DSMW: a distributed infrastructure for the cooperative edition of semantic wiki documents

Hala Skaf-Molli, Gérôme Canals, Pascal Molli

DSMW : a distributed infrastructure for the cooperative edition of semantic wiki documents

Hala Skaf-Molli
skaf@loria.fr

Gérôme Canals
canals@loria.fr

Pascal Molli
molli@loria.fr

LORIA – INRIA Nancy-Grand Est
University of Nancy
Nancy , France

ABSTRACT

DSMW is a distributed semantic wiki that offers new collaboration modes to semantic wiki users and supports dataflow-oriented processes. DSMW is an extension to Semantic Mediawiki (SMW), it allows to create a network of SMW servers that share common semantic wiki pages. DSMW users can create communication channels between servers and use a publish-subscribe approach to manage the change propagation. DSMW synchronizes concurrent updates of shared semantic pages to ensure their consistency.

Categories and Subject Descriptors

H.4 [**Information Systems Applications**]: Office Automation

General Terms

Algorithms, Reliability

Keywords

Semantic Wiki, Replication, Distribution

1. COLLABORATIVE EDITING

Semantic wikis allow to create and edit collaboratively semantically annotated documents. Compared with other collaborative systems, semantic wikis do not support offline work or multi-synchronous editing [1]. In existing semantic wikis, every change in a page is immediately visible for both end users and semantic engines. However, in many cases it is necessary to change multiple pages before making them visible. Existing semantic wikis cannot prevent users to access, navigate or query inconsistent pages. Moreover, the lack of multi-synchronous support prevents users to work isolated [3] and also prevents semantic wikis to support dataflow oriented processes. We propose an approach in which semantic wiki pages are replicated over an overlay of semantic wiki servers. Changes issued on one server are local but can be published to other servers. Remote servers can subscribe to these changes, pull them and integrate them to their local pages. Changes propagation remains under the control of the users. Concurrent changes on a page issued by different servers are handled by a merge procedure.

Using this approach, users can alternate between isolated periods of work and synchronization sequences with remote servers. They can introduce changes to multiple pages before to atomically render these changes public. They can choose when to incorporate, or not, remote changes. In addition, the approach can be the basis for implementing processes in which flows of semantic wiki pages can traverse a network of semantic wiki servers. Each wiki server can implement one or several steps of a particular process for the creation and maintenance of semantic pages.

DSMW is used in the *CyWiki* Lorraine University project. *CyWiki* is an infrastructure for the collaborative and assisted transformation of textual content into formal knowledge. The transformation process is a decentralized process in which human agents and automatic agents (text-mining, classification agents) collaborate to build knowledge units. This knowledge can then be used to query and make reasoning about the content. The application domain of the project is education.

2. DSMW APPROACH

DSMW is an extension to Semantic MediaWiki (SMW) [2]. It allows to create a network of SMW servers that share common semantic wiki pages. DSMW manages the propagation and the integration of changes issued on one SMW server to remote servers on the network. The system ensures the consistency of the whole set of replicated pages. DSMW users can create and edit semantic wiki pages in isolation on a single SMW server. Then they can publish part or all of their own changes by pushing them to DSMW public feeds, and they can subscribe to any remote public DSMW feeds, pull changes from remote servers and integrate them to the local pages.

The DSMW extension adds two main features to SMW: an optimistic replication algorithm, and an ontology to manage changes, publication and integration of changes sets.

Page replication in DSMW is handled by a dedicated replication procedure. Since semantic annotations are embedded in page content in SMW, DSMW replicates only page contents and there is no need to deal with annotations. DSMW uses the Logoot [5] algorithm to synchronize concurrent changes and guarantee the consistency of the shared pages. The propagation technique is publish-subscribe: changes issued on one server can be published by pushing them to one or several *pushFeeds*. Remote servers can subscribe to these feeds by connecting *pullFeeds* to existing remote pushFeeds. Then, they can pull changes and integrate them to

the local pages. Hereafter a brief description of the operations related to replication.

Save: the SMW save operation has been extended to build patches. Patches represent changes to a page as a sequence of elementary operations. Patches are computed by the Logoot algorithm as a diff between the current and the new version of the page. Logoot uses a special index to determine absolute insertion and deletion positions of the elements in a page. Once computed, a patch is applied to the local page and logged.

CreatePushFeed: creates a named channel to publish local changes. The content of the feed is specified by a semantic query. All pages in the query result belongs to that channel, meaning that changes on these pages will be published through that feed. A page can belong to several channels.

Push: the push operation computes the set of patches for a given pushFeed to form a ChangeSet. A changeSet is the ordered set of all patches logged for all the pages belonging to that pushFeed. Once computed, the changeSet is added to the feed and can then be pulled by remote servers.

CreatePullFeed: creates a named channel to pull remote changes. A pullFeed is connected to one single remote pushFeed, so a pull feed is defined by the URL of the remote server and the pushFeed name.

Pull: the pull operation downloads all the pending changeSets published in the pushFeed connected to a given pullFeed. Patches extracted from the changeSets are then locally applied in the order they appear. To do so, Logoot uses the absolute positions computed during the patch creation to insert or delete page elements.

The DSMW ontology represents all the concepts of DSMW: changes, change sets, push and pull feeds. This ontology makes possible the querying of the current state of the wiki and its complete history using SMW semantic queries. For instance, queries can extract *the list of unpublished changes* or *the list of published changes on a given channel* (see [4] for more details).

3. SYSTEM DEMONSTRATION

The demonstration scenario will focus on two main use-cases: *The knowledge aggregation* corresponds to the use of a DSMW server to aggregate wiki pages and knowledge from multiple sources. The demonstration example will be the following:

1. a first DSMW server holds semantic wiki pages about hotels in a city. Hotels are described with various properties and their location in the city relatively to well-known places (e.g. the train station, the main square),

2. a second DSMW server holds semantic wiki pages about touristic information in the city. It describes sites of interest and cultural events with various properties and their location in the city, relatively to well known places.

3. a third DSMW server subscribes to the public push feeds of the two previous, and regularly pull them. It then holds semantic wiki pages on both hotels and touristic information and their location in the city, and maintain these pages consistent with the original sources. This server can answer queries that cannot be evaluated on the original sources, typically to find an hotel close to a particular site of interest.

The knowledge validation steps corresponds to the use of one or several DSMW servers to implement a validation process: prior to rendering public a set of semantic wiki pages, it can be desirable to validate their content by users or by running non-regression tests. The scenario will be based on the same hotel-tourist example. It consists in adding a fourth DSMW server that will serve as a public front-end for querying the hotel-tourist knowledge base. The aggregation server will then serve to combine the original sources and validate the new knowledge base. This validation is done by users running tests by evaluating a fixed set of queries whose results are known and should not change. Once validated, changes are propagated to the fourth server and are thus accessible to end-users. This validation step ensures the consistency and the stability of the final knowledge base.

4. CONCLUSION

We have presented DSMW , a collaborative tool to support multi-synchronous collaboration over semantic wiki documents. DSMW can serve as a basis to implement dataflow oriented processes to organize the collaborative authoring of semantic wiki pages. It provides a way to develop semantic wiki documents similar to developing software using a distributed version control system like Git or Mercurial. DSMW is released as an extension to MediaWiki and run on any platform supporting Apache/MySql/PHP. The DSMW extension is available at our site `http://dsmw.org`, where a video of the system in use is also available. We continue the development and research on DSMW. Research concerns divergence awareness in DSMW and the analysis of the social networks built by the collaborative editing.

5. REFERENCES

[1] P. Dourish. The parting of the ways: Divergence, data management and collaborative work. In *4th European Conference on Computer Supported Cooperative Work*, 1995.

[2] M. Krötzsch, D. Vrandecic, M. Völkel, H. Haller, and R. Studer. Semantic wikipedia. *Journal of Web Semantic*, 5(4):251–261, 2007.

[3] P. Molli, H. Skaf-Molli, and C. Bouthier. State Treemap: an Awareness Widget for Multi-Synchronous Groupware. In *Seventh International Workshop on Groupware - CRIWG*. IEEE Computer Society, 2001.

[4] C. Rahhal, H. Skaf-Molli, P. Molli, and S. Weiss. Multi-synchronous collaborative semantic wikis. In *10th International Conference on Web Information Systems Engineering - WISE '09*, volume 5802 of *LNCS*, pages 115–129. Springer, October 2009.

[5] S. Weiss, P. Urso, and P. Molli. Logoot : a scalable optimistic replication algorithm for collaborative editing on p2p networks. In *International Conference on Distributed Computing Systems (ICDCS)*. IEEE, 2009.