Vers une personnalisation de la navigation par l'apprentissage de profils utilisateurs.

Khaled Khelif, Yassine Mrabet, Rose Dieng-Kuntz

Vers une personnalisation de la navigation par l'apprentissage de profils utilisateurs

Khaled Khelif[1], Yassine Mrabet[1], Rose Dieng-Kuntz[1]

[1] INRIA Sophia Antipolis Méditerranée
2004, route des lucioles 06902 Sophia Antipolis - FRANCE
{khaled.khelif, yassine.mrabet. rose.dieng}@sophia.inria.fr

Résumé : L'exploitation des interactions utilisateurs-sites Web peut jouer un rôle important pour l'amélioration de la navigation dans le futur Web. Dans une mesure plus particulière, dégager et reconnaître les profils des internautes à partir de ces données peut aider les navigateurs et les sites Web à personnaliser les sessions utilisateurs tout en recommandant des ressources spécifiques. Nous présentons à travers ce papier une solution de reconnaissance de profils basée sur les technologies du Web sémantique. Cette approche tire ses avantages de l'utilisation des ontologies, des annotations sémantiques sur les ressources Web et d'un moteur d'inférence et d'un moteur de recherche sémantique.

Mots-clés : Détection de profils, ontologies, annotations, navigation dans le web sémantique

1 Introduction

Afin d'introduire notre travail sur l'apprentissage des profils des utilisateurs, nous commençons d'abord par présenter le contexte et les motivations de ce travail pour ensuite essayer de le situer dans l'état de l'art de ce domaine.

1.1 Contexte et motivations

Ce travail rentre dans le cadre du projet européen Sealife (Schroeder et al., 2006). Le but de ce projet est la conception et le développement d'un navigateur sémantique pour le domaine des sciences de la vie. Ce navigateur permettra de faire le lien entre le Web existant et les infrastructures émergentes liées à la eScience. Un cas d'utilisation du navigateur Sealife consiste à faire le lien entre les informations contenues dans les pages Web des sites biomédicaux et d'autres sources de connaissances secondaires (ontologies, flux RSS, etc.). L'implémentation de ce cas d'utilisation permettra de montrer la possibilité de fournir automatiquement des informations additionnelles sur les ressources visitées par l'utilisateur tout en utilisant la sémantique contenue dans les ressources du Web. Afin d'atteindre ce but, le navigateur devra en premier lieu reconnaître le profil de l'utilisateur pour retrouver la source secondaire appropriée.

Le scénario de test de ce cas d'utilisation sera réalisé sur le site Web Neli[1], une librairie dédiée à l'investigation, le traitement et la prévention des maladies infectieuses.

Dans cet article, nous proposons une approche générique permettant d'aider à la personnalisation de la navigation sur le Web en général et de réponde à ce cas d'utilisation en particulier.

1.2 Détection des profils

Depuis des années, nous avons remarqué une augmentation significative des travaux sur la détection des profils et sur la navigation guidée par l'usage ; les buts des ces travaux sont différents : (i) prédire la future requête http afin d'optimiser le réseau [11], (ii) conseils pour le commerce électronique (Spiliopoulou et al., 1999) (Cooley et al., 1999) (Buchner & Mulvenna, 1999), (iii) prédire les comportements futurs des utilisateurs (Shahabi, al., 1997) (Nasraoui et al., 2000), (iv) fournir des mots clés contextuels pour faciliter la recherche d'informations (Ahu et al., 2004), et (v) personnaliser la navigation et fournir des recommandations aux utilisateurs (Honghua & Bamshad, 2002).

Une approche de détection de profils repose sur trois piliers :

- une méthode de construction du profil courant de l'utilisateur, soit à partir des fichiers de logs, soit à partir des informations fournies par des agents logiciels (navigateurs, proxies, etc.).
- une liste de profils modèles pour aider à la prédiction des profils des utilisateurs.
- une méthode permettant de tester la similitude entre le profil courant de l'utilisateur et les modèles de profils prédéfinis.

Dans la majorité des cas et des domaines, il est difficile de trouver des profils modèles déjà définis car cette tâche nécessite beaucoup d'efforts de la part des experts, ce qui nous amène à parler de construction, d'enrichissement et de classification de motifs de profils à partir des observations sur la navigation des utilisateurs.

Les travaux proposés jusqu'ici couvrent assez bien ces différents besoins mais n'exploitent pas la sémantique contenue dans les différentes ressources. (Cooley et al., 1999), (Bamshad et al., 2000) et (Nasraoui et al., 2000) définissent le profil comme étant une collection de pages Web, (Ahu et al., 2004) utilise des dictionnaires et des thesaurus (limités sémantiquement) pour détecter des informations dans les pages, et, (Honghua & Bamshad, 2002) utilise les ontologies pour annoter les ressources mais ne repose pas sur les raisonnements et les inférences que nous pouvons faire avec ces modèles. Ces différents travaux ont pour but de fournir des informations pouvant intéresser un groupe d'utilisateurs et de permettre de leur faire des recommandations.

Le but de notre travail est de fournir une approche qui (i) permet de proposer non seulement des documents mais aussi de trouver aussi les ressources d'un domaine particulier pouvant intéresser des utilisateurs qui ont la même activité professionnelle,

[1] http://www.neli.org.uk/IntegratedCRD.nsf/NeLI_Home1?OpenForm

(ii) repose sur les technologies du Web sémantique comme RDF, RDFS, OWL, SPARQL[2] et des mécanismes de raisonnement basés sur le moteur sémantique Corese (Corby et al., 2004), et (iii) est générique et indépendante du domaine.

Cet article décrit : une description générale de notre approche, une brève description du système SUPROD (Semantic User PROfile Detector) et de son architecture, une vue sur le processus global défini et enfin une discussion sur la pertinence des résultats obtenus.

2 Notre proposition

Dans cette section, nous présentons l'approche générique que nous avons proposé pour la détection des profils des utilisateurs en se basant sur leurs traces de navigation.

2.1 Une architecture et un système : SUPROD

Ici, nous commençons par présenter l'architecture globale du système SUPROD que nous avons développé pour valider notre approche (Fig.1).

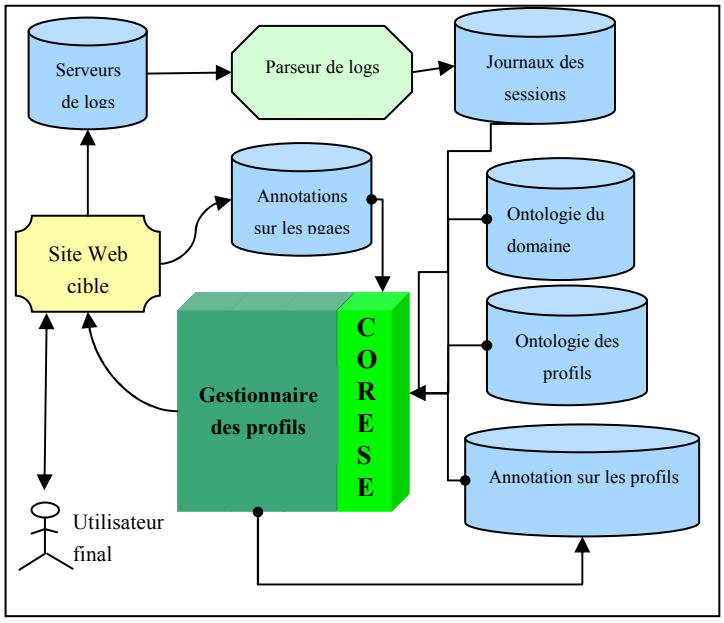

Fig.1 – Architecture de SUPROD

[2] http://www.w3.org/TR/rdf-sparql-query/

- Serveurs de logs : contiennent l'information générée par les serveurs Web décrivant les traces de navigation des utilisateurs dans les sites.
- Annotations sur les pages : annotations décrivant le contenu sémantique des pages web. La création de ces annotations est guidée par une ontologie.
- Les journaux de sessions : journaux décrivant le comportement des utilisateurs pendant une session.
- Ontologie du domaine : décrit le domaine d'intérêt du site Web.
- Ontologie des profils : décrit les profils professionnels du domaine en question.
- Annotations sur les profils : ces annotations sont générées automatiquement par le gestionnaire des profils pour décrire le profil de l'utilisateur tout en se basant sur son comportement sur le site (pages visitées + temps passé sur les pages).
- CORESE (COnceptual REsource Search Engine) : moteur sémantique utilisé dans le système SUPROD. Il permet (i) de charger des ontologies, des annotations et des règles d'inférences, (ii) de raisonner sur les annotations en se basant sur les ontologies et les règles, et (iii) d'interroger ces différentes ressources en utilisant le langage SPARQL.
- Le gestionnaire des profils : ce module recueille toutes les informations, calcule les niveaux d'intérêt des utilisateurs, détermine les similarités entre les profils et les classes suivant le processus que nous dériverons dans le reste de l'article. Il se base sur CORESE pour raisonner sur les annotations existantes et pour en générer de nouvelles.

Comme exemples utilisés dans nos expériences :

- Pour l'ontologie du domaine, nous avons choisi le réseau sémantique de UMLS (Humphreys & Lindberg, 1993) qui peut être considérée comme une ontologie biomédicale.
- Pour le serveur de logs, nous nous sommes basés sur les traces fournies par le serveur hébergeant le site Neli.
- Les annotations sur les pages de Neli ont été générées automatiquement en utilisant le système MeatAnnot (Khelif et al., 2007), qui, à partir de documents textuels génère une annotation basée sur UMLS.
- Pour l'ontologie des profils, puisque notre domaine d'intérêt est le domaine biomédical, nous nous sommes reposés sur la classification proposée par le «Canadian Human Resource and Skills Development Organism » pour les biologistes et sur la classification des médecins proposée dans (Dieng-Kuntz et al., 2004).
- Une session d'un utilisateur est définie comme étant un accès provenant de la même adresse IP tel que la durée du temps écoulé entre deux accès consécutifs est inférieure à un seuil prédéfini (pour nos expériences, nous avons choisi un seuil de 25 minutes).

2.2 Description détaillée de l'approche

Comme cela a été mentionné précédemment, nous utilisons les fichiers de logs des serveurs pour suivre la navigation de l'utilisateur sur le site. Partant des données extraites concernant l'utilisateur, le système génère un journal pour la session utilisateur contenant diverses informations telles que la longueur de la session, l'adresse IP, les URLs visitées, le nombre de visite de chaque page et le temps passé sur chaque document. La première version du profil d'usage étant ainsi construite, le système commence alors la construction du profil utilisateur.

A ce niveau, la question qui se pose est : Qu'est ce que le profil utilisateur ?

Une réponse « directe » est de dire qu'il s'agit de l'ensemble des objets qui intéressent l'utilisateur. Seulement, ces objets n'ont pas la même importance pour l'utilisateur, et en plus, même si nous arrivons à définir tous ces objets, le profil utilisateur peut évoluer que ce soit en ajoutant de nouveaux centres d'intérêts, en en supprimant certains ou encore en modifiant leur degré d'importance. C'est pour cela que nous pouvons dire que le profil utilisateur est flou, mais nous pouvons néanmoins affirmer qu'il tourne autour d'un modèle central qui le définit au mieux. C'est ce modèle que nous souhaitons définir.

Etant donné qu'il est presque impossible de définir des modèles personnalisés pour chaque utilisateur, nous tentons de le faire pour des groupes d'utilisateurs sachant que chaque utilisateur sera, par la suite, rattaché au groupe qui lui correspond le mieux. Pour ce faire, l'utilisation des techniques de filtrage collaboratif est nécessaire puisque nous avons besoin de méthodes de classification appropriées pour construire des profils agrégés pertinents. Mais avant cela, il est nécessaire d'avoir une représentation uniforme des profils utilisateurs et des profils modèles.

Notre point de départ est le profil d'usage contenant l'ensemble des URLs visitées durant la session de l'utilisateur. Ce profil nous permet de construire deux fonctions d'appartenance définissant les caractéristiques de l'utilisateur : la première concerne le profil utilisateur de niveau domaine. Elle prend en entrée les ressources du domaine (dans notre cas, les concepts de l'ontologie) qui ont été détectées dans la collection de pages extraites du profil d'exploration sur le web, et elle renvoie les degrés d'intérêt calculés pour les éléments du domaine correspondants. La seconde fonction concerne le profil utilisateur niveau exploration. Elle prend en entrée les URLs visitées par l'utilisateur et renvoie le degré d'importance calculé pour chacun de ces documents.

Désirant obtenir des solutions génériques, nous avons choisi d'utiliser des annotations RDF pour représenter ces fonctions. Les données RDF représentées ci-dessous montrent la définition d'un profil basé sur les annotations : la première annotation RDF décrit la fonction de niveau domaine et la deuxième décrit la fonction de niveau exploration.

Cet exemple représente une session identifiée par SESSION_0 ainsi que la fonction de profil, identifiée par SESSION_0_PFUNC, qui lui est associée. Cette fonction représente aussi bien le profil de niveau domaine grâce à la propriété 'defineResource' que le profil de niveau exploration grâce à la propriété 'defineDoc'. Dans cette session, l'utilisateur a exprimé son intérêt pour trois documents relatifs aux trois concepts de l'ontologie : 'Cardiology', 'Cardiac' et 'Heart'.

Définition du profil de niveau domaine :

```
<rdf:Description rdf:id="SESSION_0_PFUNC">
 <rdf:type rdf:resource="#PFUNC"/>
 <bmp:defineResource>
    <rdf:Description rdf:about="#SESS_0_PFUNC_RC0">
           <bmp:hasInput rdf:resource="&drc;Cardiology"/>
           <bmp:hasValue>26</bmp:hasValue>
    </rdf:Description>
  </bmp:defineResource>
  <bmp:defineResource>
    <rdf:Description rdf:about="#SESS_0_PFUNC_RC1">
           <bmp:hasInput rdf:resource="&drc;Heart"/>
           <bmp:hasValue>18</bmp:hasValue>
    </rdf:Description>
  </bmp:defineResource>
 <bmp:defineResource>
    <rdf:Description rdf:about="#SESS_0_PFUNC_RC2">
           <bmp:hasInput rdf:resource="&drc;Cardiac"/>
           <bmp:hasValue>11</bmp:hasValue>
    </rdf:Description>
  </bmp:defineResource>
</rdf:Description>

<rdf:Description rdf:id="#SESSION_0">
        <bmp:hasProfileFunction rdf:resource="#SESSION_0_PFUNC"/>
</rdf:Description>
```

Définition du profil de niveau exploration :

```
<rdf:Description rdf:about="#SESSION_0_PFUNC">
<bmp:defineDoc>
    <rdf:Description rdf:about="#SESS_0_PFUNC_DOC0">
           <bmp:hasInput rdf:resource="URL0 "/>
           <bmp:hasValue>13</bmp:hasValue>
    </rdf:Description>
  </bmp:defineDoc>
  <bmp:defineDoc>
    …….
  </bmp:defineDoc>
……
</rdf:Description>
</rdf:RDF>
```

Les poids des ressources de niveau exploration (pages web) sont calculés en fonction du temps passé sur chacun des documents et le nombre d'accès à ces documents. Les poids des ressources de niveau domaine sont calculés en fonction du

poids sémantique des termes reconnus comme étant des concepts ainsi que le temps passé sur le document. Ainsi, pour une ressource de domaine 'res' citée dans un ou plusieurs documents visités durant la session utilisateur, le degré d'intérêt pour la ressource 'res' est définie ainsi :

$$level(res) = \sum_{d \in D} \left(\frac{o_d(res)}{N_d} \times ts_d \right) \qquad (1)$$

level(res) est le degré d'intérêt, o_d(res) est le nombre d'occurrence de la ressource de domaine 'res' dans le document d, N_d est le nombre total de concepts détectés dans le document d, ts_d est le temps passé sur la page d (exprimé en secondes) et D est l'ensemble de tous les documents visités durant la session de l'utilisateur.

De plus, une vue sémantique de cet intérêt doit prendre en compte le fait que si l'utilisateur est intéressé par 'mycrobiology' avec un degré 'n', il/elle est aussi intéressé par 'biology' avec un degré 'm' qui sera inférieur à 'n'. Nous avons donc propagé le degré d'intérêt aux concepts ancêtres du concept d'origine 'res' à travers le lien de subsomption. Dans notre application, nous avons divisé le degré d'intérêt par des puissances de 2 proportionnelles à la distance hiérarchique moyenne entre un concept donné et le concept source d'intérêt. Ce calcul se fait par l'utilisation d'une règle dédiée du moteur Corese qui exploite la propriété RDFS 'subClassOf'. La figure 2 montre un exemple de la propagation des degrés d'intérêt, à partir des concepts d'intérêt originaux en bas de l'ontologie.

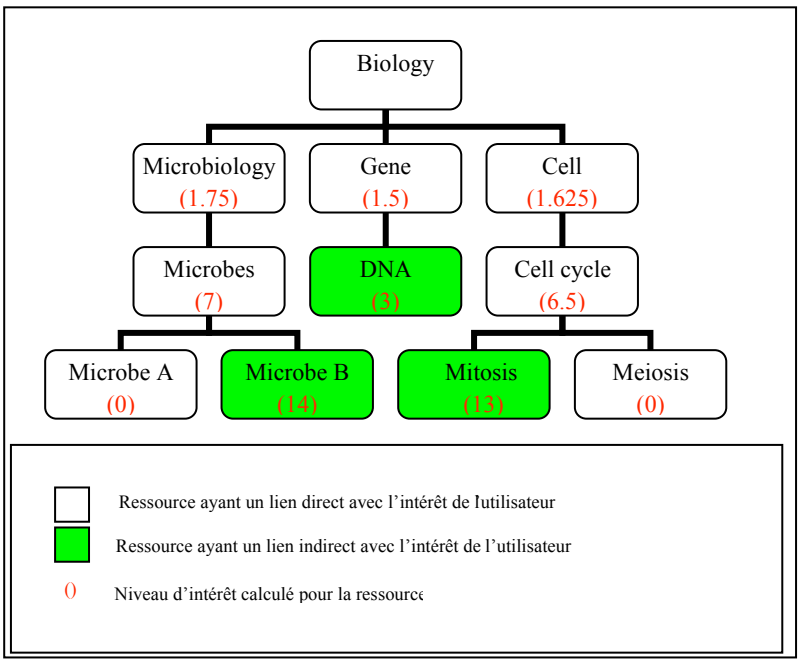

Fig.2 - Exemple de propagation des niveaux d'intérêt

Ainsi, le degré d'intérêt pour un concept donné 'c' ayant un ensemble de concepts fils 'Ch_c' est :

$$level(c) = \sum_{i \in Ch_c} \left(\frac{level(i)}{2^{dth(s)-dth(i)}} \right) \qquad (2)$$

dth(x) est la profondeur du concept fils dans l'ontologie.

Dans l'exemple de la figure 2, nous avons considéré l'ontologie comme étant un arbre (sans subsomption multiples). Dans les autres cas (plusieurs chemins entre les concepts), afin de s'assurer que chaque concept est utilisé une seule fois pour le calcul du degré d'intérêt, nous considérons la moyenne des degrés obtenus à partir des différents chemins possibles.

2.2.1 Le calcul de la similarité entre les profils

La méthode proposée permet d'extraire les principales caractéristiques concernant les centres d'intérêt de l'utilisateur et ainsi de définir un contexte d'usage beaucoup plus précis.

Une fois le profil de niveau domaine défini, le système étudie la similarité du profil courant avec ceux existants dans la base d'annotations de profils. La similarité entre deux profils dépend essentiellement de deux facteurs : (1) le nombre de leurs ressources communes et (2) la proximité de leurs degrés d'intérêts. Plus encore, la variation du degré d'intérêt pour une ressource considérée importante doit avoir plus d'influence sur les résultats de la comparaison. Il est important de noter que nous ne voulons pas nous baser sur la similarité entre les URLs visitées puisque les utilisateurs peuvent être intéressés par les mêmes sujets mais ne pas être intéressés par les mêmes pages. En effet, notre but est de classer les documents selon des profils établis par des comparaisons sémantiques, et ce afin d'obtenir un système de recommandations fondé uniquement sur des données sémantiques.

Nous définissons la similarité ainsi :

$$Sim(p_u, p_k) = \max \left(0, \frac{\sum_{r \in R_{p_u, p_k}} l_{p_u}(r) \times \left(1 - V_{p_u, p_k}(r)\right)}{\sum_{r \in R_{p_u, p_k}} l_{p_u}(r)} \right) \qquad (3)$$

$$V_{p_i, p_j}(r) = \frac{\left| l_{p_j}(r) - l_{p_i}(r) \right|}{l_{p_i}(r)} \qquad (4)$$

P_u : le profil observé de l'utilisateur durant la session
P_k : le profil modèle numéro k
$L_p(r)$: le degré d'intérêt pour la ressource 'r' dans le profil 'p'
R_p : l'ensemble des ressources définies dans le profil 'p'

$$R_{p_i,p_j} = R_{p_i} \cap R_{p_j} \qquad (5)$$

Ces mesures permettent au système de déterminer les premiers profils modèles qui correspondent à la meilleure valeur de similarité et qui excèdent un certain seuil (ce seuil sera optimisé dans la phase d'évaluation). Par ailleurs, une fois la session fermée, le profil construit sera utilisé pour améliorer les classes de profils.

La classification des profils est nécessaire et ce pour trois raisons : (1) le système ne peut pas garder indéfiniment chaque nouvelle session comme étant un nouveau modèle profil, (2) cette classification engendrera une classification des documents qui va être utilisée pour des fins de recommandation, et (3) le système a besoin de connaître les profils qui ont le plus de ressources en commun pour pouvoir les lier à une activité professionnelle (docteur, biologiste, ingénieur, etc.) ou leur recommander des documents pertinents.

Au début, les profils sont classés selon le pourcentage de ressources qu'ils ont en commun. Nous avons ainsi défini cinq intervalles de 50-60% de ressources en commun à 90-100% dans lesquels les groupes de profils (définis par un ensemble commun de ressources) sont triés par leurs scores de similarités.

Une fois que le profil de l'utilisateur (ayant un ensemble R_u de ressources) est associé au profil le plus similaire (ayant un ensemble R_p de ressources) avec une similarité S, le système procède par cette façon :

$$pcr = MIN\left(\frac{\left|R_u \cap R_p\right|}{\left|R_u\right|} \quad , \quad \frac{\left|R_u \cap R_p\right|}{\left|R_p\right|} \right) \qquad (6)$$

Où pcr est le pourcentage de ressources communes.

Si (pcr<0.5)

- Pas de classification de profils : le profil de l'utilisateur est simplement rajouté à la base de profils modèles (le nombre de profils de sessions stockés est limité et une procédure de nettoyage est lancée pour supprimer les modèles rarement utilisés).

Sinon :

- fusionner les deux profils en prenant l'intersection des ensembles de ressources les définissant (nous affectons le niveau d'intérêt moyen aux ressources et aux documents).

- affecter le profil obtenu à sa classe correcte, en tenant en compte le pcr déjà calculé.
- comparer le profil obtenu à ceux qui existent. Prendre en compte les résultats de comparaison, fusionner et réaffecter les modèles aux différents intervalles déjà définis.

2.2.2 Pertinence des annotations sur l'intérêt de l'utilisateur

Afin de déterminer la pertinence des annotations sur les centres d'intérêt de l'utilisateur, nous utilisons les deux principes suivants :

(i) Le temps passé par l'utilisateur sur chaque page est considéré ; un temps d'exclusion minimum (resp. maximum) est choisi pour considérer uniquement les pages sur lesquelles l'utilisateur passe « un temps régulier ». Nous avons choisi 5 secondes comme temps minimum et un temps maximum proportionnel à la taille du texte scientifique de la page (1 sec. par dix caractères imprimables).

(ii) Pour la dernière page visitée, nous utilisons une approximation calculée comme suit :

$$T_s(LP) = Sim\left(LP_p, U_p\right) \times T_{\max}(LP) \quad (7)$$

Avec :

T_s: temps passé sur la page.
LP: dernière page visitée durant la session utilisateur.
Sim: La fonction de similarité définie pour les modèles de profil.
LP_p: Profile de niveau domaine de la dernière page.
U_p: Le profil utilisateur observé durant la session.
T_{max} (p): Temps maximum permis pour la page p.

3 Evaluation de l'approche

Afin de valider notre approche, nous avons utilisé des logs réels du site Neli (les traces d'un mois de visites : ~ 40000 lignes de logs après nettoyage) et nous avons défini des modèles de profils sur lesquels nous avons testé le système SUPROD. Cette phase nous a permis de valider (i) notre représentation des profils, (ii) la cohérence des annotations générées, et (iii) la mise en échelle de l'approche sur de vraies données.

Une fois notre base de profils modèles construite, nous avons testé notre système de recommandation auprès de cinq utilisateurs, en leur fixant le scénario suivant :

- Choisir un centre d'intérêt dès le départ (une maladie, un gene, etc.) ;
- Naviguer sur le site Neli pour se documenter sur ce centre d'intérêt ;
- Après cinq minutes de navigation, regarder les recommandations proposées par SUPROD et dire si elles sont pertinentes ou non (en

donnant une note entre 0 et 5) par rapport au centre d'intérêt fixé à l'avance.

La table 1 récapitule les résultats de cette validation qualitative.

Table 1. Evaluation des recommandations proposées par SUPROD

Note d'évaluation	0	1	2	3	4	5
Nombre de pages recommandées par note	3	2	3	6	4	4
Proportion parmi les pages proposées	13.64%	9,09%	13.64%	27,27%	18,18%	18,18%
Moyenne	**36,37%**			**63,63%**		

Nous remarquons ainsi que le système fourni de bons résultats deux fois sur trois (les notes entre 3 et 5). Les mauvaises recommandations peuvent être expliquées par (i) le choix de nos variables (seuil de similarités entre profils, temps d'exclusion,…) qui peut s'avérer non optimal, (ii) la taille de la base de profils modèles (une quinzaine) construite à partir des logs, et (iii) la qualité et la mise à jour des annotations des pages. En effet, en regardant les pages ayant eu de mauvaises notes, nous avons remarqué qu'elles sont, soit trop générales (i et iii), soit focalisées sur un thème proche du thème initial mais ne répondant pas au besoin de l'utilisateur (ii), soit mal annotées ou carrément vides (iii).

4 Discussion et travaux antérieurs

Dans cet article, nous avons présenté une approche générique et indépendante du domaine pour la détection des profils des utilisateurs sur le Web. Les profils générés peuvent être utilisés pour (i) la personnalisation de la navigation des utilisateurs, (ii) la recommandation de documents, (iii) la découverte des activités professionnelle des utilisateurs, etc. Cette approche est implémenté dans le système SUPROD et a été testée dans le domaine biomédical à travers le site Neli.

Comparant aux autres méthodes de détection de profils (Ahu et al., 2004) (Honghua & Bamshad, 2002)(Cingil et al., 2000), notre approche est fondée seulement sur les ontologies, les annotations sémantique et un moteur d'inférence. Ces composants garantissent la généricité de SUPROD puisque il pourra être appliqué sur n'importe quel site.

Comme perspectives à ce travail, plusieurs améliorations peuvent être apportées. Pour l'instant, nous nous utilisons que la relation de subsomption pour propager le niveau d'intérêt à travers les concepts de l'ontologie pour une session de l'utilisateur, la prise en compte des relations sémantiques du domaine présentes dans l'ontologie ne peut qu'affiner et améliorer la détection du profil.

Par ailleurs, dans la phase de classification, nous avons proposé de prendre en compte les ressources communes entre deux profils ; une étude peut être menée pour tester le fait de prendre l'inclusion, l'union et l'intersection entre les ressources et

ainsi faire un choix pertinent. L'intervention des experts est nécessaire pour valider les profils modèles, faire le lien avec les activités professionnelles et proposer des nouveaux modèles. Enfin un évaluation à plus grande échelle nous permettra s'optimiser tous nos algorithmes.

Remerciements

Ce travail est financé par le projet européen Sealife (IST-2006-027269). Nous remercions Patty Kostkova et son équipe du City eHealth Research Center (City University) qui s'occupe de la gestion du site NELI.

Références

KHELIF K., DIENG-KUNTZ R. & BARBRY P. (2007). An ontology-based approach to support text mining and information retrieval in the biological domain, Journal of Universal Computer Science (JUCS), Vol. 13, No. 12, pp. 1881-1907.

CORBY O., DIENG-KUNTZ R. & FARON-ZUCKER C. (2004). Querying the Semantic Web with the CORESE engine. In Proceedings of ECAI'2004, Valencia, Spain, IOS Press, p.705-709.

SCHROEDER M., BURGER A., KOSTKOVA P., STEVENS R., HABERMANN B. & DIENG-KUNTZ R. (2006) Sealife: A Semantic Grid Browser for the Life Sciences Applied to the Study of Infectious Diseases. Proceedings of HealthGrid 2006, 120:167—78.

DIENG-KUNTZ R., MINIER D., RUZICKA M., CORBY F., CORBY O., ALAMARGUY L. (2006). Building and Using a Medical Ontology for Knowledge Management and Cooperative Work in a Health Care Network. Computers in Biology and Medicine, Volume 36, Issues 7-8, 2006

HUMPHREYS B. & LINDBERG D. (1993). The UMLS project: making the conceptual connection between users and the information they need. Medical Library Association 81(2): 170.

AHU S., BAMSHAD M. & RIBIN B. (2004). Inferring User's Information Context: Integrating User Profiles and Concept Hierarchies. Proceedings of the 2004 Meeting of the international Federation of Classification Societies, Chicago.

HONGHUA D. & BAMSHAD M. (2002). Using Ontologies to discover Domain-Level Web Usage Profile. The second Semantic Web Mining Workshop at ECML/PKDD.

COOLEY R., BAMSHAD M. & STRIVASTAVA J. (1999) Data preparation for mining World Wide Web browsing patterns. Journal of knowledge and information systems.

BAMSHAD M., COOLEY R. & SRIVASTAVA J. (2000). Automatic personalization based on web usage mining. Communications of the ACM Vol 43. No. 8.

NASRAOUI O., FRIGUI H., JOSHI A. & KRISHNAPURAM R. (2000). Mining Web Access Logs Using Relational Competitive Fuzzy Clustering. Intemational Journal on AI Tools.

SCHECTER S., KRISHNAN M., & SMITH M.D. (1998). Using path profiles to predict HTTP requests. In Proceedings of the Seventh International WWW Conference, Brisbane, Australia.

SPILIOPOULOU M. & FAULSTICH L.C. (1999). WUM: A Web Utilization Miner. In proceedings of EDBT Workshop WebDB98, Valencia, Spain.

BUCHNER A. & MULVENNA M.D. (1999). Discovering internet marketing intelligence through online analytical web usage mining. SIGMOD Record 4, 27.

CINGIL l., DOGAC A. & AZGIN A. (2000). A broader approach to personalization. Communications of the ACM Vol.43, NO. 8.

SHAHABI C., ZARKESH A.M., ADIBI J. & SHAH V. (1997). Knowledge discovery from user Web-page navigation. In Proc of workshop in research issues in Data Engineering, England.