



*Citation for published version:*

Cayzer, S 2004, 'Semantic Blogging : Spreading the Semantic Web Meme' Paper presented at XML Europe 2004, Amsterdam, Netherlands, 18/04/04 - 21/04/04, .

*Publication date:*

2004

*Document Version*

Early version, also known as pre-print

[Link to publication](#)

## University of Bath

### General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

### Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.



## Semantic Blogging: Spreading the Semantic Web Meme<sup>1</sup>

Steve Cayzer  
Digital Media Systems Laboratory  
HP Laboratories Bristol  
HPL-2004-48  
March 25, 2004\*

E-mail: [steve.cayzer@hp.com](mailto:steve.cayzer@hp.com)

semantic web,  
RDF, knowledge  
management,  
blogging

This paper is about semantic blogging, an application of the semantic web to blogging. The semantic web promises to make the web more useful by endowing metadata with machine processable semantics. Blogging is a lightweight web publishing paradigm which provides a very low barrier to entry, useful syndication and aggregation behaviour, a simple to understand structure and decentralized construction of a rich information network. Semantic blogging builds upon the success and clear network value of blogging by adding additional semantic structure to items shared over the blog channels. In this way we add significant value allowing view, navigation and query along semantic rather than simply chronological or serendipitous connections. Our vision is to use semantic web tools and ideas to help move blogging beyond communal diary browsing to rich information sharing scenarios. We have built a simple prototype as an illustration of this vision. Our semantic blogging prototype demonstrates schema driven views, new navigation modalities and richer query. It shows how a semantic blog can be used for informal knowledge management, and is set in the bibliography management domain. Our design was, broadly, to augment a blog with a metadata pipeline, with import, export and storage/access mechanisms. Three semantic behaviours (view, navigation and query) were built over this base. This work was performed as part of the SWAD-E (Semantic Web Advanced Development Europe) project, which provides targeted research, demonstrations and outreach to help semantic web technologies move into the mainstream of networked computing. We believe that we have contributed a number of useful things to this project. Firstly, a prototype that can be used to illustrate and assess a semantic web approach. Semantic web values are covered partly by the existing demonstrator, partly by stories one can tell around it, and partly by extensions that we (and others) are planning to build. Secondly, it appears that the demonstrator has more than just illustrative power. We (and others) see in semantic blogging the basis of a genuinely useful tool for applications whose scope extends far beyond bibliography management. Thirdly, a set of lessons for the deployment of useful RDF (Resource Description Framework) tools. These include the tension between RDF the model and RDF the syntax, the use of RDF for configuration and personalisation, and the importance of rich and interesting metadata. Finally, semantic blogging appears to be a promising base for outreach and publicity; we have had positive interest from individuals, start-ups, corporations and the press. We conclude this paper by looking forward to ways in which the semantic blogging theme might mature.

\* Internal Accession Date Only

<sup>1</sup> XML Europe, 18-21 April 2004, Amsterdam

Approved for External Publication

© Copyright Hewlett-Packard Company 2004

# Semantic Blogging: Spreading the Semantic Web Meme

*Keywords:* Semantic Web, RDF, Knowledge Management, Blogging

## **Steve Cayzer**

Research Engineer, Hewlett-Packard

Hewlett-Packard

Stoke Gifford

Bristol

UK

[steve.cayzer@hp.com](mailto:steve.cayzer@hp.com)

[http://www-uk.hpl.hp.com/people/steve\\_cayzer](http://www-uk.hpl.hp.com/people/steve_cayzer)

## *Biography*

Steve is a research engineer at Hewlett-Packard's (HP) laboratories in Bristol, England. He is interested in the intersection of semantic web technologies and machine learning techniques, such as automated classification and metadata enrichment. He also has a semantic blog.

---

## **Abstract**

---

This paper is about semantic blogging, an application of the semantic web to blogging.

The semantic web promises to make the web more useful by endowing metadata with machine processable semantics. Blogging is a lightweight web publishing paradigm which provides a very low barrier to entry, useful syndication and aggregation behaviour, a simple to understand structure and decentralized construction of a rich information network. Semantic blogging builds upon the success and clear network value of blogging by adding additional semantic structure to items shared over the blog channels. In this way we add significant value allowing view, navigation and query along semantic rather than simply chronological or serendipitous connections.

Our vision is to use semantic web tools and ideas to help move blogging beyond communal diary browsing to rich information sharing scenarios. We have built a simple prototype as an illustration of this vision.

Our semantic blogging prototype demonstrates schema driven views, new navigation modalities and richer query. It shows how a semantic blog can be used for informal knowledge management, and is set in the bibliography management domain. Our design was, broadly, to augment a blog with a metadata pipeline, with import, export and storage/access mechanisms. Three semantic behaviours (view, navigation and query) were built over this base. This work was performed as part of the [SWAD-E](#) (Semantic Web Advanced Development Europe) project, which provides targeted research, demonstrations and outreach to help semantic web technologies move into the mainstream of networked computing. We believe that we have contributed a number of useful things to this project. Firstly, a prototype that can be used to illustrate and assess a semantic web approach. Semantic web values are covered partly by the existing demonstrator, partly by stories one can tell around it, and partly by extensions that we (and others) are planning to build. Secondly, it appears that the demonstrator has more than just illustrative power. We (and others) see in semantic blogging the basis of a genuinely useful tool for applications whose scope extends far beyond bibliography management. Thirdly, a set of lessons for the deployment of useful [RDF](#) (Resource Description Framework) tools. These include the tension between [RDF](#) the model and [RDF](#) the syntax, the use of [RDF](#) for configuration and personalisation, and the importance of rich and interesting metadata. Finally, semantic blogging appears to be a promising base for outreach and publicity; we have had positive interest from individuals, start-ups, corporations and the press.

We conclude this paper by looking forward to ways in which the semantic blogging theme might mature.

---

## Table of Contents

---

### [1. Introduction](#)

### [2. Background](#)

#### [2.1 The Semantic Web](#)

#### [2.2 The SWAD-E project](#)

#### [2.3 Blogging](#)

#### [2.4 Semantic Blogging](#)

### [3. The Semantic Blogging Demonstrator](#)

#### [3.1 Domain Choice](#)

#### [3.2 Information Modelling](#)

#### [3.3 Enabling components](#)

##### [3.3.1 Platform](#)

##### [3.3.2 RDF storage](#)

##### [3.3.3 Metadata View](#)

[3.3.4 Metadata Edit](#)

[3.3.5 Metadata Import](#)

[3.3.6 Metadata Export](#)

[3.4 Semantic Capabilities](#)

[3.4.1 Semantic View](#)

[3.4.2 Semantic Navigation](#)

[3.4.3 Semantic Query](#)

[4. The added value of semantic blogging](#)

[5. To the future](#)

[6. Conclusions](#)

[Footnotes](#)

[Acknowledgements](#)

[Bibliography](#)

## 1. Introduction

Blogging appears to be taking the world (wide web) by storm. When the US president incumbent has a blog (<http://www.georgewbush.com/blog/>), when the BBC runs articles on blogging (<http://news.bbc.co.uk/1/hi/magazine/3078541.stm>) and when Microsoft FrontPage offers an easy way to create blogs (<http://www.microsoft.com/presspass/press/2003/jun03/06-10FrontPage2003XMLPR.asp>) one feels that blogs have finally reached the mainstream. Of course, much of the power of blogging comes from its ability to syndicate and share information using XML metadata. One format for such metadata is the RSS (RDF Site Summary) 1.0 standard [\[RSS 1.0\] \[1\]](#) which is based on RDF [\[RDF\]](#), the language of the semantic web. And one of the postulated advantages of the semantic web is that RDF encoded metadata opens up more avenues for intelligent data processing, enabling information to be shared, enriched, annotated and integrated more effectively. Hence, we [\[2\]](#) believe that *semantic blogging* might allow blogging to move into the realm of rich information management. It is this possibility that we describe in this paper, and that we continue to explore in our research.

This paper reports on the motivations, design and lessons learnt from an initial foray into semantic blogging. It is organised as follows. Firstly, the key concepts are briefly reviewed. Next, our prototype is described and the semantic web 'added value' identified. Finally, we speculate on some directions that semantic blogging might take in the future.

## 2. Background

## 2.1 The Semantic Web

We do not attempt to provide a comprehensive overview of the semantic web here. It suffices to say that the semantic web attempts to do for machine processable data what the world wide web did for human readable documents. Namely to transform information processing by providing a common way that data can be accessed, linked together and understood. To turn the web from a large hyperlinked book into a large interlinked database. As the semantic web vision starts to mature, we expect to see a number of beneficial effects. For example, the semantic web can unlock access to data currently hidden away in databases, freeing that data to be accessed by applications and tools across the globe. And making appropriate metadata (artists, locations, dates) available in machine interpretable form means that we can build applications which actively process this information - collect it, analyze it, filter it, correlate it, link it and apply it to the task at hand.

In our requirements specification [\[SWAD-E Requirements\]](#) we enumerated three key values for a semantic web approach:

- 1 **Data representation:** The foundation of the semantic web is a common format, [RDF \[RDF\]](#), to represent data. A critical feature of [RDF](#) is the use of web URIs to provide items with a well-defined place in the global namespace. This allows many sorts of data (property values of objects, relationships between objects, value annotations) to be represented uniformly and allows data from multiple locations to be combined without accidental clashing of property names or structure mismatches.
- 1 **Semantics:** The aspiration of the semantic web is to be able to express meaning. The schema [\[RDF\]](#) and ontology [\[OWL\]](#) layers of the semantic web begin to do this; for example, to say whether two terms are distinct, equivalent or whether one term is a subset of another. This capability allows a data source to expose its conceptual model explicitly in machine processable form thus providing the automation of more sophisticated processing and better integration of data from different sources. Of course, the semantic web does not require some standardized global upper ontology to function, remaining decentralized so that data sources are free to mix and match terms from different ontologies.
- 1 **Webness :** The critical innovation of the semantic web is to put both these values into a web framework. This is manifested in deceptively simple ways such as the use of URIs, enabling an agent to discover the ontology associated with a data source, the development of decentralized ontologies and the combination of terms from different ontologies.

## 2.2 The SWAD-E project

So the semantic web promises to make the web more useful by endowing metadata with machine processable semantics. And much of the basework for this endeavour has already been laid. We have a language, [\[RDF\]](#) backed by a precise semantics [\[RDF Semantics \]](#) and a series of logical formalisms [\[RDFS\]](#), [\[OWL\]](#) which allow us to bring increasing expressivity and power to inferencing over this metadata. We have a number of mature frameworks [\[Jena\]](#),

**[Redland]** and some commercial activity (see <http://www.w3.org/2004/01/sws-testimonial>). But there is still much misunderstanding and resistance, antagonism even, from the developer and wider community **[Shirky]**. Therefore, while much vital foundational work remains to be done, there is clearly an equally important need. A need for education and outreach. A need for tools which make it easy for XML developers to work with **RDF**. And a need to make what is sometimes called (tongue-in cheek) 'The Semantic Web (version 1.0)' a reality. It is with this need in mind that the Semantic Web Advanced Development (SWAD; **[SWAD]**) project was chartered. The SWAD-Europe project (**SWAD-E**; **[SWAD-E]**) aims to support this initiative in Europe, providing targeted research, demonstrations and outreach to ensure Semantic Web technologies move into the mainstream of networked computing. Semantic blogging is one of the open demonstrations contributed by Hewlett-Packard to the **SWAD-E** programme.

## 2.3 Blogging

Blogging, or web logging **[Essential Blogging]**, can be summarised as "push button publishing for the people", to borrow Evan Williams' phrase (see <http://blogger.com/>). Technically, the innovations are trivial, a way to update your web page through simple forms, text messages and even email. Socially, however, the effect is rather significant, leading to a much higher rate of publishing of timely information in the form of easy-to-digest "snippets". A second technically simple but powerful mechanism is two-way links (both who I am talking about and who is talking about me). Add to this a mechanism to publish blog "headlines"; and you have the basework for a powerful publishing, syndication and aggregation tool with the ability to create a decentralized, rich information network.

Blogs have, from their inception, supported mechanisms over and above free text search. Most blogs divide their posts into categories and dates, many blogs provide mechanisms for navigation along these axes, and some even allow syndication to be filtered in this way ("subscribe to this category"). **RDF** enables this metadata to be backed by a rich conceptual scheme, to be augmented with extra metadata, and to be linked more effectively with metadata from other blogs. These are the sorts of capabilities we want to explore with semantic blogging.

While we want to extend the blogging metaphor, we also want to preserve its key values, especially its simplicity. We want to build on blogging's proven potential for publishing, syndication & discovery, and community formation.

## 2.4 Semantic Blogging

In our applications survey **[SWAD-E Analysis]**, we noted that semantic web technologies are well suited to tasks where a user community is incrementally publishing structured and semantically rich (categorized and cross-linked) information. We also noted that blogging is a very successful paradigm for lightweight publishing, providing a very low barrier to entry, useful syndication and aggregation behaviour, a simple to understand structure and decentralized construction of a rich information network.

So why should we bring these successful, but distinct, paradigms together? In essence, semantic blogging builds upon the success and clear network value of blogging by adding additional semantic structure to items shared over the blog channels. In this way we add significant value allowing view, navigation and query along semantic rather than simply chronological or serendipitous connections. Our semantic blogging demonstrator design emphasises three key behaviours:

- 1 **Semantic View:** Semantically enriched blog metadata enables context sensitive, schema driven views of the blog content (over and above fixed templates).
- 1 **Semantic Navigation:** Semantically enriched blog metadata enables new blog navigation modalities (over and above unlabelled links).
- 1 **Semantic Query:** Semantically enriched blog metadata enables richer query and discovery mechanisms (over and above free text search).

In our demonstrator (described next) we implemented a number of features not usually found in blogs. Firstly, we enriched blog metadata using arbitrary vocabularies (this is in fact a key feature of RSS 1.0 [\[RSS Modules\]](#)). Secondly, we used this metadata to provide navigation and query interfaces onto the blog. Thirdly, we encapsulated information snippets (which we call items) into the blog, such that each blog entry is an annotation for an item. These items carry their own rich metadata which can be navigated over and searched for. Finally, we experimented with using RDF for [UI](#) configuration, allowing run time control of the blog look and feel. We also envisage a number of functionalities which, while not implemented in the current prototype, might form the basis for future extensions. For example, aggregation of multiple semantic blogs would allow a more powerful demonstration of the semantic capabilities. Ontology linking technologies such as [SKOS](#) [\[SKOS\]](#) can be used to align the conceptual schemes from different blogs. And RDF allows rich typing of item-item links, allowing the construction of a semantic, multi-layered network.

We should note here that the use of the term 'semantic' emphasises the use of semantic web technology to enable these behaviours. In the current instantiation of the demonstrator, the new capabilities are enabled primarily by using rich metadata, and require little in the way of actual semantic machinery. However, in each case the behaviour can be further extended by adding inferencing over a semantic model. A simple example would be subcategory inferencing for semantic query. We prepare our blog for such possibilities by encoding its metadata in [RDF](#).

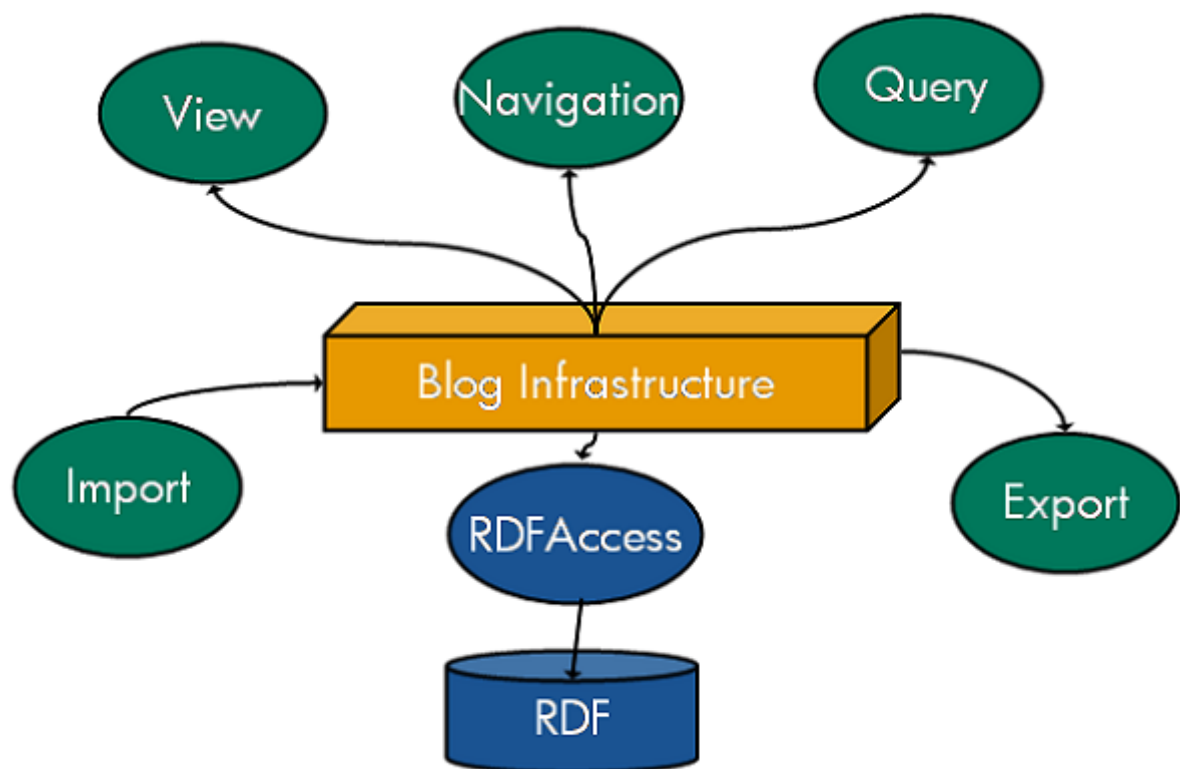
### 3. The Semantic Blogging Demonstrator

In order to test our intuitions about semantic blogging, we built a simple prototype [\[SWAD-E Demonstrator\]](#), which is available both as an online demo and for download. In this section we describe, at a high level, the design of this prototype. More details can be found in the lessons learnt report [\[SWAD-E Lessons Learnt\]](#). We start by describing the choice of domain and the information model used. We then describe the enabling infrastructure. Finally,



we show how the semantic capabilities we wish to demonstrate (view, navigate, query) were built over this infrastructure.

We have aimed with this demonstrator to develop a tool that is simple, useful, extensible and illustrative. Simple, because it should be easy to learn and to use. Useful, because it should do something that users actually want, efficiently and reliably. Extensible, because although we ground the requirements in the bibliographic domain, we expect it to be reusable for other semantic blogging applications. And illustrative, because we wish to incorporate features that demonstrate the values of the semantic web approach without losing the attractive features of blogging.



**Figure 1: Semantic Blogging Demonstrator: Architecture**

(metadata view/edit not shown)

**Figure 1** shows the basic architecture of our demonstrator. As we can see, the semantic blog is built over an existing blogging platform. As shown in the bottom part of the figure, the blog is semantically enabled by providing it with an [RDF](#) backend. An [RDFAccess](#) component provides an interface to this metadata store, while import and export complete the blog metadata pipeline. Metadata view and edit functionalities are also provided (but not shown in the diagram). The semantic view, navigation and query functionalities, which demonstrate the semantic web values, exploit this basic infrastructure.



## Semantic Blogging Demonstrator

The demonstration platform for HP Labs' semantic blogging research programme

Browsing Category: / (Show all items) (Semantic View: record card | table)

**RDF Model and Syntax** [Permalink]  
Category: *semantic-web/RDF*  
Thu Oct 09 17:13:48 BST 2003

The RDF Model & Syntax was first described [here](#) before being raised in status to a W3C recommendation.

This is the first paper I read about RDF, I found it an excellent introduction. There are more recent introduc... More ...

Posted by: *stecay*  
Metadata: [Record Card](#) | [XML format](#) | [N3 format](#) | [Browsauce](#) | [Edit metadata](#) | [Comments \[0\]](#) | [Trackbacks \[0\]](#) | [Add Trackback](#)

**RDF Model & Syntax** [Permalink]  
Category: *semantic-web/RDF*  
Thu Oct 09 15:35:05 BST 2003

Authoritative spec for RDF model & syntax

Posted by: *wiken*  
Metadata: [Record Card](#) | [XML format](#) | [N3 format](#) | [Browsauce](#) | [Edit metadata](#) | [Comments \[0\]](#) | [Trackbacks \[0\]](#) | [Add Trackback](#)

**Thesaurus mapping** [Permalink]  
Category: *semantic-web/ontologies/*  
Thu Oct 09 15:32:48 BST 2003

Search:

**March 2004**

Sun	Mon	Tue	Wed	Thu	Fri	Sat
	1	2	3	4	5	6
7	8	9	10	11	12	13
14	15	16	17	18	19	20
21	22	23	24	25	26	27
28	29	30	31			

< Feb Today Apr >

**Site Links**

- What IS it?
- Download
- Add a new entry
- Semantic Navigation
- Query By Entry
- Query By Item
- Steve Cayzer Homepage
- Main Blog
- Contact Me

**Supporting Material**

**Figure 2: Semantic Blogging Demonstrator: User View**

(example links to semantic capabilities highlighted)

The blog itself is presented to the user as follows in [Figure 2](#). This figure shows a standard blog interface - it is important that the user introduction to semantic blogging should be as familiar as possible. So the blog entries are displayed as separate boxes on the page, you can browse them individually or by category just as you would expect from a normal blog. The semantic capabilities are accessed by links on the sidebars ("Query by Entry/Item") and on the items themselves ("Record Card", "Edit metadata").

### 3.1 Domain Choice

Our demonstrator is grounded in the bibliographic management domain.

Although traditional bibliographic management deals mainly with static categorisations, the needs of a small group collectively exploring a domain exhibit a more dynamic, community based flavour. Here is a task which is characterised by a need to share small items of information with a peer group in a timely, lightweight manner. This information should be easily publishable, easily discoverable and easily navigable. It should be simple to enrich the information with annotation, either at the point of delivery or later. The information should be archived in a commonly understood way for effective post-hoc retrieval. It should be possible to be notified, in a timely way, of new items of interest. We believe that a combination of blogging and semantic web technologies offers a promising approach to this problem. Blogging provides low barrier publishing, a simple shared conceptual model, and a mechanism for natural,

dynamic community formation. The rich structure provided by semantic metadata enables improved view, navigation and query capabilities.

In addition, the need we are exploring is not one well served by current tools [[SWAD-E User Study](#)]. Although it is, of course, simple to integrate a personal bibliography and (say) a word processor, the current tools do not facilitate the sharing of metadata between small groups. This is, in part, due to the weaknesses of current bibliography standards when it comes to representing rich community annotations, and in part due to the lack of tools to make use of these annotations. The semantic blogging demonstrator is largely an attempt to answer the second need, but it also recognises the need of small groups to 'roll their own' classification and annotation schemes and yet make these annotations available for use both inside and outside their community. We have built our demonstrator with the ultimate aim of creating a community sharing tool.

Whilst bibliography management is an important task in the research community, it could be seen as a niche application in the wider community. However, the semantic blogging tools and approaches are just as applicable to dissemination and management of other content such as business documents or news items. Generalizing the results to related areas should be straightforward.

## 3.2 Information Modelling

Using a blog as a bibliographic management tool immediately forces us to take a novel viewpoint. Blog entries are separate entities. Bibliographic items are separate entities. *They are not the same*. For example, an entry might be my comment on a particular paper. The author of the entry (the comment) is different to the author of the item (the paper). Similarly, it is possible that two entries on a blog (or even different blogs) might reference the same paper. This enables us to link the two entries while retaining their identities and provenance. In order to model this within the demonstrator we took the view that

### **(blog) entries contain (bibliographic) items**

We took a number of other less significant information modelling choices. For example, the standard [RSS 1.0](#) [[RSS 1.0](#)] metadata was enriched using various modules [[RSS Modules](#)]. The [RSS](#) feed also contains full bibliographic item metadata in more novel (for blogs) vocabularies. Other properties, drawn from different vocabularies, could easily be added to the metadata, using the standard mechanisms of [RDF](#).

For generating bibliographic metadata, we used Michel Klein's BibTeX to [RDF](#) converter [[BibTeX-2-RDF](#)]. In this scheme, people are modelled as separate resources, giving a framework for asking queries such as "Who has blogged a paper written by X?".

Blog entries are assigned categories modelled using Thesaurus Interchange Format, [TIF](#) (Thesaurus Interchange Format) [[TIF](#)] [[3](#)]. This is a schema primarily designed for thesauri, yet

one that maps well to our domain. The central idea is that a topic tree can be built using topics (called concepts in [TIF](#)) and 'broader'/'narrower' relationships. Each topic is identified by a set of indicator terms, one of which is the preferred term. This allows us to build in a (very simple) natural language mapping from blog entries to our concept tree (this mapping is implemented in our prototype as described later). An alternative would be to maintain a simpler topic tree representation, but to tie the topic names to a natural language ontology such as WordNet [\[WordNet\]](#). Although intuitively appealing, such an approach has two limitations. Firstly, it is somewhat more complex than the approach we took. Secondly, a general purpose resource such as WordNet fails to capture the precise semantics of topics within specialist and highly individual domains.

Schemata were designed to control the [UI](#) (User Interface). Our general intention was to make these schemata as simple as possible. That is, rather than trying to anticipate all future needs, or architecture in very fine grain control of presentation, the ontologies would only govern simple and coarse grain characteristics, such as properties to show/hide, view type (eg table, record card) and so on. One subtlety that is worth mentioning is that we use a choice of 'opt-in' or 'opt-out' styles for displaying properties. The opt-in style means that *only* properties explicitly named in the configuration file will be shown. This style works well for display types like tables, where one would like to specify in advance the number of columns and their bound properties. The opt-out style, on the other hand, shows all properties *except* those specifically named. This is good for views like record card (and edit) where one wants to display all the metadata for an entry (except perhaps the 'internal use only' metadata). Such an approach ties in well with the 'open world' nature of [RDF](#) - that is, one does not have to anticipate all metadata that might be relevant for an entry. Finally, a design-time decision was to hide statements whose object is a resource in edit view. This is because editing a resource URI in a free text field is not generally an option we want to provide for the user.

## 3.3 Enabling components

### 3.3.1 Platform

Our semantic blogging demonstrator is built over the Java based blog platform [blojsom](#) [\[blojsom\]](#) and uses Jena [\[Jena\]](#) for its [RDF](#) capabilities. The other area in which we made 'platform' choices was that of the bibliographic schema [\[BibTeX-2-RDF\]](#), although it is not our intention to constrain users to one particular schema. There is no reasons why semantic bloggers could not describe their resources using a different standard, or even their own ontology.

### 3.3.2 RDF storage

[RDFAccess](#) is the control point to the [RDF](#) data store which backs the semantic blog. This is effectively an API which allows access to various storage implementations (memory, file access and database). Queries, additions and modifications to [RDF](#) data are handled by this

component. Tied into this component is the [RSS](#) generation (ie [RSS](#) feeds should query this database to get the metadata for each blog item rather than generating it themselves, separately from and possibly in conflict with the [RDF](#) store).

In our demonstrator, the implementation of the [RDF](#) store was a single file containing metadata for all blog items. This approach does mean that the store becomes vulnerable to small metadata errors (which are quite easy to generate during development). An [RDF](#) database would be less vulnerable to such errors, but brings its own drawbacks, such as infrastructure requirements and performance hits. Clearly the single file approach is also not scalable to very large blogs. In our next iteration we expect to migrate to a more robust solution. It is worth mentioning however that the solution worked well over the lifetime of the project. Because each blog entry carries with it only a modest amounts of metadata, the [RDF](#) store grows quite slowly as the blog content increases.

### 3.3.3 Metadata View

An important capability is to view the metadata behind each blog entry. This is accomplished in Jena by using a specialised writer (eg [N3](#) (Notation 3) or [RDF/XML](#)). In addition, third party metadata viewers such as [brownsauce](#) [\[Brownsauce\]](#) can easily be invoked, since given the [RDFAccess](#) functionality described above, we can use a simple [HTTP GET](#) to retrieve the [RDF](#) relevant to any individual blog entry. An (abbreviated) example of the metadata behind a single blog entry is shown in [Figure 3](#). Note that this (blog) entry metadata includes (bibliographic) item metadata. This is data that would not normally be present in an [RSS](#) feed, yet is relevant and useful for a bibliographic blog.

```
<rdf:RDF
  xmlns="http://purl.org/rss/1.0/"
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
  xmlns:bibtex="http://www.ontoweb.org/ontology/1#"
  xmlns:semblog="http://jena.hpl.hp.com/semblog#"
  xmlns:dc="http://purl.org/dc/elements/1.1/" >
<rdf:Description rdf:about="http://jena.hpl.hp.com/~stecay/downloads/semblo
<rdfs:label>O. Lassila</rdfs:label>
<bibtex:name>O. Lassila</j.0:name>
<rdf:type rdf:resource="http://www.ontoweb.org/ontology/1#Person"/>
</rdf:Description>
<rdf:Description rdf:about="http://jena.hpl.hp.com/~stecay/downloads/semblo
<bibtex:author rdf:resource="http://jena.hpl.hp.com/~stecay/downloads/sem
<bibtex:url rdf:resource="http://citeseer.nj.nec.com/article/lassila98res
<rdfs:label>O. Lassila, R. Swick. Resource Description Framework (RDF) mo
<bibtex:text>O. Lassila and R. R. Swick. Resource Description Framework (
<bibtex:title>Resource Description Framework (RDF) model and syntax speci
<bibtex:author rdf:resource="http://jena.hpl.hp.com/~stecay/downloads/sem
<rdf:type rdf:resource="http://www.ontoweb.org/ontology/1#Misc"/>
</rdf:Description>
<rdf:Description rdf:about="http://jena.hpl.hp.com/~stecay/downloads/semblo
<rdfs:label>R. Swick</rdfs:label>
<rdf:type rdf:resource="http://www.ontoweb.org/ontology/1#Person"/>
<bibtex:name>R. Swick</j.0:name>
```

```

</rdf:Description>
<rdf:Description rdf:about="http://jena.hpl.hp.com:3030/blojsom-devt/blog/s
  <rdf:type rdf:resource="http://purl.org/rss/1.0/item"/>
  <rdf:type rdf:resource="http://jena.hpl.hp.com/semblog#blogItem"/>
  <rdfs:seeAlso rdf:resource="http://jena.hpl.hp.com:3030/blojsom-devt/blog
  <semblog:contains rdf:resource="http://jena.hpl.hp.com/~stecay/downloads/
  <semblog:hasConcept rdf:resource="http://jena.hpl.hp.com/topic/semantic-w
  <dc:creator>stecay</dc:creator>
  <dc:date>2003-10-09T17:13:48BST</dc:date>
  <dc:identifier>http://jena.hpl.hp.com:3030/blojsom-devt/blog/semantic-web
  <dc:subject>http://jena.hpl.hp.com/topic/semantic-web/RDF/</dc:subject>
  <description>The RDF Model & Syntax was first described here before b

This is the first paper I read about RDF, I found it an excellent introductio
  <link>http://jena.hpl.hp.com:3030/blojsom-devt/blog/semantic-web/RDF/?per
  <title>RDF Model and Syntax</title>
</rdf:Description>
</rdf:RDF>

```

**Figure 3: Semantic Blog metadata**

### 3.3.4 Metadata Edit

It is also important to be able to change the item metadata once it has been entered. We provide a simple, schema-driven and password-controlled metadata edit form to accomplish this. Note that there is a difference between editing the item (bibliographic) metadata, editing the entry (standard [RSS](#)) metadata and editing the entry itself. We currently allow access to both entry and item metadata, leaving entry content editing to standard blog mechanisms. One possibility for the future would be to conflate blog content and metadata (there is only metadata; the blog entry content is created directly from the metadata).

### 3.3.5 Metadata Import

Blog entries are easily created using our SemBlogIT! utility. The idea is not novel; similar functionality is provided by MovableType bookmarklets, and a 'blog this page' utility is available on the Google toolbar. So, if you want to blog your current site, one click produces an entry form populated with a title, link to the page of interest, and any text you have highlighted. You then add any other details you want, choose a category and post the entry. The posting process is password controlled.

We have added two semantic capabilities to SemBlogIT!

Firstly, a way to attach items to blog entries. So, if you wanted to blog about a bibliographic item (say, a paper), you can enter metadata about that paper. An 'add bibliographic item' popup provides space for bibliographic fields such as author and title. Alternatively, prepared metadata (such as that produced by an online BibTeX to [RDF](#) converter [\[BibTeX-2-RDF\]](#)) can be entered directly. In either case, the blog entry will be marked as containing the bibliographic item. Incidentally, the popup dialog is itself schema-controlled, so that the choice of fields to



enter is determined by the user.

Secondly, the *Category Chooser* functionality. After submission, the demonstrator suggests a number of categories for a blog entry. This is implemented using an thesaurus, represented in **TIF [TIF]** format. This thesaurus essentially represents a hierarchy of concepts, each with a set of indicator terms (think of keywords) one of which is the preferred term for that concept. The category chooser client goes through the text for a blog entry, stemming words and matching them with the indicator terms. If a term is matched, then that concept is suggested. This is a form of ontology driven assisted metadata creation.

### 3.3.6 Metadata Export

Most blogs support metadata export in the form of **RSS** . The demonstrator uses the **RDF** flavour of **RSS [RSS 1.0]** . This format provides the usual **RDF** mechanisms for extension and vocabulary mixing. Thus, the **RSS** feed contains information over and above what is specified by the **RSS** 1.0 specification; for example, relevant bibliographic metadata linked to the requested blog entries. Although we have not implemented an aggregator, such enriched feeds could in principle be collected and community versions of the semantic services (described below) provided.

Use of **RDF** in an export feed leads to consumer issues. While the XML serialization is an essential component of **RDF** , and has undoubtedly contributed to the success of **RDF** so far, the interoperability it brings carries with it a risk. The risk is that people will depend on some 'canonical' XML serialization of an **RDF** model. Syntax constraints and accepted XML profiles can be used to circumvent these issues but such measures can pose a limitation for **RDF** expressiveness.

## 3.4 Semantic Capabilities

Recall that our aims are to demonstrate three capabilities: semantic view, semantic navigation and semantic query. Within the context of bibliographic management we have demonstrated these capabilities in the following ways.

- 1 **Semantic View:** Schema driven blog views include 'record card' format (particularly appropriate for bibliographic items) and summary tables.
- 1 **Semantic Navigation:** A schema driven navigator window provides a category tree browser and a simple facet-type interface.
- 1 **Semantic Query:** A schema driven query window allows queries over user-selected metadata.

Each capability is intended to illustrate some aspect of semantic web added value, as we describe in the following sections.

### 3.4.1 Semantic View

Given we have bibliographic items in the blog, we would like to be able to view these in a useful way. We provide two views. The first is a 'record card' view, in which the bibliographic items are described using a sort of indented record card. This enables one to distinguish between the blog entry and the bibliographic items that the entry is about. An alternative way of viewing blog entries is a summary table. This shows a list of entries contained in this particular category, with summary information (creator, date, concept and title) for each.

Like semantic edit, semantic view is schema driven. There is a semview preferences file (expressed in [RDF](#)) which, while simple, does allow for named views. We used two named views (record and table) but one could also define different views for each user. The required view is selected at runtime through a URL parameter.

In semantic view we use [RDF](#) configuration files to control the view. Because the metadata is encoded in [RDF](#), different properties (and subproperties thereof) can be shown in different ways. Implicit details can be inferred. And extra, relevant data (using different vocabularies) can be added, at creation time or later, without breaking anything. Related data (for example, resources pointed to by item metadata, such as the author of a paper) can also be linked to and displayed.

### 3.4.2 Semantic Navigation

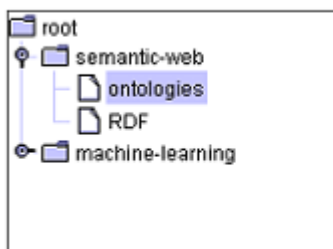
The demonstrator supports two forms of navigation (over and above standard blog navigation modalities):

- | **Tree Browsing** Here, a java applet renders a tree representation of the hierarchical categories. The tree itself is generated from an [RDF](#) schema (the blog category hierarchy is represented in [TIF \[TIF\]](#)).
- | **Facet browsing** Although this is not strictly facet navigation, the idea here is that data can be selected by choosing an intersection of features. For example, one might choose all entries written by "stecay" on a concept (or subconcept) of the "semantic web".

The navigation interface is controlled by an [RDF](#) formatted configuration file. So one can choose which properties appear in the facet browser, and which values can be chosen for them. Thus, navigation is really a constrained query. Once the user has chosen a set of navigation options (or equivalently submitted a query) the relevant blog entries are displayed in a results frame. An example of navigation can be seen in [Figure 4](#).



## Tree Browsing



## Facet Browsing

dc:creator

rss:title

semblog:hasConcept

dc:description



# Semantic Navigation

(Return to main blog)

## Thesaurus mapping [\[Permalink\]](#)

Category: [semantic-web/ontologies/](#)  
Thu Oct 09 15:32:48 BST 2003

Semantic problems of thesaurus mapping. This has relevance to my thoughts about the thesaurus interchange format and ontology linking in the semantic blogging project.

Posted by: [wikon](#)

Metadata: [Record Card](#) | [XML format](#) | [N3 format](#) | [Brownsauce](#) | [Edit metadata](#) | [Comments \[0\]](#) | [Trackbacks \[0\]](#) | [Add Trackback](#)

## Thesaurus mapping [\[Permalink\]](#)

Category: [semantic-web/ontologies/](#)  
Thu Oct 09 15:31:28 BST 2003

Semantic problems of thesaurus mapping. Just bookmarking this, it looks highly relevant to semantic portals

Posted by: [stecay](#)

**Figure 4: Semantic Blogging  
Demonstrator: Semantic Navigation**

One of the downsides to our navigation is that there is only one usefully navigable, hierarchical field - category. It is useful to query by author or by date (eg month) but these fields are by their nature flat. A more rewarding navigation experience would be using truly faceted fields. For example, environmental organisations may be organised along facets like geographic region, legal status and activity type. We are building faceted navigation into our next [SWAD-E](#) demonstrator, semantic portals [\[Semantic Portals\]](#).

Each semantic navigation component uses [RDF](#) in a different way. The tree browser takes an ontology and provides a navigable [UI](#). In this case, the ontology is a [TIF](#) encoding of the categories, but it need not be. Anything that could be represented in a suitable format ( region, organisational hierarchy, legal status) could also be navigated in this way. The facet navigation works as a constrained semantic query, which we shall describe next.

### 3.4.3 Semantic Query

The demonstrator implements two forms of semantic query. The first, Query By Entry, presents a set of fields, corresponding to blog entry metadata. The dialog looks like the facet navigator, except that the user can enter unconstrained text in each field. Again, the exact form of the display is schema driven. Once the user submits the query, the relevant entries are retrieved

from the blog database on the basis of their metadata. The results are piped into semantic view - hence, one can choose a preferred format (such as record card or table form).

In some sense, this is similar to normal blogs, which readily allow filtering of blog entries by author, date or category (all forms of metadata). But there are a number of differences. Firstly, the query interface allows the user to specify at run time what properties will be used in the search (author and category are examples, but one might also search on title or even URL). Secondly, the properties we can query on can be drawn from any of the vocabularies used in the blog, not just the standard RSS fields. Thirdly, the query can be inference enabled (though this is not currently implemented), using the ontology behind the metadata to provide a richer search. Finally, we can also search for metadata about items, not just about blog entries. That is what Query By Item does.

Query By Item allows a search for blog entries about a certain paper (or about papers on a certain topic, or by a certain person...). The query works exactly the same as before, except that the entries returned are those *about* the items of interest. The important thing here is that the focus of the search has changed - the blog entries are no longer the primary entities, the items are. Blog entries can thus be thought of as annotations attached to an item, and our search returns items of interest together with their annotations.

## 4. The added value of semantic blogging

The point of semantic blogging is to use semantic web principles to extend the power of blogging. We believe that the demonstrator, though simple, has achieved this aim in a number of ways:

- | An initial step towards semantic blogging involves simply using the capabilities of [RDF](#) to add metadata from different vocabularies to the blog entries.
- | In our case, we extended the metaphor further by using the blog as an annotation mechanism over an information network. In this case, the information consisted of bibliographic items, but this is just an example.
- | Semantic metadata enables richer view, navigation and query mechanisms. In each case, there is also potential for aggregate, community services.
- | [RDF](#) can also be used in other ways to improve the blogging experience. For example, we use it to assist the creation process (the category chooser) and to control the [UI](#) (semantic view).
- | Further use of semantic metadata might include community query, concept linking and the construction of rich information networks.

## 5. To the future

Although it is only a prototype, the demonstrator has attracted significant, and positive, interest from individuals, corporations and journalists. In particular, there are a number of groups that have expressed interest in building on our demonstrator work, and we look forward to working with these groups as semantic blogging technology evolves. In addition, we (Hewlett-Packard) plan to build an internal semantic blog that will enable us to test our assumptions about the day to day utility of such an approach. A (modified) semantic blog is also one of the input paths to our ongoing semantic portals demonstrator [\[Semantic Portals\]](#). In this project, the use of [RDF](#) for configuration is taken further, as the presentation depends not just on user supplied preferences, but also the type of data being displayed. Facet navigation is also a key part of the semantic portal [UI](#).

There are two key research areas that remain to be fully explored. Firstly, building an aggregator of semantic blogs enables some of the community services discussed above to be implemented and evaluated. Secondly, the use of ontology linking allows the decentralized creation and merging of different peers' conceptual models (eg categorisation schemes).

The demonstrator [\[SWAD-E Demonstrator\]](#) is available both as an online demo and for download; we welcome further feedback and enquiries.

## 6. Conclusions

We conclude that semantic blogging is a useful metaphor, both to illustrate semantic web values and to provide a hook for a wider community to interact with the semantic web. We believe that our semantic blogging prototype is of interest for these reasons. In this paper, we have provided an overview of its architecture and evaluated our design choices, for the benefit of anyone building similar systems. We have outlined the advantages of a semantic approach, some of which can be demonstrated using the implemented prototype.

We started this project with two aims. To build a working, useful demonstrator, and to provide a focal point for discussion. We have partially achieved the first aim, building a successful prototype which we expect to mature into a useful tool. We have exceeded our expectations on the second, attracting unsolicited interest from industry, academia and the press. It seems semantic blogging has potential to be a genuinely useful technology. We look forward to the semantic blogging theme maturing in many, perhaps even unexpected, ways in the future.

## Footnotes

1. There are alternate expansions of the acronym 'RSS' but this expansion is the generally accepted one for the RSS version that we use, RSS1.0.
2. Throughout this paper, the plural form "we" is used to refer variously to the author (Steve Cayzer), the implementers (author plus William Kong), the designers (implementers plus Paul Shabajee), or the entire project team (designers plus Dave Reynolds and Ian Dickinson). We also take this opportunity to acknowledge the myriad contributions from others in Hewlett Packard, our [SWAD-E](#) partners and the wider community.
3. The interested reader should be aware that [TIF](#) has now been replaced by [SKOS](#) (Simple Knowledge Organisation System) [\[SKOS\]](#).

## Acknowledgements

This work was undertaken as part of the [SWAD-E](#) project. The author would like to thank members of the semantic web research group at Hewlett-Packard for their helpful comments on this manuscript.

## Bibliography

### **[BibTeX-2-RDF]**

*BibTeX-2-RDF translator*, by Michel Klein <http://www.cs.vu.nl/~mcaklein/bib2rdf/> (uses ontology available at <http://ontoweb.aifb.uni-karlsruhe.de/>)

### **[blojsom]**

*Blojsom*, a Java based blogging platform <http://blojsom.sourceforge.net/>

### **[Browsersauce]**

*Browsersauce*, an RDF Browser <http://browsersauce.sourceforge.net/>

### **[Essential Blogging]**

*Essential Blogging: Selecting and Using Weblog Tools* Cory Doctorow, Rael Dornfest, J. Scott Johnson, Shelley Powers, Benjamin Trott, Mena G. Trott, 2002 O'Reilly

### **[FOAF]**

*The friend of a friend project* <http://www.foaf-project.org/> FOAF Vocabulary Specification (RDFWeb Namespace Document 16 August 2003; Dan Brickley & Libby Miller): <http://xmlns.com/foaf/0.1/>

### **[Jena]**

*Jena*, a semantic web toolkit <http://www.hpl.hp.com/semweb/jena.htm>

## [OWL]

*OWL Web Ontology Language Overview* W3C Recommendation 10 February 2004  
<http://www.w3.org/TR/2004/REC-owl-features-20040210/>

## [Redland]

*Redland*, an RDF Application Framework <http://www.redland.opensource.ac.uk/>

## [RDF]

*RDF/XML Syntax Specification (Revised)* W3C Recommendation 10 February 2004  
<http://www.w3.org/TR/2004/REC-rdf-syntax-grammar-20040210/>

## [RDFS]

*RDF Vocabulary Description Language 1.0: RDF Schema* W3C Recommendation 10 February 2004 <http://www.w3.org/TR/rdf-schema/>

## [RDF Semantics ]

*RDF Semantics* W3C Recommendation 10 February 2004  
<http://www.w3.org/TR/2004/REC-rdf-mt-20040210/>

## [RSS 1.0]

*RSS 1.0 Standard* <http://web.resource.org/rss/1.0/> RSS1.0 Specification:  
<http://web.resource.org/rss/1.0/spec> RDF Rich Site Summary - history and resources:  
<http://www.oasis-open.org/cover/rss.html>

## [RSS Modules]

*RDF Site Summary 1.0 Modules* <http://web.resource.org/rss/1.0/modules/>

## [Semblog Vision]

*Semantic Blogging Demonstrator - The Vision*  
<http://jena.hpl.hp.com/~stecay/downloads/blogTalk.pdf>

## [Shirky]

*The Semantic Web, Syllogism, and Worldview* [http://www.shirky.com/writings/semantic\\_syllogism.html](http://www.shirky.com/writings/semantic_syllogism.html) See summary of discussion at <http://www.poorbuthappy.com/ease/semantic/>

## [SKOS]

*Simple Knowledge Organisation System* A deliverable of SWAD-E workpackage 8.1 <http://www.w3c.rl.ac.uk/SWAD/deliverables/8.1.html> SKOS core schema <http://www.w3c.rl.ac.uk/2003/11/21-skos-core> SKOS mapping schema <http://www.w3c.rl.ac.uk/2003/11/21-skos-mapping>

## [SWAD]

*Semantic Web Advanced Development* <http://www.w3.org/2000/01/sw/>

## [SWAD-E]

*Semantic Web Advanced Development - Europe* <http://www.w3.org/2001/sw/Europe/>

## [SWAD-E Analysis]

*Semantic Web Applications - Analysis and Selection* Dave Reynolds , Steve Cayzer, Ian Dickinson, Paul Shabajee  
[http://www.w3.org/2001/sw/Europe/reports/open\\_demonstrators/hp-applications-selection.html](http://www.w3.org/2001/sw/Europe/reports/open_demonstrators/hp-applications-selection.html)

## [SWAD-E Demonstrator]

*Semantic Blogging Demonstrator*. The demonstration platform for HP Labs' semantic blogging research programme <http://jena.hpl.hp.com:3030/blojsom-devt/blog/> Online demo instructions: <http://jena.hpl.hp.com/~stecay/downloads/script.html> Download: <http://jena.hpl.hp.com:3030/blojsom-devt/download.jsp>

### [SWAD-E Lessons Learnt]

*SWAD-Europe deliverable 12.1.4: Semantic Blogging - Lessons Learnt* Steve Cayzer [http://www.w3.org/2001/SW/Europe/reports/demo\\_1\\_report/](http://www.w3.org/2001/SW/Europe/reports/demo_1_report/)

### [SWAD-E Requirements]

*SWAD-Europe: Semantic Blogging and Bibliographies - Requirements Specification* Steve Cayzer, Paul Shabajee  
[http://www.w3.org/2001/sw/Europe/reports/open\\_demonstrators/hp-requirements-specification.html](http://www.w3.org/2001/sw/Europe/reports/open_demonstrators/hp-requirements-specification.html)

### [Semantic Portals]

*Semantic Portals* part of SWAD-Europe Deliverable 12.1  
[http://www.hpl.hp.com/semweb/portal.htm#Semantic Portals](http://www.hpl.hp.com/semweb/portal.htm#Semantic%20Portals) Requirements Specification: <http://www.w3.org/2001/sw/Europe/plan/workpackages/live/esw-wp-12.1.html>

### [SWAD-E Thesaurus]

*SWAD-E Thesaurus activity* <http://www.w3c.rl.ac.uk/SWAD/thesaurus.html> SWAD-E Workpackage 8 description: <http://www.w3.org/2001/sw/Europe/plan/workpackages/live/esw-wp-8.html> RDF Thesauri: <http://www.w3c.rl.ac.uk/SWAD/rdfthes.html>

### [SWAD-E User Study]

*SWAD-E User Study on Bibliographic Management* Paul Shabajee  
[http://www.w3.org/2001/sw/Europe/reports/open\\_demonstrators/hp-requirements-specification.html#sec-userStudy](http://www.w3.org/2001/sw/Europe/reports/open_demonstrators/hp-requirements-specification.html#sec-userStudy)

### [TIF]

*Thesaurus Interchange Format* Paper: "Modelling Thesauri for the Semantic Web" <http://www.w3c.rl.ac.uk/SWAD/thesaurus/tif/deliv81/final.html> Schema: <http://www.w3c.rl.ac.uk/SWAD/thesaurus/tif/tif.html> Now replaced by SKOS

### [WordNet]

*WORDNET - a lexical database for the English language* <http://www.cogsci.princeton.edu/~wn/> RDF Schema: <http://xmlns.com/2001/08/wordnet/>