Semantic Search for Enterprise 2.0*

Alexandre Passant
DERI, NUI Galway
alexandre.passant@deri.org

Philippe Laublet
LaLIC, Univ. Paris-Sorbonne
philippe.laublet@paris-sorbonne.fr

John Breslin, Stefan Decker
DERI, NUI Galway
firstname.lastname@deri.org

ABSTRACT

In this paper, we describe how Enterprise 2.0 can benefit from lightweight semantics for information integration, enabling a better and easier search process for the end-users.

Keywords

Enterprise 2.0, Semantic middleware, SIOC, Semantic Wikis

1. MOTIVATIONS

Enterprise 2.0 [5] describes the next generation of information management and collaboration tools being used in organisations and is considered as *"the use of emergent social software platforms within companies, or between companies and their partners or customers"*[1]. Yet, while it eases the publishing process, searching for relevant information in these environments can be a complex task. Indeed, distributed and heterogenous data is constantly generated by end-users from various sources (blogs, wikis, RSS feeds ...), leading to information fragmentation problems. Furthermore tagging practices introduce ambiguity and heterogeneity issues, while the plain-text nature of these tools makes query answering a complex task. In this paper, we present a case-study on how Semantic Search can be enabled in Enterprise 2.0 ecosystems[2]. While research on information integration and middleware architectures for enterprises has already been studied from various perspectives [1] [4], our focus is (1) to enable it using lightweight semantics and (2) to take into account the social aspects of Enterprise 2.0, both in terms of producing and using semantic annotations by end-users.

*The work presented in this paper has been funded in part by Science Foundation Ireland under Grant No. SFI/08/CE/I1380 (Lion-2).

[1] `http://andrewmcafee.org/blog/?p=76`

[2] The case-study described in this paper took part at Electricité de France R&D – `http://rd.edf.fr` – during a three-years period where various Enterprise 2.0 services have been deployed.

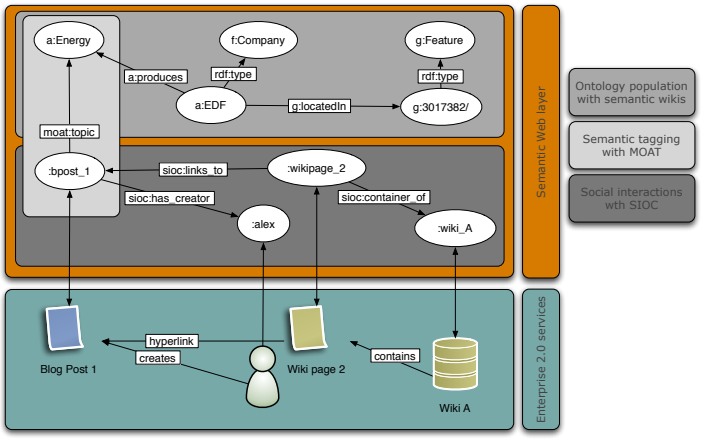

Figure 1: Semantics for Enterprise 2.0

2. SEMANTIC ENTERPRISE 2.0

To solve the previous issues, our proposal and implementation is based on a middleware process comparable to the *RDF bus* architecture defined by Berners-Lee[3]: semantic annotations are produced from existing tools, being then interlinked and queried thanks to RDF(S)/OWL and SPARQL. In addition, enhancements of the various tools must be as lightweight as possible in order to not disturb users in their existing habits. The first step to achieve this interoperability between Enterprise 2.0 components is to rely on common data models to represent socially-created information from various sources. Hence, we rely on SIOC [2] and plugins for original tools that automatically produce SIOC-based RDF data from these sources. In our use-case, more than than 17,000 instances of `sioc:Post` and related subclasses were exported from blogs, wikis and RSS feeds, created by a total of 79 users, providing a fist layer of common semantics wherever the data comes from. In order to model the data contained in some of these items in a machine-readable way, we extended the original wiki system to provide semantic modeling capabilities. Our wiki add-on [7] provides the ability to define form-based templates and to map them to classes and properties of lightweight ontologies. Hence, by simply editing wiki pages and filling related forms, more than 300 instances (and related properties) have been collaboratively created and maintained by end-users, bridging the gap between plain-text wiki pages and RDF triples about the enti-

[3] `http://www.w3.org/2005/Talks/1110-iswc-tbl/#(24)`

Figure 2: Overview of our Semantic Search engine for Enterprise 2.0

ties described in these wiki pages. The system also features auto-completion capabilities in order to interlink RDF data from various pages thanks to shared URI. Finally, in order to link SIOC-based data to ontologies instances, we relied on semantic tagging capabilities with MOAT [6]. Combined together, these three independent steps help to model a unified graph of semantic annotations on the top of existing Enterprise 2.0 applications (Fig. 1): wikis are used to manage ontology instances, that are linked to Web 2.0 items thanks to MOAT, those items and their relationships being modeled with SIOC.

3. ENABLING SEMANTIC SEARCH

As RDF data is then created thanks to various distributed services and as a matter of efficient querying, our approach is based on a ping service and a central triple-store: services that produce semantic annotations send a ping to the store service each time data is created or updated, with this data being instantaneously stored. The architecture was designed using SPARQL / SPARUL languages and protocols, enabling standard ways to (1) integrate new data in the triple-store and (2) connect querying and browsing services on the top of it. Among the various services we built, one of them is a dedicated semantic search engine which allows users to find information by concept from a given keyword, using all available sources of information (blogs, wikis ...) and relying on the proposal defined by [3]: bridging the gap between text-based search and concept search, and retrieving information about this concept rather than simply documents. To achieve this goal, the first step is to identify the relevant entity from a given keyword: this is done by retrieving all instances containing that keyword as a label or an associated tag (via MOAT). For instance, a user searching for *"France"* will get the suggestions *"Association des Maires de France"*, *"France"* and *"Electricité de France"*. Once the entity has been identified, the engine lists – in a single page – all information that is known about the entity, *i.e.* (Fig. 2) (1) the various tags related to this instance, (2) the main wiki page (i.e. the one used to create the instance), (3) the related wiki pages (i.e. wiki pages about entities linked to the main one) and (4) semantically-tagged content. It is important to consider that tagged content is then retrieved regarding the entity and not only the keyword, hence solving heterogeneity issues. Furthermore, these different steps are achieved thanks to SPARQL queries that are completely hidden to end-users, while the ping system allows real-time information integration from various distributed sources.

4. CONCLUSION

In this paper, we described how Enterprise 2.0 systems can be enhanced thanks to Semantic Web technologies, enabling a better search experience while hiding the underlying RDF(S)/OWL and SPARQL complexity to end-users. While we have mainly considered Web-based applications in this use-case, other data sources such as desktop applications can be considered in the overall approach.

5. REFERENCES

[1] Jūrgen Angele and Michael Gesmann. Data integration using semantic technology: A use case. In *RuleML2006*, 2006.

[2] J.G. Breslin, A. Harth, U. Bojars, and S. Decker. Towards Semantically-Interlinked Online Communities. *2nd European Semantic Web Conference*, May 2005.

[3] Ramanatgan V. Guha, Rob McCool, and Eric Miller. Semantic Search. In *WWW2003*, pages 700–709. ACM Press, 2003.

[4] Alexander Maedche, Boris Motik, Ljiljana Stojanovic, Rudi Studer, and Raphael Volz. Ontologies for Enterprise Knowledge Management. *IEEE Intelligent Systems*, 18(2):26–33, 2003.

[5] Andrew P. Mcafee. Enterprise 2.0: The dawn of emergent collaboration. *Management of Technology and Innovation*, 47(3), 2006.

[6] Alexandre Passant and Philippe Laublet. Meaning of a tag: A collaborative approach to bridge the gap between tagging and linked data. In *Proceedings of the WWW 2008 Workshop Linked Data on the Web (LDOW2008), Beijing, China*, Apr 2008.

[7] Alexandre Passant and Philippe Laublet. Towards an interlinked semantic wiki farm. In *3rd Semantic Wiki Workshop (SemWiki2008)*, May 2008.