with the URL and extracts the keywords (“XML”, “JAVA”, and “SUN”).

2) The application creates JottingPerspective1 (JP1), using the extracted keywords ("XML", "SUN", "JAVA").
   2a) With these keywords the application maps JP1 to topics in the shared digital library topic ontology to the topic “XML”.

3) The application creates JottingPerspective2 (JP2) using the extracted keywords and the description Daniel entered.

4) The application creates a new topic UserTopic 1 (UT1) with the extracted keywords of J1 using the name “SEKT” Daniel had entered. The keywords for this new topic are now "SEKT","XML", "SUN", and "JAVA". The topic will be displayed with the name "SEKT". If the user didn’t enter a topic description a summary of all documents belonging to the new topics is generated to make a description for the topic available.
5) The new topic "SEKT" is added to Daniel's profile.
6) Daniel finds another web page he wants to remember, so he creates a jotting for this web page, using the topic "SEKT".
   6a) The application creates a new Jotting (J2), including the URL and the extracted keyword ("WEB").
   6b) The application creates JottingPerspective4 with the owner being Daniel, using Daniel’s description, and setting its visibility to public.
   6c) The application aids Daniel to add JottingPerspective 4 (JP4) to the UserTopic 1 (UT1) based on similar keywords found in the web page and in the keyword bucket of UT1. The application adds the keyword "WEB" to UT1.
   6d) The application creates JottingPerspective3 (JP3), with the extracted keyword ("WEB") and the library as owner.
   6e) The application maps JP3 to topics in the shared digital library topic ontology, using the topic “SUN”.
7) The user Joe copies the Jotting J2 from user Daniel. The only thing he changes is the topics describing the Jotting for Joe. He uses his own topic "PROJECT" and the shared topic “SUN”.
   7a) The application creates the new JottingPerspective5 (JP5). The only extracted keyword is "WEB", and the names of the topics entered by Joe "PROJECT" and “SUN”.
   7b) The application creates the new topic UserTopic 2 (UT2). The topic will be displayed with the name "PROJECT" and has the extracted keyword "WEB", and the topic name entered by Joe "PROJECT". Additionally the topic “SUN” is assigned to the jotting, as Joe has told the application to do so.
   7c) The topic UT2 is added to Joe's profile.

Notes: The indirection using a jotting perspective is necessary to allow different views on the same jotting. By allowing different views on the same jotting, it is possible that different users can assign different topics to the same jotting. Therefore different users can keep differing views on the same jotting. The use of the DILIGENT [7] process could be used to merge these differing world views into one, when a new updated shared ontology is distributed (step “local update”).

A keyword bucket for jotting perspectives is not necessary. It might only be helpful for retrieving a jotting, when a user looks for it. Usually this could also be done by using the keyword buckets of topics and jottings.

The descriptions of the topics provide important information for users navigating through the ontology, and assigning correct topics to jottings. This feature should not be underestimated because it provides a simple way for a user to tell which topic they believe would best describe a jotting, thus disambiguating topics. A topic’s name on its own does not give the user sufficient information to disambiguate between topics, especially semantically close topics.

8 References

1. Terziev, I., Kiryakov, A., and Manov, D., SEKT deliverable D1.8.1 Base upper-level ontology (BULO) Guidance,

2. SEKT project brochure, http://www.sekt-project.com/project/SEKT_brochure.pdf, last accessed on 8th February 2005.

3. Alsmeyer, D., Cheung, A., Kings, N., Skinner, F., Thurlow, I. and Warren, P., DBT Digital Library Initial Analysis, July 2004.

4. Cheung, A., D11.2.2 State-of-the-Art Report on the Use of Semantic and Collaborative Technologies in Digital Libraries, December 2004.

5. Davies, J., Kings, N., and Thurlow, I., SEKT deliverable D5.1.1 State-of-the-Art Report on knowledge access, July 2004.

6. Goncalves, M.A., Fox, E.A., Krowne, A., Calado, P., Laender, A.H.F., da Silva, A.S., Ribeiro-Neto, B. : 'The Effectiveness of Automatically Structured Queries in Digital Libraries', *JCDL'04*, June 7–11, 2004, Tucson, Arizona, USA.

7. Sure, Y., Tempich, C., and Vrandecic, D., SEKT deliverable D7.1.1 SEKT Methodology: Initial Guidelines, 2004.