SWooki: Supporting Disconnection in a Peer-to-peer Semantic Wiki

Hala Skaf-Molli, Pascal Molli, Gérôme Canals

LORIA – INRIA Nancy-Grand Est
Nancy Université
skaf,molli,canals@loria.fr

RESUME
SWooki est un wiki sémantique P2P. Un réseau SWooki est formé d'un ensemble de serveurs wiki sémantiques dans lequel chaque serveur héberge une copie des pages wiki ainsi que les annotations sémantiques. Cette architecture permet de supporter une alternance de période de travail connectées et déconnectés. Durant une déconnexion, plusieurs copies d'une même page peuvent être modifiées sur différents sites de manière parallèle et donc diverger. Au moment de la reconnexion, ces pages sont synchronisées. SWooki utilise un mécanisme de réplication optimiste pour assurer la cohérence des pages wiki et des dépots RDF. La démonstration présente les différentes fonctionnalités du système

MOTS CLES : Wiki sémantique, pair-à-pair, réplication, déconnexion.

ABSTRACT
SWooki is a P2P semantic wiki. A SWooki network is a set of interconnected semantic wiki servers. Each server hosts a replica of semantically annotated wiki pages. The replicated architecture allows to support mobile work. Mobile work is characterized by the alternate of connection and disconnection phases. During disconnection phase, the different replicas can be modified in parallel so they diverge. Later, upon reconnection they are synchronized. SWooki uses an optimistic replication mechanism to maintain the consistency of the replicated wiki pages and the replicated RDF repositories. This demonstration shows various functionalities of SWooki.

CATEGORIES AND SUBJECT DESCRIPTORS: D.2. [Software Engineering] : Distribution, Maintenance, and Enhancement.

GENERAL TERMS: Design, Performance, Reliability.

KEYWORDS: Semantic wiki, peer-to-peer, replication, disconnected work.

INTRODUCTION

Wikis are very popular collaborative writing tools. They allow users connected to the web to concurrently access and edit wiki pages. In spite of their fast success, wiki systems have some drawbacks. They suffer from bad search and navigation [14], it is not easy to find information in wikis. They have also scalability, availability and performance problems and they do not support mobile works and atomic changes [15, 6]. To overcome these limitations, wiki systems have evolved in two different ways: semantic wikis and peer-to-peer wikis.

Semantic Wikis [14, 10, 8, 2] integrate semantic annotations within wiki page to improve the wikis structures, the search and the navigation between pages. This makes the wiki content machine-readable allowing content extraction for reuse in external applications.

Peer-to-Peer wikis take the natural advantages of a P2P network i.e. faults-tolerance, better scalability, infrastructure cost sharing and better performances. The basic idea is to replicate wiki pages on the peers of a P2P network. The main problem is maintaining the consistency of replicas. Existing P2P wikis [15] use optimistic replication algorithms to ensure data consistency.

Semantic wikis, as classical wikis, suffer from scalability, availability and performance problems and they do not support mobile works and atomic changes [4]. To overcome these limitations, we propose a peer to peer extension for semantic wiki called SWooki [8, 9]. SWooki combines the advantages of both P2P wikis and Semantic wikis.

SWOOKI APPROACH

SWooki is a P2P semantic wiki. A SWooki network is P2P network of autonomous semantic wiki peers that can dynamically join and leave the network. Every peer hosts a local copy of all wiki pages and their associated semantic

France is located **in** [Europe]
The capital of France is [Paris]

Figure 1 **:** Editing of wiki page "France"

France is located **in** [locatedIn :: Europe]
The capital of France is [hasCapital :: Paris]

Figure 2 **:** Editing of semantic wiki page "France"

data. Semantic data are embedded in wiki text as in Semantic MediaWiki (SMW) [5]. The hyper links between wiki pages are typed. For instance, a link between the wiki pages ”France” and ”Paris” may be annotated by a user as ”capital”. Figures 1 and 2 show the example of a wiki page and the corresponding semantic wiki page.

These annotations express semantic relationships between wikis pages. Semantic annotations are usually written in a formal syntax so they are processed automatically by machines and they are exploited by semantic queries. For the end-user, a semantic wiki page will appear as a normal wiki page with a facts box at the end of the page. The facts box contains the page's semantic data.

Replication and Consistency

Replication of wiki pages and RDF repository allows: (i) off-line editing i.e. even when disconnected, a peer can continue editing, this is mandatory for mobile work, (ii) transactional changes i.e. by allowing off-line work, it is possible to generate a change that concerns multiple wiki pages (iii) local requests execution i.e. as all data are local, all requests can be processed locally. The main issue is maintaining the consistency of the replicas.

In SWooki, when a peer updates its local copy of data, it generates the corresponding operation. This operation is processed in four steps:

- It is executed immediately against the local replica of the peer,
- it is broadcasted through the P2P network to all other peers,
- It is received by the other peers,
- it is integrated to their local replica. If needed, the integration process merges this modification with concurrent ones, generated either locally or received from a remote server.

To synchronize data, SWooki adpats the linear text P2P synchronization algorithm detailed in [7] to take in consideration the semantic data as detailled in [11]. SWooki synchronization algorithm ensures the CCI (Causality, Convergence and Intention Preservation) consistency model [12] on the wiki text and the semantic data.

SWooki use cases

SWooki supports *disconnected work and transactional changes* and *Ad-hoc Collaborative Editing*. Adding off-line capabilities to web applications is currently a major issue. Current technologies for adding off-line capabilities to web applications focus on Ajax applications. However, the off-line mode of these web applications does not provide all features available in the on-line mode. This can be an obstacle for a wiki system. For instance, the off-line mode of the wiki allows navigation but it does not allow editing. A P2P semantic wiki tolerates naturally disconnected work by means of an integrated merge algorithm. With such system, it is possible to travel with a complete wiki system on a laptop, make changes off-line and re-synchronize with the P2P network as soon as an Internet connection is available. The off-line mode enables also transactional changes. Users can work disconnected if they lack internet connection or if they decide to disconnect directly from the user interface. While disconnected, a user can change many semantic wiki pages in order to produce a consistent change. The optimistic replication algorithm in P2P semantic wikis forces all changes to commute according to the CCI consistency [12] . Consequently, the concurrent execution of several transactions is always equivalent to a serial one. Thus, a consistent state is always produced.

Ad-hoc Collaborative Editing is derived from the *disconnected work* . Imagine several mobile wiki users have a meeting. Unfortunately, there is no Internet connection available in the meeting room. Therefore, they decide to set up an ad-hoc network within the meeting room. A P2P semantic wiki is able to propagate changes within the ad-hoc network and allows collaborative editing just for these off-line users. Of course, when the meeting is finished and users return to their organizations, their semantic wiki systems will re-synchronize with the whole P2P network.

Implementation

SWooki is the first peer-to-peer semantic wiki. It is implemented in Java as servlets in a Tomcat Server. This prototype is available with a GPL license on sourceforge at `http://sourceforge.net/projects/wooki` and it is also available online at: `http://wooki.loria.fr/wooki1`.

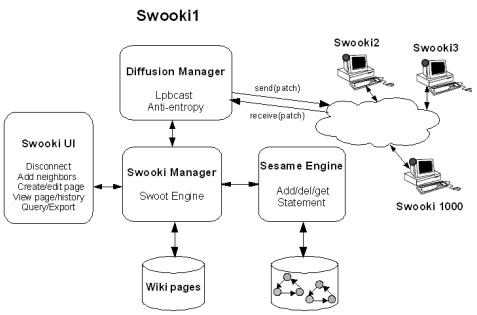

Figure 3 **:** SWooki Architecture

A SWooki server is composed of the following components:

User Interface. The SWooki UI component is basically a regular Wiki editor. Its only difference with a regular wiki is that it sends to the SWooki manager patches containing operations. So this component is able to generate patches of SWooki operations. It handles also concurrency awareness.

SWooki Manager. The SWooki manager implements the SWooki algorithms. Its main method is *Integrate(Patch)* that calls the *Receive()* algorithm for all operations contained in the patch.

Sesame Engine. We use Sesame 2.0 [1] as RDF repository. Sesame is controlled by the SWooki manager for storing and retrieving RDF triples. We used a facility of the Sesame interface to represent RDF triples as multi-set [11]. This component allows also generating dynamic content for wiki pages using queries embedded in the wiki pages. It provides also a feature to export RDF graphs.

Diffusion Manager. The diffusion manager is in charge to maintain the membership of the unstructured network and to implement a reliable broadcast. Membership and reliable broadcast of operations are ensured by an implementation of the Lpbcast algorithm [13]. This algorithm ensures that all connected sites receive messages and that there is no partition in the P2P network. Of course, disconnected sites cannot be reached. So we added an anti-entropy mechanism based on [3]. The anti-entropy algorithm selects randomly a neighbor in the local table of neighbors and sends a digest of its own received messages. The receiver returns missing messages to the caller.

Demonstration Overview

In our demonstration, we will show various functionalities of Swooki such as sequential editing, concurrent editing, off-line work and local query evaluation. For the purpose of the demonstration, we need three Swooki servers, called Swooki1, Swooki2 and Swooki3. These servers form a P2P network. They share wiki pages about the movie '"Iron Man"

Sequential editing In this scenario, only the server Swooki1 modifies the page "Iron Man". The user of Swooki1 adds a new line to her local copy of the page, the line contains semantic data. Upon page saving, the modification is propagated to servers Swooki2 and Swooki3. After modification integration, the local copy of the wiki page and the RDF repository at each server are identical.

Concurrent editing. In this scenario, both servers Swooki1 and Swooki2 modify concurrently the same line of the page "Iron Man". The concurrent modifications are propagated to both servers, a conflict is detected and is tagged in the page content as "produced by a merge" thanks to concurrency awareness mechanism The user of Swooki1 solves the conflict, her modification is propagated to other servers. At the end, all servers converge towards the same value for the wiki page and the RDF repository.

Mobile work. In this scenario, the sever Swooki3 is deconnected from the network. The user of Swooki3 modifies her copy of the "Iron man" page. The demonstration shows that this modification is local, it is not propagated to other servers. The modification is propagated to other servers at Swooki3 reconnection by using anti-entropy mechanism. At the end, all servers converge towards the same value for the wiki page and RDF repository.

Local Query evaluation In this scenario, we show how the same query can be executed locally at every server, without any data transfer thanks to semantic data replication.

Conclusion

SWooki is a P2P Wiki system in which replicated wiki pages can be accessed and edited on any server of the network. This architecture allows disconnected works. SWooki contributes to three main issues: optimistic replication of wiki page content, semantic data replication and concurrency awareness. The SWooki system is fully operational. Its first release can be freely downloaded from `http://wooki.sourceforge.net/`

We are evaluating our system from both the performance and the user point of view. To validate the scalability of the approach, we are conducting an experiment on GRID 5000, the french national grid infrastructure. We also plan a usability study, in particular to get feedback about our concurrency awareness mechanism.

Our work points out some open issues : security, access control to pages and replication overlay. Currently, all pages and semantic data are replicated on all servers. Our objective is to develop a new replication system in which users can control replica creation based on a social network: created pages and added semantic data are replicated only to trusted neighbors.

BIBLIOGRAPHIE

1. Broekstra, J., Kampman, A., and van Harmelen, F. Sesame: A Generic Architecture for Storing and Querying RDF and RDF Schema. In *The Semantic Web - ISWC 2002: First International Semantic Web Conference, Sardinia, Italy, June 9-12, 2002*, 2002.

2. Buffa, M., Ereteo, G., Faron-Zucker, C., Gandon, F., and Sander, P. SweetWiki: A Semantic Wiki. *Journal of Web Semantics, special issue on Web 2.0 and the Semantic Web*, 6(1), february 2008.

3. Demers, A., Greene, D., Hauser, C., Irish, W., Larson, J., Shenker, S., Sturgis, H., Swinehart, D., and Terry, D. Epidemic algorithms for replicated database maintenance. In *PODC '87: Proceedings of the sixth annual ACM Symposium on Principles of distributed computing*, pages 1–12. ACM, 1987.

4. Du, B., and Brewer, E. A. Dtwiki: a disconnection and intermittency tolerant wiki. In *WWW '08: Proceeding of the 17th international conference on World Wide Web*, pages 945–952, New York, NY, USA, 2008. ACM.

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

6. Morris, J. DistriWiki:: a distributed peer-to-peer wiki network. *Proceedings of the 2007 international symposium on Wikis*, pages 69–74, 2007.

7. Oster, G., Urso, P., Molli, P., and Imine, A. Data Consistency for P2P Collaborative Editing. In *Proceedings of the ACM Conference on Computer-Supported Cooperative Work - CSCW 2006*, Banff, Alberta, Canada, Nov. 2006. ACM Press.

8. Rahhal, C., Skaf-Molli, H., and Molli, P. Swooki: A peer-to-peer semantic wiki. In *The 3rd Workshop: The Wiki Way of Semantics- SemWiki2008, co-located with the 5th Annual European Semantic Web Conference (ESWC), Tenerife, Spain*, June 2008.

9. Rahhal, C., Skaf-Molli, H., and Molli, P. SWooki: Un Wiki Sémantique sur réseau Pair-à-Pair. *Ingénierie des Systémes d'Information*, 14(1), 2009.

10. Schaffert, S. IkeWiki: A Semantic Wiki for Collaborative Knowledge Management. *1st International Workshop on Semantic Technologies in Collaborative Applications (STICA06), Manchester, UK*, 2006.

11. Skaf-Molli, H., Rahhal, C., and Molli, P. Peer-to-peer semantic wikis. In *DEXA'09: 20th International Conference on Database and Expert Systems Applications*, Linz, Austria, August 2009.

12. Sun, C., Jia, X., Zhang, Y., Yang, Y., and Chen, D. Achieving Convergence, Causality Preservation, and Intention Preservation in Real-Time Cooperative Editing Systems. *ACM Transactions on Computer-Human Interaction*, 5(1):63–108, march 1998.

13. Th., Guerraoui, R., Handurukande, S. B., Kouznetsov, P., and Kermarrec, A. M. Lightweight probabilistic broadcast. *ACM Transactions on Computer Systems (TOCS)*, 21(4):341–374, November 2003.

14. Völkel, M., Krtözsch, M., Vrandecic, D., Haller, H., and Studer, R. Semantic wikipedia. In *WWW '06: Proceedings of the 15th international conference on World Wide Web*, pages 585–594, New York, NY, USA, 2006. ACM Press.

15. Weiss, S., Urso, P., and Molli, P. Wooki: a p2p wiki-based collaborative writing tool. In *Web Information Systems Engineering*, Nancy, France, December 2007. Springer.