

Intégrer le comportement de l'utilisateur dans les documents Web

Integrating users behaviour in Web documents

N. Singer¹

S. Trouilhet²

¹ Centre Universitaire Jean-François Champollion,
Avenue Georges Pompidou 81100 Castres
nicolas.singer@univ-jfc.fr

² IRIT – site de Castres
Avenue Georges Pompidou 81100 Castres
trouilhe@irit.fr

Résumé

Même avec les nombreux outils proposés, la recherche de documents sur le Web reste un acte individualiste. Nous proposons un système multi-agent, Mawa, qui mutualise les expériences des utilisateurs en utilisant leurs parcours pour reconstituer les documents Web.

Cet article présente la modélisation choisie pour représenter les parcours sur le Web. La représentation de ces parcours évolue dynamiquement avec un renforcement ou une atténuation des forces d'interconnexion des pages. Ce sont les interactions des agents du système qui mènent à la constitution de documents à hyperliens virtuels.

Le système utilise l'activité des utilisateurs et n'implique aucune modification du comportement de ce dernier.

Mots Clef

Documents Web, sessions de navigation, hyperliens virtuels, système multi-agent.

Abstract

Research of documents on the Web remains an individualistic act. We propose a system that collects individual experiences to help future users. Mawa system agents are spread through the Web and interact to build virtual Web documents using older navigations.

This article presents our modelisation to represent navigation on the Web. The representation of these navigations evolves dynamically with reinforcement or diminution of the relation forces between pages. We insist on the description of the interactions between agents and on how new documents are created in introducing virtual hyperlinks.

Our system analyzes the user activity and implies no change of its behaviour.

Keywords

Web documents, navigation sessions, virtual hyperlinks, multi-agent system.

1 Introduction

Le Web permet à tout un chacun de publier de l'information sur n'importe quel thème, et de connecter cette information à la base de connaissance globale que constitue la toile. Il n'est pas rare que la même information soit traitée différemment par plusieurs auteurs. C'est la mise en commun de ces travaux qui rend virtuellement disponible une connaissance de grande valeur sur le sujet, très coûteuse à trouver en dehors du Web.

La grande difficulté du chercheur d'information est alors de passer d'un potentiel supposé de connaissance disponible (qui est la raison de sa recherche) à une connaissance effective qui passe par la consultation de nombreuses pages reliées par des liens.

Le Web propose des outils pour orienter l'utilisateur dans ses parcours mais cette métaconnaissance du Web ne garantit pas à l'internaute l'exhaustivité de l'information obtenue [6], [8]. Rien ne l'empêche de penser que quelqu'un d'autre, qui a peut-être passé plus de temps à chercher ou qui a eu plus de chance, a trouvé d'autres pages de meilleure qualité traitant du même sujet ou y apportant un complément important.

Nous sommes persuadés qu'une des façons d'atténuer l'insatisfaction de l'internaute est de le faire automatiquement bénéficier des recherches des autres, en lui présentant des documents réorganisés à partir des navigations de ceux qui l'ont précédé.

Une étude du comportement des utilisateurs du Web montre en effet qu'ils ont tendance à se contacter régulièrement (dans des forums de discussion par exemple ou par des questions posées dans la presse écrite) afin d'obtenir des renseignements sur la localisation de telle ou telle page traitant d'une information donnée.

Notre méthode se base sur le fait que des personnes ayant effectué un début de parcours commun entre des pages

partagent potentiellement un même centre d'intérêt ; l'expérience des uns peut profiter aux autres.

Nous proposons un système permettant d'évaluer et de quantifier les navigations des utilisateurs du Web. Les données ainsi obtenues sont utilisées pour mettre en évidence des ensembles dynamiques de pages.

Ces ensembles constituent un document plus grand qui correspond à toutes les informations trouvées par une collectivité ayant eu temporairement un même centre d'intérêt.

2 Le système de reconstitution de documents Mawa

Notre travail rejoint les recherches faites sur les documents virtuels. Un document virtuel se qualifie comme un document qui ne possède pas d'état persistant et dont les instances sont calculées dynamiquement en fonction de la demande [2], [7].

Dans des environnements comme le Web, la composition dynamique permet d'adapter le document pour une meilleure compréhension d'une classe d'utilisateurs. Dans [4] par exemple, les auteurs personnalisent les documents Web, en proposant plusieurs stratégies de lecture aux utilisateurs. Le processus s'effectue en trois étapes : **sélectionner** puis **organiser** parmi un ensemble de fragments d'informations, ceux qui semblent pertinents par rapport à la spécification de l'auteur ou par rapport à l'objectif du lecteur. Ces deux étapes sont complétées par l'étape d'**assemblage** qui crée une stratégie de lecture.

Cependant, cette adaptation demande de représenter le savoir-faire des auteurs ainsi que des stratégies de recomposition liées à la nature du document et de l'application envisagée.

C'est également le cas pour les travaux de [3] qui se servent des réseaux sémantiques pour faciliter la navigation et la recherche d'informations. Un réseau sémantique basé sur des liens entre concepts se superpose au réseau hypertexte. Des liens verticaux sont alors créés pour relier les deux niveaux.

Nos choix sont différents car notre stratégie ne repose sur aucune métaconnaissance ou ontologie ; ceci afin de pouvoir s'appliquer à n'importe quel document Web. Pour cela, nous nous basons sur les navigations des utilisateurs. Les parcours sont mis en commun pour organiser les pages des documents.

Les ressources sont totalement distribuées sur Internet et manipulées par un grand nombre d'utilisateurs. Pour les collecter, nous avons conçu un système multi-agent Mawa (assistant Web multi-agent), composé d'agents répartis sur le réseau aussi bien sur les postes client que serveur [1].

Tout agent possède des connaissances (représentation de l'environnement, accointances), des compétences et des facultés de communication.

Les agents se représentent l'environnement comme un graphe orienté dans lequel les noeuds sont les pages Web. Ils manipulent les connexions de pages sous la forme d'un quintuplet

$$\langle pi, pf, s, t, f \rangle$$

- *pi* et *pf* sont les **références des pages** reliées (*pi* pointe *pf*).

- *t* est le **type du lien**. Ce type peut-être *interne* (référence à une partie du même nœud), *organisationnel* (chemin d'accès relatif qui définit la composition du document), *d'indexation* (chemin d'accès relatif avec un point de branchement) ou *référentiel* (chemin d'accès absolu). Nous classons ces différents types dans la catégorie des types *réels*. En revanche, un lien qui n'existe pas mais qui est ajouté par le système de façon à recomposer un document est de type *virtuel*.

- *f* est la force du lien. Initialement il s'agit du nombre de fois où le lien a été emprunté par les utilisateurs, pondéré par les actions de l'utilisateur sur *pf*. Sa valeur évolue au fur et à mesure des interactions entre agents, soit positivement, soit négativement (sa valeur peut-être donc devenir négative). Ce nombre est interprété comme la **force d'interconnexion** des deux pages.

- *s* représente la **session de navigation**, c'est-à-dire une unité de navigation cohérente par rapport à un centre d'intérêt. Ce centre d'intérêt est resté généralement inchangé pendant la navigation. Quand il varie, cela est dû à la nature des pages consultées qui peuvent détourner l'utilisateur de sa recherche initiale. La caractérisation de cette session est un des éléments qui permet de faire émerger un comportement collectif.

Ces quintuplets représentent des parcours de longueur 1 entre deux pages. Un parcours de longueur *n* est un ensemble de *n* quintuplets consécutifs.

Les communications entre agents sont des envois de quintuplets $\langle pi, pf, s, t, f \rangle$.

Initialement, les agents ne se connaissent pas : un agent nouvellement ajouté n'a aucune accointance ; il apprend l'existence d'autres agents et se fait connaître au fur et à mesure des navigations.

Il existe deux types d'agents : des **agents assistants**, installés sur les postes clients et des **agents connecteurs**, installés sur les serveurs.

L'assistant a le comportement suivant : il observe l'activité de l'utilisateur. Pour un nœud consulté, il détermine les mots-clés qui le caractérise, les temps de consultation ainsi que le chemin pour arriver et repartir du nœud.

L'assistant est décomposé en quatre parties : l'analyste, l'écouteur, le cartographe et l'interface (figure 1).

L'analyseur prend en charge l'analyse syntaxique du nœud courant (marqueurs, hyperliens). Par exemple, il repère les liens organisationnels pour recomposer le document.

L'écouteur observe l'activité sur le nœud courant : clics de souris sur les hyperliens, temps de consultation des pages, demandes d'impression.

Le cartographe met à jour les séquences de quintuplets correspondant à l'activité de l'utilisateur.

L'interface, quant à elle, présente les séquences de quintuplets.

L'agent connecteur prend en compte la session de l'assistant afin de la fusionner avec les sessions antérieures.

La figure ci-dessous illustre les différents éléments de Mawa, leur emplacement et leurs échanges.

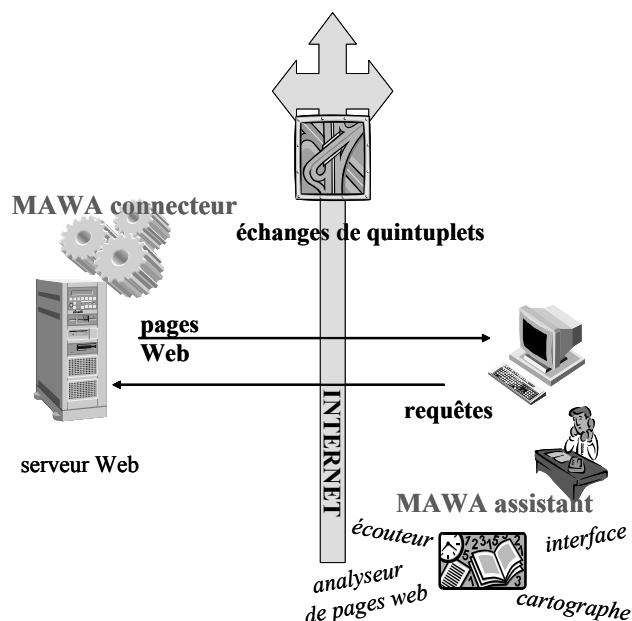


figure 1 – Architecture de Mawa

La section qui suit décrit les deux types de sessions manipulées par le système : les sessions utilisateurs et les sessions collectives, ces dernières étant obtenues par appariement des premières entre elles.

C'est la section 4 qui explique comment sont construites les sessions utilisateurs.

3 Caractérisation des sessions

La session est l'élément central de notre étude. C'est en regroupant les parcours en sessions que se composent les documents Web.

3.1 Session utilisateur

Du point de vue de l'agent assistant, une session peut-être caractérisée d'au moins trois façons : soit du point de vue temporel, soit du point de vue sémantique, soit du point de vue comportemental.

Dans le premier cas, une session est une unité temporelle de navigation, c'est-à-dire un ensemble de parcours situé dans un intervalle de temps repérable.

L'intervalle de temps est dit repérable s'il est isolé des autres, c'est-à-dire s'il se produit une rupture de navigation avant et après cet intervalle. Ainsi, si x est l'intervalle de temps moyen entre deux chargements de pages alors toute demande de pages séparée de la précédente par un temps bien supérieur constitue une rupture.

Dans le cas sémantique, une session est un ensemble de parcours dont toutes les pages sont liées par au moins un même thème (elles traitent de la même information). La caractérisation de ce type de session nécessite l'analyse du contenu des pages et de leurs balises, afin de faire émerger les points communs ou les ruptures de sens entre pages. Elle nécessite aussi la modélisation et l'acquisition de connaissances liées au domaine. Nous avons cherché à rendre cet aspect secondaire car nous ne voulons pas que le système dépende d'un critère qui ne peut être automatisé.

Dans le cas comportemental, ce sont les actions de l'utilisateur qui caractérisent une session. Celui-ci pourrait expliciter à son agent, de façon interactive, le début ou la fin de celle-ci (en cliquant dans une barre d'outils par exemple). Pour arriver au même résultat, nous préférons ne pas solliciter l'utilisateur et analyser de façon automatique ses actions (comme une saisie manuelle d'Url).

Idéalement, les différents types de caractérisation devraient être utilisés de concert de façon à ce que le repérage du début et de la fin d'une session soit le plus précis possible. En effet c'est le regroupement de parcours en session qui sert de base de travail à nos agents.

On parlera de session partielle pour désigner une session dont on connaît le début (le premier parcours) mais pas la fin. Ce type de session est typiquement une session en cours de construction.

3.2 Session collective

Du point de vue de l'agent connecteur, une session est un objet construit (que nous appelons *session consolidée*) à partir des sessions des agents assistants.

Cette construction se fait par appariement des sessions qui se ressemblent de façon à obtenir des sessions étendues (ou au contraire restreintes).

L'agent connecteur stocke aussi le nombre de sessions utilisateurs ayant participé à la construction de la session. Ce nombre, appelé *force d'une session*, permet de normaliser la force des liens d'une session (les valeurs des forces d'interconnexion sont ainsi ramenées entre 0 et 1).

Notons donc que l'agent assistant caractérise une session selon un mode individuel, tandis que l'agent connecteur tente d'intégrer ces expériences individuelles dans la mémoire collective.

La session de l'agent connecteur est le document virtuel correspondant à un comportement collectif de consultation.

3.3 Sessions apparentées

Un des rôles de l'agent connecteur est de déterminer si les objectifs d'un utilisateur sont les mêmes que ceux d'un utilisateur précédent. Pour cela, il part du principe que deux sessions de navigation qui se ressemblent correspondent à deux utilisateurs qui partagent le même but.

Dans quel cas un connecteur estime-t-il que deux sessions se ressemblent ?

Règle d'appariement :

Soit deux sessions de navigation, s et s' . Soit S et S' l'ensemble des pages consultées respectivement de ces deux sessions avec $|S| \leq |S'|$.

On dira que s et s' sont apparentées si et seulement si

$$|S \cap S'| \geq |S - S \cap S'|$$

Ainsi, une session portant sur un plus petit nombre de pages qu'une autre session, lui est apparentée, si elle a plus de pages communes avec elle que de pages différentes (figure 2).

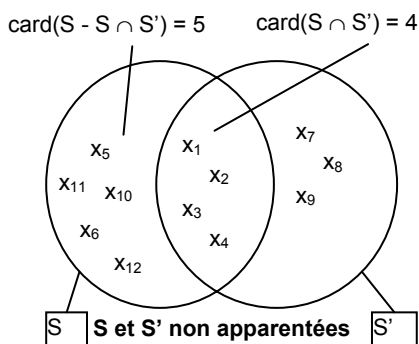
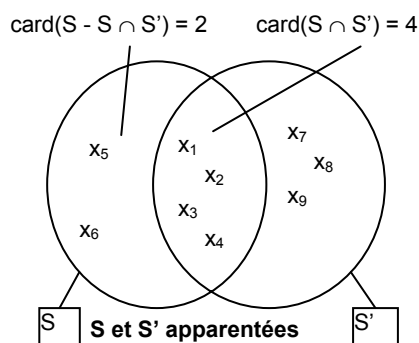


figure 2 – Comparaison de sessions

Il peut y avoir **fusion de deux sessions apparentées** : dès lors que l'agent connecteur juge qu'une session utilisateur

est apparentée avec une (ou plusieurs) des sessions consolidées, il fusionne la première avec la (ou les) deuxième(s), et ne conserve que le (ou les) résultat(s).

Cette fusion se fait selon la *règle de consolidation* suivante :

Soit s et s' deux sessions. La session s'' , fusion de s et de s' , est construite en trois étapes :

1) Pour tout quintuplet de la session s , donc de la forme $\langle x, y, s, t, f \rangle$:

a) s'il existe un quintuplet de la forme $\langle x, y, s', t', f' \rangle$ et si $f+f' > 0$, alors

$\langle x, y, s'', t'', f+f' \rangle$ est créé

avec si $t = \text{virtuel}$ et $t' = \text{virtuel}$

alors $t'' = \text{virtuel}$

sinon $t'' = t$

b) s'il n'existe pas de quintuplet $\langle x, y, s', t, f' \rangle$ et si $f > 0$ alors

$\langle x, y, s'', t, f \rangle$ est créé.

1) Pour tout quintuplet de la session s' , donc de la forme $\langle x, y, s', t, f' \rangle$:

s'il n'existe pas de quintuplet $\langle x, y, s, t, f' \rangle$ et si $f' > 0$

alors $\langle x, y, s'', t, f' \rangle$ est créé.

2) La force de s'' est égale à la somme de la force de s et de s' .

Ainsi, les parcours de longueur un, non communs aux deux sessions sont intégrés dans le résultat de la fusion, dans la mesure où la force de leur lien est strictement positive (règle 1b et 2).

Les parcours communs de longueur un, sont intégrés dans le résultat de la fusion si la somme des forces des liens est strictement positive. La nouvelle force du lien est cette somme (règle 1a).

Le type du lien n'est « virtuel » que si les parcours étaient virtuels dans les deux sessions, sinon il est de type « réel ». Autrement dit un parcours perd sa virtualité dès qu'un utilisateur l'a suivi.

Notons enfin que cette règle permet aussi la fusion avec une session vide (une session vide est une session ne contenant aucun parcours).

Ce cas particulier est utilisé lorsqu'une session utilisateur ne peut s'apparier à aucune session consolidée ; elle devient alors une session consolidée en fusionnant avec la session vide (ce mécanisme est explicité dans la section fonctionnement du système).

3.4 Ajout de liens virtuels

Lors de la phase de consolidation, l'agent connecteur peut décider d'ajouter dans la session résultat, des nouveaux parcours. Ces parcours sont des raccourcis qui permettent de relier directement deux pages d'une même session.

La création d'un lien virtuel obéit à la *règle d'ajout de lien virtuel* suivante :

Soit une session consolidée s . Pour tout parcours de longueur $n > 1$ caractérisé par n quintuplets consécutifs de la forme :

$$\langle x_1, x_2, s, t_1, f_1 \rangle \langle x_2, x_3, s, t_2, f_2 \rangle \dots \langle x_n, x_{n+1}, s, t_n, f_n \rangle$$

si $f_n > \max(f_1, f_2, \dots, f_{n-1})$

alors le quintuplet $\langle x_1, x_{n+1}, s, \text{virtuel}, f_n \rangle$ est créé.

Cela signifie que si deux pages sont reliées par un parcours de longueur supérieure à un où le dernier lien est le plus fort, alors il est possible d'éviter les pages intermédiaires pour passer de l'une à l'autre. Ce raccourci permet, par exemple, d'éviter de repasser par une page correspondant à la réponse d'un moteur de recherche.

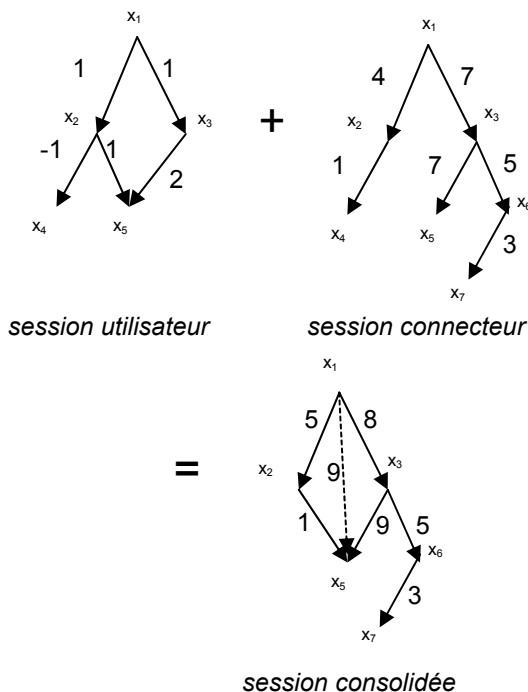


figure 3 – Fusion de deux sessions

La figure 3 donne un exemple de fusion entre une session utilisateur et une session consolidée. Les deux sessions peuvent fusionner car elles ont cinq documents en commun contre deux qui ne le sont pas.

Dans le résultat de la fusion, le lien de x_2 à x_4 disparaît car sa force devient nulle. Un lien virtuel est créé de x_1 à x_5 car la force du lien (x_3, x_5) devient supérieure à celle du lien (x_1, x_3) .

Appréciation de la force des liens

La valeur de la force d'un lien s'apprécie en la normalisant, c'est-à-dire en la ramenant à la force de sa

session. Ainsi un lien de force f dans une session de force k a une valeur normalisée égale à f/k .

Cette valeur est pour la majorité des liens comprise entre zéro et un (les liens de force négative sont éliminés de la session).

Plus la valeur est proche de un, plus la force du lien est importante. Inversement, si le lien a une force normalisée proche de zéro, c'est que peu des personnes ayant réalisé la session l'ont suivi ou que ce lien est brisé (la page cible du lien n'existe pas) et est en voie d'élimination.

Quelques liens peuvent avoir une force normalisée supérieure à un. Ce sont des liens régulièrement suivis, pour lesquels un groupe d'utilisateurs a montré un intérêt particulier (décelé par son activité sur les pages). En effet, ces liens sont le résultat de la consolidation de sessions utilisateurs où ils avaient une force strictement supérieure à un (voir plus bas la construction d'une session utilisateur).

Enfin la normalisation de la force des liens permet de comparer une session à une autre.

A titre d'exemple si la force de la session connecteur de la figure 3 est de 7, voici la représentation normalisée de la session consolidée.

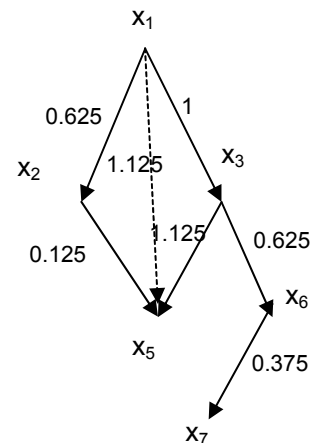


figure 4 – Session consolidée normalisée

4 Fonctionnement du système

Nous allons décrire dans cette section le protocole de mise au point des sessions consolidées, et par conséquent des documents virtuels.

Construction d'une session utilisateur

Au fur et à mesure des consultations de documents, l'agent assistant construit une session de navigation, c'est-à-dire un ensemble de parcours de longueur un (représentés chacun par un quintuplet). La force de cette session est égale à 1.

Sa construction s'effectue comme suit :

Soit s la session en cours. Pour tout lien suivi, soit x_1 la page d'origine et x_2 la page de destination.

La force f' du lien suivi est déterminée comme suit :

- si le lien est brisé, alors $f' = -1$;
- si le lien a pu être suivi (x_2 existe et est accessible) alors $f' = 1$;
- optionnellement, si le lien a pu être suivi et que l'utilisateur effectue une action montrant un intérêt particulier lors de la consultation de x_2 , alors $f' = \text{bonus de l'action}$.

S'il n'existe pas de quintuplet de la forme $\langle x_1, x_2, s, t, f \rangle$ alors $\langle x_1, x_2, s, \text{réel}, f' \rangle$ est créé, sinon le quintuplet devient $\langle x_1, x_2, s, t, f+f' \rangle$.

Notons qu'un parcours n'est ajouté que lorsqu'un lien hypertexte a été effectivement suivi. L'utilisation de l'historique de navigation (le retour arrière) et la saisie manuelle d'un URL par exemple, font certes passer d'une page à une autre, mais ne sont pas considérés par les agents comme des parcours.

Parmi les actions observables automatiquement et qui peuvent renforcer ou abaisser la force d'un lien, nous avons retenu :

- l'impression de la page (bonus)
- un temps de consultation plus important que pour les autres pages de la session (bonus)
- un temps de consultation très faible suivi d'un retour arrière (malus)

Nous sommes conscients qu'il ne s'agit que d'un ensemble minimal de ce qui peut être utilisé pour caractériser la force des liens, mais il ne nécessite aucune intervention de l'utilisateur.

Pourront se greffer ultérieurement des modules d'analyse établie en collaboration avec des ergonomes qui permettront de faire remonter une information plus précise.

L'agent assistant pourrait aussi décider d'agir selon un mode interactif. Il questionnerait alors l'utilisateur à certains moments ou lui laisserait la possibilité de noter une page dans un sens positif ou négatif.

Le fonctionnement du système est de toute façon préservé quel que soit le potentiel d'analyse des agents assistants. Peuvent en effet cohabiter des agents rudimentaires, qui ne font que compter le nombre de fois où un lien est suivi, et des agents plus perfectionnés, qui étudient de façon interactive ou automatique le comportement de l'utilisateur sur une page.

Echanges de sessions partielles entre assistant et connecteur

Soit s la session en cours de construction. Pour chaque parcours $\langle x, y, s, t, f \rangle$ ajouté à la session, l'agent assistant

contacte l'agent connecteur chargé de la page y et lui communique l'ensemble des quintuplets de la session s . Cet ensemble constitue une session partielle.

L'agent connecteur, sur réception de cette session, extrait toutes les sessions consolidées avec lesquelles elle est apparentée en appliquant la *règle d'appariement*. Il porte le résultat de cette extraction à la connaissance de l'agent assistant.

Utilisation de la session consolidée

La présentation de la session consolidée et le traitement de la page à afficher (la page y) peuvent être indifféremment effectués par l'agent assistant ou connecteur.

Par exemple, l'agent connecteur peut modifier la page en ajoutant des liens :

- soit des liens virtuels qui partent de la page y dans une des sessions apparentées à s ,
- soit des liens réels menant à y dans une des sessions apparentées à s . L'utilisateur peut ainsi *remonter* sur des pages dont il n'aurait pas eu connaissance.

La force des liens peut être mise en valeur en adoptant une présentation graphique appropriée : les liens forts peuvent être rendus avec une police ou une graisse plus importante.

L'agent assistant peut, de son côté, ouvrir des fenêtres supplémentaires pour présenter l'ensemble des sessions apparentées à s sous la forme de cartes.

L'ensemble des pages d'un document virtuel apparaît alors (un peu comme dans la rubrique *plan du site* de nombreux sites Web) et un simple clic permet de se rendre à n'importe quelle page.

Consolidation des sessions

Quand l'agent assistant a terminé sa session, il envoie un message à tous les connecteurs qui sont devenus ses accointances pendant la construction de s . s quitte alors son statut de session partielle ; elle est transmise en tant que session terminale et peut être utilisée pour consolider les sessions connecteurs.

Les agents connecteurs utilisent alors leur ensemble de sessions apparentées avec s pour les consolider.

Soit S' l'ensemble des sessions apparentées à s .

- si S' est vide alors une nouvelle session consolidée est créée par fusion de s avec la session vide en appliquant la *règle de consolidation*.
- sinon s fusionne avec chaque élément de S' par application de la *règle de consolidation*. Le résultat est un ensemble de sessions S'' qui remplace S' . S'' est à son tour consolidé par

application de la *règle de consolidation en chaîne*.

La *règle de consolidation en chaîne* est nécessaire car après le calcul de S'' , cet ensemble peut contenir deux sessions apparentées.

En effet la consolidation de deux sessions, en les modifiant, peut avoir pour effet que la règle d'appariement les déclare apparentées.

Règle de consolidation en chaîne :

Soit un ensemble de sessions S . Tant qu'il existe dans S deux sessions apparentées s et s' :

- 1) s'' est le résultat de la fusion de s et de s' ,
- 2) s et s' sont retirées de S ,
- 3) s'' est ajoutée à S .

Echanges entre connecteurs

L'ensemble de quintuplets qui caractérise une session, n'est pas forcément mémorisé par un seul connecteur.

Dans le cas général, il est au contraire réparti sur plusieurs agents connecteurs oeuvrant sur des serveurs Web différents.

Un connecteur mémorise la partie de session qui *traverse* son serveur : pour chaque session s , il stocke l'ensemble des quintuplets de la forme $\langle x, y, s, t, f \rangle$ où *au moins* x ou y sont des pages hébergées sur son serveur.

On dira que le connecteur fait autorité sur cette partie de la session.

Par conséquent, afin d'avoir accès à l'intégralité d'une session, un connecteur doit contacter ses accointances, et leur demander les parties manquantes. Les accointances possédant les parties manquantes sont connues dans la mesure où une session est un ensemble de parcours liés les uns aux autres.

Une accointance de premier niveau se définit comme suit :

soit une partie s_c de session s sur laquelle un connecteur c fait autorité. Les accointances de premier niveau de c pour s sont l'ensemble des connecteurs faisant autorité sur les pages de s_c non hébergées sur le serveur de c .

Ainsi, si $\langle x, y, s, t, f \rangle$ est un quintuplet de s_c et si x (ou y) n'est pas hébergé par le serveur Web de c alors le connecteur faisant autorité sur x (ou y) est une accointance de premier niveau de c pour s .

La méthode utilisée par un connecteur pour obtenir l'intégralité d'une session est alors la suivante :

soit une session s dont on veut obtenir tous les quintuplets. Soit S l'ensemble de ces quintuplets connus par le connecteur. Soit A l'ensemble des accointances de premier niveau du connecteur pour s .

Pour tout élément c de A ,

1) un message est envoyé à c demandant en retour la liste S_c des quintuplets de s sur lesquels c a autorité ainsi que l'ensemble A' des accointances de premier niveau de c pour s

2) $A = A \cup A'$

3) $S = S \cup S_c$

A la fin, S contient l'ensemble des quintuplets de la forme $\langle x, y, s, t, f \rangle$.

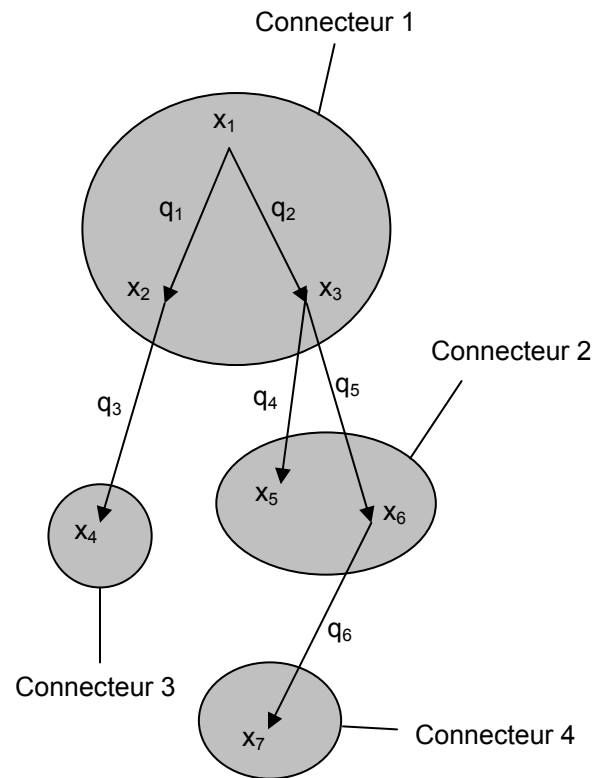


figure 5 – Répartition d'une session entre plusieurs connecteurs

La figure 5 donne un exemple de session répartie entre plusieurs connecteurs.

Pour le connecteur 4, qui ne connaît de la session que le quintuplet q_6 , l'obtention de l'intégralité de la session se fait en trois étapes :

- 1) Contact du connecteur 2, obtention des quintuplets q_4 et q_5 , ajout du connecteur 1 dans les accointances concernant la session
- 2) Contact du connecteur 1, obtention des quintuplets q_1 , q_2 et q_3 , ajout du connecteur 3 dans les accointances concernant la session

- 3) Contact du connecteur 3, aucune obtention de quintuplets car cette branche de la session se termine sur la page x_4

La session complète est donc composée de $\{q_1, q_2, q_3, q_4, q_5, q_6\}$.

5 Conclusion

Nous avons proposé un système de composition de documents virtuels basé sur le comportement des utilisateurs. Ce système spécifie le document virtuel en donnant l'ensemble des informations nécessaires pour composer le document réel.

A ce titre il répond aux trois problèmes du processus de composition : la *sélection*, l'*organisation* et l'*assemblage* de fragments. Ces trois points sont résolus par la notion de *session* qui correspond à une mise en commun d'une série d'expériences de navigation.

Une session est une *sélection* de parcours, *organisés* en fonction de la façon dont les utilisateurs les ont empruntés et *assemblés* selon leur force.

La mise en œuvre de ce système repose sur la présence d'agents à la fois sur les clients Web et sur les serveurs. Nos travaux précédents sur Mawa [5] ont donné lieu à un prototype d'agent connecteur. Ce noyau doit être complété par la réalisation de l'agent assistant.

Nous devons à présent vérifier l'émergence effective de centres d'intérêt de navigation au travers de la mutualisation des parcours observés.

Notre modèle va être dans un premier temps expérimenté dans le cadre clos et thématique d'un intranet : nous pensons mettre en évidence des parcours types reliés à des catégories d'utilisateurs, soit selon notre modélisation, l'agrégation des sessions utilisateurs en quelques sessions collectives.

Nous souhaitons poursuivre ces tests dans le cadre de l'Internet mais en nous limitant à des serveurs reliés entre eux. Nous commencerons par des serveurs Web universitaires de Midi-Pyrénées.

Disons enfin que si la constitution de documents virtuels est une aide appréciable pour l'utilisateur, c'est aussi une information précieuse pour le producteur de contenu et le gestionnaire de site web. Elle lui permet d'auditer les consultations et d'évaluer l'intégration de son site ou de son œuvre à l'intérieur d'un ensemble plus vaste d'informations.

Bibliographie

- [1] Ferber J., *Les systèmes multi-agents*, InterEditions 1995.
- [2] Gruber T. R., Vemuri S. and Rice J. 1997. Model-based virtual document generation. *International Journal*

of Human-Computer Studies, Vol. 46, No. 6, pp. 687–706.

[3] Habrant J., Corbel A., Girardot J.J., Les réseaux sémantiques comme outils d'aide à la navigation sur le web, *Conférence Interfaces Homme Machine 99*, Montpellier, France, novembre 99.

[4] Iksal, S., et al., Semantic composition of special reports on the Web: A cognitive approach, in *Hypertextes et Hypermédia H2PTM'01*, P. Balpe, et al., Editors. 2001, Hermès. p. 363-378.

[5] Singer N., Trouilhet S., An Advanced Interface for Information Access on a Large Scale System: MAWA for the Web, *7th International Conference on Information Systems Analysis and Synthesis (ISAS 2001)*, Orlando, USA, July 2001.

[6] Vasudevan V., Palmer M., On Web Annotation: Promises and Pitfalls of Current Web Infrastructure, *HICSS '99*, 1999.

[7] C. Watters and M. Sheperd, Research issues for virtual documents, *Workshop on Virtual Document, Hypertext Functionality and the Web*, Toronto, 1999.

[8] Zeiliger R., Reggers T., Baldewyns L., Jans V., Facilitating Web Navigation: Integrated Tools for Active and Cooperative Learners, *International Conference on Computers in Education'97*, Charlottesville, VA., 1997.