

<p>III</p> <p>Groupe Information- Interaction-Intelligence</p>				<p>Ed1</p>
<p>RAPPORT FINAL</p> <p><i>PLATEFORME ET APPLICATIONS DE VALIDATION DU PROTOCOLE « WebOP »</i></p> <p><i>Projet WEB-OF-PEOPLE</i></p>				
<p><i>Ce document ne peut-être communiqué ou dupliqué par quelque moyen que ce soit sans autorisation conformément au contrat entre les parties.</i></p>				
	<p>Nom</p>	<p>Service</p>	<p>Date</p>	
<p>Auteur(s)</p>	<p>Jean-Marc SAGLIO</p>	<p>ENST</p>	<p>04/03/05</p>	

Liste de diffusion			
Destinataire	Service	Pour action	Pour info.
Alain LEGER alain.leger@rd.francetelecom.com +33 (0)2 99 12 42 23	FT/Division R&D/Tech/EASY	x	
Michel PLU michel.plu@rd.francetelecom.com +33 (0)2 96 05 36 98	FT/Division R&D/Tech/EASY	x	

FICHE D'EVOLUTION

Édition	Date	§	Action	Nature de la modification
0	04/03/2005		C	
1	09/03/2005		M	quelques explicitations

§ = Chapitre(s) ou paragraphe(s) modifié(s) Action = C : Création, M : Modification, S : Suppression

GESTIONNAIRE DU DOCUMENT

Nom	Service	Téléphone	Adresse messagerie
Jean-Marc SAGLIO	ENST	+33 1 4561 8062	saglio@enst.fr

1 OBJET

Ce rapport final a pour objet, selon la demande de FT R&D, de

<<

- mettre en évidence les intérêts et les qualités particulières des résultats obtenus en les illustrant chaque fois que possible à partir d'exemples reproductibles sur les prototypes livrés,

- expliquer, justifier et motiver les choix techniques significatifs effectués

- dresser un bilan, éventuellement en listant des modifications recommandées (les erreurs faites à ne pas refaire), et les perspectives d'études ultérieures.

>> (email de M.Plu)

Son plan en découle :

1. rappel des notions-clés et de l'architecture générale du « web-of-people »
2. résultats obtenus, qualité, intérêt
3. choix techniques
4. bilan et perspectives

On trouvera en annexes la liste détaillée des livraisons et celle des documents connexes produits hors de ce contrat.

2 NOTIONS-CLES ET ARCHITECTURE GENERALE DU WEB-OF-PEOPLE

2.1 Des « weblogs » sémantiques

Un weblog – ou « blog » est logiquement un journal constitué séquentiellement de courts articles (« billets », « brèves », « petites annonces », « fiches », « posts » ...) chacun distingué par une URL quasi permanente (« permalink »), ayant tous un même descriptif (des « méta-données ») standard (« RSS » ou « Atom » ou autre), et présentés, dans un « canal » unique (une URL), en ordre rétro-chronologique (newest first).

Chaque « post », par méta-donnée ou par lien hypertextuel HREF dans son texte, peut référencer une autre ressource du web, en particulier un post dans un autre blog.

Les fonctions associées à chaque « post » une fois qu'il est publié peuvent être :

- Le compteur d'accès en lecture
- L'envoi par un lecteur de ce post à un tiers
- Le commentaire directement envoyé à l'auteur et attaché au post selon les modalités choisies par l'auteur
- Le commentaire écrit sur un autre blog que l'auteur est invité à récupérer (sur demande – « *trackback* » – ou automatiquement – « *pingback* ») pour l'attacher à son post.

Les méthodes associées au blog entier peuvent être :

- L'archivage périodique
- L'agrégation d'autres blogs par leur « *feed* » (format de syndication)

Les weblogs élémentaires n'ont pas d'autre sémantique que le descriptif – Atom ou RSS – de leur flux de posts tous identiquement formatés. On parle de « weblog sémantique » dès lors que les posts du weblog peuvent être indexés sémantiquement par un référentiel conceptuel (hiérarchie ou réseau de termes, carte de topiques, ontologie de classes d'objets) explicite sur le site du weblog. Il n'y a pas de modèle standard pour de tels référentiels, seulement le RDF ou les Topic Maps du W3C.

2.2 Au service de « documentalistes du world-wide-web »

Si les weblogs peuvent servir à l'auto-publication de textes ou créations d'auteurs sur le world-wide-web, ils peuvent aussi servir ceux qui ne se considèrent que comme des « documentalistes » dont l'activité principale est de sélectionner des ressources du web dignes d'intérêt, de les annoter, de les décrire et indexer sémantiquement et de les re-présenter dans des formats rendant plus facile leur découverte – par des moteurs de recherche sémantique – et leur réutilisation éventuelle.

Cette activité est dans la continuité de celle de la gestion de marque-page - « *bookmarking* » - pour laquelle les navigateurs offrent différentes méthodes complémentaires :

- Ajouter/déplacer/modifier/supprimer marque-page dans sa collection personnelle
- Envoyer le marque-page par email à un tiers (notification)
- Archiver/publier sa collection (par exemple avec le XML Bookmark Exchange Language – XBEL)

Un *bookmarking* plus avancé, c'est à dire annotant, indexant et publiant à l'intention des autres « documentalistes » ou confrères ou abonnés ou autres partenaires, peut facilement être réalisé par l'intégration d'un navigateur et d'un éditeur de weblog sémantique.

Restent les difficiles questions de l'hétérogénéité des référentiels sémantiques et des conflits possibles dans les intentions de communication.

2.3 Coopérant dans un réseau de « pairs »

« Pairs » signifie à la fois clients et serveurs les uns pour les autres mais aussi égaux dans leurs rôles et autonomes – chacun devant pouvoir fonctionner sans le réseau et celui-ci devant pouvoir fonctionner sans que tous les pairs soient nécessairement présents.

Ce qui est l'enjeu principal d'un tel réseau c'est la **confiance mutuelle** qui, loin d'être acquise a priori comme elle le serait entre les membres d'un club ayant satisfait aux conditions d'admission, est toujours en construction/destruction de telle sorte que

- les liens entre les pairs (aux « nœuds » du réseau) sont de qualité et d'intensité variables
- le partage des ressources documentaires (posts et termes/topiques/classes sémantiques) n'est pas restreint à l'alternative public v. privé, mais peut être défini par son extension (une liste partielle de pairs)
- chaque pair crée une vue dynamique « intuitu personae » de ses ressources pour chaque autre pair ou liste d'autres pairs

Ces deux dernières caractéristiques ne sont généralement offertes dans le web que par des sites construits comme des portails et adossés à des bases de données sachant gérer les droits d'accès et les vues dynamiques.

Heureusement les systèmes multi-bases de données dans lesquels chaque base n'a qu'un propriétaire mais offre des droits de connexion et d'accès contrôlé par le propriétaire à tous les autres usagers (les propriétaires des autres bases) ont permis, depuis vingt ans, d'en établir les modèles fonctionnel et opérationnel.

Chacun restant « maître chez soi » le développement du partage, c'est à dire l'ouverture pour d'autres de l'accès à ses créations personnelles, ne peut se faire que par un mécanisme de type messagerie dont le principe est la « notification » à certains autres de certains « événements » intervenus « chez soi », c'est à dire des créations, modifications, suppressions de ressources dans sa base personnelle.

Ce que les autres font des notifications qu'ils reçoivent est ou libre (cas de l'email) ou partiellement contraint par un protocole de « communication honnête » [13], protocole qui peut varier d'une communauté de pairs à l'autre, mais pour lequel un minimum semble nécessaire pour aider au développement du partage.

2.4 Architecture du « web-of-people »

Sa définition est détaillée dans un document [2] faisant partie de la livraison du lot 1 du contrat.

Disons pour résumer qu'il s'agit d'une architecture P2P où chaque pair est propriétaire d'une base personnelle – appelée « personal knowledge base » (PKB) – de posts et de topiques, base qu'il fait évoluer librement avec les seules contraintes suivantes :

- Respect d'un schéma RDF de PKB pour les propriétés des posts et des « topiques » par lesquels ils sont indexés, mais aussi pour les droits d'accès à ces propriétés donnés aux autres pairs dans le réseau
- Utilisation d'un langage de requête intra- comme inter-PKB ayant une puissance d'expression pour des requêtes de chemin – « path queries » – au moins égale à celle du langage du LDAP
- Utilisation d'un service de notifications formatées RDF pour une liste minimale d'événements de modification de sa PKB à son initiative ou en réponse à une notification reçue d'un autre pair
- Respect du même schéma RDF de PKB également pour les propriétés inter-PKBs additionnelles – dites « sociales » – des posts et des topiques qui permettent de garder la mémoire suffisante des notifications envoyées (dès qu'une mise à jour est faite dans sa PKB), ou reçues (dès que celles-ci sont retirées de la boîte de réception où elles étaient mises en attente de traitement).

L'architecture est compatible avec des mémoires plus redondantes et, peut être, versionnées ou historisées, avec des schémas de propriétés plus complexes, avec des langages de requête et des services de notification plus expressifs, avec des services additionnels de « médiateurs ». Un des buts du projet « WebOP » était très précisément de réfléchir sur une spécification minimale ouverte à des extensions « implementor defined » mais assurant la base d'une communication conduisant à un partage de mémoire volontaire, confiant mais aussi cohérent.

3 RESULTATS OBTENUS, QUALITE, INTERET

3.1 Spécification générale [2]

La spécification d'architecture, de schéma d'information, des services de requête et de notification a fait l'objet de nombreuses relectures entre les deux équipes ENST et FT R&D, pour finalement être stabilisée dans sa version 2 release 6 c.

Outre la clarification des concepts ci-dessus mentionnés, cette spécification présente

- le vocabulaire RDFS pour un schéma minimal de PKB ajoutant aux propriétés *dc*: des topiques et aux propriétés *atom*: des posts des propriétés *webop*: pour
 - o décrire un index hiérarchique dans lequel les nœuds sont des topiques, liés par la propriété *webop: subTopicOf* et les feuilles des posts attachés aux premiers par la propriété *webop: hasTopicOf* ce qui permet de donner aux posts la sémantique de la hiérarchie de topiques (qui peut elle-même prendre la sémantique d'un classement documentaire ou d'un thésaurus)
 - o décrire par les propriétés *webop: accessibleBy*, *webop: inaccessibleBy* et *webop: referencingAllowed* autant de vues personnelles sur une PKB que de lecteurs externes à cette PKB désignés individuellement ou par listes
 - o décrire les liens inter-PKBs, entre topiques comme entre posts, gardant la mémoire des recommandations/intégrations comme des abonnements/commentaires exprimés par des messages de notification spontanée (proactive) ou suscitée (réactive)
- le vocabulaire RDFS pour décrire le format des messages entre pairs, notamment pour introduire la demande – optionnelle – associée à toute notification de *feedbackTo* (option permettant d'implémenter un « *trackback* » garanti par un médiateur)
- l'enveloppe SOAP pour les services de requête et de notification du WebOP en cas de déploiement du réseau web-of-people sur une plateforme multi-serveurs.

Cette spécification a fait aussi corrélativement deux choix d'architecture

- distinguer la gestion des personnes (les « pairs ») de la gestion de leurs ressources, ce qui revient à ajouter une base « annuaire » aux bases PKBs (ce qui n'empêche nullement d'avoir dans chaque PKB des propriétés à valeur literal égale à un identifiant de pair dans l'annuaire *webop*)
- ne pas identifier un pair à un serveur ip, mais seulement à une entrée dans l'annuaire *webop* (ce qui permet la cohabitation de pairs « ip-autonomes » avec des pairs hébergés dans un même serveur ip – « plate-forme *webop* » mono-serveur)

Par contre elle n'a défini que dans leurs caractéristiques générales les techniques d'authentification et de chiffrement applicables (page 33 de la spécification).

Pour ce qui est du « chaining » la spécification générale a laissé ouverte la question des méthodes applicables en réponse à des requêtes qui voudraient profiter des liens d'intégration entre topiques pour « attraper » des réponses provenant de plusieurs PKBs. Cette question fait l'objet de recherches toujours en cours dans la communauté des chercheurs en BD semi-structurées distribuées. Les cycles, qui sont la preuve d'équivalence reconnues par les pairs, mettent une sérieuse limitation aux possibilités de définir des algorithmes optimaux pour la collecte inter-bases. La spécification générale laisse donc aux applications le choix de la méthode de réponse pour ce qui est de la formulation des sous-questions « à suivre » :

1. seulement par le pair interrogé (« à la » LDAP v1)
2. par propagation par chacun des pairs interrogés (« à la » LDAP v2)

Par contre elle fixe les règles pour la « transitivité » des visibilitées en fonction de la propriété attachée à chaque topique *webop:referencingAllowed* (fin §6.2 page 27).

3.2 Prototype de validation [6]

Au delà de la validation de la spécification par une première application offrant à chaque pair du réseau Web-Of-People les trois grandes fonctions

1. gestion des notifications reçues et modification automatique de sa PKB dans certains cas de réponses prédéfinies
2. requêtes d'exploration des autres PKBs dans la mesure de leur accessibilité au pair qui les explore et avec « chaining » éventuel
3. gestion de sa propre PKB et envoi automatique de notification au fur et à mesure des changements susceptibles de concerner d'autres pairs

ce prototype en a bien démontré l'interdépendance en quelque sorte « cyclique » dans un processus d'activité documentaire collaboratif (cf. figures 3 et 4, pages 12 et 13 de la spécification générale)

Différents scénarii ont pu valider la définition de différentes étapes pour l'établissement d'une communication « honnête » :

- mise en accessibilité d'un topique, sans ou avec recommandation ciblée
- intégration d'un topique externe accessible, sans ou avec acceptation d'abonnement
- abonnement du topique externe « accepteur » à tout nouveau post publié dans le topique qu'il a intégré
- création d'un post en commentaire d'un post reçu par abonnement (« personnalisation »)

Ce dernier mécanisme a été très contraint dans le prototype de validation, comme il l'était dans le « SomeOne » de FT R&D. La spécification générale ne l'imposait pas.

Le scénario qui se termine par une intégration réciproque et le choix d'un *webop: subjectIndicatorRef* commun, et des commentaires croisés sur les posts de l'un ou de l'autre, c'est à dire par une mise en quasi-équivalence de topiques de 2 PKBs différentes, a pu être validé.

Des scénarii spécifiques ont été définis (dans la spécification fonctionnelle de l'application de validation [3] annexe 2, page 46 et +) pour mettre en évidence l'existence et les moyens des réponses possibles anti-spam ou anti-broadcasting, dont l'automatisation reste bien sûr à spécifier dans les applications.

3.3 Interface navigationnelle zoomable [9]

Le but de ce prototype était de présenter les résultats d'une ou plusieurs requêtes du service WebOP dans une interface graphique plutôt que sous forme de listes même multiples et bien ordonnées en synopsis.

Compte tenu des problèmes à la fois théoriques (propagation de sous-requêtes en suivant la découverte des liens d'intégration inter PKBs) et pratiques (grande combinatoire des choix de présentation graphique) on a choisi de ne spécifier qu'une application de visualisation mono-vue et de navigation différenciant clairement les séquences intra- et inter-PKBs (« focusing » v. « searching »).

La manipulation effective de jeux de tests construits à partir des données de l'Open Directory Project nous a permis de préférer certaines dimensions d'affichage des topiques parmi toutes celles possibles : le « poids » exprimé en nombre de posts et l'« actualité » exprimée comme le minimum de jours (ou de semaines ou...) écoulés depuis la création du topic ou de son dernier post attaché, sont préférés parce que les plus sûrement variables donc dispersant le mieux les topiques comme des étoiles dans le « starfield ». Mais il pourrait se faire qu'un fort développement des liens inter-PKBs dans le web-of-people produise des dispersions plus manifestes selon les dimensions « réputation » ou « magnitude ». Pour cette raison ces choix doivent rester « implementor defined ».

L'intérêt – provisoire – du prototype d'interface livré aura été de présenter la trilogie fonctionnelle : (1) gestion des notifications reçues, (2) exploration des PKBs extérieures – « univers » externes, (3) modification de sa PKB avec notifications automatiques, sous un aspect plus proche des exigences des utilisateurs de la trilogie classique : (1) gestion des méls reçus, (2) navigation ou recherche sur le web, (3) modification de sa bibliothèque de bookmarks ou de posts en blog.

4 CHOIX TECHNIQUES

Il faut bien distinguer les choix faits dans la spécification générale (lot1) de ceux faits dans le développement des deux applications prototypes : plate-forme de validation multi-serveurs (lot3) et interface de navigation zoomable (lot2). Pour ces derniers il y a eu deux types de choix techniques :

4.1 Choix de spécification

Ceux-ci ont été essentiellement orientés par l'exigence d'interopérabilité dans le cadre des standards du W3C :

- SOAP comme protocole obligatoire – entre serveurs distincts – pour les services de requête et de notification
- RDF comme format obligatoire pour les contenus échangés, méta-données descriptives de message ou de sélection de ressources

ou par les usages établis dans les applications « populaires » : annuaires, gestionnaires de mails, de bookmarks, weblog .. :

- classement ensembliste des topiques avec une simple sémantique de « dossiers » (modèle hiérarchique + alias ou raccourcis)
- langage « navigationnel » pour parcourir des forêts d'arbres avec les liens inter-branches ou inter-arbres
- « permalinks » pour les posts qui, une fois publiés, ne peuvent plus, par principe, qu'être commentés dans de nouveaux posts (« personnalisés » selon le vocabulaire WebOP)

4.2 Choix de développement

Ceux-là ont été plus circonstanciels.

Le choix, dans la plate-forme de validation, de Jena – logiciel avec « copyleft » – comme base de données RDF (utilisation de son interface de requêtes exécutées en interne comme des suites de requêtes SQL présentées à une base MySQL stockant des triplets) procède plus d'une volonté expérimentale et démonstrative que d'une volonté de spécification, qui reste, à notre avis « implementor defined » : Jalios – logiciel avec « copyright » – aurait pu, comme dans SomeOne, être choisie comme base de données RDF, ou bien, dans une autre direction, MySQL aurait aussi bien pu être utilisée directement pour stocker les PKBs et les interroger selon une interprétation relationnelle de leur commun schéma.

Dans cette même plate-forme de validation, l'architecture logicielle a été inspirée de celle de SomeOne pour ce qui est du « wrapper » de base de données, et quelques autres composants ont pu partiellement être réutilisés. Mais ces choix ne relèvent que des perspectives de développement propres à FT R&D.

Le choix, dans cette plate-forme, du « chaining » à la LDAP v1, fait que le client qui envoie une requête au serveur webop, s'il veut une réponse globale, doit, après réception de la réponse provenant du pair auquel il avait adressé sa requête, formuler la ou les requêtes « à suivre » vers les pairs qui ont été désignés par le précédent (s'il y est autorisé par *webop: referencingAllowed*), il peut ainsi « naviguer » de PKB en PKB.

Le choix, dans l'interface de navigation, des objets à projeter (les topiques), des dimensions visibles (« poids », « actualité », « racine », ...) et cachées (« réputation », « magnitude », « profondeur », ...) de l'espace de projection graphique (« starfield »), et des méthodes de zoom (« range slider ») ont été faits autant pour illustrer nos recherches récentes [10] que pour suggérer le vaste choix des possibles en matière de navigation inter-PKBs (distinguant les moments de navigation intra-PKB – « focusing » – des moments de transfert inter-PKBs – « searching », comme dans une métaphore de navigation « inter-galactique »).

Le choix technique d'Infovis © comme toolkit graphique est apparu heureux pour ce qui a été de la spécification des affichages et des interactions possibles et, également, pour le développement qui a pu être réalisé dans un temps relativement court. Mais il présente les limites du genre « starfield display » avec « range sliders ». D'autres modèles d'interface compatible avec le zooming géométrique ou sémantique existent et pourraient être préférés. L'intérêt du modèle « starfield » était de s'obliger à projeter les topiques dans un espace de dimensions qu'il a fallu bien spécifier. La sémantique de la visualisation et du zoom associé est ainsi directement liée à la sémantique universelle des dimensions valables pour toutes les PKBs

5 BILAN ET PERSPECTIVES

Le prototype « WebOP » offre aujourd'hui de larges possibilités de démonstration des services que le « web-of-people » imaginé par FT R&D pourrait apporter au www qui n'est aujourd'hui que partiellement sémantique et très peu coopératif (au contraire il démontre tant de malveillances possibles que la confiance pour un usage coopératif « honnête » a besoin de la mise en place de protocoles nouveaux et mieux contrôlables par tout un chacun).

Il ouvre la possibilité de tester et de garder les traces de scénarii d'échange sophistiqués.

Il met en évidence ce qui pourrait être un standard émergent pour des formats et des protocoles d'échange entre « pairs » dans des réseaux très élargis.

Le souci de sa qualité n'a porté que sur les définitions d'architecture générale, d'interopérabilité, de fonctionnalité, de complétude et de simplicité des services offerts.

Il est très sûrement perfectible sur les aspects ergonomiques (graphisme et interactivité) et opérationnels (confidentialité, robustesse, performance).

L'ENST et, plus précisément, les chercheurs du Groupe III pourraient être intéressés dans plusieurs développements complémentaires :

- simulation en respectant le modèle « WebOP » de scénarii d'échange impliquant de nombreux « pairs » jouant différentes stratégies et de nombreuses « ressources » puis analyse des résultats en terme de topologie du réseau des connaissances (J-L. Dessalles)
- introduction de composants de traitement des textes pour permettre par un médiateur des analyses de distance/proximité entre posts et donc entre topiques voire entre pairs (F. Yvon)
- développement d'une interface navigationnelle zoomable plus « sémantique » (E. Lecolinet, S.Pook)
- développement d'applications conçues sur des modèles de stockage différents (purement relationnel) et/ou avec une intégration différente des trois fonctionnalités principales (browsing inter-PKBs et création de posts « blogués » à la manière de « bookmarks ») (J-M.Saglio)
- recherche sur l'ordonnancement optimal des « chaining » de requête dans différentes hypothèses restrictives de la spécification générale « webop » (T.Abdessalem)

D'autres groupes du Département INFRES de l'ENST pourraient être intéressés par l'étude plus détaillée des aspects confidentialité et sécurité.

6 DETAIL DES LIVRAISONS

Toutes ces livraisons, plus les rapports d'avancement et les documents annexes, se trouvent sur le site <http://www.infres.enst.fr/people/saglio/etudes/WebOfPeople/share.html> en accès réservé ENST et FT R&D

- [1] Ta Tuan Anh, *spécification de la plate-forme « WebOP »*, premier rapport lot1
- [2] Ta Tuan Anh, *spécification de la plat-forme « WebOP »*, spec-lot1-v2_6c.pdf
- [3] Dao Viet Phuong, *spécification fonctionnelle de l'application de validation du « WebOP »*, spec-Fonction-v14.pdf
- [4] Dao Viet Phuong, Ta Tuan Anh, *spécification architecturale de l'application de validation du « WebOP »*, spec-lot3-v0.pdf
- [5] Dao Viet Phuong, *spécification logicielle détaillée de l'application de validation du « WebOP »*, WebOPValidLog.pdf
- [6] Dao Viet Phuong, *package de l'application de validation du « WebOP »*, dossiers WEBOP2 et 3
- [7] Phan Nghiem Long, Saglio Jean-Marc, *spécification de l'application "interface navigationnelle zoomable"*, spec-lot2-v2.01.pdf
- [8] Phan Nghiem Long, *spécification logicielle détaillée de l'application « interface navigationnelle zoomable »*, WebOPAppletLog.pdf
- [9] Phan Nghiem Long, *package de l'application "interface navigationnelle zoomable"*, dossier WEBOP3.0x

7 AUTRES DOCUMENTS

- [10] T-A. Ta et J-M. Saglio. *An Object Zooming Model for the Semantic Web*. "ICTTA". Damas (Syrie), avril 2004.
- [11] T-A. Ta, J-M.Saglio et M. Plu. *An architecture based on semantic weblogs for exploring the Web Of People*. "ICEA SWWC workshop". Valencia (Espagne), août 2004.
- [12] O. Barillet, JP. Château, JM. Deborde, C. Sénot et JM. Saglio. *Weblogs et syndication de ressources web*. Rapport de recherche, École Nationale Supérieure des Télécommunications, ENST2005D001, N. 2005D001, janvier 2005
- [13] J-L. Dessalles, *Evolution et émergence de la communication*, thème de plusieurs publications : <http://www.infres.enst.fr/~jld/papiers/emmercom.html>