Knowledge Representation for Web Navigation

Simon Jupp[1], Robert Stevens[1], Sean Bechhofer[1], Yeliz Yesilada[1], and Patty Kostkova[2]

[1] School of Computer Science, University of Manchester, Oxford Road, Manchester, UK, M13 9PL.
[2] City eHealth Research Centre, City University, London, UK

Abstract. Representations of domain knowledge range from those that are ontologically formal, semantically rich to those that are ontologically informal and semantically weak. Representations of knowledge are important in many tasks, one of which is the support of *travel* around information spaces through the identification and linking of concepts in a field. In this paper we explore how representations of ontologically informal, semantically weak domain knowledge as captured by the Simple Knowledge Organisation System (SKOS) can enable a system to take advantage of the large number of existing ontological representations to support semantic linking of Web based information and thus facilitate information travel.

1 Background

We present an exploration of how background knowledge of biomedicine can be represented to support the task of semantic linking of documents on the Web. Bioinformatics and related disciplines rely upon Web based resources for information gathering. Information can be gathered from the Web *via* document retrieval using search, and document navigation *via* hypertext links embedded into those documents. Despite the obvious success of the Web, it is not without its limitations. Web users currently rely on search engines to retrieve documents and once retrieved the information is presented for interpretation by humans only. These documents may contain hypertext links to other documents on the Web, but little information is given about the semantics of the link between two resources.

Some of these limitations are being addressed with the use of Semantic Web technologies [13]. The Semantic Web is an extension of the existing Web where information is published in a representation that is interpretable by computers. It is hoped a Semantic Web will provide an infrastructure to improve the way we gather information from the Web.

Two key Semantic Web technologies endorsed by the World Wide Web Consortium (W3C)[3] are the Resource Description Framework (RDF)[4] and the Web

[3] http://www.w3.org/
[4] http://www.w3.org/RDF/

Ontology Language (OWL)[5]. RDF provides a base vocabulary for describing resources and *ad-hoc* relationships between them. OWL is an extension of RDF and provides a vocabulary for building ontologies. Ontologies are used to encode knowledge about a particular domain in the form of the entities and the relationships between them. An ontology language like OWL has well defined semantics, this facilitates computational interpretation of statements expressed in OWL. It is hoped ontologies will provide the content to annotate documents on the web, this *Semantic Markup* provides a mechanism for computers to interpret a document's contents.

Whilst ontologies offer great promise, the library and information sciences have a history of using knowledge artefacts, known as Knowledge Organisation Systems (KOS), to classify and index documents [22], [28], [21]. The KOS provide support for document retrieval and navigation applications. These KOS can vary from simple dictionaries to more complex structures like thesauri and controlled vocabularies, which introduce structure to the knowledge in the form of associative relationships between concepts. The W3C have recently published the Simple Knowledge Organisation System (SKOS)[6], a standard vocabulary for representing KOS like structures. SKOS has a serialisation into RDF that facilitates the use of SKOS in Semantic Web applications. OWL and SKOS may appear similar at first, but have both been developed to fulfill different purposes. Choosing which to use depends largely on application requirements.

The Sealife project[7] seeks to develop a series of browsers in the context of the Semantic Web and Semantic Grid [11]. The Semantic Web/Grid offers an infrastructure for large scale *in silico* science *via* a large number of computational services. The Semantic Web/Grid settings and applications, however, need to be combined with the continuing presence and use of numbers of Web documents describing knowledge about biology. Ontologies and controlled vocabularies potentially provide great benefits for describing and exploring the data in these document resources as well as in the more usual avenues of annotation of data and its subsequent analysis [3, 29]. The Sealife browsers aim to use these vocabularies and ontologies as descriptions of knowledge in the life sciences to flexibly manage the dynamic inter-linking of these documents and services. In this way, a Sealife browser can couple standard modes of Web usage to the emerging Semantic Web/Grid infra-structure.

The work presented here relates specifically to gathering the background knowledge, in the form of ontologies and controlled vocabularies, to support document navigation. We investigated the use of a KOS over a formal ontology in providing background knowledge to support document navigation in a Semantic Web browser like Sealife. The Conceptual Open Hypermedia SErvice (COHSE) application is an implementation of a Sealife browser. We extended COHSE to support SKOS style representation of knowledge; COHSE had previously worked solely with OWL ontologies. We discuss the advantages of using SKOS over OWL

[5] http://www.w3.org/TR/owl-features/
[6] http://www.w3.org/TR/skos-reference/
[7] http://www.biotec.tu-dresden.de/sealife/

in this application scenario and demonstrate the application in travelling the web for resources on infectious disease using a KOS developed in SKOS.

2 Semantic Web Browsing

Navigation *via* hypertext is still the mainstay of the current Web [32]. Yet the author owned and unary links of standard HTML frequently neither offer the link sources nor targets needed by a particular group. The ability to browse documents on the web *via* hyperlinks embedded in text is still a fundamental part of the information gathering process used by biologists. As successful as hypertext is, it is not without its limitations [5]:

Hard Coding: Links are hard coded into the HTML source of a document.
Ownership: Ownership of the page is required to place links in pages.
Legacy: A link target can be deprecated leaving invalid links on pages.
Unary targets: The current web links are restricted to point-to-point linking; there is only one target for a link.

Consider this simple example; A document on the Web contains information about `Polio` and the term `Polio` is marked up as a hypertext-link to another document on the Web. This link is hard coded and is unary to the other document, the target of this link is chosen and controlled by the owner of these documents. If the document had some *Semantic Markup*, from a Knowledge Base (KB) about human diseases, a Semantic Web browser could interpret its content and offer services (or links) relating to `Polio`. By having background knowledge about `Polio`, the browser also identifies `Polio` as having related terms such as `Polio disease`, `Polio virus`, `Polio symptoms`, `Polio immunisation`, and `Polio treatment`, again it could offer appropriate service based on these terms. Having this background knowledge available to the Web browser offers potentially new and beneficial avenues for exploration of information relating to `Polio`.

This example is just one of many potential uses of Semantic Web technologies. Despite its potential, uptake of semantic technologies on the web has been slow due to the large cost associated with developing KBs [12] and the subsequent cost of providing *Semantic Markup* to existing content on the web. The number of KBs in the form of ontologies and vocabularies is growing considerably, this is especially true in medicine and the life sciences [3]. To address the issues of adding *Semantic Markup* manually, a range of semantic web applications emerged that used ontologies to automatically provide this markup dynamically, at browse time. The main goal of these applications was to use ontologies to identify a document's content, and offer appropriate services dynamically. Some notable examples of these early Semantic Web browsers are Magpie [9], KIM [23], Piggy Bank [14], AkTive-document [16], GoPubMed[8] and COHSE [5]. The Sealife project seeks to further this research on Semantic Web

[8] http://www.gopubmed.com/

browser, Sealife has added new functionalities to both GoPubMed and COHSE and investigated the use of Semantic Web browsers with use cases from the life sciences.

3 The COHSE System

Conceptual Open Hypermedia supports the construction of hypertext link structures built using information encoded in ontologies [6]. Dynamic linking, supported by ontologies, offers a mechanism to help overcome some of the restrictions outlined above. The Conceptual Open Hypermedia SErvice (COHSE) system (COHSE is available *via* here[9]) enhances document resources through the addition of hypertext links (see Figure 1). These links are generated based on a mapping between terms found in the document and lexicons available from the ontology. Links can have multiple targets based on the type of concept identified (see Section 6). In addition, the structure of the ontology facilitates navigation to further targets based on sub/super concepts asserted in the ontology. Figure 1 shows a simplified view of the COHSE architecture, the browser interacts with the Knowledge Service and a Resource Manager *via* the COHSE agent and DLS.

The COHSE architecture has been demonstrated in several fields, including the GOHSE [2] system, which was an application in bioinformatics that used the Gene Ontology (GO) [30] as an ontology. GOHSE would, for example, highlight GO terms in a document; the user could then see the GO definition for that term and be offered a menu of links that would perform either a PubMed Search; take the user to the AmiGO browser[10]; browse GO associations[11] *etc.*. The Sealife project is now looking to extend this work and provide the background knowledge that integrates many of the ontologies and other knowledge resources being developed in biomedicine, to aid query by navigation to both scientists and health care professionals in the study of infectious diseases.

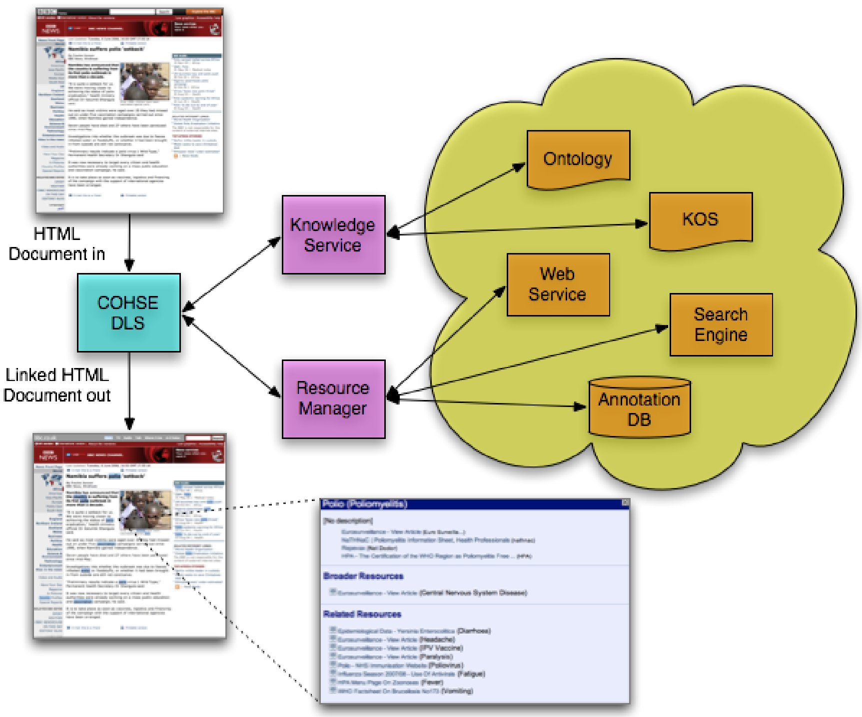
4 Navigation requirements: Ontology or Vocabulary?

Much work has already been published on the requirements for a Semantic Web browser in the context of document navigation on the web [5], [4], [10]. Here we focus on three key requirements of a KB to provide the background knowledge to support the task of navigation in a Semantic Web browser like COHSE:

1. The KB should provide rich lexical support for mapping terms in web documents to terms in the KB.
2. The KB should provide support for representing the relationships between the terms. In particular the ability to generalise or specialise a given term.
3. The KB should be flexible enough to accommodate data from the many other types of KBs that already exist or are under development.

[9] http://cohse.cs.manchester.ac.uk
[10] http://amigo.geneontology.org/cgi-bin/amigo/go.cgi
[11] http://www.geneontology.org/GO.annotation.shtml

Fig. 1. COHSE architecture: Architecture of the COHSE system showing how a plain web document is processed, the DLS uses the Knowledge Service and Resource Manager to add hyperlinks to documents and provide new link targets.



Ontologies, like those represented in OWL, have been used to meet the requirements of navigation in a Semantic Web browser. However, by restricting these applications to ontologies, we are unable to take advantage of semantically weaker structures, like thesauri, that also contain knowledge in structure that have proven useful for navigation in the library sciences. On deeper inspection we believe that semantically weaker structures actually provide a better "fit" for the kind of navigation we want to support with COHSE. This is not to say that we reject the use of formal ontologies or OWL within the COHSE system. Rather, we have introduced an additional level of abstraction into the knowledge model that allows us to implement the underlying knowledge structure using whichever formalism we require (OWL, SKOS, etc.) while presenting a unified interface to the user.

OWL is currently a standard recommendation by the W3C for representing ontologies on the web. OWL provides a language for defining formlised conceptual models. It comes with well defined semantics that enable precise interpretations of these models by logic based reasoners – reasoning can then be used to help build and maintain consistent ontologies. These features are key for the development of a Semantic Web that supports automated machine-to-machine processing of data. OWL enables the data to be expressed explicitly and with little or no representational ambiguity. However, such an approach comes at a cost. OWL ontologies are complicated to build and maintain and require various levels of expertise in not only the ontology's subject matter, but also certain aspects of philosophy and logic [24].

Given the fact that Semantic Web browsers are intended for human interaction with the data, the relationships between concepts in the KB need not be as strict as those represented in OWL. For example, consider navigating documents following the specialisation of some concept. Whilst the strict sub-super class relationships in OWL can support this navigation, the semantics of this relationship are not necessary for the task. Restricting these applications to OWL ontologies alone means that modeling specialisations of concepts, that we as humans may find completely intuitive, becomes difficult or even impossible in OWL. In OWL, a subclass relationship means each instance of the subclass is also an instance of the superclass and as such, is a strong statement about two classes of instances. Specialisation and generalisation relationships need not be describing this particular relationship between instances.

The examples given for *polio* above are a good example of this point. For a further demonstration of this point we can look at a terminology that is commonly used in medicine – the Medical Subject Headings (MeSH) [20]. MeSH is not an ontology, it is the National Library of Medicine's (NLM) controlled vocabulary thesaurus which is used to index articles from many of the worlds leading biomedical journals. It consists of sets of terms organised in hierarchical structures that permits searching at varying levels of specificity. The semantics of the parent/child relationships between terms are not formally defined, and are simply considered broader/narrower. For example, the semantics of `A hasNarrower B` simply means that users interested in `A` might also be interested in the more specialised topic `B`. The MeSH terms found under `accident` include kinds of accidents—as expected *e.g.* `Traffic accidents`, but also `Accident prevention`. This is not a good ontological distinction, but a valid one in the context of navigation and retrieval.

The contrast in semantics means that conversions from MeSH into OWL (with the broader/narrower relationships represented as subclass axioms) are not possible without easily misinterpreting the intended semantics. From a navigation point of view, these weak semantics are acceptable and mean the background knowledge can contain relationships between terms that may be unpalattable to represent ontologically, but perfectly reasonable when we make the semantics of the relationship similar to that of `broader/narrower`. The looser notions of broader/narrower as found in vocabularies or thesauri provide the user with weaker statements amounting to "this entity has something to do with another more specialised/generalised entity". This enables us to create knowledge bases that contain the kinds of relationships between terms, that we as humans find more intuitive in a context of navigation and retrieval.

One could argue at this point that a possible solution would be to simply introduce new vocabulary into OWL that represents the `broader/narrower` relationships, but this introduces a requirement that the application machinery be aware of these, resulting in a non-generic solution. SKOS does provide a recognised standard for representing these relationships often found in KOS (`broader/narrower/related`). Although these relationships may not have the precise semantics that comes with OWL's relationships, in this context, the looser interpretation is more appropriate to the task in hand. By enabling COHSE to work with both OWL and SKOS style representations, we lose nothing and benefit from being able to introduce a wider range of knowledge bases that do not convert into OWL ontologies in a straight-forward manner.

4.1 The Simple Knowledge Organisation System

The Simple Knowledge Organisation System (SKOS) is a proposed standard for representing and publishing classification schemes, thesauri, taxonomies and subject heading systems on the web. It is currently under development as part of the W3C Semantic Web Deployment Group (SWDWG) and a SKOS specification has been published as a W3C last call [12].

The SKOS model can be used to structure and represent knowledge artefacts that contain statements about concepts and the relationships between them. Thesauri, taxonomies, classification systems, subject heading and other similar artefacts are considered to be different types of concept schemes, and often share many features in common. These features are primarily in the form of a lexical resource along with some semantic relationships between each resource. The semantic relationships between resources are typified by synonym, hypernym, hyponym, antonym, broader, narrower, related and so on.

One of the main goals of SKOS is to provide a simple and extensible model that can be used to express these kinds of relationships between resources. SKOS is designed to be extensible and modular. Central to SKOS is the core vocabulary deemed sufficient to represent the common features found in concept schemes. A concept can be considered any unit of cognitive thought, each concept has a set of properties which include the lexical forms that describe it and its relationship to other concepts. The class `skos:Concept` has instances which are the resources being described in the concept scheme. Each resource will typically have some preferred label (`skos:prefLabel`), some alternative labels which are considered synonyms (`skos:altLabel`) and a definition (`skos:definition`). Concepts are organised into hierarchies using broader-narrower relationships, or linked via associative relationships.

Returning to the requirements for background knowledge in a navigation system, we find that SKOS fulfills all three. SKOS has support for representing a rich set of varied lexical information, which is useful for mapping term in documents to concepts from the SKOS. The core semantic relationships in SKOS allow us to express the informal relationships held between concepts that are

[12] http://www.w3.org/TR/swbp-skos-core-guide/

deemed suitable for navigating a knowledge space. As SKOS provides a minimal set of features to support the task in COHSE, it can be used as a representation for knowledge bases that do not convert readily to OWL.

5 Gathering the background knowledge

For a Sealife browser such as COHSE to be useful across such a diverse subject as biology we need a system to rapidly collect the current available ontology-like resources together, and represent them using standard Semantic Web technologies. The biomedical domain already has a rich collection of vocabularies and ontologies such as MesH [20] and all the many other vocabularies held within UMLS [17], GALEN [26], the OBO ontologies[13], to name but a few. The languages used to represent these resources vary considerably, and can range from simple taxonomy languages through to rich, formal logic based languages such as OWL.

Converting existing terminologies into representation suitable for Semantic Web applications can be difficult. Before SKOS, OWL was deemed the standard for representing knowledge artefacts on the Web. Conversions from the MeSH, GALEN and OBO formats into OWL exist, but this is non-trivial as many of these formats do not have a precisely defined semantics like OWL. Care must be taken when converting to OWL from formats where the semantics are weaker than those found in OWL. In contrast, the weak semantics of SKOS make it possible to convert a wide range of artefacts into SKOS, making the artefacts available to SKOS aware applications. The conversion is potentially *lossy*, especially when converting from OWL to SKOS, but where it is simply the structure that is necessary for the task and not the semantics that can drive sophisticated inference, such loss is acceptable. In addition, of course, with such a conversion, the original artefact still exists in its original representation. Providing both SKOS and OWL representations of knowledge bases is perfectly reasonable and should be encouraged.

Conversions into SKOS already exist for the MeSH thesaurus [31]. For the Sealife project, we produced a range of converters for many representations including the OBO format[14]; OWL; and the vocabularies in the UMLS (The OBO to SKOS converter is available online[15]).

5.1 OBO to SKOS conversion

Many of the ontologies and controlled vocabularies being developed in the life sciences are done through the OBO foundry [1]. The OBO foundry provide a principled set of requirements for building ontologies in the OBO format, they also provide a set of relationships that OBO developers are encouraged to use. These relationships do not have the strong model-theoretic semantics of OWL,

[13] http://obo.sf.net
[14] http://www.geneontology.org/GO.format.shtml
[15] http://www.cs.man.ac.uk/~sjupp/skos/index.html

but they are precise and described in natural language in the Relations Ontology (RO) [27]. The OBO language does not have the same level of expressivity as OWL, but it does benefit from many features that are not available in OWL, such as rich annotation support. Conversion from the OBO format into OWL exist [19] making integration of OBO ontologies with Semantic Web applications possible. There are currently over 54 OBO ontologies being actively developed, providing a great coverage of domain terminology used in biology. Many of these terminologies benefit from having rich labeling support and textual description for each term. Providing a SKOS representation of the OBO ontologies along with an OWL version means more Semantic Web applications that are SKOS aware can benefit from this rich terminological resource. In the use case (Section 6) we also demonstrate how having a SKOS style representation of OBO ontologies enables developers to reuse sections of OBO terminologies when building their own vocabularies in SKOS.

Here we present the outlines for a conversion of the OBO format into SKOS. When converting OBO ontologies into SKOS we can take the relationships defined in the RO and map them to the semantically weaker notions found in SKOS to assert `broader`, `narrower` and `related` relationships between SKOS concepts. Here is an example of the conversion one might make when mapping ontological properties to SKOS properties.

- OBO_REL:part-of -> skos:broader (e.g. finger part_of hand)
- OBO_REL:contains -> skos:narrower (e.g. skull contains brain—ignoring the cabity for now)
- OBO_REL:adjacent-to -> skos:related (e.g. nuclear membrane adjacent-to cytoplasm)

Another advantage when converting properties from ontologies to SKOS is the ability to assert the inverse. Consider an ontology where `Nucleus partOf cell`, from an ontological point of view this implies that each and every Nucleus is `partOf` some `Cell`. The inverse, however, is not true, every Cell does not `hasPart Nucleus`. When converting to a SKOS model we can assert the inverse using the `broader` property to say that `Nucleus` has a broader term called `Cell`, which is reasonable. When navigating around documents about cells, the system could then also provide links to documents about nuclei—users interested in cells are often also interested in nuclei.

6 Use Case Scenario

One example application for the COHSE Sealife browser is to provide dynamic hyper-linking of resources from the National electronic Library of Infection (NeLI) [15] portal (http://www.neli.org.uk) to other related resources on the web. NeLI is a digital library bringing together the best available on-line evidence-based, quality tagged resources on the investigation, treatment, prevention and control of infectious disease. Many documents on the NeLI site contain few, if any, hyperlinks to other resources on the web. It would take a large curational effort and cost to manually mark up and maintain these pages with links to other web resources. In addition to this problem, NeLI has a range of users; different link targets are required based on the kind of user browsing the NeLI site. This example shows how the SKOS representation affords the navigation we need and also shows that building a new vocabulary in SKOS can be done rapidly by reusing portions of a variety of controlled vocabularies that have already been converted into SKOS (e.g. MeSH and OBO).

The ability to identify user groups is important. Users can range from members of the public, molecular biologists, to clinicians and family doctors. Each group has a different perspective on the bio-medical domain, and is therefore interested in different kinds of information. By providing alternative vocabularies for different users, the system can identify link sources relevant to that user and also provide multiple targets to relevant web resources. Table 1 shows four different user groups, some questions they might want answering and the different kinds of target sites a Sealife browser would offer them based on user type [18].

User Group	**Question**	**Targets Sites**
Family Doctor (GP)	Tuberculosis drugs and side effects?	British National Formulary (BNF)
Clinicians	Tuberculosis treatments guidelines?	Public Health Observatories (PHO)
Molecular Biologists	Drug resistant tuberculosis species?	PubMed
General Public	What is tuberculosis?	Health Protection Agency (HPA) or the NHS direct online website.

Table 1. This table shows some of the different user groups that visist the NeLI site, they all have different types of question they want to answer about the same concept. The table outlines different targets COHSE would provide to the user based on the user group they belong.

The system is demonstrated with a simple use case involving a news site linking to NeLI. News sites are often the first to report on disease outbreaks *via* news feeds. Consider the scenario where a traveller is planning a trip to Namibia, only to find an article on the BBC website[16] about a recent outbreak of Polio. COHSE can provide links to relevant resources that had not been included by the original author. Such resources could include information about the polio virus, its effect on humans, vaccination information and also geographical information about the local area. A family doctor, in contrast, might use a vocabulary skewed to the their interests to link through to sites on drugs, details of symptoms and clinical presentations, treatment and local hospital facilities *etc.*.

Figure 2 shows the COHSE system in action. The first image shows the original BBC article, this page has the links provided by the original author of

[16] http://www.bbc.co.uk

the page, these include links to related news stories and also a link to the World Health Organisation. Whilst these links may be useful to a user, it is easy to imagine a wide range of related information one might want to get to from this page. In this scenario the user wants information on polio vaccination. The BBC article contains no direct link so the user would be forced to use a search engine. Querying `polio vaccination` against a search engine would typically bring a wide range of documents that contain the word `polio` and/or `vaccination`, the user must manually filter through these documents to find the information they require. The search engine will not discriminate between news article, fact sheets or clinical guidelines, neither will it provide any indication as to who are the world authorities on vaccination information. All of this must be processed by the user which takes time and effort. There is also the problem that the search engine will miss documents that do not mention `polio` directly. The documents might mention `poliomyelitis`, which is the clinical term for the disease caused by polio virus. These documents may be vital to the user but missed because of the language used in the query.

Fig. 2. COHSE in action: Three screen shots showing COHSE adding hyperlinks to a news article, providing multiple link targets and finally the new page with more concepts highlighted.

The COHSE system can help the user in this scenario by identifying the key concepts in the documents and offering relevant links from the current site based on related terms and synonyms. When viewing the BBC article through the COHSE portal the document is first processed by the COHSE DLS (see Section 3). In this scenario the COHSE Knowledge Server (KS) has been loaded with a SKOS vocabulary that contains terms relating to infectious disease (the vocabulary was developed at NeLI and is represented in SKOS[17] [7]). The DLS identifies concepts in the document based on queries against the KS, the original document is returned to the user with the concepts highlighted and hyperlinked. The term `Polio` in the document has been identified as a synonym of the term `Polio Virus` and highlighted. The user can now select this new hyperlink to see what link targets the COHSE system is offering.

When the `polio` link is selected, the term is dynamically queried against the COHSE Resource Manager (RM). In this scenario the RM has been loaded with five web-services. These web services have been selected by experts from NeLI as being authoritative sites on information about infectious diseases. The term polio along with broader, narrower and related terms from the concept scheme are queried against the RM with the most relevant results returned as link targets to the user. In this example we see that `Polio Vaccination` is related to the term Polio via the `narrower` relationship. `Polio Vaccination` when queried against the RM returned links to a document from NeLI with guidelines about polio immunisation. The second screenshot (Figure 2) shows the links box offered to the user with links relating not only to polio, but to all the related terms found in the concept scheme. As the user continues to navigate around the web, the COHSE system continues to identify concepts in web pages and offer appropriate hyperlinks.

One important aspect to guiding navigation in this way is that the modality is kept, having to switch between windows and search engine results breaks the modality and can interfere with the task at hand [8]. We believe that having a selection of relevant links on hand in the current window adds value to the user experience.

The infectious disease vocabulary used as background knowledge for the NeLI use case has been developed in SKOS [7]. There already exist a wide range of vocabularies that cover the domain of infectious disease, these include the OBO Disease ontology[18] and MeSH. NeLI would like to reuse terminology from both to drive the COHSE system on their site, but they also want to extend these vocabularies with specific terms from their own internal vocabulary. NeLI do not wish to model the domain of infectious disease ontologically, they need a flexible vocabulary that contains the domain concepts and information about synonyms and related concepts, which can be used to drive navigation around their site and others on the Web. SKOS meets their requirements, and by having SKOS representation of both the MeSH and the OBO Disease ontology they are able to integrate these existing resources, and extend them with their own vocabulary for use within the COHSE system. Having to combine these three vocabularies in OWL would not only involve violating the OWL semantics, but is simply

[17] http://www.cs.man.ac.uk/~sjupp/ontologies/NeLI-demo.xml
[18] http://diseaseontology.sourceforge.net/

unnecessary for the task. The NeLI vocabulary not only demonstrates the use of SKOS in driving Semantic web browsers like COHSE, but also shows how a knowledge base with weak semantics like SKOS makes it easier to integrate a wide range of differing resources in applications where weaker semantics are acceptable.

7 Discussion

The advent of the Semantic Web/Grid has brought large computational resources to bear upon knowledge intensive sciences such as biology. Despite this, user facing tools that support document orientated tasks are not yet available within this new paradigm for computing. Semantic technologies can be applied to this problem to overcome some of the limitations of the current Web's navigational structure. We have argued that a representation with weaker semantics, such as SKOS, will enable us to exploit the wealth of ontologies, thesauri and other types of knowledge organisation schemes already existing within the biomedical domain.

Representations with stricter semantics are not always suitable for representing artefacts such as MeSH. KOS such as MeSH were created, not to make ontological descriptions of biological reality, but to aid navigation and retrieval of information about biomedicine.

The nature of formal ontologies can sometimes make it difficult to express relationships between concepts that experts from the domain would expect to find under some circumstances, such as in information retrieval tasks [25]. Thesauri are suited to represent the way words and language are used in the field. The Sealife project will demonstrate how the effort and cost associated with building the kinds of rich formal ontologies that are represented using OWL, can feed into other knowledge artefacts, like thesauri, vocabularies, classification schemes *etc.*. in SKOS, which are then available for use in different application scenarios.

Acknowledgements

Funding by the Sealife project (IST-2006-027269) for Simon Jupp is kindly acknowledged. Yeliz Yesilada was supported by Sun Microsystems Laboratories.

References

1. The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration. *Nat Biotech*, 25:1251–1255, 2007.
2. Sean Bechhofer, Robert D. Stevens, and Phillip W. Lord. Gohse: Ontology driven linking of biology resources. *J. Web Sem.*, 4(3):155–163, 2006.
3. Olivier Bodenreider and Robert Stevens. Bio-ontologies: current trends and future directions. *Briefings in Bioinformatics*, 7(3):256—274, 2006.
4. Peter Brusilovsky. Methods and techniques of adaptive hypermedia. In *Adaptive Hypertext and Hypermedia*, pages 87–129. Kluwer Academic Publishers, 1996.

5. Leslie Carr, Sean Bechhofer, Carole Goble, and Wendy Hall. Conceptual Linking: Ontology-based Open Hypermedia. In *WWW10, Tenth World Wide Web Conference*, Hong Kong, May 2001.
6. H. Davis, I. Heath W. Hall, G. Hill, and R. Wilkins. Towards an Integrated Information Environment with Open Hypermedia Systems. In *ECHT '92, Proceedings of the Fourth ACM Conference on Hypertext*, pages 181–190, Milan, Italy, 1992. ACM Press.
7. G. Diallo, P. Kostkova, G. Jawaheer, S. Jupp, and R. Stevens. Process of Building a Vocabulary for the Infection Domain. In *21st IEEE International Symposium on Computer-Based Medical Systems*, 2008.
8. A Dix, J Finlay, G Abowd, and R. Beale. *Human-Computer Interaction*. Prentice-Hall, 3rd edition, 2003.
9. Martin Dzbor, John Domingue, and Enrico Motta. Magpie: Towards a semantic web browser. pages 690–705, 2003.
10. Martin Dzbor, Enrico Motta, and John Domingue. Magpie: Experiences in supporting semantic web browsing. *Web Semant.*, 5(3):204–222, 2007.
11. Carole Goble and David De Roure. The grid: an application of the semantic web. *SIGMOD Rec.*, 31(4):65–70, 2002.
12. B. M. Good, E. M. Tranfield, P. C. Tan, M. Shehata, G. K. Singhera, J. Gosselink, E. B. Okon, and M. D. Wilkinson. Fast, cheap and out of control: A zero curation model for ontology development.
13. J. Hendler. Science and The Semantic Web. *Science*, page 24, Jan 2003.
14. D. Huynh, S. Mazzocchi, and D. Karger. Piggy bank: Experience the semantic web inside your web browser. In In, E. Motta, V. R. Benjamins, and M. A. Musen, editors, *International Semantic Web Conference*, 2005.
15. Patty Kostkova, Gemma Madle, Julius R. Weinberg, and Jane Mani-Saada. Agent-based up-to-date data management in national electronic library for communicable disease. *SI "Applications of intelligent agents in health care"*, pages 103–122, 2003.
16. Vitaveska Lanfranchi, Fabio Ciravegna, and Daniela Petrelli. Semantic web-based document: editing and browsing in aktivedoc. In *in: Proceedings of the 2nd European Semantic Web Conference*, pages 623–632. Springer, 2005.
17. McCray AT. Lindberg DA, Humphreys BL. The unified medical language system. *Meth Inform Med*, 32:281–291, 1993.
18. G. Madle, P. Kostkova, J. Mani-Saada, and A. Roy. Lessons learned from Evaluation of the Use of the National electronic Library of Infection. *Health Informatics Journal*, 12:137–15, 2006. Special Issue, Healthcare Digital Libraries.
19. Dilvan A. Moreira and Mark A. Musen. Obo to owl: a protege owl tab to read/save obo ontologies. *Bioinformatics*, 23(14):1868–1870, July 2007.
20. Stuart J. Nelson, Douglas Johnston, and Betsy L. Humphreys. Relationships in Medical Subject Headings. In Rebecca Bean, Carol A.; Green, editor, *Relationships in the organization of knowledge*, pages 171–184. Cluwer Academic Publishers, 2001.
21. National Library of Medicine. Medical subject headings: main headings, subheadings, and cross references used in the index medicus and the national library of medicine catalog. 1st ed, 1960.
22. T. Peterson. Introduction to the Art and Architecture Thesaurus. Oxford University Press, 1994.
23. B. Popov. Kim_sap_iswc168.pdf (application/pdf-objekt) kim - semantic annotation platform.

24. Alan Rector, Nick Drummond, Matthew Horridge, Jeremy Rogers, Holger Knublauch, Robert Stevens, Hai Wang, and Chris Wroe. Owl pizzas: Practical experience of teaching owl-dl: Common errors & common patterns. In *14th International Conference on Knowledge Engineering and Knowledge Management EKAW 2004*, pages 63–81, 2004.
25. Alan Rector, Robert Stevens, and Nick Drummond. What causes pneumonia? the case for a standard semantics for "may" in owl. In *OWL: Experiences and Direction (OWLED) workshop series*, 2008.
26. J Rogers and AL Rector. The galen ontology. *Medical Informatics Europe 1996*, pages 174–178, 1996.
27. Barry Smith, Werner Ceusters, Bert Klagges, Jacob Kohler, Anand Kumar, Jane Lomax, Chris Mungall, Fabian Neuhaus, Alan Rector, and Cornelius Rosse. Relations in biomedical ontologies. *Genome Biology*, 6(5):R46, 2005.
28. Dagobert Soergel, Boris Lauser, Anita Liang, Frehiwot Fisseha, Johannes Keizer, and Stephen Katz. Reengineering thesauri for new applications: the AGROVOC example. volume 4(4) of *Journal of Digital Information*. Oxford University Press, March 2004.
29. Robert Stevens, Chris Wroe, Phillip Lord, and Carole Goble. Ontologies in bioinformatics. In Stefan Staab and Rudi Studer, editors, *Handbook on Ontologies in Information Systems*, pages 635–657. Springer, 2003.
30. The Gene Ontology Consortium. Gene Ontology: Tool for the Unification of Biology. *Nature Genetics*, 25:25–29, 2000.
31. M. van Assem, V. Malaisé, A. Miles, and G. Schreiber. A method to convert thesauri to skos. In Y. Sure and J. Domingue, editors, *ESWC06 Vol. LNCS 4011*, pages 95–109. Springer, 2006.
32. Harald Weinreich, Hartmut Obendorf, Eelco Herder, and Matthias Mayer. Off the beaten tracks: exploring three aspects of web navigation. In *WWW '06: Proceedings of the 15th international conference on World Wide Web*, pages 133–142, New York, NY, USA, 2006. ACM.