



## Semantic Blogging and Bibliography Management<sup>1</sup>

Steve Cayzer, Paul Shabajee<sup>2</sup>

Digital Media Systems Laboratory

HP Laboratories Bristol

HPL-2003-130

July 3<sup>rd</sup>, 2003\*

This paper sets out an approach which we call *semantic blogging*. We start from the observation that blogging is a highly popular and effective approach to information sharing. We then assert that certain ideas taken from the semantic web research programme can enrich and extend the blogging paradigm. We describe what we mean by semantic blogging, and why this approach is beneficial. We are building a demonstrator, set in the context of small group bibliography creation and management, which will illustrate the advantages of our approach.

\* Internal Accession Date Only

<sup>1</sup> BlogTalk Proceedings, 2003

<sup>2</sup> Graduate School of Education and ILRT, Bristol, UK

© Copyright Hewlett-Packard Company 2003

Approved for External Publication

# Semantic Blogging and Bibliography Management

[Steve Cayzer](#), HP Laboratories, Bristol, UK

[Paul Shabajee](#), Graduate School of Education and ILRT, Bristol, UK

## Abstract

This paper sets out an approach which we call *semantic blogging*. We start from the observation that blogging is a highly popular and effective approach to information sharing. We then assert that certain ideas taken from the semantic web research programme can enrich and extend the blogging paradigm. We describe what we mean by semantic blogging, and why this approach is beneficial. We are building a demonstrator, set in the context of small group bibliography creation and management, which will illustrate the advantages of our approach.

## Proposal

This proposal is part of the Semantic Web Advanced Development for Europe ([SWAD-Europe](#)) research effort. The brief is to develop a demonstration application which illustrates the advantages of a semantic web approach. A basic intuition of our thinking can be gained by imagining a utility that enables a user to aggregate the latest blogs *about a particular topic*, e.g. Java the island.

Our approach involves attaching *semantics*, or meaning, to web markup (*metadata*). To understand this process, it is helpful to consider an everyday example. English speakers tend to attach semantics to the word "cat", whereas the Viennese equivalent "katze" might simply be seen as an sequence of letters; an opaque symbol. The [semantic web](#) defines standards for attaching meaning to symbols in a way that computers can process. The symbols are mapped to concepts and relationships, which are formally described using an [ontology](#). Using these principles, we are building blogging tools which associate symbols with appropriate concepts, and which use this understanding to provide enhanced services. This is what we mean by *semantic blogging*.

We have chosen blogging because it is already used to great effect in our domain of interest, which is the sharing of small items of information. It provides a low barrier to entry for personal web publishing and yet, through automatic syndication and aggregation mechanisms, the blogs are accessible to a wide community. Blogs have a simple to understand structure and yet links between blogs and items (so called *blogrolling*) support the decentralized construction of a rich information network. While we propose extending 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.

We enable semantic blogging by annotating our blogs with meaningful symbols. This *semantic structure* has two key effects:

- **Rich Query:** Semantically enriched blog metadata enables new subscription, discovery and navigation behaviours.
- **Rich Structure:** Access to semantic markup enables both richer annotation and sharing of higher level structures (like categorisation schemes), encouraging peer commentary and recommendation activity.

There is some movement in the blogging community towards what we call semantic blogging. For example, some blog commentators envisage [connecting blogs using semantic links](#), and the [Topic Exchange](#) activity is a step towards the use of shared ontologies. These developments indicate that there is a real need for the capability that we are proposing.

To demonstrate the benefits of semantic blogging, we are building a tool set in the context of small group bibliographic management. This domain allows us to test our intuitions about semantic blogging in a concrete and grounded setting. It is also an area where existing tools are [generally unsatisfactory](#) for the purposes of small groups. We believe that semantic markup can add real value in this domain.

We want our demonstrator to illustrate semantic web principles by providing genuinely useful functionality to bloggers.

## Paper

In this paper, we outline our vision of semantic blogging. We start by defining the semantic web, and its relevance to blogging. We provide a motivating scenario which demonstrates the kinds of functionality provided by a semantic blog. Next we describe the semantic blogging paradigm in greater depth. Finally we provide an outline of the demonstrator we are building.

## The Semantic Web

A simple description of the [semantic web](#) is that it is an attempt to do for machine processable data what the world wide web did for human readable documents. That is, 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.

In relation to the blogging world, the most important notion is that of semantic search and navigation. Once aggregators "understand" the difference between Java the language and Java the island, and that your blog category 'Java' is a specialisation of my category 'Programming Languages', they can provide semantically relevant results rather than just syntactic matches. Moreover, these results can be sorted, filtered and presented in a meaningful fashion.

## Illustrative Scenario

Bob and his peers are interested in semantic blogging. Bob does a [Google](#) search for relevant papers, and finds some that look interesting. After having read a few, he posts the details on his semantic blog, using a low cost method such as copy & paste or drag & drop. He categorises the item, rates it and adds a free text comment. The default, chronological structure of his blog makes a useful 'reading diary'. He uses a *topic hierarchy* to navigate across his archive.

To effectively access his peers' archives, Bob performs a *community query* for an interesting sounding paper and receives a summary table with the comments and rating of his peers on that paper. Next, he performs a community query for 'papers like this' which generates no hits. He chooses the 'generalize this query option' and this time there are some related papers. Again they are presented in summary form with title, topic and rating (this can be customised) and he follows links to interesting looking papers to examine his peers' comments.

For one of the followed links, he accesses the abstract and downloads the PDF. He creates his own blog entry (using a *blog this* bookmarklet option) which automatically copies all the metadata (title, author, topic etc) created by his peer. It also creates a link between his (new) blog entry and the peer's. He can also, if he wishes, 'bulk import' peers' blog entries from the summary table. He can now add his own comment and rating, and recategorise the item if required. Finally, each bibliographic item also has a list of *citation links* - papers cited by, and citing, this paper. Bob may follow these links if he wishes to find other interesting bibliographic items, and possibly import them too.

Bob is interested in keeping up to date with semantic blogging papers. He sets up a *community alert* for the semantic blogging topic, which provides an update on any new community blog entries in this category. He sets up a web page to display a summary of these new semantic blogging entries.

Later, when writing a paper, Bob browses his bibliographic blog for papers with a topic (or supertopic) of semantic blogging and again gets a summary table. He can filter this table using other metadata (eg rating) or unstructured data (i.e. free text). He can also augment the table using a community query. Once his is happy he has the right subset of papers, he exports the data to a [BibTex](#) file for use with his [L<sup>A</sup>T<sub>E</sub>X](#) paper.

## Semantic Blogging

It is worth examining in more detail why we are bringing two successful, but distinct, paradigms (blogging and the semantic web) together. We believe that there are compelling reasons to combine the two. The rich structure and query properties enabled by the semantic web greatly extends the range of blogging behaviours, and allows the power of the metaphor to be applied in hitherto unexplored domains. Small group bibliographic management is a concrete example of a task that illustrates the benefit from the combined paradigm. 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 an ideal solution to this problem - blogging for low barrier publishing, a simple shared conceptual model, and a mechanism for natural, dynamic community formation - semantic web for rich structure, which enables richer community annotation, and rich query, which enables more powerful discovery and navigation.

We believe that the semantic blogging approach will have the following benefits:

### Rich Query

- **Subscription** Blogging as it stands provides a convenient time-based channel structure. However the channels must be subscribed to independently and the filtering within these channels is limited to free-text search. By using the semantic blogging approach we make it possible to subscribe to categories such as "what's going on in field x". Aggregation need

not be bound to a single feed, and can do semantic filtering within these feeds. It is thus possible to define a subscribed category which contains a subset of items from a number of channels.

- **Discovery** This capability means the discovery of new, potentially useful, sources of information. Blogging discovery mechanisms include directories, blogrolling, item-item hyperlinks and informal routes (such as email). The semantic web offers capabilities that allow an even richer discovery capability. For example, bibliographic items can be related via an ontology, and thus a user can query a wide community (and many channels) for papers "in the semantic web category". If items are associated with a consistent (or at least corresponding) identifier, users can also ask for blog entries "about this item".
- **Navigation** Current navigation within blogs is largely chronological, and hence not well suited for non time-based data, such as bibliographic records. We want to enable users to browse their own blog (and others') for such data in an intuitive manner (for example, across a user-defined topic hierarchy). In addition, semantic links will allow navigation to other items "agreeing with" (or "disagreeing with") this one. These links are an alternative to explicit citation-type links and thus provide a community based network based around ideas and discussions.

### Rich Structure

- **Structured Entries** Blog entries can be marked up with a variety of meaningful concepts (like author and topic) which can be used both to discover semantically related information, but also to navigate the information in a variety of useful ways. Such annotations may also be supplied with context data, such as provenance. Because we index items by URI, we enable multiple, provenanced annotations to be integrated across a community.
- **Shared Ontologies** By exploiting the semantic web ontology layer we are able to not only represent rich topic hierarchies for classifying citations, but also to link and share these topic sets across communities. Thus different communities can use distinct classification schemes and yet the data can be shared across the same infrastructure and potentially the relationships between terms in the different schemes can be explicitly represented to allow cross-community search.

### The Demonstrator

We are building a demonstrator which tests these intuitions. Our aim is for the demonstrator to be 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. It should be deployable. 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 advantages of the semantic web approach (semi-structured data, semantics and webness) without losing the key advantages of blogging (low effort publishing, easy subscription and decentralized discovery).

We expect to produce a tool which, in general terms, allows a community to effectively manage their bibliographic data, and to harness the power of the group for discovery, recommendation and collective learning. Specifically this means that the demonstrator will exhibit lightweight capture of bibliographic data, rich discovery and navigation mechanisms, useful presentation of the relevant information, and good integration with other tools. In short, an application that is genuinely useful for bibliography management.

### Conclusions

In this paper we have described how semantic web principles can be used to enrich the blogging paradigm. We have shown how the enriched semantic blogging metaphor may have many benefits for the blogging community. We are currently building a demonstrator which will illustrate some of these benefits.

## References

[SEMWEB]

*W3C Semantic Web activity*  
<http://www.w3.org/2001/sw/>

[USER\_STUDY]

*User Study* - part of:  
*SWAD-Europe: Semantic Blogging and Bibliographies - Requirements Specification*  
[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)

[BibTeX]

*LaTeX: A Document Preparation System* by Leslie Lamport, 1986, Addison-Wesley.  
*BibTeXing* ( [btxdoc.tex](#) ), by Oren Patashnik, February 1988, (BibTeX distribution).  
<http://www.ecst.csuchico.edu/~jacobsd/bib/formats/bibtex.html>