

# AN APPROACH TO CONNECT WEB-BASED COMMUNITIES

John G. Breslin, Stefan Decker, Andreas Harth

*Digital Enterprise Research Institute, National University of Ireland, Galway, University Road, Galway, Ireland  
{john.breslin, stefan.decker, andreas.harth}@deri.org*

## ABSTRACT

Online communities are islands that are not interlinked, where complementary discussions can exist on disparate systems but it is difficult to exploit this available distributed information. A Semantically Interlinked Online Community (SIOC) can enable efficient information dissemination across such communities through the use of an ontology which will be created to model the concepts identified in discussion methods. Data instances of these concepts will be provided from a community site interface, allowing connections between local and remote instances. The connections between communities will be weighted for searching and matching purposes. SIOC is a prerequisite for a search engine that will answer questions rather than providing links to possibly relevant information.

## KEYWORDS

Online communities; semantic web; ontologies; forums.

## 1. INTRODUCTION

Online community sites such as bulletin boards and archived mailing lists allow improved communication and interactive contact within a community, by providing an online collaboration space for members to find and contribute certain interest-related or regional information (Wellmann and Gulia, 1999).

However, most online communities are hosted on standalone sites that cannot be interconnected due to application and interface differences. Parallel discussions on interrelated topics may exist on a number of sites that are not linked. There is a huge amount of related information that could be harnessed across such online communities, from similar member profile details to common-topic discussion forums.

Online communities have already evolved from standalone forums to online social networks (O'Murchu et al., 2004). In the future, the trend is moving towards networked social semantic desktops (Decker and Frank, 2004). Table 1 shows this evolution and the associated ontologies that are required.

Table 1. Placement of SIOC in the evolution of community discussion methods

| Traditional web | Semantic web                         |                         |
|-----------------|--------------------------------------|-------------------------|
|                 | Ontology                             | Application             |
| Forum           | <b>SIOC</b>                          | Semantic Forum          |
| Social Network  | <b>SNO (Social Network Ontology)</b> | Semantic Social Network |
| ...             | ...                                  | ...                     |
| Desktop         | <b>Desktop Ontology</b>              | Semantic Social Desktop |

This paper will present an approach to develop a Semantically Interlinked Online Community (SIOC), which will identify the main concepts in online communities and enable the linking of these concepts by modelling the relationships between them.

We will begin by describing the problems with connecting discussions and people in online communities. Then we will present an approach that can both maximize the usage of available information and enable the location of relevant information in online communities. This solution has three parts: to identify and model the concepts found in online discussion methods, to allow the creation of explicit and implicit connections between sites by both humans and computers, and to create a data infrastructure between different community sites.

## 2. MOTIVATION

At present, many isolated communities can exist that are discussing complementary topics, and information on these community sites cannot be harnessed correctly by search engines that are limited to syntactic matching, e.g., by message keyword on bulletin boards. Information is being repeatedly requested across separate sites, and people are wasting time searching for relevant community information by waiting for answers that have already been posted elsewhere. For example, a forum post on community site A may be closely related to an archived mailing list message on site B, but a forum search will not represent this link. Linking individual posts with others is possible on the HTML level, but on the Semantic Web content has to be made accessible to machines.

In a typical scenario, a user is searching the current web for information on installing broadband on a Linux-based PC in their house in Galway. There may be a post discussing local ISPs on a bulletin board dedicated to Galway that references (on the HTML level) both a Usenet post comparing broadband modems and a mailing list post detailing how to install broadband on Linux. At present, the user would have to traverse three sites to find the relevant information. In the future, by making use of SIOC as illustrated in Figure 1, a search for broadband on the Galway bulletin board could immediately return the relevant text from the interlinked Usenet and mailing list posts.

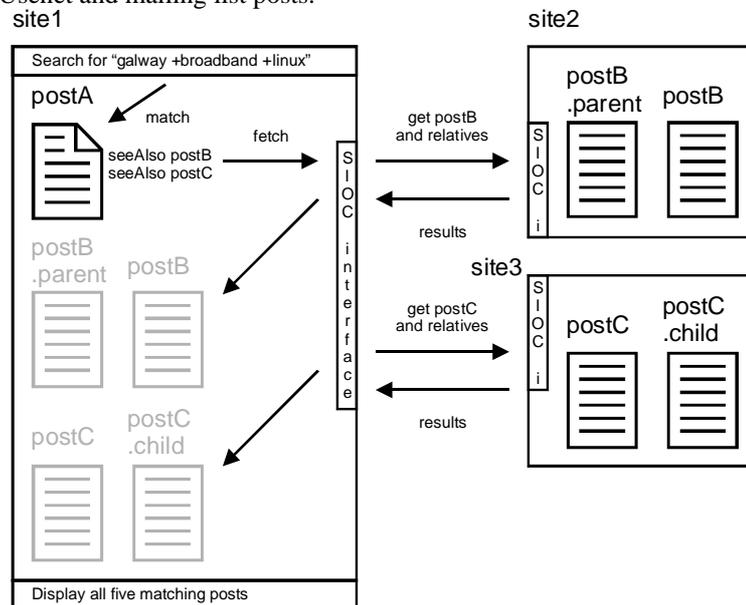


Figure 1. Cross community search using the SIOC interface

## 3. SOLUTIONS

The research approach for SIOC involves three parts:

- The creation of an ontology by identifying and modelling the concepts in community sites. Semantic Web ontologies can be used to enrich community sites and to make the underlying information available to both humans and software agents.
- A method for identifying and quantifying connections between community sites. The format of concept instances will be machine readable, allowing automatic detection and extraction of explicit and implicit connections respectively between people and related forums on separate sites.
- An interface to query and exchange data instances of the ontologised concepts. This interface will allow community sites to make messages, events, user profiles, etc. available in a common format for cross-community querying and storage.

### 3.1 Ontology Construction and Refinement

An analysis will be performed on existing discussion systems to identify the concepts and properties to be modelled in the SIOC ontology, e.g., people, groups, posts, forums, locations, events, categories, etc.

There is the issue of whether one should link and reuse existing ontologies, or use mappings to an entirely new ontology and therefore require more intelligent applications. If a mapping is provided, there is more flexibility but algorithms are required to perform efficient mapping and data transformation.

Where possible, terms will be reused from ontologies such as FOAF, Dublin Core and RSS 1.0, and new terms will be defined to allow integrated access to different formats exported by the various discussion systems and to describe how core concepts can be linked to each other; e.g., on bulletin board systems, forums and posts will have parent containers that are not defined in mailing lists.

### 3.2 Creating Connections

SIOC will enable the linking of community sites in a machine-interpretable format. Automatic connections between community sites can be made in various ways: e.g., we can infer a connection if the same person posts to different sites, or if one forum is explicitly linked to a forum at another site. By also allowing users to manually create links between different concept instances, the result is a useful network of community sites.

In an explicit method of interlinking, communities can be connected through varying orders of `rdf:seeAlso` links (perhaps derived from links on the HTML level), which can be defined by site administrators, forum moderators or post creators. They may also be connected through a usergroup's or user's explicit social network.

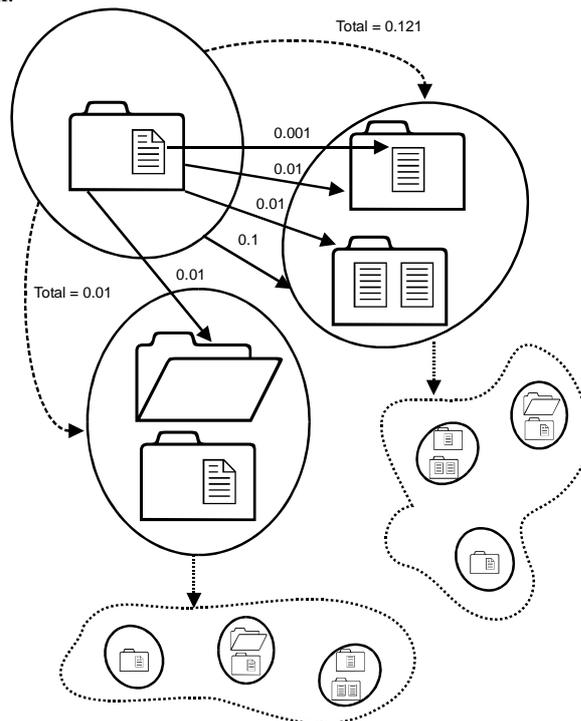


Figure 3. Sample weighting of connections between community sites

These links can have various weightings, for example, an explicit linking of sites could be 10 times greater than a linking of forums (and 10 times greater than a linking of posts). Similarly, a link between sites could be 10 times more important than one between usergroups, and the usergroup link could also be 10 times more important than a single user-to-user link. These individual weighting components can be summed to determine an overall site-to-site value, which can be used to determine which sites should be used in cross-community searches (for example, in Figure 3, this may be sites with a ranking greater than 0.1).

Since most communities do not have explicit connections, SIOC will require a method for creating implicit connections, inferring or deriving some similarity or links between forums, communities, etc. For example, this may focus on concepts, terms, hierarchies, common people or groups of common posters. If an explicit usergroup link does not exist, it could be implicitly determined by a majority of users linking to users in another usergroup.

### 3.3 Data Infrastructure

**Sites.** In the context of SIOC, it is appropriate to extend our notion of community sites. If we say "sites", we mean all sorts of community tools from forums to blogs. All sites should implement a common interface that allows access to the underlying data in a semi-structured format via the SIOC ontology. For that matter, it is not relevant whether the data comes from a site in a network of community sites, from a node in a P2P network, or from a web service.

**Warehousing or Query Interface.** For a site to access data from other sites, there are two options that will be investigated: warehousing of data one level or two levels away from a community into a local database, or a query interface to a site's data by extending the HTTP GET access interface to allow it to perform queries on RDF data.

**Virtual Integration.** For use with SIOC, we propose a database architecture known as "virtual integration" or answering queries over views. Data is fetched on demand when a query arrives. All necessary transformations (schema mapping, etc.) are carried out in this process. The query is translated, sent to all sources, and the resulting RDF is translated back into the caller's ontology.

## 4. CONCLUSIONS

This paper presented an approach to develop a Semantically Interlinked Online Community (SIOC), by identifying the main concepts in online communities, modelling the relationships between them, and enabling the linkage of concept instances. SIOC offers many future benefits to online communities, including:

- Once there exists enough sites that have richer query facilities to instances of SIOC data, then these different sites can be interlinked (for example, to export a set of user accounts for use by another site).
- Applying reasoning facilities (such as those provided by OWL-S time) to make use of events descriptions,
- Visualization of scheduling information of individuals or groups of people,
- Representations of where people related to a certain topic are located geographically.

Interlinking community concepts into a coherent representation such as SIOC will enable more sophisticated applications and therefore will result in more efficient information dissemination in communities.

## REFERENCES

- Decker, S. and Frank, M., 2004. The Networked Semantic Desktop. *WWW Workshop on Application Design, Development and Implementation Issues in the Semantic Web*. New York, USA.
- Lara, R. et al, 2004. An Evaluation of Semantic Web Portals. *Proceedings of the International Conference in Applied Computing (IADIS AC2004)*. Lisbon, Portugal.
- O'Murchu, I. et al, 2004. Online Social and Business Networking Communities. *In Proceedings of the Workshop on the Application of Semantic Web Technologies to Web Communities (ECAI 2004)*. Valencia, Spain.
- Smarr J., 2004. Technical and Privacy Challenges for Integrating FOAF into Existing Applications, *Proceedings of the 1st Workshop on FOAF, Social Networks and the Semantic Web (FOAF Galway)*. Galway, Ireland.
- Staab, S. et al., 2000. Semantic Community Web Portals, *In Computer Networks*, Vol. 33, Nos. 1-6, pp. 473-491.
- Wellman, B. and Gulia, M., 1999. Virtual Communities as Communities: Net Surfers Don't Ride Alone. *In Communities in Cyberspace*, pp. 167-194.