

SYFAX : Un système de cache Web Sémantique pour des communautés Distribuées

Habib SMEI*, **Mesaac MAKPANGOU****, **Abdelmajid BEN
HAMADOU*****

* ISET SFAX -Laboratoire LARIS, Habib.smei@isetsf.rnu.tn

** Laboratoire INRIA PARIS, Mesaac.makpangou@inria.fr

*** ISIMS SFAX - Laboratoire LARIS Abdelmajid.benhamadou@isimsf.rnu.tn

RÉSUMÉ: *Le Web est une gigantesque base de données utile pour les chercheurs et les décideurs. Toutefois, l'absence d'informations sur la sémantique des documents Web ne permet pas d'exploiter efficacement les informations disponibles sur le Web. En particulier, les programmes d'indexation et d'extraction automatique des méta-données sont très inefficaces, comparer à ce qu'un indexeur humain peut obtenir grâce à la connaissance de la sémantique des documents. Cet article présente SYFAX, un système de gestion de caches sémantiques et portails communautaires. Un cache communautaire est constitué d'une part, d'un réseau d'entrepôts d'informations pertinentes pour une communauté d'utilisateurs ; d'autre part, un système d'extraction et de gestion de méta-données attachées aux documents partagés par la communauté cible. Les documents peuvent être produits par les utilisateurs eux-mêmes ou provenir du Web. Les méta-données, elles peuvent être spécifiées par les utilisateurs, par des experts humains, ou par des outils automatiques d'indexation et d'annotation. Nous discutons les bénéfices que les communautés d'utilisateurs coopérants peuvent tirer du système SYFAX.*

ABSTRACT: *The Web is a gigantic useful data base for the researchers and the decision-makers. However, the absence of information on the semantics of the Web resources doesn't permit to exploit the available information efficiently. In particular, the programs of indexing and automatic extraction of Meta data are inefficient compared with that a human indexer can get. This article presents SYFAX, a semantic cache management and communal portal. A communal cache is constituted on the one hand of warehouses containing judicious information for a community of users; on the other hand, a system of extraction and management of Meta data attached to the documentations shared by the community target. These resources can be produced by the users themselves or come from the Web. Meta data can be specified by the users, by human experts, or by automatic indexing tools. We debate the earnings that the cooperating user communities can pull the SYFAX system.*

MOTS CLÉS : Web ; sémantique ; méta-données ; recommandations ; profilage ; notification ; indexation ; entrepôt.

KEY WORDS: Web; semantic; Meta data; recommendations; streamlining; notification; indexing; warehouse.

1. Introduction

Le Web constitue aujourd'hui le principal véhicule de partage d'informations à l'échelle mondiale. Il permet aux utilisateurs répartis d'accéder à une gigantesque base de données, concernant des sujets très variés. De nombreux outils de recherche et de filtrage (moteurs de recherche, systèmes de recommandation, portails communautaires) ont été proposés pour faciliter l'accès aux informations disponibles sur le Web. Aucun de ces outils ne permet à une communauté d'utilisateurs, de disposer d'un véritable système d'informations dédié. Si nous considérons par exemple la communauté de personnes impliquées dans l'enseignement supérieur et la recherche dans un pays, elles ont besoin de partager des informations relatives à leur domaine : programmes d'enseignement, bibliographies, notes de cours, tutoriaux, résultats d'enquêtes, rapports divers, etc. Pour construire à partir d'informations disponibles sur le Web les systèmes d'informations performants dédiés à des communautés d'utilisateurs, on a besoin d'un outil qui soit capable de capturer les besoins de chaque communauté et de déterminer la pertinence des informations par rapport à ces besoins. Il est aussi nécessaire de renforcer la coopération interne grâce notamment au partage d'expertises disponibles au sein de la communauté. Cet article présente SYFAX, un système qui permet d'offrir à chaque communauté d'utilisateurs, la vue d'un **cache Web sémantique distribué**, qui a pour but d'améliorer la découverte et l'accès aux documents Web pertinents pour cette communauté d'utilisateurs. SYFAX assure la fouille du Web, l'extraction des données sémantiques (i.e, méta-données) concernant les documents et les utilisateurs, et le stockage des documents et de leurs méta-données associées. SYFAX offre aux utilisateurs des mécanismes de coopération afin de leur permettre de partager aussi bien leurs expériences que les informations à leur disposition. L'étude de cas devant guider la conception et la validation du système SYFAX est celle de la communauté universitaire. L'accent mis sur cette étude de cas ne signifie pas que SYFAX est dédié à ce cas. La méthodologie proposée n'est pas restreinte au contexte pédagogique, elle est tout à fait applicable à toute autre communauté d'utilisateurs. Le reste de l'article s'organise comme suit. Le paragraphe 2 présente le cas du système d'informations pédagogiques et explicite les motivations de ce travail à la lumière des besoins spécifiques à la communauté des enseignants et chercheurs. L'architecture du système SYFAX est décrite au paragraphe 3. Ce paragraphe identifie aussi les différents rôles des composants et décrit les interfaces ainsi que de la première maquette qui a été réalisée dans le but de valider quelques composants du système SYFAX. Le paragraphe 4 donne un aperçu des systèmes comparables et explicite les points qui distinguent SYFAX de ces systèmes. La conclusion expose les perspectives de développement du système.

2. Le cas de systèmes d'informations pédagogiques

Les nouvelles technologies de l'information et de la communication deviennent de plus en plus présentes dans notre vie. Elles ont une influence considérable sur les

individus, sur leurs méthodes de travail et de prise de décision. Elles offrent des moyens efficaces de traitement, de communication et d'échange rapide d'informations. Ces possibilités de traitement, de communication et de partage d'informations sont très utiles dans les domaines où les habiletés de visualisation et d'échange automatique d'informations jouent un grand rôle. L'activité pédagogique est l'un des secteurs primaires qui peut pleinement profiter de ces nouvelles technologies, et ce afin d'améliorer les moyens pédagogiques et les outils d'apprentissage. Cela impose bien évidemment une modification des méthodes de travail des enseignants et plus globalement, une modification de la mentalité et du mode de fonctionnement des établissements du système éducatif. En effet, Les membres de la communauté éducative doivent apprendre à mettre en commun les ressources pédagogiques propres à chacun pour que toute la communauté s'en profite. Les enseignants doivent améliorer les méthodes et les moyens pédagogiques qu'ils utilisent en permettant l'accès à des présentations interactives facilitant les possibilités d'apprentissage et de compréhension chez l'étudiant. D'autre côté, la communication entre les différents acteurs de la formation s'avère aussi primordiale et ce dans le but de discuter les contenus, d'échanger les documents et les idées, de diffuser les ressources numériques et audiovisuelles mises à disposition des membres de la communauté éducative. En ce qui concerne les apprenants (étudiants), ils sont de plus en plus appelés à devenir des acteurs actifs qui construisent leurs apprentissages en interaction avec un environnement ouvert et riche d'informations. Mais ce n'est pas parce qu'il navigue sur Web qu'un étudiant apprend. Bien au contraire, la surabondance et l'absence de structuration de l'information ne permettent pas une exploitation efficace des ressources qui y sont contenues. Il devient, donc, de plus en plus nécessaire d'offrir à l'étudiant un environnement structuré et riche de documents pertinents et bien répertoriés, ainsi que des outils nécessaires à sa formation. Mais il faut que le contenu soit contrôlé et continuellement supervisé par les enseignants et les experts de l'éducation. On assure ainsi, une qualité de ressources sur le plan de l'efficacité et de la pertinence à la disposition des membres de la communauté. De ce fait, la conception d'un système d'information pédagogique utilisant les NTIC s'avère une nécessité afin de favoriser l'implémentation de nouveaux outils et services pédagogiques, et de créer un environnement rassemblant des acteurs divers (chercheurs, enseignants, étudiants, ...). Un tel système peut être réalisé en exploitant les ressources pédagogiques présents déjà dans le Web, mais aussi en exploitant d'autres ressources issues des acteurs spécialistes dans le domaine d'enseignement.

C'est dans ce cadre, que se situe le système SYFAX, qui permet de créer un environnement coopératif, assurant une participation active aux membres de la communauté éducative, en offrant des services leur permettant de participer à l'amélioration du système et d'utiliser des services de coopération et d'échange d'idées. Plus précisément, la construction d'un système d'information pédagogique efficace exige en premier lieu le support d'une interface de stockage de documents que les auteurs peuvent utiliser pour rendre accessibles leurs documents (cours, articles, sujets d'examens ou autres). Ce service doit permettre de construire un

entrepôt de ressources pédagogiques hautement disponible et facilement exploitable. En second lieu, il faut un outil capable d'extraire les ressources pertinentes qui existent déjà dans le Web pour les injecter dans le système d'informations. Enfin, on a besoin de disposer de mécanismes de coopération pour faciliter les échanges et le partage d'expertise au sein de la communauté.

3. Système SYFAX

L'objectif du système SYFAX est de fournir à une communauté (i.e., universitaire) un portail facilitant le partage des documents pertinents aux membres de cette communauté. Pour découvrir les documents pertinents pour une communauté, SYFAX repose sur la collaboration de ses membres (en leurs mettant à disposition un système d'annotation pour donner leurs avis sur les documents) et des systèmes de fouille du Web personnalisés en fonction des attentes de cette communauté. Chaque membre contribue en mettant à disposition les informations qu'il juge pertinentes, alors que les systèmes de fouille aspirent des documents disponibles sur le Web qui intéressent cette communauté. Les systèmes de fouille peuvent donc être vus comme des membres experts de la communauté, qui auraient la charge de faire de la veille sur le Web pour le compte de la communauté entière. L'ensemble de documents produits par les membres (normaux ou experts) constitue le système d'informations partagées par la communauté. SYFAX vise à faciliter la publication, la découverte et l'accès à ces documents par tous les membres. Cela conduit aussi à créer des communautés virtuelles d'utilisateurs partageants les mêmes goûts et centres d'intérêts.

3.1 Architecture globale

SYFAX est basé sur un modèle Client/Serveur dans lequel chaque utilisateur est un **client SYFAX** qui collabore avec d'autres clients via le **Serveur SYFAX**.

Un **client SYFAX** est assimilable à un assistant personnel ; il tourne chez chaque utilisateur du système SYFAX. Un utilisateur est, soit un membre de la communauté qui publie et accède à des documents intéressant la communauté, soit expert qui aspire des documents publics disponibles sur le Web et les met à la disposition des autres membres. Le client SYFAX propose aux utilisateurs des interfaces de recherche, de gestion de documents ainsi que des outils de filtrage collaboratif. Les **serveurs SYFAX** gèrent les documents et méta-données qui sont associées. Ils coopèrent pour offrir aux utilisateurs de SYFAX l'abstraction d'un cache Web sémantique distribué performant aussi bien en terme de latence d'accès que de pertinence des résultats de recherche. Ils offrent aux utilisateurs des outils de recherche et de filtrage adaptés. Le filtrage se base essentiellement sur les avis des utilisateurs. En effet, chaque utilisateur peut donner son avis (aspect annotation) sur chaque document accédé. Les annotations des utilisateurs pour un document donné peuvent être consultées par les utilisateurs et faire même l'objet de nouvelles annotations. La figure (Figure 1) illustre l'architecture répartie du système SYFAX. Dans cet exemple, on a deux serveurs SYFAX. Sur chacun d'eux, on a un utilisateur

simple, qui peut réaliser des fonctions de gestion de documents (recherche, ajout, modifications,...) et un expert qui peut lui-même effectuer des opérations de gestion de document et qui est chargé d'aspirer les documents intéressants du Web pour le publier dans SYFAX.

3.1.1.1.1.1 *Architecture d'un serveur SYFAX*

Un serveur SYFAX doit permettre la réalisation d'un ensemble de fonctionnalités, notamment la recherche et la présentation des informations ; la collaboration inter-utilisateurs ; et l'offre d'une qualité de service satisfaisante concernant la disponibilité de documents, la pertinence des résultats de recherche, la latence des notifications d'événements intéressants aux utilisateurs et la latence d'accès aux informations. Pour réaliser ces différentes fonctionnalités, un serveur SYFAX est constitué de cinq services : un entrepôt de documents, un service de recherche d'informations, un service de veille et d'alerte des utilisateurs, et un portail de gestion d'interactions avec les utilisateurs. La figure 2 illustre les interactions entre ces différents composants.

Entrepôt de documents : Ce composant stocke les documents accessibles aux

utilisateurs de la communauté. L'entrepôt de données est alimenté par deux classes d'utilisateurs :

- Les auteurs déposent les documents qu'ils souhaitent rendre accessibles aux membres de la communauté.
- Les experts qui fouillent le Web et récupèrent des nouveaux documents se rapportant aux sujets traités.

Service de recherche : Il permet à tout utilisateur de rechercher des informations dans l'entrepôt. Ce service gère une base des méta-données associées aux documents disponibles dans l'entrepôt de son serveur. La nature des méta-données dépend des besoins de la communauté concernée. Pour le système d'informations pédagogiques, les informations telles que le titre, le résumé, le public visé, les enseignements concernés font partie des méta-données associées à chaque document. Le service de recherche exploite la base de méta-données ainsi que les profils utilisateurs pour fournir des résultats pertinents aux requêtes des utilisateurs connus du système. Lorsqu'une recherche est infructueuse sur un serveur SYFAX, on relaie la requête aux autres serveurs de la communauté qui l'exécutent et retournent les résultats au serveur ayant initié la requête.

Service de veille et d'alerte : Ce service a la charge de surveiller la dynamique de l'entrepôt et d'alerter les utilisateurs qui le souhaitent lorsque les événements qui les intéressent surviennent. Les événements considérés peuvent être de natures diverses, par exemple le dépôt d'un document traitant d'un sujet donné, sa mise à jour ou sa suppression ou toute autre opération pouvant être utile pour les membres de la communauté. Ce composant intègre aussi un système de recommandation basé sur un filtrage collaboratif pour les membres de la communauté. Tout utilisateur qui le souhaite peut donner son avis sur tout document au moment de la consultation. Le système propose deux types d'avis : le premier est relatif à la pertinence du document par rapport au profil de l'utilisateur alors que le second donne l'appréciation de l'utilisateur quant à la qualité du document. L'idée directrice ici est que si un document est jugé pertinent par un utilisateur par rapport à son profil, il peut être considéré comme pertinent pour les utilisateurs ayant des « profils proches » de celui de cet utilisateur. Si de plus un tel document est considéré comme de bonne qualité par un expert reconnu, c'est un document qu'on peut recommander à ceux que cela intéresse. Enfin, ce service offre aussi des abstractions de coopération inter-utilisateurs (forums, espace de travail partagé, tableau virtuel partagé) aux membres qui souhaitent créer des groupes restreints. Au total, le service de veille et d'alerte offre aux utilisateurs quatre interfaces pour : les enregistrements et la gestion des profils d'utilisation des utilisateurs ; les annotations des documents ; les souscriptions des utilisateurs ; les créations des abstractions de coopération inter-utilisateurs et la gestion de groupes restreints.

Le portail d'accès (ou l'Interface Utilisateur) : Ce composant gère les interactions avec les utilisateurs du système SYFAX. C'est-à-dire que les utilisateurs ne voient pas les différents composants ci-dessus, mais une seule interface offerte par le système global. Ce composant se charge ensuite de répercuter les requêtes sur les différents composants.

3.1.1.1.2 *Prototype SYFAX*

Nous avons développé un premier prototype comprenant les principaux composants du serveur SYFAX. Nous utilisons des outils de développement libres (open source) à cause de leurs disponibilité et leurs flexibilité. Nous utilisons un serveur Web APACHE comme serveur de stockage. Il s'agit du serveur le plus utilisé dans le monde Internet et Intranet. Il assure une grande flexibilité d'utilisation et aussi une grande sécurité. Il possède des nombreuses fonctionnalités dont la possibilité de définir une configuration spécifique à chaque fichier ou répertoire partagé, ainsi que de définir des restrictions d'accès grâce aux fichiers htaccess. Pour la gestion des bases de données, nous utilisons MYSQL qui est un système de gestion de bases de données puissant. Il s'agit d'un système permettant de mettre en place des bases de données relationnelles très complexes tout en offrant un niveau de fiabilité irréprochable ainsi que des bonnes performances. Le langage des requêtes utilisé est le PHP qui est un langage pour le Web dynamique, très facile à l'interfacer avec MYSQL. Le prototype actuel comprend le service d'inscription des ressources et de Gestion de l'entrepôt de documents ainsi que l'outil de recherche.

Le Service d'inscription des ressources permet les opérations de stockage, de suppression et de mise à jour des documents de l'entrepôt. Il permet aussi d'inscrire les méta-données correspondant dans la base sémantique des méta-données

La gestion de l'entrepôt repose sur un ensemble de fonctions permettent de :

- Ajouter des documents (l'auteur du document)

Les ajouts des documents et des méta-données sont effectués par les utilisateurs enregistrés dans le système. Lors de l'enregistrement, l'utilisateur remplit un formulaire (informations sur le profil et le contexte d'utilisation) et reçoit un compte et un mot de passe qui l'identifie pour toutes les opérations futures. Au moment du dépôt, une fois l'auteur s'est identifié auprès du système, on lui propose un formulaire afin qu'il complète les informations descriptives du

document (par exemple, le titre du document, les mots-clés associés, ...)

- Modifier des documents (l'auteur du document) : Une interface de mise à jour est offerte aux auteurs désirant modifier le document ou les informations descriptives. De même l'utilisation de cette interface requiert un compte et mot de passe.
- Supprimer des documents (l'auteur du document) : La suppression d'un document se fait par son propriétaire. Le système met à jour la base de description des documents.
- Ajouter des mots clés pour un document : Tout utilisateur peut proposer des mots clés pour un document, néanmoins, le système ne les prendra en conséquence que s'ils sont proposés par plusieurs utilisateurs (Le système n'ajoute effectivement un mot clés à l'ensemble des mots clés du document que si ce mot a été proposé par plusieurs clients).

Le service de recherche permet à un utilisateur à partir d'une requête de rechercher une information dans l'entrepôt. Ce système gère une base de méta-données sémantiques associées aux documents disponibles dans l'entrepôt de la communauté. C'est en se basant sur cette base de méta-données que le système de recherche fournit des réponses pertinentes aux requêtes des utilisateurs.

L'interface de recherche offre à l'utilisateur plusieurs paramètres qu'il doit spécifier au moment de la requête. Parmi les paramètres supportés, on trouve :

- Les mots clés : l'utilisateur précise l'ensemble des mots clés dans sa recherche, l'interface lui propose en effet de choisir les mots qui doivent être obligatoirement présents, les expressions complètes, etc
- Le Titre du document, en effet l'utilisateur peut proposer le titre du document s'il le connaît ; D'autres options sont disponibles aussi, comme la langue du document, son type,

...

Une fois le résultat de la recherche est affiché, l'utilisateur peut choisir l'affichage du document ou simplement son résumé.

4. Systèmes comparables

Dans ce paragraphe nous présentons des systèmes dont les fonctionnalités sont proches de celui de SYFAX ou intégrant des outils ou modules traités par notre système. La description donnée de ces systèmes n'est pas exhaustive et ne constitue pas une étude critique approfondie de leurs caractéristiques et leurs limites

4.1 Moteurs de Recherche

Un moteur de recherche est composé d'un robot, d'une base de données et d'un agent. A partir d'une page initiale comprenant des liens hypertexte, le robot parcourt récursivement toutes les pages atteignables et collecte les informations pertinentes. Les informations collectées sont ensuite stockées dans la base de données du moteur

pour être exploitées par l'agent lors des interrogations.

ARCHIE fut le premier moteur de recherche développé dans le but de faire l'inventaire des bases de données toutes les nuits [23]. Depuis, plusieurs moteurs de recherche ont été développés. Les plus connus d'entre eux sont Yahoo [24], Google [25], [AltaVista](#) [26], [Voilà](#) [27], [Fast Search](#) [28], [HotBot](#) [29], [Northern Light](#) [30], [GO](#) [31], Lycos [32], Excite [33], et Nomade [34]. Pour les utilisateurs, le principal problème que posent les moteurs de recherche est l'importance du « bruit » dans les réponses, c'est-à-dire des informations trouvées qui n'intéressent pas les utilisateurs. Ce problème est dû à l'absence d'informations concernant d'une part la sémantique des documents indexés et d'autre part, le contexte d'utilisation des informations recherchées. Pour tenter de remédier à ce problème, certains robots exploitent des balises qui sont souvent présentes dans les sources HTML des documents. Par exemple, la balise "keyword" permet de spécifier les mots clés d'un document. Cette solution a toutefois un impact limité sur les performances. Premièrement, les balises ne sont pas présentes dans une majorité des documents. En particulier, un grand nombre d'utilisateurs reposent sur des outils de génération automatiques et omettent d'insérer des balises. En second lieu, les éléments de description les plus pertinents dépendent de la communauté des utilisateurs d'un document ; aujourd'hui, les moteurs de recherche ne disposent pas d'informations sur le contexte d'utilisation. SYFAX propose, un mécanisme de recherche à deux niveaux. Au premier niveau, des systèmes de fouille personnalisés selon les besoins des communautés cibles, rendent un service comparable à celui des robots dans les moteurs de recherche existants. De plus ces systèmes de fouille participent à l'analyse et la génération des méta-données exploitables par la recherche de second niveau. Au second niveau, on exploite les méta-données ce qui permet de répondre avec plus de pertinence aux requêtes des utilisateurs.

4.1.1.1.1.1 Portails communautaires

Les portails communautaires [20, 21, 22] recensent les informations sur des thématiques bien déterminées, comme par exemple le sport, la musique, les voyages, la gastronomie, etc. Les portails communautaires permettent de réunir des gens qui partagent la même passion et leur offrent des informations qui correspondent bien à leurs attentes. Cependant, les utilisateurs de ces portails sont des simples consommateurs d'information et de services. Les utilisateurs ne contribuent pas à améliorer les services proposés. Une participation active des utilisateurs permettrait d'enrichir la base de données des portails par l'ajout de nouveaux documents ou des annotations qui pourraient servir à d'autres internautes partageant le même portail.

Parmi les portails communautaires comparables à SYFAX, on peut citer Spédago, LEARN-NET, et Nautilux. Spédago [35] est un portail pédagogique permettant d'héberger des cours dans le domaine informatique ; il intègre aussi un module de recherche dans l'arborescence des catégories, ou par mot-clé. LEARN-NET (LEARNing NETwork for Teachers and Trainers) [36] est un projet supporté par le programme SOCRATES de la Communauté Européenne, qui vise la mise en place d'un réseau d'apprentissage collaboratif pour les formateurs et les enseignants. Un autre système à vocation pédagogique est Nautilux [37], un serveur

Internet/Intranet à vocation pédagogique. Il comporte un mini portail des ressources pédagogiques, et un forum de discussion. SYFAX se distingue de ces différents systèmes en permettant une participation active des utilisateurs (publication des documents, annotations, coopération inter-utilisateurs) et par le support d'un module de recommandation de documents que le système juge pertinents.

4.1.1.1.1.2 Systèmes de recommandation

Les systèmes de recommandations constituent une autre classe de solution pour aider à la recherche d'informations dans le Web [9, 17, 19]. Un système de recommandation met aux prises trois types d'entités de base : les producteurs qui élaborent des propositions et suggestions pour des utilisateurs ; les consommateurs qui sont les utilisateurs désirant recevoir des recommandations du système ; et le module de calcul de recommandations dont le rôle est de déterminer quelle ressource doit être envoyée à quel utilisateur. Les systèmes de recommandations se basent principalement sur les profils utilisateurs qui sont des structures de données décrivant les centres d'intérêts des utilisateurs dans l'espace des objets à recommander. Une fois une telle structure construite, on peut l'utiliser soit pour filtrer les objets disponibles (on parle alors de filtrage basé sur le contenu), soit pour recommander à un utilisateur des informations ayant donné satisfaction à d'autres utilisateurs de profil semblable au sien (on parle alors de filtrage collaboratif). Deux exemples de systèmes de recommandation qui se rapprochent le plus du système SYFAX sont Pharos et Firefly.

Pharos [38] est un système de recommandation qui permet à des groupes de personnes d'évaluer et de se recommander des pages Web pertinentes selon leur domaine d'intérêt. Il permet le partage d'annotations sur les documents Web dans les forums (aussi appelé des canaux). Les annotations sont des meta-données structurées que n'importe qui peut attacher à tout document Web. Pharos calcule une recommandation personnalisée pour chaque document qui est la synthèse des annotations à ce document. La personnalisation est accomplie de deux manières. Premièrement, Pharos calcule automatiquement la corrélation entre les membres d'un canal et donne plus de poids aux gens qui ont de profils semblables à celui du demandeur. Deuxièmement, si on connaît d'autres canaux, on peut indiquer à Pharos la situation du profil par rapport à ces canaux. Les bases d'annotations sont exploitées de façon collaborative. Pour affiner l'utilisation de ce service, des algorithmes de synthèse produisent pour chaque membre du groupe des recommandations personnalisées de pages intéressantes.

Firefly est un système de recommandation de sites, de films, de groupes musicaux basé sur les travaux de Pattie Maes et de son équipe. Ce service est désormais associé à My yahoo, service de diffusion personnalisé d'information du service Yahoo. L'utilisateur donne le maximum de renseignements sur ses goûts musicaux, ses goûts cinématographiques ou ses sites web préférés. Le service est censé pouvoir lui conseiller d'autres sites, groupes musicaux, etc. en fonction des goûts des autres utilisateurs du service. Le service peut aussi mettre en relation les personnes ayant des goûts communs, permettant ainsi la constitution de communautés virtuelles. L'objectif est de recréer des communautés d'utilisateurs de l'Internet en favorisant

l'échange d'information et le contact entre personnes ayant les mêmes centres d'intérêts. Les systèmes de recommandations permettent de réduire le bruit dans les résultats de recherche. Toutefois, les systèmes de recommandation existants présentent eux aussi des limites. En particulier, la capture de profil d'un utilisateur n'est pas aisée ; le contexte d'utilisation est insuffisamment pris en compte ; la classification des documents est souvent manuelle, ce qui conduit à restreindre de façon excessive l'ensemble de documents accessibles aux utilisateurs.

5. Conclusion et perspectives

Construire au dessus du Web des systèmes d'informations dédiés à des communautés d'utilisateurs données est un des défis majeurs à relever pour faciliter et rendre plus efficace la recherche d'informations sur le Web. Cela favorise le partage et la collaboration entre les utilisateurs appartenant à une même communauté. SYFAX propose à la fois une méthodologie et des outils pour construire et exploiter des systèmes d'informations dédiés à des communautés d'utilisateurs. La méthodologie préconisée consiste à attacher des méta-données aux informations partagées par la communauté. L'association des méta-données à tout document est un élément clé dans l'évolution vers le Web sémantique, rendant ainsi plus facile la découverte des informations pertinentes en fonction du contexte d'utilisation. SYFAX préconise aussi de laisser à des systèmes de fouille et/ou des utilisateurs experts le soin d'associer des méta-données aux documents partagés par la communauté. Nous avons présenté dans ce papier l'architecture du système SYFAX et ainsi qu'un premier prototype dont le but est de valider l'architecture proposée. A court terme, nous visons la terminaison du prototype actuel. En particulier, la coopération inter-serveurs et le composant de veille et d'alerte doivent être prototypés. Pour ce faire nous comptons utiliser des protocoles de coopération comme le WebDAV[40], un des protocoles qui permet la création d'entreprises virtuelles, et qui capable de servir une large gamme d'applications collaboratives.

Toute cette première phase vise à construire un système d'informations pédagogiques déconnecté du Web. Par la suite, nous nous intéressons aux outils qui pourraient être intégrés dans un système de fouille du Web pour enrichir notre système d'informations pédagogiques.

Bibliographie

- [1] Christine Michel. Evaluation de systèmes de recherche d'information comportant une fonctionnalité de filtrage par des mesures internes, Thèse de doctorat, Université Lumière, Lyon 2, Institut de la communication, 1999
- [2] Marti A. Hearst. Interfaces for Searching the Web. <http://www.sciam.com/0397issue/0397hearst.html>
- [3] Joe Barker. Principe de fonctionnement des méta-moteurs, leurs limites, sélection des plus pertinents d'entre eux. Avril 2000, www.indicateur.com/Doc/etude_these.htm.
- [4] Quentin Delacroix. Un système pour la recherche plein texte et la consultation hypertexte de documents techniques. thèse de doctorat en informatique, 1999 [QUELLE UNIVERSITE]

- [5] Dr. Sheila McIlraith . conférence : Mobilisation du Web sémantique avec des services Web compatibles DAML. Knowledge Systems Laboratory ,Stanford University Stanford, CA. le 31 mai 2001
- [6] [Steve Lawrence](#) and [C. Lee Giles](#). Searching the Web: General and Scientific Information Access. IEEE Communications, Volume 37, Number 1, pp. 116-122, 1999.
- [7] Recommendations Sharing Forum for Internet Communities, <http://webtools.dyade.fr/pharos/> ; [8] Media Me trix. <http://www.mediametrix.com/>
- [9] Huberman, B. A. & Adamic, L. A. Evolutionary Dynamics of the World Wide Web <http://www.parc.xerox.com/istl/groups/iea/www/growth.html>
- [10] Dublin Core. The Dublin Core: A Simple Content Description Model for Electronic Resources (1999); <http://purl.oclc.org/dc/>
- [11] Alexander Lebedev. Best search engines for finding scientific information in the Web. Moscow State University, <http://www.chem.msu.su/eng/comparison.html>
- [12] Brigitte Trousse. Illustrations pour l'aide à la recherche d'informations sur le Web Colloque des 3 et 4 décembre 1999- École normale supérieure.
- [13] Michel Jaczynski, Rushed Kanawati, Brigitte Trousse. Systèmes de recommandations, http://www.inria.fr/rapportsactivite/RA98/aid/resul_man.html
- [14] J. Ben Schafer, Joseph A. Konstan, and John Riedl. E-Commerce Recommendation Applications. Proceedings of the Data Mining and Knowledge Discovery conference. Kohavi, Provost (2001).
- [15] Olivier Andrieu. Trouver l'info sur le Web. Edition : [Eyrolles](#) octobre 2001, ISBN : 2-212-11000-6.
- [16] Philippe Lefèvre. La recherche d'informations, du texte intégral au thésaurus. Edition Hermes. 2000, ISBN : 2-7462-0173-9
- [17] P. RESNICK, H. R. VARIAN. Recommender systems.«Communications of the ACM », numéro 3, volume 40,1997, pages 56-58.
- [18] G. KAHN. Natural Semantics. « Proceedings of STACS'87 », Computer Science n247, Springer-Verlag, Berlin, février 1987
- [19] U. SHARDANAND, P. MAES. Social Information Filtering: Algorithms for Automating Word of mouth. In « CHI'95: Mosaic of creativity », ACM, pages 210-217, address Denver, Colorado, mai, 1995.
- [20] Acier.Org : <http://www.ffa.fr/homepage.htm>;
- [21] Le portail d'Usinor : <http://www.usinor.com/francais/>
- [22] Community Web Portals References: <http://www.ics.forth.gr/~christop/Portals.html>
- [23] Le serveur Archie du CRU : <http://www.cru.fr/serveurs/archie.html>
- [24] Yahoo [<http://www.yahoo.com>] ; [25] Google [<http://www.google.com>]
- [26] [AltaVista](#) [<http://www.av.com>] ; [27] [Voilà](#) [<http://www.voila.fr>]
- [28] [Fast Search](#) [<http://www.Alltheweb.com>] ; [29] [HotBot](#) [<http://www.Hotbot.com>],
- [30] [Northern Light](#) [<http://www.Northernlight.com>] ; [31] [GO](#) [<http://www.Go.com>],
- [32] Lycos [<http://www.Lycos.fr>] ; [33] Excite [<http://www.Excite.com>],
- [34] Nomade [<http://www.Nomade.fr>] ; [35] Spédago [<http://spedago.unice.fr>]
- [36] LEARN-NETT [<http://www.ulb.ac.be/project/metrie/learnett1.htm>]
- [37] Nautilux [<http://www.lenautlux.net>] ; [38] PHAROS [<http://webtools.dyade.fr/pharos/>]
- [39] FIREFLY [<http://www.firefly.com>] ; [40] WEBDAV [<http://www.webdav.org>]