Let the Citizen Speak: A demand-driven e-government portal using Semantic Web Technology

Erhat Ilgar
erhat.ilgar@ordina.nl

John Oldenhuizing
Zenc, The Hague
john.oldenhuizing@zenc.nl

Peter Mika
Vrije Universiteit, Amsterdam
pmika@cs.vu.nl

Introduction

Opening up a catalogue of governmental information and service offerings through the Web is just as much an organizational challenge as a technological one. While in practice governments operate as a loose network of organizations with significant control over their own information provisioning through the Web, citizens expect 'the government' to provide coherent answers even to questions that cut across the internal lines of authority. Further, when citizens speak, they want to be heard: this means that they cannot be sent from site to site, presented with irrelevant information or information that is in a form which they don't understand and cannot directly act upon.

Semantic technologies have been recently considered as a solution to address the problems of matching 'supply, supply and demand' in e-Government. Web-based ontology languages (RDF/OWL) that have been designed to operate in distributed environments are widely expected to help in bridging the gap between the decentralized mode of operation of the government and its citizens' need for centralization when it comes to answering questions related to their actual problems.

The application of ontologies and ontology-based metadata as used by the British and German governments for example, covers a number of problems on the supply side of the desk. Ontologies of various expressivity (taxonomies, controlled vocabularies) improve the interchange of syntactic metadata (e.g. time, format, creator, rights) and cluster e-Government documents along key topics. Although these initiatives are still based mostly on XML technology, when employing the full power of web-based ontology languages the list of topics can be (1) extended locally and (2) adapted over time without necessarily breaking interoperability with global versions or previous versions, respectively.

While addressing problems on the supply side, these ontologies -designed for use by government employees to annotate data during information production- provide little relief in matching supply and demand. They reflect the way in which the government conceptualizes its domain, both in the terms chosen and the relationships represented. Such an ontology, however, does not capture the language of the citizens who are querying the information supply.

A crucial point we argue in this paper is that it is not feasible nor desirable to manually engineer a 'citizens ontology' in the same way one would go about engineering an e-Government ontology. The government is not in a position to create an ontology for the people nor would it be possible to maintain such an ontology due to the volatility of the demand side and the resulting ontological drift.

We propose that use of techniques collectively labelled as *emergent semantics* might show a way out of this dilemma. Instead of manually engineering an ontology for a community, emergent semantics proposes to observe ontologies emerging from the way in which people interact with an information system (Mika 2005). Instead of being a contract of the majority with limited validity and in constant need of maintenance, the resulting ontologies circumvent the problem of ontological drift by dynamically tracking the changing ways in which the people conceptualize their domain.

In this paper we show a simple case of applying the idea of emergent semantics to matching sources on the supply side of e-government as well as matching supply and demand. This case highlights how an existing e-government portal could benefit from semantic technology.

The e-government portal overheid.nl

Overheid.nl[1] is the main access point to government services in the Netherlands, maintained by the Dutch government agency ICTU on behalf of the Ministry of the Interior and Kingdom Relations (MinBZK). The website provides browsing and search facilities over the content found in nearly 1300 governmental websites in the Netherlands.

Since the deployment of overheid.nl release 3 in 2003, the ICTU has been looking for ways to improve the website under the motto "no wrong door": every navigation path should lead to relevant information - the sooner the better. The current site is less than ideal from this respect. Browsing, for example, begins by choosing the type of information (e.g. government organizations, topics, products, laws etc.) or focus group. This choice leads to separate, largely overlapping but mostly unconnected taxonomies. For example, there is no connection between the government topic Nature and Environment and the related product category Nature

[1] `http://www.overheid.nl`

and Landscape. The search facility is also loosely connected to these hierarchies and returns a flat list of webpages that simply contain the exact keywords the user typed in.

Clearly, a way needed to be found to integrate the taxonomies of the supply side as well as to relate them to search questions (queries) posed by the visitors of the website through the search interface. The solution required us in the first step to translate the existing XML schemata into an RDF representation. The resulting ontology contains the three main hierarchies used in the current website (product categories, topics and organizations) as well as a list of products, laws and dossiers. The categorization of products according to topics was also imported.

Once in a common semantic format, the ontology was enriched in two ways. First, some relationships were entered manually, for example dossiers were assigned to topics as well. Second, looking for a way to enrich the ontology with the demand side's viewpoint we used an analysis from a previous Zenc project covering the search log of the portal for the period 2003-medio 2004. The conclusion of that analysis was that all 1,7 million queries collected could be reduced to only about 4,000 unique and relevant keywords (see (Titulaer, Peters, & Oldenhuizing 2005)). (This lexical analysis also yielded a set of synonyms and spelling variations for these terms.) These search terms were then assigned to those dossiers, organizations, laws or products that would provide the most relevant information for a given search term. While in this case this assignment was carried out manually, it would be clearly possible to take into account what the users have found relevant from the results of the search.

The ontological classes of the three hierarchies were thus linked through shared keywords, which are terms that were suggested by the visitors themselves. The advantage of such an emergent ontology mapping lies in its flexibility. Should the users vocabulary or conceptualization change, the terms' changing relevancy to certain classes would also immediately effect the relationships between classes. For example, depending on the time of the year, the relevancy of certain terms changes: while the query "tax" should return the link to tax forms when issued around time of the year when tax declarations are due, while it might be more relevant to budget issues at a time when the government is discussing tax policy for the coming fiscal year.

Prototype implementation and Future Work

We have implemented a prototype system to demonstrate the added value of the enriched ontology when applied to the content of the existing portal (Ilgar 2005). Browsing using this new portal exploits the links that were added to the ontology so that the result of navigation is not any more dependent on the point of entry (e.g. laws can also be found by browsing for products) and the number of potential navigation paths are multiplied. Search has been re-implemented in a way that keywords are not only matched against the title or text of documents, but also against the search terms in the ontology, their synonyms and spelling variations, returning related entities of different kinds.

We also expect that semantic technology would have a positive effect on the maintainability of the website in the long run. In particular, ontologies can be easily extended by third party information providers without breaking their compatibility with overheid.nl. This is not possible with XML Schema, where the validity of documents break when additional entities or attributes are added. Further, overheid.nl itself can also more easily change the ontology in compatible ways: categories can be easily merged, made subclasses of each other etc. and these changes are immediately reflected in the way items are organized on the website.

The Dutch government is also in the process of adapting a variation of Dublin Core[2] for government metadata.[3] In the future, the concept identifiers in the overheid.nl ontology can also serve as subject identifiers in these descriptions, which would shift the task of categorizing information from the editors of the website to the authors of the content. Together with a mechanism for the automatic discovery of new content (for example, through the use of RSS technology) the process of integrating new content into the website could be fully automated.

At the same time emergent semantics provide the flexibility for the supply-side ontology to co-develop with the changing language and interests of the visitors of the website. We see a future role for emergent semantics also in other kinds of applications where mappings between citizen's speak and the conceptualization of experts needs to be established. The application of this idea to urban planning is described in (Titulaer, Peters, & Oldenhuizing 2005).

References

Ilgar, E. 2005. De Semantic Web technologien toegepast op overheidsinformatie (Case: www.overheid.nl). Master's thesis, Vrije Universiteit, Amsterdam. Available online at `http://www.zenc.nl/publicaties/index.php`. In Dutch.

Mika, P. 2005. Ontologies are us: A Unified Model of Social Networks and Semantics. In *Proceedings of the 4th International Semantic Web Conference (ISWC)*, volume 3729 of *Lecture Notes in Computer Science*.

Titulaer, R.; Peters, R.; and Oldenhuizing, J. 2005. Matching demand and supply in government information Opening the eGovernment service catalogue. Unpublished, available on request.

[2] `http://dublincore.org/`

[3] `http://www.advies.overheid.nl/attachment.db?2733`