How to build a Snippet Manager

Steve Cayzer, Paolo Castagna

Hewlett-Packard Laboratories, Filton Road, Stoke Gifford, Bristol BS34 8QZ
{steve.cayzer, paolo.castagna}@hp.com

Abstract. In our research group, there is a need to capture, organize and share resources associated with a domain of exploration. We are building a tool for this task, based on previous experience in the knowledge management domain. In this position paper, we present our thoughts on what works (and what doesn't work), together with details of our initial implementation.

1 Introduction

The snippet manager idea is not a new one [1]. It refers to the idea of a small peer group capturing 'snippets' of information in a lightweight manner, categorizing and sharing them. There have been a number of approaches to this problem over the years; in this paper we present our opinions on what works, and what doesn't work. We also define the scope – even the ideal snippet manager would not be a panacea for knowledge management generally. Rather, it is a useful tool (or at least a useful concept) for a specific task. We have started to implement a tool using the principles outlined here, and we present some design details. We also introduce the success factors that we intend to adopt, and that we hope will be more generally useful.

2 The snippet manager problem domain

Knowledge management is defined widely, for example achieving a global sharing of knowledge within a company [2]. However, our domain of interest is rather more tightly defined:

"In [my group] we frequently circulate items of interest (such as news articles, software tools, links to Web sites, and competitor information). We call them ***snippets****, or information nuggets, we would like to store, annotate, and share. Email is not the ideal medium for these tasks; its transient nature means the snippets are effectively lost over time. Yet the risk from using a more formal process, like a centralized database, is that it is both cumbersome to use (a barrier to entry) and overly rigid in its data model (not amenable to storing different types of information). Our need illustrates what I call decentralized, informal knowledge management…"* [3]

The point here is that the domain allows us to make some simplifying assumptions. Firstly, the group is small, often co-located. Members can 'pop round' to discuss ideas or gather for an informal discussion. In the particular group we are designing for, there are regular, weekly meetings. So we assume that information flow is unhindered, and that conflicts or inconsistencies can be quickly ironed out. We need not rely on snippet manager as the sole conduit for communication. A small group also makes the job of converging on a domain model much easier. We do not assume that we will get the model(s) right first time, but the first pass should be good enough to get general buy-in, and improvements can occur by means of incremental evolution.

Secondly, the users are technically literate. This means that they are likely to quickly get to grips with a new tool, and may be motivated to make some small changes to behaviour for sufficient added value (a good example of this would be the use of 'graffiti' writing on PDAs). However, getting the balance right is not always easy; we note that in our (internal) semantic wiki, people hardly ever use the supplied wiki syntax to add RDF metadata. This may be due to the lack of instant, direct reward for adding metadata, a point which we are trying to address in our work.

Finally, the domain of interest is tightly focused, even more so than that envisaged by Cayzer [3]. Our laboratory is interested in a myriad of semantic web related topics, but the user group for the current incarnation of the snippet manager is specifically and actively looking at a particular topic, that of enterprise information management.

2.1 Use cases

What is it, then, that the snippet manager is expected to achieve? We are in the early stages of this project, but we have engaged the user community and gathered some initial use cases in order to inform our guiding principles, a few of which are described here. We reiterate that these principles are relevant for our domain of interest – small group, tightly focused, technically literate researchers. We don't expect the principles to necessarily generalize to the whole of knowledge management, or to the web at large. However, we believe that our problem domain is sufficiently common for these principles to be of value for the semantic web community.

Easy Capture
"I need a way to collect evidence (web pages, PDFs, emails, forum posts), and to categorise parts of these so they can be linked together for post-hoc search."
"I need a way of annotating resources with evidence - 'why have you written this'"
It should be ludicrously simple to collect snippets, using familiar methods such as bookmarklets or email. Snippet Manager should also pull in snippets from other sources (eg intranet databases) and handle provenance.

Editable Ontologies

I need to create a category or classification for a new area of interest. Now, I need to add new companies, products, documents or links and tag them with this classification. I want to create relationships between these instances - e.g. competitor links.
Our users will certainly want to change the ontologies on the fly. Although this sounds like a tall order, in our case both the structure of the ontologies (effectively taxonomies) and the nature of the changes (adding/removing/renaming a node) can simplify the implementation enormously. Of course the UI for such changes may not be trivial.

Export
I want a regular alert showing the results of a web search for a topic. [OR I want to produce a report that shows all relevant products or technologies for a given topic]
From a technical point of view, the ability to export (meta)data in a standard, machine readable way is a future-proofing mechanism, intended to prevent the portal becoming yet another information silo. From a user point of view, export in a *human-readable* form is equally important.

Web Application
Our experience with the early snippet manager prototype [1] taught us that there is considerable reluctance to download software, let alone to standardize on it across a group. In addition, the snippet manager should be integrated into a users' normal work pattern. For our group, this suggests a web application such as a portal.

Immediate Feedback
There should be an instant reward for the user who adds metadata. The community aspect should be (from the user's point of view) a beneficial side effect.

4 Implementation Details

We have used the semantic portal [4] idea and codebase to provide a browsing interface over the group's snippets. Essentially, this portal uses the metadata to drive a facet browser, so that users can find what they are looking for using a variety of search paths. We are building simple capture modalities such as bookmarklets, mail processors and web forms; and importers for other systems such as blogs, technical reports, people databases and the group's official wiki. For export, we plan RSS feeds, email alerts, customizable reports and a SPARQL[5] interface for programmatic access.

4.1 Success Factors

We have previously built several semantic web applications whose primary function was to demonstrate a particular aspect of the technology. Our focus here is to build a tool. Therefore the simplest way of assessing its success is to measure its usage:

1. At what rate is new content added to the snippet manager?
2. What proportion of the user group use the snippet manager as a day to day tool
3. How well does the snippet manager integrate with other tools in use?
4. How often is the snippet manager consulted for information or report generation?
5. What is the satisfaction level of the users?
6. How quickly can new user requirements be integrated into the tool

These measures are largely qualitative in nature. Yet they get to the heart of what of means to build a semantic web tool for personal, and group, productivity. We intend to assess our work using these criteria.

5 Related Work

Simile's Semantic Bank [6] is a snippet repository that lets you persist, share and publish data collected by individuals, groups or communities. Data capture is accomplished using a Firefox extension called Piggy Bank [7], and the information is accessed using a faceted browser. It is probably the closest system in philosophy to ours, but there are some important differences. Firstly, Semantic Bank is intended to be a general purpose, potentially global scale snippet repository. This means that there are significant research challenges in making the ontologies both sufficiently compact and understandable. In snippet manager we chose to have a small number of tightly focused facets. Secondly, our aim is to allow both the gathering of snippets and the linking of these snippets with data from other sources.

The broader idea of a semantically enabled website is explored in a number of public portals, notably the Semantic Web Community Portal [8] and MindSwap [9], both of which use metadata for filtering and querying. As the number of items increases, the value of our faceted browsing approach becomes more apparent. There are other public portals such as Ontaria [10] and SchemaWeb [11], which are primarily intended for browsing ontology data.

Many people use their weblog as a knowledge management tool and we think that structuring the content of a post by adding some metadata could be useful for a group of people. But the chronological view that weblog gives to the content not always it the best solution to let users move through information. Our solution to this, which we call semantic blogging [3], uses metadata guided views, such as record cards or tables. A similar approach has been taken by the structured blogging community [12]. A more subtle point is the information model, in which the blog entry is no longer the primary object. Rather, the information item (such as web page, report or person) which is being blogged about takes centre stage. The blog entry is an annotation attached to this item. Armed with this perspective, the semantic blog becomes a useful personal knowledge management tool, and a source of data for the snippet manager.

Wikis are also interesting tools for collaboratively building knowledge, and there are examples [13, 14] that use metadata to enhance navigation and to provide multiple views. In some ways the snippet manager idea is similar (although our data entry mechanism is different); however we integrate information from a number of sources. Just like blogs, wikis are a valuable source of data for the snippet manager.

6 Conclusion

In this position paper, we have outlined our thoughts on what it would take to build a snippet manager for small group domain-focused knowledge sharing. We have shared some design principles which we hope will prove generally useful. We have also explained how we are going about building a system using these ideas. We have high hopes that our user-centred approach will function less as an interesting demo and more as a genuinely useful tool.

References

1. Banks, D., Cayzer, S., Dickinson, I., Reynolds, D. The ePerson Snippet Manager: a Semantic Web Application. Hewlett-Packard Laboratories Technical Report HPL-2002-328 (2002): http://www.hpl.hp.com/techreports/2002/HPL-2002-328.html
2. Davenport, T. H. and Prusak, L. Working Knowledge: How Organizations Manage What They Know. (1997) Harvard Business School Press.
3. Cayzer, S. Semantic blogging and decentralized knowledge management. Communications of the ACM 47, 12 (Dec. 2004), 47-52.
   DOI= http://doi.acm.org/10.1145/1035134.1035164
4. Reynolds, D., Shabajee, P., Cayzer, S., Steer, D. Semantic Portals Demonstrator - Lessons Learnt. Sept 2004. http://www.w3.org/2001/sw/Europe/reports/demo_2_report/
5. Prud'hommeaux, E., Seaborne, A. SPARQL Query Language for RDF. W3C Working Draft 21 July 2005 http://www.w3.org/TR/rdf-sparql-query/
6. The Semantic Bank project http://simile.mit.edu/semantic-bank/
7. Huynh, D., Mazzocchi, S., Karger. D. Piggy Bank: Experience the Semantic Web Inside Your Web Browser. Submitted to 4th International Semantic Web Conference (ISWC 2005)
8. The Semantic Web Community Portal http://beta.semanticweb.org/
9. The MindSwap Group http://www.mindswap.org/
10. The Ontaria project http://www.w3.org/2004/ontaria/
11. SchemaWeb: RDF Schemas directory http://www.schemaweb.info/
12. Structured blogging http://structuredblogging.org/
13. Aumueller, D.Semantic authoring and retrieval within a Wiki. Demonstration track, 2nd European Semantic Web Conference (ESWC 2005). http://wiki.navigable.info
14. Tazzoli, R., Castagna, P., Campanini, S. E. Towards a Semantic Wiki Wiki Web. 3rd International Semantic Web Conference, Poster Track (ISWC 2004).
    http://platypuswiki.sourceforge.net/