

Rapport scientifique d'étape du Projet CEDRE RORAX Réplication Optimiste et Réparation Automatique de documents XML N° 05 Sci F3 /L3

Hala Skaf-Molli (1), Maître de Conférences
Hala Naja-Jazza (2), Maître de Conférences

¹ Université Henri Poincaré, Nancy1,
LORIA Campus Scientifique, BP 239 Vandœuvre-lès-Nancy, France,
skaf@loria.fr

² Université Libanaise, Faculté des Sciences, Branche 3, Kobbé, Tripoli, Liban
hjazzar@ul.edu.lb

1 Contexte général

Le travail coopératif à travers des réseaux informatiques s'impose de plus en plus dans le monde académique et industriel. Les équipes sont en effet distribuées à travers l'espace, mais aussi le temps et les organisations. Les tendances récentes de l'Internet penchent vers un espace d'information global au sein duquel plusieurs auteurs interagissent pour éditer un document. Ce modèle contraste avec le modèle précédent où un auteur unique écrivait pour un ensemble de lecteurs. Malgré la montée des besoins en applications collaboratives, il est surprenant de constater combien les systèmes commerciaux répondent mal à ces nouvelles demandes. Bien souvent, les utilisateurs sont réduits à s'échanger leurs documents par mail et évitent de travailler en parallèle de peur de générer des conflits d'édition. Des alternatives existent comme les wikis ou les gestionnaires de configuration. Les gestionnaires de configuration restent pour l'instant majoritairement destinés aux programmeurs et ne touchent pas le grand public. Les Wikis ont une audience bien plus grande mais supportent mal l'édition en parallèle des pages wikis. Plus récemment, un acteur majeur de l'Internet a publié un éditeur collaboratif en ligne ³ permettant à plusieurs utilisateurs d'éditer en ligne le même document. Bien sûr, cette initiative représente un progrès visible pour le grand public. Mais elle ne permet pas de répondre à tous les besoins de l'édition coopérative ^[1].

³ <http://docs.google.com/>

[1] Noël, S., Robert, J.M.: Empirical study on collaborative writing: What do co-authors do, use, and like? Computer Supported Cooperative Work **13**(1) (2004) 63–89

De manière générale, pour supporter le travail coopératif, il est nécessaire de résoudre trois problèmes majeurs ^[2]: la cohérence des données partagées, la gestion de la conscience de groupe et la coordination des différents acteurs à travers un procédé d'édition collaborative. Aujourd'hui quasiment aucun système d'édition coopérative ne prend en compte ces trois dimensions ^[3]. Notre objectif est de développer un éditeur coopératif qui puisse supporter les trois dimensions de la coopération.

2 Travail réalisé

Le projet RORAX s'intéresse au maintien de la cohérence de données répliquées et plus précisément aux données XML. Nous avons travaillé durant cette année sur les points suivants:

Synchroniseur XML Dans un premier temps, nous avons participé à l'élaboration d'un algorithme de synchronisation. Cet algorithme prend en entrée deux documents XML représentés par des arbres ordonnés. Contrairement aux algorithmes existants, notre algorithme s'intègre dans un environnement générique. Il permet de synchroniser des documents XML, des documents texte et plus généralement n'importe quel type de données. Le résultat de ce travail est publié dans [1].

Réparation automatique La synchronisation d'un document XML peut produire un document sémantiquement incohérent. Par exemple, la synchronisation d'un document XML muni d'une DTD peut produire un document qui ne satisfait pas cette DTD. Nous avons proposé des mécanismes de réparation automatique pour rétablir la cohérence de répliques de documents XML corrompues par réconciliation. Nous avons commencé par identifier les différentes formes d'incohérences produites par notre algorithme de réconciliation de données XML. Le résultat de ce travail est détaillé dans [2]. Ce travail a montré qu'il n'est pas possible d'assurer la cohérence sémantique de manière automatique. Un état de cohérence doit être validé par un acteur humain.

Procédés coopératifs Nous pensons que les procédés d'édition collaborative peuvent être utilisés pour assurer la cohérence sémantique. Par exemple, un procédé de validation permet une validation du document par les différents auteurs ce qui permet d'améliorer la qualité du document et d'arriver à une cohérence sémantique établie par les participants. Nous avons commencé par analyser les procédés d'édition collaborative utilisés dans l'environnement Wikipédia ⁴. Le résultat de ce travail est détaillé dans [3].

⁴ www.wikipedia.org

[2] Godart, C., Molli, P., Oster, G., Perrin, O., Skaf-Molli, H., Ray, P., Rabhi, F.: The toxicfarm integrated cooperation framework for virtual teams. *Distributed and Parallel Databases* **14** (2004)

[3] Adler, A., Nash, J.C., Noël, S.: Evaluating and implementing a collaborative office document system. *Interacting with Computers* **18**(4) (2006) 665–682

D'une part, ces procédés sont définis de manière informelle et d'autre part, ils reposent sur une architecture centralisée. Ceci a pour conséquence, une mauvaise tolérance aux pannes tolérant aux pannes et des performances obtenues au prix de solutions coûteuses. Aujourd'hui, les systèmes pair à pair permettent la diffusion de contenu à très large échelle sans nécessiter de solutions coûteuses ^[4]. Malheureusement, ces systèmes ne permettent pas la gestion des procédés. De plus, un système P2P diffuse un contenu qui ne peut pas être changé dans le temps. Nous voulons utiliser le potentiel de ces systèmes non seulement pour la distribution de contenu mais aussi pour l'édition de contenu.

Thèse en cours Nous avons démarré en novembre 2006 une thèse sur cette problématique en co-direction entre l'Université libanaise et l'Université Henri Poincaré, Nancy1. L'objectif de la thèse est la gestion de procédés d'édition coopérative dans un réseau pair à pair.

3 Nos perspectives

L'étape suivante consiste à pouvoir déployer des procédés d'édition coopérative dans le cadre d'un système d'édition collaborative sur réseau P2P. L'intégration des procédés dans une architecture P2P pose des problèmes scientifiques généraux ^[5]: Comment définir ces procédés ? comment sont stockés ces procédés en l'absence d'un serveur central ? Comment instancier ces procédés ? Comment les exécuter ^[6]? Ces problèmes peuvent être résolus de manière simple en nommant un coordinateur au sein du réseau P2P pour gérer les procédés. Cette solution n'est évidemment pas acceptable pour un réseau P2P symétrique ^[4] où les différents nœuds du réseau jouent un rôle identique. Nous envisageons donc de répliquer les procédés sur les différents nœuds du réseau. Le problème est alors d'assurer la cohérence du procédé lui-même. Il est important d'assurer que la transition entre les états du procédé est propagée sur tous les répliques. Dans les systèmes distribués, ce problème est généralement résolu en utilisant les algorithmes de consensus. Malheureusement, ces algorithmes sont définis pour des systèmes distribués statiques et ne sont pas compatibles avec les contraintes d'un réseau P2P. Certains algorithmes de réplication optimiste ^[7] sont compatibles avec les

[4] Androutsellis-Theotokis, S., Spinellis, D.: A survey of peer-to-peer content distribution technologies. *ACM Computer Surveys* **36**(4) (2004) 335–371

[5] Skaf-Molli, H.: P2p intercoop: P2p-based decentralised inter-organizational cooperation - motivations and requirements. Technical report, INRIA (2006)

[6] Yildiz, U., Tamisier, T., Skaf-Molli1, H., Feltz, F.: Decentralized execution of p2p collaborative processes. In: Journée de recherche de l'AIM Association Information et Management Innovation et Systèmes d'Information, Sceaux, France (2006)

[7] Oster, G., Urso, P., Molli, P., Imine, A.: Data consistency for p2p collaborative editing. In: Proceedings of the 2006 ACM Conference on Computer Supported Cooperative Work, CSCW 2006, Banff, Alberta, Canada, November 4-8, 2006, ACM (2006)

contraintes d'un réseau P2P mais n'assurent pas un niveau de cohérence suffisant pour garantir l'exécution correcte d'un procédé d'édition. La convergence comme critère de correction n'est pas suffisante pour assurer l'exécution correcte d'un procédé répliqué. Notre objectif est donc de proposer un algorithme de réplication optimiste adapté à la réplication de procédés collaboratifs sur un système d'édition collaborative déployé sur un réseau P2P.

4 Le bilan

Nous pouvons résumer les résultats obtenus au terme d'une année de coopération dans le cadre du projet RORAX du programme CEDRE par:

- Deux articles publiés dans deux conférences internationales ICEIS2007 [1] et EISWT2007 [4].
- Deux rapports de recherche [2, 3].
- Un site Web du projet RORAX disponible au: www.loria.fr/~skaf/pmwiki/pmwiki.php/Main/Rorax.
- Une thèse en co-direction entre l'Université Henri Poincaré, Nancy1 et l'Université Libanaise a démarré en Novembre 2006. La thèse est intitulée *les procédés d'édition coopératifs sur les réseaux pair à pair*. Le sujet de la thèse ainsi que la convention de co-direction sont joints au dossier.

References

1. Oster, G., Skaf-Molli, H., Molli, P., Naja-Jazzar, H.: Supporting Collaborative Writing of XML Documents. In: Proceedings of the International Conference on Enterprise Information Systems - ICEIS 2007, Funchal, Madeira, Portugal (2007)
2. Skaf-Molli, H., Naja-Jazzar, H., Molli, P.: Inconsistency of xml documents during cooperative editing. Technical report, INRIA (2006)
3. Rahhal, C., Naja-Jazzar, H., Skaf-Molli, H.: Procédés d'édition coopérative dans l'environnement wikipédia. Technical report, INRIA (2006)
4. Skaf-Molli, H., Ignat, C.L., Rahhal, C., Molli, P.: New Work Modes for Collaborative Writing. In: International Conference on Enterprise Information Systems and Web Technologies- EISWT 2007, Orlando, Florida (2007)