“Iceberg”: un modèle de diffusion restreinte pour un réseau de confiance entre pairs

Annie Danzart, Vincent Oria, Trung Tien Phan Quang, Jean-Marc Saglio

▶ **To cite this version:**

Annie Danzart, Vincent Oria, Trung Tien Phan Quang, Jean-Marc Saglio. “Iceberg”: un modèle de diffusion restreinte pour un réseau de confiance entre pairs. 2006. hal-00102151

HAL Id: hal-00102151
https://hal.science/hal-00102151

Preprint submitted on 29 Sep 2006

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

“Iceberg”: un modèle de diffusion restreinte pour un réseau de confiance entre pairs

Annie Danzart (ENST Paris), Vincent Oria (New-Jersey IT), Phan Quang Trung Tien (IFI Hanoi), Jean-Marc Saglio (ENST Paris)

danzart@enst.fr vincent.oria@njit.edu pqttien@ifi.edu.vn saglio@enst.fr

ABSTRACT

Publishing worldwide appears quickly dangerous to beginners in web logging due to its “share to all or share to none” model, and is very far from the gossiping model used in P2P (neighbor to neighbor) networks; this project explored a restrictive publishing model, called the "*iceberg visibility model*", based on self-defined views offered (with associated RSS feeds) to self-defined groups of self-registered readers (eventually subscribers to RSS feeds).

Categories and Subject Descriptors

D.3.3 [**Programming Languages**]: Language Contructs and Features – *abstract data types, polymorphism, control structures.*

Keywords

Semantic web, Peer-to-Peer Systems, Really Simple Syndication (RSS), database views.

1 MOTIVATION

Première remarque : Dans l’univers de la publication sur le web, Wikipedia distingue le “web profond” [Deepweb] (ou caché, ou invisible) du “web de surface”, le seul visible de la plupart des moteurs de recherche. Pour cette encyclopédie libre le “web profond” comporte non seulement les ressources dynamiques i.e. construites par une application en réponse à une requête largement paramétrable, mais aussi les ressources dont l’accès est limité par des règles selon l’identité du lecteur. Ainsi souvent seuls des résumés, des fiches signalétiques ou d’autres descripteurs, sont accessibles publiquement (et gratuitement), alors que les contenus complets ne le sont pas. Publier ne signifie pas nécessairement donner une visibilité totale.

Deuxième remarque: Le web d'hier n'avait que très peu d'écrivains comparativement au nombre des lecteurs, pour plusieurs raisons liées soit à la difficulté d'écrire elle-même, soit aux risques et aux responsabilités de la publication.LIENHYPERTEXTE. Depuis l'explosion des “blogs”, toutes les difficultés ont semblé disparaître: chacun peut devenir, avec des outils simples, unique écrivain et unique responsable de la publication sur “son” site. Mais le modèle simpliste de publication « pour le monde entier » est-il vraiment celui que chaque blogueur souhaite ? Certains qui ne cherchent qu’une communication « entre amis » ne se le sont-ils pas fait imposer par les fournisseurs d’accès et de services internet limitant la variété des rôles possibles ? Le qualificatif « entre amis » n’est souvent que la simple traduction de l’appartenance à la communauté virtuelle des utilisateurs d’un même serveur, dans laquelle la visibilité est totale et les recommandations tous azimuts. La demande de contrôle de son propre réseau de contacts et de contrôle de la diffusion de ses créations ou annotations personnelles de contenu fait éclater ce modèle simpliste. Le besoin de construction de réseaux où les personnes peuvent décider librement de « quels contenus partager avec qui » a été décrit dans plusieurs études récentes (par ex. le « web of people » [PABV03]) et des systèmes pour y répondre sont en cours d’expérimentation (par ex. le [SomeOne] de Wanadoo ou le [Yahoo! 360°] de Yahoo).

Notre recherche est centrée sur le développement de modèles et d'outils bien adaptés à ces écrivains qui cherchent à construire librement leurs réseaux de pairs en partageant les contenus qu’ils créent, selon la confiance mutuelle qu’ils s’accordent et peuvent remettre en cause. Elle ne situe leur coopération ni dans l’écriture à plusieurs mains des mêmes pages, comme dans le cas des "wikis", ni dans l’acceptation de « workflows » éditoriaux, comme dans les "spips". Elle admet l’autonomie et la responsabilité totale des écrivains, comme dans le cas des blogs.

Nos spécifications générales en découlent :

1. Il n'y a pas d'annuaire du réseau. Chaque écrivain se dévoile à ses lecteurs uniquement par la partie de son site HTTP qu’il leur rend visible avec édition automatique de "fils" RSS/XML[HHL04] organisés par thèmes (blog sémantique). La notification de nouveautés par SMTP/POP est facultative.
2. Chaque lecteur doit s'afficher d'abord comme écrivain (nous ne nous intéressons pas aux écrivains de seuls commentaires sur le site des autres) avec un « web de surface » visible pour tous (la « racine » de son site).
3. Chaque écrivain décide ensuite à discrétion

   a) pour la présentation externe des objets qu'il souhaite échanger, d’une hiérarchie thématique de dossiers dans laquelle le plus profond correspond aux échanges les plus rares et à la plus grande confiance.

   b) pour chaque branche thématique, des différents niveaux de confiance pour les lecteurs qu’il invitera et des groupes ayant visibilité à différents niveaux de confiance dans ces différentes branches thématiques (un lecteur pourra être invité dans plusieurs groupes)
4. Chaque lecteur, à l’intérieur de la vue qui lui sera donnée par chacun de ses pairs, pourra :

   a) faire une simple visite, sans rien garder dans sa propre mémoire

   b) garder dans sa mémoire l'URI de la collection d'objets qui l'intéresse (dossier ou union de dossiers) en l'incluant (souscription au fil RSS/XML) dans un de ses propres dossiers

   c) ”répondre” à un objet qu'il a lu chez un pair en en créant un nouveau dans sa propre mémoire.

Cet article a pour objet de détailler ces spécifications pour définir un modèle de diffusion contrôlée basée sur la confiance entre pairs qui serait facile à adopter et à implémenter. En voici le plan : dans le §2 nous définirons le schéma d'une collection personnelle de contenus organisée pour cette diffusion contrôlée vers les pairs, dans le §3 nous définirons notre modèle « iceberg » et ses intéressantes propriétés. Le §4 montrera son utilisation pour accompagner les variations possibles de confiance entre pairs et nous conclurons avec le §5 en décrivant notre prototype de test et les questions théoriques restant ouvertes et explorées par ailleurs.

2 BASE DE CONNAISSANCE PERSONNALISEE (PKB)

Chaque pair est propriétaire d'un domaine personnel de contenus dont il est l'unique auteur. Mais ces contenus peuvent être de types variés (notes personnelles, courriels, articles de son blog, etc..)et les exportations vers d'autres formats fréquents. De ce fait, la multiplicité des copies stockées sous différents formats peut faire difficulté, tout comme l'utilisation de classement dans de multiples hiérarchies de dossiers. Aussi le stockage et le classement dans une mémoire personnelle unique peut apparaître très utile. Appelons, comme dans [TSP04], « personal knowledge base » (PKB) une telle mémoire et définissons-en le schéma minimal.

2.1 Contenus définis comme des "billets"

Appelons "billets" l'abstraction commune aux "notes" personnelles, marque-pages (bookmarks), posts de blogs et messages du courrier électronique. Précisons formellement les attributs descriptifs selon la spécialisation des contenus :

Note::=(texte | hypertexte) "court" -- contenu court

Brève::=(Note, titre)

Billet::=(Brève ,date ,auteur)

Commentaire ::= (Billet, référence billet commenté)

Bordereau ::= (Billet, référence à pièce jointe -- FileName)

Signet|Bookmark ::= (Billet, référence à ressource décrite -- URL)

Fiche ::= (Bordereau | Signet)

Bulletin|Post ::= (Billet, permalink[Permalink]) – dans un Blog ou un Forum

Blog ::= {Bulletin*} -- séquence temporelle de bulletins, en général du même auteur

Mél ::= (Billet, emailDestinataire+)

Mél_bordereau ::= (Bordereau, emailDestinataire+)

Mél_signet ::= (Signet, emailDestinataire+)

Billet_reponse ::= (Billet, référence à Billet antécédent)

2.2 Classement hiérarchique par dossiers thématiques ou « topiques »

L'ontologie personnelle d'un pair peut être un graphe complexe.

Mais dans le cas d'une simple diffusion, une vue externe de structure hiérarchique (comme pour les dossiers de signets, de méls, etc..) peut suffire. Nous montrerons plus loin que cette structure est également adéquate pour gérer des "niveaux de confiance" pour les dossiers et sous-dossiers dans chaque branche thématique. Du fait que les dossiers servant à classer les billets ont aussi une qualité sémantique (on fait généralement des classement thématiques) nous parlerons de « topiques » plutôt que de dossiers.

Nous dirons que des billets sont « attachés » à un « topique » quand ils sont classés dans le dossier étiqueté par ce « topique ».

L'accès autorisé à un topique donne l'accès en lecture à tous les billets qui lui sont attachés. Autrement dit, « ouvrir » un topique c'est autoriser la lecture de tous ses billets. Cette règle permettra d'associer éventuellement à un topique un « fil RSS/XML » sur tous ses billets.

Figure 1 : Une PKB, arbre de topiques personnels

2.3 Liens inter-PKBs

Ayant posé que chaque pair du réseau a une PKB dans laquelle il est seul écrivain et administrateur, voyons comment chaque pair peut gérer des liens entre des objets (topiques ou billets) de sa PKB et ceux stockés dans les PKBs des autres pairs.

Nous dirons qu'un topique personnel « inclut » un topique d'un autre pair si ce topique externe est considéré par le créateur du topique personnel comme un sous-topique intéressant (et ses billets comme des billets à « bookmarquer »). Cette opération d'inclusion revient à l'« agrégation » d'un contenu « syndiqué ».

Nous dirons qu'un billet personnel « répond » à un billet externe s'il y fait référence pour le commenter ou « y répondre ».

Avec ces deux liens inter-PKB, créés exclusivement dans chaque PKB par son propriétaire, les PKBs forment un système multi-base dans lequel chaque pair peut librement agréger à sq propre mémoire la mémoire de tout (topiques et billets) ce qui lui est rendu visible par les autres.

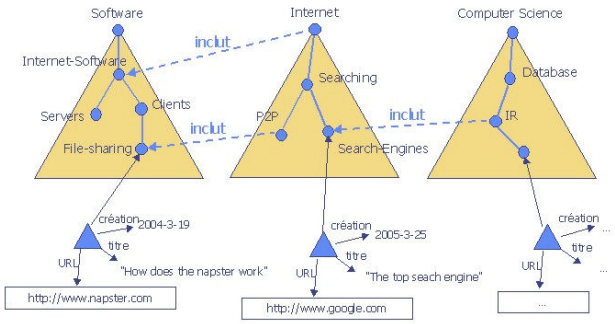

Figure 2 : Un système multi-PKBs

3 MODELE ICEBERG

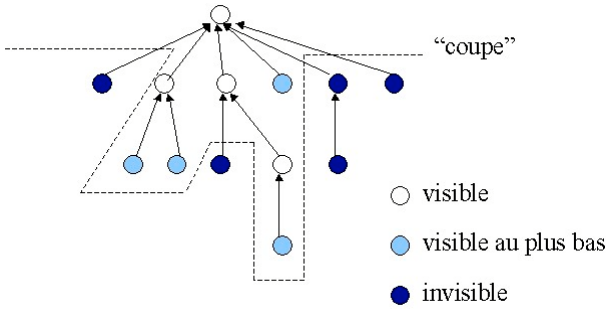

Figure 3 : Coupe d'une PKB

Il nous faut donc définir un modèle de visibilité, c'est à dire de contrôle par chaque pair de l'accès en lecture par les autres pairs, qui soit simple à mettre en œuvre et suffisamment dynamique pour suivre les fluctuations de la confiance mutuelle entre les pairs.

Nous appelons « iceberg » un modèle de vues séparant la partie visible (« haute ») de la partie invisible (« basse ») de l'arbre des topiques par une « coupe » de cet arbre, telle que chaque branche soit « coupée » à une certaine profondeur, séparant au-dessus les topiques visibles des topiques invisibles au-dessous.

Pour qu'un pair « existe » dans le réseau il faut que sa racine soit visible de tous, c'est à dire publique, pour que l'existence de l'arbre ne puisse être cachée. Cela n'interdit pas que d'autres topiques « inférieurs » soient également publics. Un topique visible par un nombre limité de pairs est dit de visibilité « restreinte ». Un topique invisible à tous sauf à son propriétaire est dit « privé ».

3.1 Théorèmes

Définition : Une coupe se définit par la liste des nœuds visibles au plus bas.

Théorème1 : Toute liste de nœuds constituée de 0 ou 1 nœud par branche est une "coupe".

Preuve : ou bien une branche est « coupée » en dessous du nœud figurant dans la liste, ou bien il faut remonter à la racine qui, elle, est toujours visible.

Corollaire : toute liste de nœuds présentant plus de 1 nœud par branche peut être réduite à 0 ou 1 nœud par branche pour devenir une "coupe".

Théorème2 : l'union de « coupes » est toujours une « coupe ».

Preuve : l'union pour chaque branche correspond au nœud visible le plus bas.

Théorème3 : l'intersection de « coupes » est une « coupe ».

Preuve : l'intersection pour chaque branche correspond au nœud visible le plus haut.

3.2 Choix d'une notation

Pour manipuler simplement les arbres de topiques on pourra utilement les identifier soit par leur chemin nominal depuis la racine (distinguished name), soit par un Dewey Decimal Classification [DDC].

4 GESTION DE LA CONFIANCE

Chaque pair établit, personnellement, librement et de façon réversible, des liens de confiance avec les autres.

4.1 Principe

Plus on est en confiance avec un pair, plus on lui "découvre" l'arbre de nos topiques en profondeur.

Notre modèle de sécurité pour chaque PKB s'inspire du modèle de Biba [CFMS95, p88] avec une « discretionary policy » de type « hiérarchie d'objets » dans laquelle la visibilité d'un objet entraîne la visibilité de tous ses ancêtres. Son intérêt est de transformer la définition de la liste des objets visibles en un choix de coupe (c'est à dire un choix de niveau de visibilité pour chaque branche de la hiérarchie). Ce niveau est ajustable dynamiquement en fonction de la confiance donnée au lecteur. Contrôler la visibilité des topiques (ou la diffusion des billets) revient à ce choix de niveau dans chaque branche thématique.

4.2 Pratique

En pratique le pair propriétaire d'une PKB prédéfinira autant de vues (i.e. coupes) que de groupes de lecteurs qu'il prévoit d'avoir.

Une seule vue sera définie pour chaque groupe, par contre un lecteur pourra appartenir à plusieurs groupes. Les vues seront identifiées par un nom (local à la PKB).

Chaque fois qu'un pair a un nouveau contact de lecteur, il doit se demander à quels groupes l'affecter. Il se peut que ce nouveau lecteur nécessite la création d'un nouveau groupe. Donc, chaque nouveau lecteur est affecté soit à un nouveau groupe, soit à un groupe existant.

S'il se trouve que la vue que l'on veut définir pour un groupe existe déjà (est déjà nommée) alors elle lui sera assignée. Si une vue n'est plus assignée à aucun groupe et qu'elle n'a pas été nommée, elle disparaît.

Le schéma global d'une PKB (schéma de définition et de contrôle de ses objets) est donné par la figure 4.

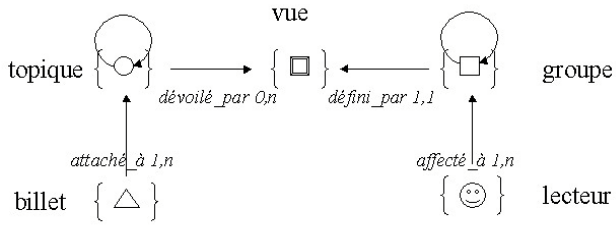

Figure 4 : Schéma global d'une PKB

Notons aussi que les topiques visibles doivent pourvoir être offerts comme des fils RSS/XML pour que les lecteurs puissent les inclure comme sous-topiques dans leurs PKBs.

4.3 Scenarii

Comment entrer en contact avec un (groupe de) lecteur(s) souhaité(s) ? Comment décourager les "free riders" (qui veulent lire beaucoup et ne rien écrire ou rendre visible) ? Comment « re-voiler » des topiques trop tôt « dévoilés » à un lecteur auquel on a eu tort de faire confiance ? Etc..

La réponse est chaque fois de (re)définir une vue « en coupe » adaptée, mais il est nécessaire d'en comprendre les conséquences dans les mémoires des pairs.

C'est pourquoi la validation du modèle ne pourra être acquise que si des scenarii de test correspondant à chacune de ces questions sont acceptés et que l'on peut vérifier que la cohérence de la mémoire multi-bases est conservée.

5 PERSPECTIVES

5.1 Perspectives pratiques

Notre premier objectif est de valider expérimentalement avec un prototype [Webographe]. Son logiciel d'installation est mis en accès libre à l'URL http://www.infres.enst.fr/webograph. Il permet de déployer des serveurs LAMP ou WAMP pour des écrivains-lecteurs agissant comme des pairs autonomes.

Le choix de la technologie standard LAMP ou WAMP a été fait pour permettre le déploiement facile du prototype. De plus le logiciel libre Mozilla Firefox a été choisi comme base de développement parce qu'il est facile de lui ajouter les extensions nécessaires.

Avec ce prototype un pair ne peut naviguer dans une PKB d'un autre pair qu'après s'être identifié comme lecteur ce qui lui construit dynamiquement la vue des seuls topiques que ce pair lui offre. Au cours de sa navigation il peut décider d'inclure dans sa propre PKB un topique externe visible en l'agrégeant par son fil RSS/XML à l'un de ses propres topiques, selon son choix. Il peut aussi décider de « répondre » à un billet en créant dans un de ses propres topiques un nouveau billet lui faisant référence. Les liens ainsi créés restent en mémoire dans sa PKB.

Les autres fonctions du prototypes peuvent être mises en œuvre localement sur la PKB du propriétaire, sans navigation préalable dans une PKB externe. Il s'agit du stockage de nouveaux billets, de l'organisation des topiques, de la définition des vues et de l'inscription des lecteurs dans ces vues/groupes.

5.2 Perspectives théoriques

Outre la question de cohérence posée plus haut (à la fin du §4. 3), la gestion de ce que l'on appelle classiquement la "grant option" doit être étudiée. Il s'agit de savoir si les solutions connues (« revoke cascading » ou « non cascading ») peuvent être appliquées dans le seul cadre de notre modèle ("pur HTTP").

Un deuxième point doit aussi être étudié : Le modèle Web-of-People auquel nous nous référons [TSP04] nécessite pour sa cohérence une notification. Pour tous les événements correspondants à, au moins, une mise à jour dans une PKB, un trigger envoie un message à tous les pairs concernés par cette mise à jour. Il nous faudra vérifier que notre modèle peut se passer cette signalisation minimale pour établir ou rompre les contacts.

6 CONCLUSION

En résumé, ce papier présente un modèle pour que, dans un réseau de pairs autonomes, chacun contrôle librement ce qu'il veut dévoiler aux autres, ce qui lui permet d'assurer librement les risques et les profits d'une communication à confiance variable.

L'implémentation d'un prototype de validation fonctionnelle de notre modèle – le « webographe » - y est présentée.

Références

[PABV03] Plu, M., Agosto, L., Bellec, P., Van De Velde, W.: The Web of People: a dual view on the WWW. In *Proc. of the 12th International World Wide Web Conference*, WWW'03 Budapest, 2003.

[TSP04] Ta, T.A., Saglio, J.M., Plu, M.: An architecture based on semantic weblogs for exploring the Web of People. In *Workshop Application of Semantic Web Technologies to Web Communities*, ECAI'04 Valencia, 2004.

[SST05] Saglio, J.M., Scholl, M.., Ta, T.A.: Efficient Query Processing in P2P Networks of Taxonomy-based Sources. In *Workshop Data Integration and the Semantic Web*, CAiSE'05 Porto, 2005.

[HHL04] Tony Hammond, Timo Hannay, and Ben Lund : The Role of RSS in Science Publishing, Syndication and Annotation on the Web. In *D-Lib Magazine* December 2004, Volume 10 Number 12, ISSN 1082-9873, http://www.dlib.org/

[CFMS95] S. Castano, M. Fugini, G. Martella, P. Samarati : Database Security. ACM Press, Addison-Wesley, 1995.

[SomeOne] http://someone.voila.com/
Layda Agosto, Michel Plu, Laurence Vignollet, Pascal Bellec: SomeOne: A cooperative system for personalized information exchange, International Conference on Enterprise Information System (ICEIS), Angers, 2003

[Yahoo!360°] http://fr.360.yahoo.com/

[DDC] http://www.oclc.org/dewey

[WAMP] http://en.wikipedia.org/wiki/WAMP

[LAMP] http://en.wikipedia.org/wiki/LAMP

[Deep web] http://en.wikipedia.org/wiki/Deep_web

[Permalink] http://en.wikipedia.org/wiki/Permalink

Notre prototype (documentation et download):

[Le Webographe] http://www.infres.enst.fr/webograph/