Mapping the Semantic Web: A Grassroots Approach

[author]

[affiliation]
[email]
[url]

Abstract. “Find what I mean, not just what I said” is perhaps what users desire most from the Semantic Web. In this paper, the author suggests that average users can help define meaning for new and existing content, and offers specific proposals to facilitate concept mapping and information discovery. In particular the need for permanently available Abstract URIs is discussed, and a possible model for a concept discovery infrastructure analogous to the Domain Name System is presented.

Keywords: SemanticWeb, taxonomy matching, information discovery, taxonomy mapping, ontology matching

1 Introduction

In the Scientific American article that popularized the term 'Semantic Web' (1), Tim Berners-Lee and coauthors offered an ambitious use case: an individual attempting in real time to choose a medical specialist by a combination of reputation, insurance coverage and location, coordinate these criteria with another trusted individual (his sister) and then schedule an confirm an appointment for a third person.

Reaching for such an ambitious goal can be of great value in guiding the development of the multiple subsystems needed for W2. In this article, however, the author presents a serious of less ambitious use cases that are nearly attainable with current technology. the author will argue that the missing piece needed is a form of Abstract URI and present a possible roadmap for implementation.

1.1 An Alternate Use Case

The protagonists of our use cases are a mother and her a small child, "Meme" and "Memia". Today she has a routine doctor's appointment including some vaccinations. Recalling from news reports that there is some issue with preservatives in vaccines, Meme checks the paper inserts before approving the shots, and holds off on the ones that have preservatives. Speaking into her smart phone, she orders: "Take a note. File under Memia - Medical. She reads off the vaccine manufacturers name and a few of the chemical preservatives noted on the insert. She presses one button to mark completion of the voice note, then adds a command: "Search related terms: vaccines, safety, mercury, autism". Thanking the nurse she returns the paperwork, and comforting Memia after receiving the preservative-free shots the two return home.

That night, Meme logs into her personal blog/organizer to find a text version of the note she recorded earlier, along with the results of a search agent displaying highly recommended articles in specific categories that matched both unusual words in her voice note and one or more of the search terms she specified. After reading several items referenced under

Medical/Pediatric/Vaccines/Tripedia
and Medical/Pediatric/Autism/Thimerisol

Meme finds that while Thimerisol was by FDA recommendation not supposed to be in vaccines any longer, in fact it is still is in the supply chain. After forwarding a recommendation for an alternative to her doctor, she decides to make public her firsthand notes as open content under the relevent categories. While Meme doesn't have the time to try to get thru to the FDA and prevent other children from receiving a potentially dangerous substance, or even to write a polished and entertaining blog article, she can at least click a single button to make her rough note document the fact of its use. [remove? Aggregators add her note as type 'First hand report" in the specified categories.] She also takes a second to click the "!" icon on the most helpful article, resulting in an incrementally higher score in relation to the keywords she searched on. Finally, she tags the article for her own reference as simply "FDA article" and automatically files it where she originally took the note, under 'Memia/Medical' along with her daughter's medical files.

A few days later Meme visits one of her favorite Aggregator sites to search for items of type 'tip' under her already configured tags of 'early math skills' and 'learning games'. She quickly finds and prints a few new useful items, one by a parent she's previously found helpful and has added to her trust/esteem/high-relevence net. Having an extra minute to spare she browses briefly on public parts of her distant cohort's blog/organizer and adds her own comment on a game she's tried in the past.

Then its time to go out for the day, and at the zoo Meme and Memia make the accidental discovery (when Memia drops her cookie) that the Blue Macaw in the South American exhibit will talk back if someone says "Uh-oh". Charmed, Meme whips out her smart phone again and records a public note with a photo under her tag 'zoo', which she has previously mapped to the Tucson Zoo.

When they get home, they log on and discover that not only had a few other families noted the same thing, but the Macaw had been owned for years by a local family, was named 'Butler', and would also respond to 'May I help you?' Reading up on some of the other animals' recent and past histories, Memia is eager to return and see more of her furry and feathered friends. Meme, an avid birder, quickly notes the recent appearance of a wild Reddish Egret visiting the Flamingo exhibit, and happiliy agrees to a return trip.

1.2 Need for Mapping to Precise Abstract Concepts for Efficiency

Meme could have performed similar tasks on the existing Web, but with significantly less efficiency. The difference between seconds, milliseconds and minutes, between 90% relevent or 90% irrelevent results, is not merely of quantitative significance. The uses of the Semantic Web in real life and real business will depend hugely on its efficiency. An activity that takes 5 minutes is not going to be integrated into routine daily tasks.

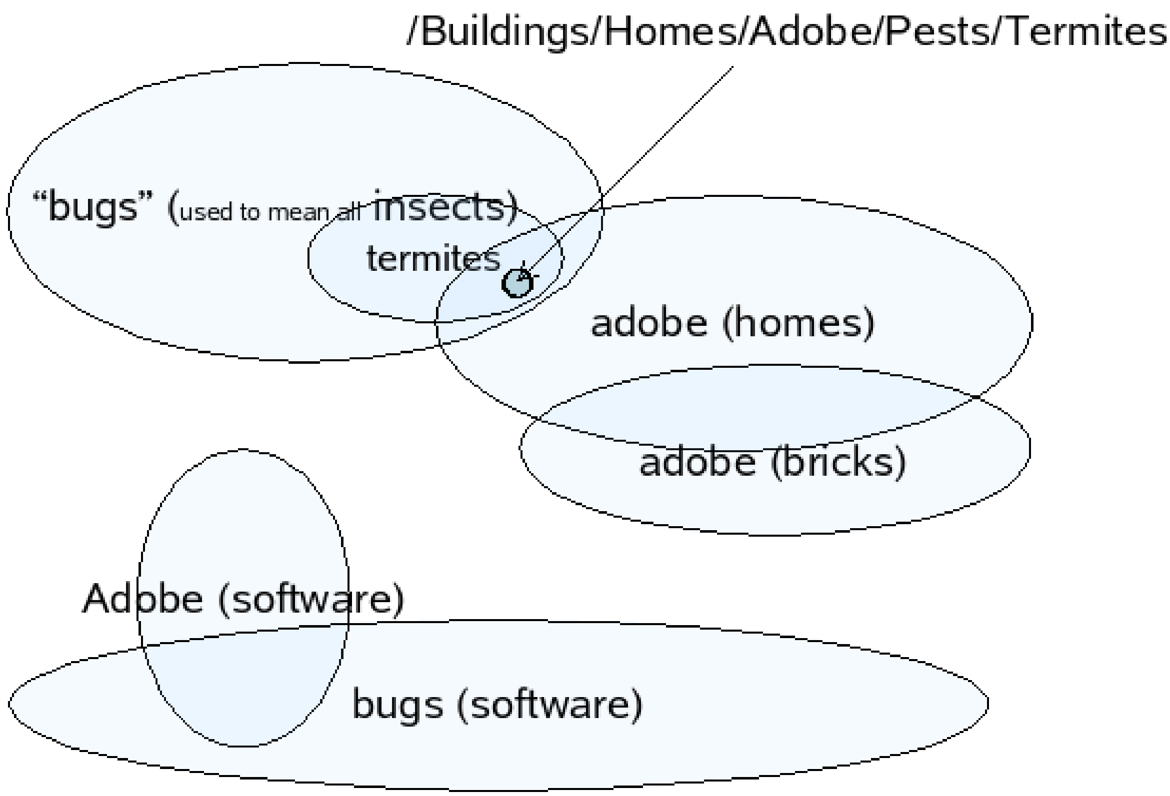
In our hypothetical scenario, Meme was able to perform rapid discovery and posting of information because she was able to map her needs and ideas to well defined abstract concepts and from there to documents matching those precise concepts. What is missing in the current Web is a widely used method of mapping to well-known abstract concepts.

2 Coordinate Systems for Information Space

Consider a concept as a point in 'Information Space' or "I-S" for short. An ontology, taxonomy or folksomy then acts as a coordinate system for some portion of I-S, and a mapping as a coordinate transformation. A particular coordinate system may magnify or distort the shape of the space, making it easy and efficient to represent a certain collection of points. The ability to transform from one representation to another allows users to search and post using a comfortable set of coordinates that is familiar and compact, yet still share information with the rest of the web.

Full text searches essentially use words in the language as the coordinates to a point. Until natural language processing is more sophisticated, any structure of the language in the query is usually lost. Unfortunately, this means that individual words are often ambiguous. An example is hardly necessary (search for 'adobe bugs' when looking for insects inhabiting mud houses and you will be disappointed to find instead errors in Adobe software; on the other hand a search for "OWL collection" returns more references to stuffed birds than it does to sets of ontologies), and with more and more areas of information added to the web ambiguity will become a greater and greater problem. Full text searching within a specified context across the entire web is a desirable capability. To achieve this a large portion of existing documents would somehow need to be related to known points in I-S.

Figure 1: words and categories as regions in information space



Such relationships need not be absolute. In 'Semantics for the Semantic Web: the Implicit, the Formal and the Powerful', Sheth et al discuss the use of probabilistic and fuzzy logic approaches to assign membership to a class: "Obviously, there are two ways of doing that [assigning membership] - manual assignment by domain experts and automatic assignment using techniques such as machine learning."(2) A third way may be to exploit the 'many eyes' of a large user base. This grassroots or market-based approach seems particularly suited to probabilistic assignment. Key to making a grassroots approach effective is the wide availability of easy, fast tools for annotation and categorization.

As TBL points out, the RDF spec already "enables anyone to define a new concept, a new verb, just by defining a URI for it somewhere on the web" (1), and many existing OWL implementations do just that. Using arbitrary URIs for identifying concepts has several drawbacks, however, at least for 'normalized' systems that other users are expected to map to. URIs are owned and may change arbitrarily. Concepts exist independent of ownership, and once referred to, ideally should not be unexpectedly altered.

Several authors have discussed this problem from different perspectives: in "Disambiguating RDF Identifiers" (3) Sandro Hawke discusses the problem of not knowing whether a particular URI is to be interpreted as a subject or a web page; and the 'GoodURIs' w3 wiki (4) discusses several desireable attributes for URIs generally and for use with RDF. Let us take a step back and examine what are the desireable attributes for addressing a concept, with the goal to make it easy for users to relate their own information to it.

A **Concept Identifier** should be...

persistent	a user referring to a concept should not have to change the reference at a later time. If a concept identifier system needs to change with time it may extend without modifying existing concepts or use versions.	The PURL project (5) addresses this to some extent, but not entirely - a concept shouldn't depend on any actual URL
compatible	the overall system of concept identifiers should be compatible with / able to include existing systems such as the Dewey Decimal system, the Library of Congress Classification system, international ZIP codes, ISBN numbers,	The info URI scheme addresses this issue (6)
available	not necessarily browseable, but in some manner the concept identifier system needs to be easily available and ideally free to use by the intended audience Rather than 'browsable' via http, the hierarchy should be easily traversable by a concept chooser tool.	dmoz.org is a good example of a freely available concept hierarchy
comprehensible	a concept identifier should be immediately comprehensible to an end user, or have an expanded form that is comprehensible, so that users can participate in mapping items to it	
structured	a URI of http://www.co-ode.org/ontologies/pizza/2005/10/18/pizza.owl#Mushroom may be comprehensible, but it has no inherent information content in the way that a URI such as http://dmoz.org/Business/Food_and_Related_Products/Baked_Goods/Pizza/Mushroom would have. The latter URI can be meaningfully operated on to sort or relate information by category.	
extensible	The URI http://dmoz.org/Business/Food_and_Related_Products/Baked_Goods/Pizza/Mushroom does not actually exist; the real Open Directory Category ends with 'Pizza'. Although a user cannot extend dmoz.org directly, there should be some way for the user to build on an existing scheme.	

As the Agile Development with Rails authors Thomas and Hansson note, "In the software world, we often ignore good ideas from the past as we rush headlong to meet the future." (7) The use of existing, direct-reference URIs as concept identifiers seems to ignore a simple lesson from the venerable Dewey Decimal system: "The main innovation and advantage of DDC is that it's an indirect, rather than a direct, reference to a book's location. " (8)

A concept URI acts as an indirect reference, as opposed to a resource URI that points to a particular, possibly transitory, page on the web.

Of course, no single concept address schema can be authoritative, but having a few good, widely-used schemas that can be easily mapped to each other would seem to be an essential tool for building and annotating the Semantic Web. Perhaps one way of building 'a' useful schema through a grassroots approach would be to harness the power of a wiki.

> Proposal: The NISO AX committee (9) add a Core Namespace of
>
> info:ofi/nam:wiki:
>
> and register to a trusted, well-established organization such as the Open Directory Project (ODP) (10), w3.org, or wikipedia.org, that is interested in managing the space. It could be initialized with the current set of ODP categories and then opened up to allow users to add subcategories subject to editing in a similar manner to wikipedia.org. *

The mere existence of a "good" concept address namespace is not sufficient, of course. For average users to participate in annotating the Semantic Web, they will need easy and ubiquitously available tools, integrated into familiar settings. In addition, efficient aggregation of user annotations may require the existence of brokers or even a central registry. (Essentially an extension of the Annotea project(11)). These issues are discussed further in the Infrastructure and Roadmap sections below.

[footnote] * It is not necessary that the info: protocol be used for creating a set of concept URIs. Another abstract-use protocol would do as well, or even a set of http: addresses with the understanding that the URIs could be valid to use as concepts even if they were no longer browsable.

3 Infrastructure for Information Discovery: Infoservers as analogue to Nameservers

or, Who Knows About X and How Do I Talk to Them?

Regardless of the particular URI used to identify a concept, an additional challenge the Semantic Web must overcome is the efficient discovery of specific information. The current most popular methods of searching the web (ie Google) depend on the search portal provider to have spidered the entire Web, and have ranked the keyword-specific relevence of each document. Since there exist numerous search portals, a huge amount of traffic is dedicated to a constant re-spidering of the entire Web by every contender. Even if a search portal wishes to focus on a specific content area, they have no efficient way to examine only relevent pages. Each must attempt to repeatedly examine a snapshot of all accessible human information - which continues to grow at a rate even faster than hardware storage capabilities.

With the advent of the Semantic Web, we should see more specialized search portals, each indexing a relatively precise and complete cross-section of the Web. These sites could query each other in a mesh similar to the current DNS system, with discovery and registration of information distributed across numerous 'infoservers'.

Proposal: A W3 working group be created to develop a WHOKNOWS protocol analogous to WHOIS(12), for the discovery of information related to a specific concept.

Infoservers, roughly analogous to DNS servers, would register resource URIs related to a specific concept URI, as well as additional information on query fields and result formats available. They would accept queries by concept URI, returning a list of resource URIs, their relations to the concept URI, and optionally interface specifications for each.

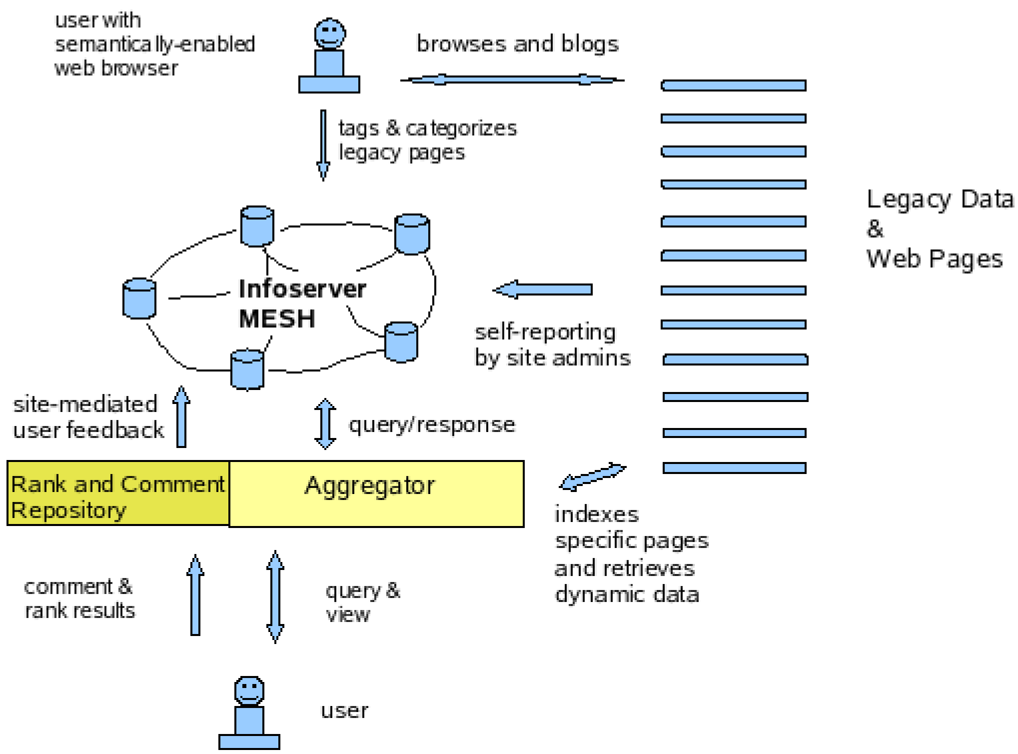
Unlike DNS servers, Infoservers would operate in a mesh with shared authority. An important principle of an extensible concept addressing system is that

Concepts exist independent of ownership

Not only can no one can own or control use of a concept, but concepts may continue to be used even after the originator of the concept has discarded it. It should not matter if a particular resource URI exists on the web, the idea defined by a concept URI should still be able to be referenced.

Infoservers would not themselves attempt to store all information related to a specific concept, and would not attempt to present summaries directly to users. Rather, they would provide a location to register data sources (resource URIs) related to a concept, and to allow any user to assert such a relation. Other servers, let us call them Aggregators, would be free to use the information provided by the Infoservers to assist them in creating a subject-specific portal site. An Aggregator might query the Infoserver mesh to discover new data sources, then retrieve and index data directly from those sources, and using its own ranking algorithms present a summary of results in response to a users query.

Figure 2. Infoserver Mesh receives RDF assertions from multiple sources about legacy pages



Infoservers as envisioned here are essentially **Annotea** servers (11) with a few additional capabilities and requirements added.

Note that several classes of users are presumed. Data server administrators are expected to be able to register details about the interface to their server (possibly with help from a form parser or other appropriate tool). Semi-technical bloggers are expected to be able to use a sufficiently user-friendly tool for choosing categories and relations, or to use pre-mapped tags provided by a site administrator. Aggregator end users, generally non-technical, are expected to be able to click an icon next to a search result to indicate simple positive or negative relevence of the result. A Trust Web similar to that discussed by Guha (13) will be needed as well, and issues of how to identify users and organizations in an open system will need to be dealt with.

3.1 Possible Requirements for an Infoserver Implementation

In order to share authority effectively, Infoservers will need to fulfull a set of requirements something like the following:

an Infoserver...

- MUST register the set of concept URIs it wishes to be responsible for with a central registry

- MUST accept simple submissions of RDF assertions that fit its registered criteria (with wide latitude for rejection of spam-type submissions/determining trusted sources). Simple assertions take the

form of a triple
(Resource URI, Relation URI, Concept URI) plus some form of ID for the submittor

- MUST respond to simple WHOKNOWS queries containing a concept URI with a summary of available data relating to the specified concept

- MUST maintain a current mirror of the full concept hierarchy, at least one Infoserver from each registered mesh and provide to users upon request.

- MAY support additional query and report fields/categories and report availability of such fields as part of the query/response

- MAY develop appropriate subject-specific templates for submitting detailed interface information for a resource, and make such templates available in response to queries.

- MAY use existing ontologies such as those available in SUMO (14) to present templates for resource description

- MAY limit the types and sources of data it will accept, and may set a Time To Live expiration time on submitted data

- MUST share data with other Infoservers in the mesh (ie others which have registered for the same or overlapping concept URIs). This may be implemented with a notification system at set intervals along with a request/response protocol for sending updates.

- MAY accept OWL queries.

Why a mesh, when it adds additional requirements as compared to a single all-knowing central repository? While a single server or network could certainly handle the CPU and traffic load, a centrally controlled system would not allow distributed development of the many subject-specific query fields and formats, or for innovation on trust, ranking and other subsystems. The goal of the proposed Infoserver MESH is to provide an information discovery infrastructure while allowing maximum independence of the individual nodes.

The Aggregators are envisioned as privately controlled and unregulated servers taking advantage of the information provided by the Infoservers. Different Aggregators may use entirely different algorithms for determining what actual data to index and present to the user, how to rank query results and how to integrate 'Semantic' data with traditional text-based indexes. Decoupling discovery, reporting, spidering/indexing, ranking and presentation as much as possible allows greater possibility for innovation. In particular specialized Aggregators can add value based on specialized knowledge of the subject domain, while still being tied into the central registry.

3.2 Use Cases and Examples

<u>Example 1: First-hand reporting on a specific event</u>
User who was present at a protest march uses a blogsite's built-in event reporting tool (with Who/What/When/Where/Why fields, plus a category chooser) to create an blog entry about their experience. The web server automatically submits the triple

Resource URI: http://blogserver.com/my/blog/entry934
Relation URI: info:ofi/nam:wiki:/Meta/Relations/First_Hand_Report
Concept URI: info:ofi/nam:wiki:/News/Current_Events/Political/Protests/Marches

to an Infoserver handling the /News/Current_Events category, along with location and time in an RDF fragment like so:

```
<rdf:RDF
xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
xmlns:rel="http://infoserver-central.something/2006/12/31/info-
relations-ns#"
>
<rdf:Description
 rdf:about="http://blogserver.com/my/blog/entry934">
 <rel:First_Hand_Report>/News/Current_Events/Political/Protests/Marches
</rel:First_Hand_Report>
<rel:Located_At>/Regional/North_America/United_States/Arizona/Localitie
s/T/Tucson</rel:Located_At>

<rel:Effective_Date>2006/01/12</rel:Effective_Date>

<rel:Asserted_By>/Users/Identified_by_Drupal_Login/blackbird2001</rel:A
sserted_By>

</rdf:Description>
</rdf:RDF>
```

Note that the /Meta/Relations category has been converted into an RDF namespace; there should be some kind of autoconversion so that the namespace itself is extensible. In this example we use general-use named relations 'Located_At' and 'Effective_Date' rather than an event-specific namespace. Other things than events can be located at or dated; as much as possible we reuse simple relations rather than branching into multiple namespaces, to facilitate searching accross different types.

The actual text of the report is not submitted, only the category, location and date. The user may have used an existing category or have created a new subcategory using a simple chooser. Requests for new categories are managed at the discretion of the receiving Infoserver; otherwise data may be registered under the nearest existing category.

<u>Example 2: Assertion about an existing site</u>

User discovers a neat site and bookmarks it into a personal category that they have previously mapped to a standard one. Their personal organizing software registers the assertion with the nearest interested Infoserver:

(personal tag: Physics of Music/Spectrograms)
Resource URI: http://nastechservices.com/Spectrograms
Relation URI: info:ofi/nam:wiki:/Meta/Relations/Related (the most basic relation...)
Concept URI:
info:ofi/nam:wiki:/Science/Technology/Acoustics,_Ultrasound_and_Vibration/Spectrograms/

A more precise user might have submitted this assertion with the relation

Relation URI: info:ofi/nam:wiki:/Meta/Relations/Has_Examples_of

While a searcher would probably receive results influenced by both assertions, the more precise entry might result in a higher ranked result given equally trusted sources.

Example 3: Registry of existing CGI interface

System administrator of a publicly available database submits the URL of the CGI interface, and using a helper tool creates a mapping between some of the fields and a standard template, adding descriptions of the non-standard fields. This mapping will be used by Aggregators, which may submit a user's query to multiple sites using different interfaces, then aggregate the results along with full-text and specific semantic categories.

Thus, to become a functioning part of the Semantic Web, it is not necessary for legacy databases to convert to a standards-compliant interface. It is only necessary for someone to create a mapping with the existing interface.

4 Implementation Notes and Roadmap

4.1 Concept URI: Some Implementation Considerations

If, in addition to the :wiki: namespace, the not-to-onerous requirement is added that each category shall have not more than 255 subcategories, and that there shall be a maximum of 64 levels of subcategory, then it will be possible to encode each concept URI in a 64-byte string or 'catcode'. Such a compact encoding lends itself to efficient implementation of searches, sorts, queries and inter-server communication. See code snippets from "Real world implementations and examples" section below.

Because of the possibility of deletions and changes, it is also essential to have a versioning mechanism. The concept URI might be expanded to

info:ofi/nam:wiki/[N.nn]/category.../

where N.nn is the version number of the schema. The active schema subject to editing would always carry an odd version number, bu at predefined intervals (perhaps monthly) the schema would be frozen and assigned an even version number. It is likely that the changes would be a small percentage of the namespace, so queries using an incompatible version number could still be attempted with a fair likelihood of success. Three bytes assigned to the version number, given monthly reversioning, allow for over a million years of updates.

For certain categories a multi-byte subcategory is desired, such as

info:ofi/nam:wiki/Location/By_GIS_Coordinates/////[latitude]/////[longitude]

The multiple slashes could indicate to the URI parser to allocate additional bytes for the category which follows. An actual GPS URI might thus read

info:ofi/nam:wiki/Location/By_GIS_Position/////3939.7N/////10506.6W

Special subcategories could also be assigned to Meta information, such as Relations and Types. User tools would allow tagging within a local context in order to present a readable and familiar interface. Default sets of tags common to vertical markets would need to be made available to jump-start implementation.

4.2 From Here to There : A Local Implementation

Although the main point of this paper is the usefulness of a semantic infrastructure, including an extensible namespace,central registry and communication protocol, such things obviously take time and committees. Actions that can be taken locally by individual users or at least website creators have no such hurdle.

Links between documents are interpreted in the PageRank world always as a recommendation of importance. However, human document authors may intend something entirely different. There already exists an attribute of the anchor tag that indicates explicit RELATION between documents, but that currently recognizes a limited set of attributes such as Next, Prev, Chapter, Section and so on. (15) It would not break most browsers and could provide spiders richer semantics if the REL attribute of an anchor tag could be extended to any semantic relation, ie

```
<A HREF=”http://someexternalurl.com” REL=”RECOMMENDED”>
```

would mean that the author of the current document does indeed recommend the remote link. But

```
<A HREF=”http://someexternalurl.com” REL=”REFUTED”>
```

would mean something quite different - the current document refutes the argument in the linked document. Further, it would be possible to explicitly relate remote documents to specific concepts without submitting to any server by using links such as the following:

```
<A HREF=”http://www.harmsy.freeuk.com/mercury.html”
REL=”RELATES_TO:/Science/Astronomy/Solar_System/Mercury/”>Mercury</A>
<A HREF=”http://www.mgoldmandds.com/detox2.htm”
REL=”RELATES_TO:/Science/Chemistry/Elements/Mercury/”>Mercury</A>
 <A HREF=”http://www.zyra.org.uk/mercury2.htm/
REL=”RELATES_TO:/Arts/Literature/Myths_and_Folktales/Myths/Roman/”>Mercury</A
>
```

Again, a concept chooser tool as part of the editor or generator would be needed to make use of such a tag sufficiently convenient. But, if it were to increase the ranking of a page and its standing as an authority, many authors probably would be willing to do the work to generate accurate relations.

4.3 Existing, Partial Real-world Implementations

The author has been “In Search of a Good Search Protocol” (16) since participating in an Information Superhighway Infrastructure grant committee in 1994. In 2003 Julian Lishev, author of the CPAN Categories.pm module (17), graciously made some modifications to include relations between categories and released it privately as RCategories. Based on this module the author has developed a small application called Abra which can function as a personal organizer or a simple CMS system for a website. It was used among other things for the grassroots support site (18) in the John Edwards primary campaign, and was linked directly from the official candidate site.

Although much of Abra needs improvement and is not particularly worth noting, the core engine which is based on 64-byte category codes (as described above under Implementation Considerations) has perhaps a few reusable bits. The power of category codes is that it becomes possible to select all items which belong to any subcategory of a given category with a single SQL statement:

```
SELECT [desired fields] from rcatdb_items,rcatdb_categories where
LEFT(itemcode,$lvl) = LEFT(catcode,$lvl) and rcatdb_categories.id = $curcat and
[other critera such as security]
```

The LEFT() comparison immediately determines if the given item's category is a subcategory of the currently active one. Being able to efficiently segregate items by category subtree in this way also makes it possible to easily apply views, security and other data model features based on category, so that more than one application can be run off of the same database and code. In fact the author uses a single installation as a personal organizer and to generate several websites and blog interfaces.

When run as an organizer, the interface displays the actual categories created by the user most prominently and in place of standard menu items. Thus the concepts meaningful to the user take the place of having to learn a set of concepts meaningful to the program author. The intention of 'Abra' is to function as 'A Brain extension' in creating a personal mapping into information space.

5 Conclusion

The Semantic Web will continue to emerge whether or not a particular set of concept URIs comes into wide use, and whether or not any concept-discovery infrastructure is created. These two technologies might however speed its acceptance and in particular allow 'regular' users to help semantically map both the existing web and new content as it is added.

Sharing information in a precise and efficient manner is fundamental not only to technological progress but to society and the economy as well. Economist Friedrich Hayek argued that the ability to coordinate dispersed knowledge and allow that knowledge to be used by others provided the central advantage of a market economy. (19) Hayek viewed the price mechanism as the method of sharing and synchronizing knowledge - but price is a blunt instrument compared to what is now possible online.

A final thought: in his classic book *Working*, Studs Terkel describes a broadcasting CEO who dictates over 100 letters before breakfast and keeps five full-time secretaries constantly busy. (20) A combination of existing technologies for voice recognition, error correction (for asynchronous transmission over lossy connections) and reasonably intelligent message handling software could put equivalent resources within reach of the average consumer. With the resulting explosion of content, we will need Semantic Web technologies more than ever.

(1)T. Berners-Lee, J. Hendler and O. Lassila. The semantic web. *Scientific American*, May 2000

(2)A. Sheth, C. Ramakrishnan and O. Thomas. Semantics for the Semantic Web: The Implicity, the Formal and the Powerful. *IJSWI*, Jan-Mar 2005

(3) S. Hawke, Disambiguating RDF Identifiers, http://www.w3.org/2002/12/rdf-identifiers/

(4) GoodURIs - ESW Wiki, http://esw.w3.org/topic/GoodUris

(5) Persistent Uniform Resource Locator project of the OCLC, http://purl.org

(6) "info" URI Scheme, http://www.info-uri.info

(7) D. Thomas, D. H. Hansson, A. Schwarz, T. Fuchs, L. Breedt, M. Clark, *Agile Web Development with Rails*, p 12

(8) “What's so great about the Dewey Decimal System?”, http://www.straightdope.com/mailbag/mdeweydecimal.html Jan 2006

(9) The OpenURL Framework for Context-Sensitive Services Standards Committee AX, http://www.niso.org/committees/committee_ax.html

(10) Open Directory Project, http://dmoz.org

(11) The Annotea Project, http://www.w3.org/2001/Annotea/

(12) WHOIS Protocol Specification, http://www.rfc-archive.org/getrfc.php?rfc=3912

(13) R. Guha. Open rating systems. Technical report, Stanford University,2003

(14) SUMO, the Suggested Upper Merged Ontology, http://www.ontologyportal.org/

(15) Basic HTML data types, http://www.w3.org/TR/html401/types.html#type-links

(16) In Search of a good Search Protocol: Concept-based and context-sensitive searching. http://abra.info/searchrfc.html

(17) Categories - Create and process categories within MySQL DB http://search.cpan.org/~jlishev/WebTools-1.27/lib/modules/Categories.pm

(18) http://winwithedwards.com

(19) B Caldwell, *Hayek's Challenge An Intellectual Biography of F.A. Hayek* as quoted at http://www.press.uchicago.edu/Misc/Chicago/091910.html

(20) S. Terkel, *Working,* p. 390 1972 The New Press.