

Mémoire

En vue de l'obtention du diplôme de
Mastère de Recherche en Sciences du
Transport et de Logistique

Proposition des méthodes de recherche
d'itinéraires pour les systèmes de
personnalisation d'information dans le
domaine du transport

Présenté par : *Amna BOUHANA*

Encadré par : *Adel MAHFLOUDHI*

Année universitaire : 2008/2009

Résumé

Informier le voyageur sur les différents modes de transport disponibles, c'est lui permettre d'accéder plus facilement aux différents services de transport en commun. L'offre de cette information, nécessite la mise à disposition des voyageurs du transport en commun d'un Système d'Information Multimodale (**SIM**). Ce dernier doit être capable de fournir aux utilisateurs une information pertinente et adaptée à ses besoins. Cependant, les sources d'information auxquelles ce système doit accéder, sont nombreuses et distribuées. La personnalisation de l'information présente une solution pour améliorer la recherche d'information sur les itinéraires et les modes de transport (informations multimodales) et les adapter aux préférences et aux besoins des utilisateurs.

C'est dans ce cadre que le travail proposé s'inscrit. Il présente deux méthodes de recherche d'itinéraires :

- La première consiste à calculer le degré de pertinence d'un itinéraire par rapport à un ensemble d'itinéraires possibles en tenant compte des critères de préférences de l'utilisateur.
- La deuxième méthode consiste à calculer un score de correspondance ou de similarité entre un itinéraire et le profil de l'utilisateur.

Ces deux méthodes contribuent à l'amélioration de processus de recherche d'itinéraires dans le domaine de transport en délivrant une information plus pertinente qui devrait satisfaire davantage les besoins de l'usager, et en particulier du transport collectif ou en commun.

Mots- clés

Personnalisation, profil utilisateur, information multimodal, transport, recherche d'itinéraire

Summary

Informing the traveller about the various ways of available transportation, allows him (her) to reach easily the various services of public transportation. The offer of this information requires provision to the customers, a Multimodal Information System (**MIS**). This last one must be capable of supplying to the users relevant information and adapted it to their needs. However, the sources of information which this system has to reach are numerous and distributed. The personalization of the information presents a solution to improve the search for information about routes and adapt them to the preferences and to the needs of users.

It is in this frame that the proposed work joins. It presents two methods for itineraries research:

The first one consists in calculating the itinerary relevance degree compared to possible itineraries set by taking into account the user preferences criteria.

The second method consists in calculating a score of correspondence or similarity between a route and the profile user.

These two methods contribute to the improvement of the itinerary process research by delivering more relevant information which should satisfy of advantage the needs of the user, and in particular the collective or in common transport.

Keywords

Personalization, profile user, multimodal information, transport, itineraries research.

Dedicaces

A mon père que (dieu ait son âme)

A ma mère Yasmina

A mes frères Momtez, Anis , Abdallah et Abderrahmane.

A ma sœur Asma

A mes meilleurs amis Moez benfadhel et Ikram katrouch.

Vous vous êtes dépensés pour moi sans compter.

En reconnaissance de tous les sacrifices consentis par tous et chacun pour me permettre d'atteindre cette étape de ma vie.

A mes oncles, tantes, cousins et cousines.

A ceux et celles qui, de près ou de loin, m'ont encouragée à la réalisation de ce travail.

A tous ceux qui m'aiment

A tous ceux que j'aime

A ceux qui me sont chers

Je dédie ce travail

Amna Bouhana

Remerciements

Qu'il me soit permis de présenter ici mes remerciements à tout un petit monde de personnes qui ont rendu possible le présent travail et qui ont contribué à son élaboration sous quelque forme que ce soit.

*Il est agréable, d'adresser mes sincères remerciements à mon encadreur Monsieur **Adel Mahfoudhi**, pour l'effort fournie, les conseils prodigués, sa patience et sa persévérance dans le suivit de ce travail.*

*Je tiens à remercier également Monsieur **Abed Mourad** et Monsieur **Makrem soui** pour leur précieuse collaboration pour la réalisation de ce travail.*

J'adresse également mes remerciements à tous mes enseignants qui m'ont donnée les bases de la science.

Enfin, mes remerciements sont adressés aux «Messieurs les membres de jury» pour l'honneur, qu'ils m'ont accordé en acceptant de juger mon travail.

Cette liste ne saurait être exhaustive, je remercie, donc, tout ceux dont les noms n'apparaissent pas sur cette page, et je m'excuse auprès d'eux.

Amna Bouhana

Liste des tableaux

Tableau 1 -les différents systèmes de personnalisation dédiés	14
Tableau 2 - synthèse des systèmes d'information transport	24
Tableau 3 - les approches basé sur différentes méthodes de personnalisation (cognitives) ...	37
Tableau 4 - Classification des méthodes de personnalisation selon le contexte utilisé et le type de filtrage adaptée de.....	43
Tableau 5 -Comparaison entre les deux systèmes.....	49
Tableau 6 -les différentes méthodes d'évaluation et leurs critères	54
Tableau 7 -Rangs des critères par rapport aux solutions d'une requête de recherche d'itinéraires.....	58
Tableau 8 -Analogie de l'importance d'un terme t dans un document d avec l'importance d'un critère i pour un itinéraire J	63
Tableau 9 -comparaison entre les deux méthodes de personnalisation.....	85

Liste des Figures

Figure 1- Architecture multi-agents du MAPIs (Petit rosé.03)	25
Figure 2- architecture générale de PerSyst (Anli.06)	26
Figure 3- Les dix grandes étapes de la méthode PERCOMOM (Brossard.08).....	28
Figure 4- Un graphe multimodal et distribué	45
Figure 5- Les phases de l'algorithme génétique	47
Figure 6- Méthodologie de développement d'SIP (Anli.06).....	53
Figure 7- Diagramme des cas d'utilisation du service de recherche d'itinéraire personnalisé	70
Figure 8- Inscription de l'utilisateur au système	71
Figure 9- Identification de l'utilisateur	71
Figure 10- Choix de l'itinéraire pour un exemple de déplacement entre une origine et une destination	73
Figure 11- La liste des itinéraires possibles	74
Figure 12- classement des itinéraires par ordre de pertinence	79

Introduction générale

Face à l'augmentation et à la complexité des déplacements, les usagers du transport en commun souhaitent disposer d'une information fiable sur l'ensemble de modes de transport qui est mis à leur disposition. Cette information multimodale est difficile à mettre en œuvre pour des nombreuses raisons : organisationnelles, économiques, juridiques et Techniques, etc.

Les sources d'information sont très nombreuses et dispersées entre les différents opérateurs de transport. Les technologies et les supports de diffusion et de présentation de l'information sont aussi multiples et en constante évolution.

L'information multimodale constitue l'un des maillons essentiel de cette problématique. Il s'agit de fournir les informations utiles et pertinentes sur les différents modes de déplacement (métro, tramway, bus, etc.) à fin, d'une part d'améliorer le confort et l'efficacité des trajets à un niveau individuel, et d'autre part de favoriser l'usage multimodal et raisonné des différents modes de transports à l'échelle collective. En effet, le choix du mode de déplacement est souvent contraint par l'accessibilité de ce dernier, son prix, sa qualité, et la régularité du service proposé. Par conséquent, un voyageur informé sur les modes de transport disponibles (le coût de déplacement, le temps de parcours dans les véhicules de transport, le temps d'attente dans les stations de correspondance, même sur les meilleurs itinéraires à prendre entre une source et une destination données) accède plus facilement aux services de transport en commun. Afin d'offrir ses informations on doit disposer d'un système d'information multimodale (SIM). Ce dernier est destiné à répondre à une demande de déplacement par le recueil et le traitement de données relatives aux différents modes de transport disponibles et à leur enchaînement. Ce système est susceptible de gérer des flux volumineux de données, distribuées et hétérogènes. Il doit être capable de fournir aux utilisateurs une information pertinente. En effet, l'usager des transports ne souhaite avoir à disposition que peu d'informations, juste celles qui l'intéressent directement, c'est-à-dire, celles qui sont adaptées à ses besoins et à ses préférences. Ces informations sont alors personnalisées, en d'autres termes, destinées à un utilisateur spécifique et non un utilisateur

générique. Par ailleurs, la personnalisation de l'information peut représenter une "valeur ajoutée" non négligeable.

Dans le domaine de l'information multimodale personnalisée, nous distinguons deux principaux domaines :

- La personnalisation de l'information d'une part,
- L'information multimodale d'autre part.

Dans ce contexte, l'état de l'art des différents domaines touchant à cette problématique a été réalisé. Ainsi, les principales approches et méthodes utilisées pour la personnalisation, ainsi que les principaux systèmes de l'information multimodale, ont été présentés et étudiés. Cette étude Bibliographique a permis de faire les constats suivants:

- La plupart des Systèmes d'Information Multimodale s'appuient sur une architecture de type Système Multi-Agents (SMA) et détiennent des sources d'information distribuées.
- Aucun SIM n'offre des itinéraires prenant en compte les activités d'attraction touristique ou de loisir préférés par les utilisateurs, ou des services spécifiques (administratifs, culturels, commerciaux, etc.) répondant aux centres d'intérêts d'un usager ; ceci peut constituer une piste intéressante pour améliorer la qualité de ces systèmes.
- Aucun système n'intègre les critères de préférences de l'usager dans le processus de recherche d'itinéraires.
- l'inexistence des approches d'évaluation des méthodes de personnalisation et du SIM.

L'objectif principal de notre travail est de proposer un système de recherche d'information personnalisé dont le but est de rechercher les itinéraires qui correspondent le mieux aux préférences de l'usager.

Pour atteindre cet objectif, nous avons suivi une démarche en trois étapes:

1. Une étude de l'état de l'art : Cette étude a permis d'identifier le domaine de la personnalisation de l'information dans le domaine d'information.
2. La proposition de deux méthodes de personnalisation :
 - la première se base sur le calcul de degré de pertinence.
 - la deuxième se base sur le calcul de degré de similarité.

3. l'application de deux méthodes proposées sur un cas d'étude dans le domaine de transport.

Le présent document est organisé en quatre chapitres :

Le premier chapitre présente le cadre général du travail. Il décrit les composantes du domaine de la personnalisation de l'information. La principale composante est le profil utilisateur. Il décrit aussi les différents axes de recherche sur l'information multimodale dans le domaine du transport, et en particulier, la personnalisation de cette information.

Le second chapitre propose les principales méthodes de personnalisation qui relèvent des travaux existants : Les méthodes cognitives ou basée sur le contenu et les méthodes sociales. Les méthodes de calcul d'itinéraires sont aussi présentées : les algorithmes de calcul d'itinéraires monocritère et multicritères. Enfin, les principaux méthodes et critères d'évaluation des systèmes d'information personnalisée (SIP).

Le chapitre 3 décline notre proposition de deux méthodes de personnalisation. La première est basée sur une approche statistique permettant de calculer le degré de pertinence d'un itinéraire. La seconde est basée aussi sur une approche statistique et mathématique qui permet de calculer le degré de similarité entre un itinéraire et le profil de l'utilisateur.

Le chapitre 4 est consacré pour l'application de deux méthodes de personnalisation dans le domaine de transport de personnes.

Enfin, nous terminons par une conclusion générale dans laquelle nous résumons l'ensemble des propositions et des résultats. Nous donnons à la fin une présentation des perspectives de recherche qui s'ouvrent à ce jour.

Chapitre1 : La personnalisation des systèmes d'information et d'information multimodale

I. Introduction

L'objectif de système d'information multimodale est de mettre à la disposition des utilisateurs, l'information sur les différents modes de transport et leurs interconnexions. Deux grands types d'information régissent le déplacement du voyageur : celles qui lui permettent de bien organiser son trajet avant son départ et celles qui lui permettent de bien exploiter ou d'optimiser son voyage grâce à une information en temps réel « en push » sur les temps d'attente, les perturbations, etc. Cette information multimodale détermine ainsi le choix modal et facilite l'usage des réseaux de transport.

L'évolution des nouvelles technologies et la banalisation d'Internet et de la téléphonie mobile ont notamment permis le développement de système d'informations multimodales et l'ont rendu dynamique et performant.

L'efficacité de ces systèmes repose sur leurs capacités à mobiliser des nombreux canaux d'informations pour fournir à l'usager une réponse personnalisée. Cependant, l'existence de plusieurs opérateurs qui gèrent ces différents modes de transport, ainsi que la diversité des sources d'information et de profils utilisateurs rendent la tâche du système d'information voyageur de plus en plus complexe. Ce constat rend nécessaire la personnalisation de ses systèmes.

La personnalisation est un axe de recherche émergent et en pleine expansion au niveau international. L'objectif principal est de délivrer à l'utilisateur une information appropriée à ses préférences, à ses centres d'intérêts et d'une manière plus générale à son profil. La personnalisation de l'information représente ainsi une « valeur ajoutée » non négligeable [Petit Rosé.03].

Le présent chapitre se décline en deux parties principales :

La première partie sera consacrée à la définition de la personnalisation ainsi qu'à la notion de profil utilisateur.

La deuxième partie relatera l'état de l'art en présentant les travaux existants dans le domaine de transport et de la personnalisation. Pour finir nous présenterons la problématique de notre travail et les objectifs de notre recherche.

II. La personnalisation de l'information : principes et concepts

Le besoin de la personnalisation est en partie dû à la profusion des données parmi lesquelles chaque utilisateur cherche des réponses pertinentes [Oulasvirta et al.08]. Cette profusion s'explique par différentes raisons parmi lesquelles nous pouvons citer :

- l'augmentation des capacités de stockage et la baisse de leur coût.
- la distribution des données via Internet, en matière de partage d'information.

Par conséquent, l'accès à une information pertinente, adaptée aux besoins et au contexte de l'utilisateur devient alors un enjeu crucial pour l'utilisateur. Ceci nécessite la conception des systèmes de recherche d'information efficaces dont le but est d'aider l'utilisateur à trouver l'information utile parmi la masse des données disponibles.

Le problème ne concerne pas la disponibilité de l'information mais plutôt sa pertinence vis-à-vis du contexte de son utilisation. Le but de ces travaux est de délivrer à l'utilisateur une information pertinente adaptée à son contexte, à ses préférences et répondant à ses besoins précis.

C'est dans ce cadre de personnalisation de l'information que notre travail s'inscrit. Il constitue un enjeu majeur pour l'industrie informatique et les opérateurs du transport. Quelque soit le contexte de système d'information, la pertinence de l'information délivrée, son intelligibilité et son adaptation aux usages constituent des facteurs clés du succès ou de rejet de ces systèmes de recherche d'information.

II.1 Définition de la personnalisation

Dans la littérature jusqu'à maintenant, il n'existe pas de définition standard portant sur la personnalisation. Les définitions les plus adoptées sont :

« La personnalisation de l'information peut être définie comme étant un ensemble de préférences individuelles, pouvant être représentées de différentes manières, qui vont être utilisées pour fournir les réponses les plus pertinentes possibles à l'utilisateur » [Domshlak et al. 07].

Une autre définition a été proposée par Petit-Rosé « personnaliser l'information consiste à s'adapter aux buts (c'est-à-dire les besoins, les raisons qui ont mené l'utilisateur à interroger le

système), aux préférences (c'est-à-dire l'ensemble des critères qui permettent de distinguer une solution d'une autre pour une même requête) et aux capacités de l'utilisateur » [Petit-Rosé.03].

Selon [Anli.06], la personnalisation est définie comme étant « la capacité d'adaptation d'un système interactif par rapport au contexte. Cette adaptation peut aussi bien porter sur le contenu que sur le contenant ».

Brossard [Brossard .08] qui se limite aux modélisations conceptuelles des applications a proposé la définition suivante : La personnalisation, c'est la capacité de fournir à un utilisateur, à chaque instant, des contenus et des services adaptés à ses besoins et à ses attentes en utilisant des interactions homme-machine appropriées. Pour se faire, cette adaptation utilise sa connaissance de l'utilisateur (caractéristiques et préférences), sa connaissance de l'environnement, ainsi que des informations sur le comportement de l'utilisateur au moment de l'interaction.

D'après ces définitions nous pouvons dire que la personnalisation est généralement définie suivant les objectifs des auteurs, chaque auteur l'adapte suivant ses besoins et selon son domaine.

Nous constatons que la personnalisation peut aussi bien porter sur le contenu (données, information, etc) que sur le contenant (présentation, plate-forme d'interaction, mode de communication, canal de communication). Nos travaux portent sur la personnalisation de contenu.

Dans la partie suivante nous allons présenter des exemples de système d'information personnalisé.

II.2 Exemples des systèmes personnalisés

Dans cette partie nous allons présenter les principaux systèmes de personnalisation existants dans la littérature. Ces systèmes sont destinés, pour les plus parts, à faciliter la navigation sur le Web. Ils permettent d'assurer l'observation de comportement de l'utilisateur ainsi que le processus de recherche et de filtrage de l'information.

Dans la littérature, on distingue deux types de systèmes de personnalisation :

- ✓ Systèmes de personnalisation dédiés
- ✓ Systèmes de personnalisation pour l'aide au développement de système d'information personnalisé.

II.2.1. Systèmes de personnalisation dédiés

Les systèmes de personnalisations dédiés se présentent sous la forme des applications dédiées à des fonctions particulières et à des objectifs bien déterminés.

Ces systèmes sont spécialement conçus pour être installés et être exécutés au niveau de la plate-forme d'interaction de l'utilisateur. Ces systèmes se présentent, généralement, sous forme d'agents logiciels pour aider la navigation sur le web. Ils assurent les fonctions d'observation du comportement utilisateur, de recherche d'information, de filtrage et de présentation de ces informations [Anli.06].

Le tableau ci-dessous présente les différents systèmes qui existent dans la littérature :

Systèmes	Recherche d'information	Type de personnalisation	Domaine d'application	Méthodes de personnalisation
IFWeb [Asnicar et Tasso .97]	γ	Contenu	Recherche documentaires	Filtrage basé sur le profil
Letizia [Lieberman .95]	γ	Contenu	Recherche sur web	TF-IDF
Web Mat [Keeble et Macredie .00]	γ	Contenu	Recherche d'information sur internet	TF-IDF
Info Sleuth [Deschaine et al. 00]	γ	Contenu	Reformulation de requête	Ontologies
Con Call [Waern et al.98]	γ	Contenu	Le système sert à rassembler, filtrer et diffuser des appels pour des conférences, des workshops, etc., sur Internet	Filtrage collaboratif
Gullvier's genie [O'Hare et O'Grady. 03]	γ	Contenu	Domaine touristique : fournir des informations touristiques sur PDA	Filtrage basé sur le profil et recommandation contextuelle
MAPIS (MultiAgent Personalized Information System) [Petit-Rozé. 03]	γ	Contenu	Domaine de transport Information voyageur (itinéraires)	Filtrage Basé sur profil : Apprentissage par renforcement
Smart Radio [Hayes et al.04]	γ	Contenu/Contenant	Domaine de la musique	Pearson

Tableau 1-les différents systèmes de personnalisation dédiés [Anli.06]

Selon [Anli et al .05], ces systèmes de personnalisation présentent cinq caractéristiques :

- ✓ Multi-application : le système peut être accessible par plusieurs applications.

- ✓ Recherche d'information : capacité du système à effectuer une recherche d'information sur Internet ou sur des bases de données hétérogènes
- ✓ Assistance à l'utilisateur : capacité du système à fournir une recommandation à partir d'un raisonnement automatique
- ✓ Gestion de profil : les méthodes utilisées par le système pour exploiter le profil de l'utilisateur.

Ces systèmes présentent l'avantage d'être directement en contact avec l'utilisateur ce qui facilite la collecte de données sur le comportement de l'utilisateur pour une personnalisation plus fine. Leur inconvénient majeur réside dans le fait que la réutilisation de ces systèmes est très limitée.

II.2.2. Systèmes de personnalisation pour l'aide au développement de système d'information personnalisée

Ces systèmes ont pour objectif de faciliter la conception des systèmes d'information personnalisée.

Ces systèmes se présentent sous la forme d'une application configurable. Ils ont pour rôle de gérer le profil de l'utilisateur et de sélectionner l'information pertinente. Parmi les systèmes les plus connus dans la littérature, on peut citer celui de [Kobsa et al .95] [Pohl et al. 97] : **BGP-MS** (Belief, Goal and Plan Maintenance System). C'est un système de modélisation de l'utilisateur qui permet de prendre en compte les buts, croyances et connaissances de l'utilisateur. Il peut être utilisé dans un serveur avec multi-utilisateurs et multi-applications.

En conclusion, on distingue deux approches pour la conception de système de personnalisation :

- La première approche consiste à fournir un système interactif qui incorpore lui-même (ad-hoc) la personnalisation.
- La deuxième approche consiste à fournir un système dédié à la personnalisation et qui interagit avec une application tierce pour la personnalisation.

De même, nous pouvons dire que la personnalisation de l'information a pour objectif de délivrer à l'utilisateur une information appropriée à ses préférences, à ses centres d'intérêts ou plus globalement à son profil. L'ensemble de ces informations est représenté dans un modèle d'utilisateur appelé souvent par « **profil** ».

II.3. Le profil utilisateur dans le cadre de la personnalisation

II.3.1. Définition de la notion de profil utilisateur

On appelle **profil utilisateur** toute structure qui permet de **modéliser** et de **stocker** les données caractérisant l'utilisateur. Ces données représentent les centres d'intérêts, les préférences et les besoins en informations de l'utilisateur ou un groupe d'utilisateurs [Zmerli.04]. Selon [Kostadinov.08], le profil utilisateur est une mise en facteur de la partie invariante des préférences de l'utilisateur.

Même si l'ensemble des approches de personnalisation s'accorde sur l'importance de disposer d'un profil utilisateur, il manque un consensus sur la typologie des connaissances le constituant ainsi que sur la manière de représenter ces connaissances. Pour la majorité des approches, le profil contient la description des centres d'intérêts et préférences de l'utilisateur.

II.3.2. Les techniques de construction des profils utilisateur

Pour la construction d'un profil utilisateur, [Kostadinov .08] a dégagé un certain nombre de caractéristique à respecter, on peut citer :

- **Le moment de construction de profil** : la construction de profil est souvent faite en dehors des sessions d'interaction entre l'utilisateur et le système. Mais il y a des approches qui proposent de construire le profil utilisateur en temps réel [Sugiyama et al.04].

- **L'implication de l'utilisateur** : les approches se devisent en deux groupes :

- un groupe qui n'implique pas l'utilisateur dans la construction de profil [Chan.00].
- un groupe qui implique l'utilisateur en lui demandant de remplir un formulaire (exemple : Yahoo) ou de corriger manuellement ses préférences [Werner et al.04].

- **Le type de sources de données utilisées pour la construction des profils** : différents types de données peuvent être exploitées pour la construction des profils utilisateurs. On peut citer, les données d'usage [Mobasher.05], les données de contenu [Michlmayr et al.07], les données décrivant la structure et l'organisation des documents (Lieberman et al.95) ou encore les données provenant de l'utilisateur [Sugiyama et al.04].

- **La manière de mise à jour des profils** : lorsque des nouvelles données sont collectées sur l'utilisateur ou sur son comportement, le profil doit être mis à jour pour tenir compte de ces données. On distingue selon ce critère des systèmes qui mettent à jour

uniquement les parties du profil utilisateur concernées par les nouvelles données et les approches qui construisent à nouveau le profil utilisateur.

Ces paramètres peuvent servir également comme critères de choix de la méthode de construction de profil la plus appropriée par rapport aux besoins des applications.

II.3.3. Représentation de profil

Dans la littérature, nous pouvons distinguer 3 types de représentation :

- **Ensembliste** : le profil y est généralement formalisé comme des vecteurs de termes pondérés [Ciro et al. 03] ou classes de vecteurs [Michlmayr et al. 07]. On peut citer la représentation naïve des centres d'intérêts et à base de mots clés, tel que le cas des portails Web, My Yahoo, Info Quest, etc.
- **Sémantique** : la représentation du profil met en évidence, dans ce cas, les relations sémantiques entre l'information et le contenu [Tamine et al.06]. La représentation est essentiellement basée sur l'utilisation d'ontologies [Boughanem et al.07] [Brossard .08] (ou réseaux sémantiques probabilistes [Lin et al .05].

Selon [Brossard.08], la notion d'ontologie est largement utilisée dans ce qu'on appelle le web sémantique dont l'objectif est de permettre une caractérisation unique des pages web afin de permettre de donner un sens à leur contenu.

Dans ce cadre l'objectif visé est de ne plus faire des recherches sur le web en fonction des mots contenus dans les pages web mais en fonction des concepts métier qui y sont associés et permettant ainsi d'avoir des recherches plus pertinentes en fonction des besoins des utilisateurs.

Dans le domaine de la personnalisation de l'information, l'ontologie qui représente le profil utilisateur est personnelle, dans le sens qu'elle reflète la vue de l'utilisateur au domaine. Elle est utilisée pour assigner une sémantique aux documents choisis par l'utilisateur et pour faciliter ainsi la recherche d'information pertinente.

- **Multidimensionnelle** : La représentation multidimensionnelle du profil s'inscrit dans une réflexion globale sur la personnalisation de l'information. En effet, le profil utilisateur est un élément clé dans le processus de recherche, la modélisation de l'utilisateur doit pouvoir capturer toutes les dimensions qui représentent l'utilisateur.

Le profil est structuré selon un ensemble de dimensions représentées selon divers formalismes. Dans la littérature, beaucoup de chercheurs ont essayé de définir un modèle de

l'utilisateur multidimensionnel [Kostadinov.08] [Anli.06] [Tamine et al.06] [Bouzeghoub et al. 07] [Boughanem et al.04].

Kostadinov [Bouzeghoub et al .04] a défini un modèle générique qui inclut tous les aspects de la personnalisation. Ce modèle comporte huit dimensions : les données personnelles, les centres d'intérêts, l'ontologie du domaine, la qualité attendue des résultats délivrés, la customisation, la sécurité et la confidentialité, le retour de préférences (feedback) et enfin les informations diverses.

Nous pouvons dire que le contenu d'un profil utilisateur ne se limite pas aux préférences sur le contenu des données. Le profil utilisateur peut également contenir des données personnelles, des préférences sur la qualité des données et sur la sécurité ou encore des préférences sur le contenant et sur le processus de la production et/ou de la livraison des données.

Une fois le profil utilisateur construit, il est utilisé par les systèmes d'information afin de fournir à l'utilisateur des données pertinentes, dans le format le plus approprié, au bon moment et au bon endroit.

La personnalisation de l'information touche plusieurs domaines, aussi bien dans le contexte des systèmes d'information d'entreprise, du commerce électronique, de l'accès aux connaissances ainsi que les loisirs. L'information relative au transport est aussi concernée.

III. L'information multimodale

Il existe plusieurs définitions pour l'information multimodale. Celle proposée par l'Atec « la fonction essentielle d'un système d'information multimodales et de fournir à l'usager de transport toutes les informations nécessaires à la réalisation de son voyage.

Cette définition vise à réduire l'incertitude des usagers sur les itinéraires, le mode de déplacement envisageable, la durée et le coût de ses déplacements selon le mode utilisé et si possible à orienter le comportement des usagers aux bénéfices d'une utilisation optimale des infrastructures et d'une priorité au transport collectif.

Ainsi, on peut dire que l'information multimodale recouvre plusieurs domaines :

- ✓ Mode de transport : transport en commun, transport individuel, parking, etc.
- ✓ Type d'information et de données : information temps réel, information statistique, information sur les horaires, sur les tarifs, sur l'itinéraire, ...
- ✓ Moment de déplacement : information avant, après ou pendant le voyage.
- ✓ Couverture géographique : suivant la longueur (distance) de voyage l'information devra porter sur l'agglomération.

- ✓ Média personnel d'information.

Ces informations ont pour objectif de réduire l'incertitude des voyageurs sur les itinéraires. Favoriser la multimodalité consiste à améliorer les services offerts afin d'aider les usagers dans leurs déplacements et de développer la complémentarité entre les modes dans le but d'optimiser la gestion du système de transport.

Afin de développer les services d'informations multimodales, divers progrès ont été réalisés. Dans les pays européens, plusieurs projets ont été lancés en vue d'améliorer la qualité d'information délivrée aux voyageurs.

A titre d'exemple, nous pouvons citer :

- La Plateforme de Recherche et d'Expérimentation pour le Développement de l'Information Multimodale (PREDIM). Ce projet est lancé en 2001 par le gouvernement français qui regroupe plusieurs projets régionaux pour générer l'information multimodale par l'intégration de l'information issue de différentes sources ; c'est le cas dans le cadre du projet I-Trans et le du projet « VIATIC MOBILITE ».
- le projet DELFI qui a pour mission de mettre en place un système d'information multimodale qui couvre l'ensemble de territoire allemand, il offre au voyageur une information multimodale porte-à-porte avant le voyage.
- aux Pays Bas, le projet PITA pour « Personal Intelligent Travel Assistant » est un projet de recherche réalisé dans le cadre du programme « Seamless Multimodal Mobility » conduit par les universités de Delft et Eindhoven. Il permet à un voyageur de planifier son trajet sur internet PITA, de télécharger sur un appareil mobile (téléphone mobile, Smartphone, PDA) et de disposer d'une assistance réel tout au long du trajet.

Il est important de signaler que la qualité de l'information joue un rôle primordial dans le développement de la multimodalité aux services de transport de personnes.

III.1.Définition de la notion multimodalité

La multimodalité : la multimodalité entre une origine et une destination donnée, c'est la mise à disposition de plusieurs moyens de transport ou chaîne de moyen de transport, à partir desquels l'usager doit faire un choix afin de satisfaire son besoin de déplacement.

L'intermodalité : c'est l'enchainement de plusieurs moyens de transport pour un déplacement entre une origine et une destination.

L'information multimodale : c'est l'ensemble d'éléments d'information sur les différents modes de transport ou chaîne de moyens de transport disponibles sur le trajet.

Les services d'informations multimodales : Ils sont diffusés par différents canaux support. Ces services fournissent des fonctions sur le Web telles que le calcul-d'itinéraire, l'affichage et la diffusion des horaires (Web, téléphone, bornes). Ils diffusent aussi des informations sur les tarifs, les réservations et sur le temps d'attente, etc.

III.2.La personnalisation dans le domaine de transport

Dans le domaine du transport collectif, l'information aux voyageurs n'a pas cessé de se développer.

III.2.1. Les causes qui ont poussé à la personnalisation de l'information dans le domaine de transport

Dans le cas du développement durable, il apparaît que la proposition d'une information aux voyageurs via un portail unique et personnalisé serait susceptible de favoriser l'utilisation des transports en commun. Cette personnalisation est susceptible de réduire la part modale de l'usage de l'automobile dans les déplacements et par conséquent, l'émission des gaz polluants. Elle permet aussi d'améliorer la qualité d'information transport afin de rendre les déplacements beaucoup plus simples. Ceci est d'autant vrai lorsque ce déplacement nécessite la combinaison de plusieurs modes de transport.

Selon [Brossard. 08] un système d'information voyageur doit :

- Informer les utilisateurs des possibilités qui leur sont offertes.
- Inciter les utilisateurs à utiliser les transports collectifs pour leurs déplacements aussi bien pour le travail que pour les loisirs.
- Fournir aux utilisateurs le meilleur itinéraire pour un déplacement donné.
- Accompagner les utilisateurs pendant leurs déplacements.
- Fournir une information fiable et actualisée
- Être simple d'utilisation afin de pouvoir être consulté par tous les types d'usagers possibles (enfants, personnes âgées, personnes avec un ou plusieurs handicaps, etc.).

Donc la personnalisation de l'information fournit un élément important pour aider l'utilisateur à prendre une décision d'emprunter le transport collectif ou pas.

III.2.2.Les systèmes commerciaux existant

Dans le domaine de l'information voyageur, il existe des nombreux systèmes commerciaux. Les systèmes qui prennent en compte un seul mode de transport et fournissent des informations simples et non personnalisées, parmi ceux-ci on peut citer :

Le site Transville (<http://www.transvilles.com>) permet, pour un voyageur, d'organiser ses déplacements dans la ville de Valenciennes et dans quelques villes environnantes.

Des systèmes qui présentent des solutions regroupant quelques éléments de personnalisation comme le site de la SNCF (<http://www.voyagesncf.com>). Ce dernier recouvre presque la totalité des transports ferroviaires en France et permet pour sa part de préenregistrer ses déplacements préférés ainsi que quelques autres préférences comme, par exemple, le type de place et le type de train afin de faciliter les achats de titres de transport par l'usager.

En matière d'accompagnement des usagers pendant leurs déplacements, la totalité de ces solutions ne dispose généralement que d'un nombre limité de services.

D'autres systèmes proposent des solutions plus personnalisées et plus développées comme celui de la société de transport de Washington (<http://www.wmdata.com>). Ce dernier permet à l'utilisateur de recevoir l'information en temps réel sur différents supports tels que le téléphone portable, l'ordinateur portable, I-pod, etc.

III.2.3 Les systèmes d'information voyageurs personnalisée :

Dans le domaine de transport collectif la mise à disposition de l'information aux voyageurs n'a pas cessé de se développer. Plusieurs systèmes, qui ont été proposés, ont pour objectif de personnaliser l'information avant, pendant et parfois après le déplacement, tout en essayant de couvrir le plus de canaux et de plates-formes d'interaction possibles. L'objectif de ces systèmes est de délivrer aux usagers une information personnalisée, de qualité et qui satisfait les préférences des utilisateurs. [Anli.06] classe les systèmes d'information transport selon quatre caractéristiques :

- **les modes de transport** : correspondant aux types de transport (train, bus, métro, tramway, etc) considérés par le système.
- **les canaux et les plates-formes d'interaction** : ils décrivent les moyens d'accès (web, wap, SMS, etc.) au système d'information et les supports d'interaction (PC, téléphone mobile, PDA, etc.) auxquels ils sont destinés.
- **les principaux services** : ils résument les informations et les applications que le système met à la disposition de l'utilisateur.

- **la personnalisation** : elle présente les modes de collecte d'information utilisateur et les types de personnalisations fournies par le système .Il existe deux types de personnalisation :

- ✓ Par rapport au contenant : telle que la plasticité et la composition de l'interface par l'utilisateur.
- ✓ Par rapport au contenu : telle que le tri d'information par rapport aux préférences de l'utilisateur.

Dans ce travail, nous nous intéressons plus à la personnalisation par rapport au contenu.

Le tableau ci-dessous regroupe les différents systèmes d'information développés pour le domaine de transport multimodal. La plupart de ces systèmes ont pour objectif de mettre à la disposition des usagers les informations relatives au transport et de résoudre la problématique de personnalisation de l'information dans le domaine de transport.

Système	Mode de transport	Canaux / Plateforme d'interaction	Principaux services	personnalisation
SNCF	TGV, Train, TER	Web, Mail, SMS / PC, Téléphone	-Recherche d'itinéraire interurbains - Réservation et achat de billets - Organisation du voyage de bout en bout (réservation voiture, hôtel, etc.) - Prévision du trafic TGV - Informations des perturbations du trafic par région	-Remplissage automatique des préférences (trajets habituels,confort, compagnons, réductions, etc.) à partir d'un questionnaire rempli lors de l'inscription - Envoi de SMS dans le cas d'une perturbation sur un trajet planifié par l'usager
RATP	Train (SNCF), RER, Métro Tramway, Bus, Marche	Web, SMS, Wap / PC, Téléphone	- Recherche d'itinéraire en île de France - Plans de proximité - Affichage des horaires	- Préférences (requêtes, choix, critères, etc.) sauvegardées pour les utiliser dans une autre

			<ul style="list-style-type: none"> - Comparaison entre temps de déplacements en TC et VP - Plans des lignes et des réseaux - Informations des perturbations 	connexion
LE PILOT	Bus, Tramway TER (SNCF), Métro, Marche	Web/Pc	<ul style="list-style-type: none"> - Recherche d'itinéraire dans Marseille et ses environs - Recherche des horaires - Evénements culturels, sportifs, etc. - Informations sur les travaux prévus 	<ul style="list-style-type: none"> - Précision des critères (plus rapide, sans métro, etc.) transports par l'utilisateur lors de la recherche d'itinéraire
TFL	Bus, Métro, Train, Tramway, Marche	WAP, SMS, Web, Mail / PC, Téléphone, PDA	<ul style="list-style-type: none"> - Recherche d'itinéraire à Londres - Tableau des départs - Informations connexes aux déplacements (Toilettes, accessibilité, etc.) - Tarification, achat et réservation de billets 	<ul style="list-style-type: none"> - Préférences (requêtes, choix, critères, etc.) sauvegardées pour les utiliser dans une autre connexion
Agent Perso	Train, Bus, Tramway, Train, Marche	Web / PC	<ul style="list-style-type: none"> - Recherche d'itinéraire dans la région Nord Pas-de-Calais - Recherche d'itinéraire thématique 	<ul style="list-style-type: none"> - Apprentissage automatique des préférences (modes, critères, etc.) de l'utilisateur
PIEPSER	Train, Bus, Tramway, Vélo, Voiture, Marche	SMS, Web, WAP / PC, Téléphone	<ul style="list-style-type: none"> - Recherche d'itinéraire en Allemagne - Horaires des lignes - Perturbations sur le trajet 	<ul style="list-style-type: none"> - Inscription au service d'information personnalisée - Proposition des alternatives en cas de perturbation
YTV	Bus, Métro,	Web /	<ul style="list-style-type: none"> - Recherche 	<ul style="list-style-type: none"> - Préférences

	Tramway, Train, Marche	PC	d'itinéraire à Helsinki - Tarification, cartes des lignes, liste des horaires	(requêtes, choix, critères, etc.) sauvegardées pour les utiliser dans une autre connexion
TISONI	Mode de transport urbain	Web / PC	- Recherche d'itinéraire au Singapour	Apprentissage automatique des modes de transport préférés par l'utilisateur - Estimation des temps de parcours en fonction du trafic
My Bus	Bus	Web, WAP / PC, Palm, Téléphone	- Recherche d'itinéraire à Seattle	- Interface adaptée par rapport à la plate-forme d'interaction

Tableau 2- synthèse des systèmes d'information transport [Anli .06]

Nous pouvons citer d'autres systèmes qui sont plus avancés dans le domaine de la personnalisation de l'information transport et qui sont développés dans le cadre du PREDIT au sein du LAMIH : AGENPERSO (AGENts logiciels PERSONnels) de [Petit-Rozé .03]. ces deux systèmes fournissent une aide personnalisée dans la préparation d'itinéraires combinant un ensemble de modes de transport.

Ils se basent sur les préférences de l'utilisateur pour lui fournir l'itinéraire correspondant à son profil. Dans ces travaux [Petit-Rosé.03] a fixé quatre critères de préférence pour fournir l'itinéraire qui correspond mieux aux préférences de l'utilisateur :

- ✓ le plus rapide
- ✓ le plus court chemin
- ✓ le moins cher
- ✓ le moins de correspondance.

Pour la personnalisation de système d'information voyageur Petit rosé [Petit rosé.03] a développé un système de personnalisation à base d'architecture multi-agent baptisée MAPIS

(Multi-Agent Personalized Information System). Le modèle MAPIS (figure 1) met l'accent sur la préparation des itinéraires multimodaux et prend en compte essentiellement le contexte identitaire.

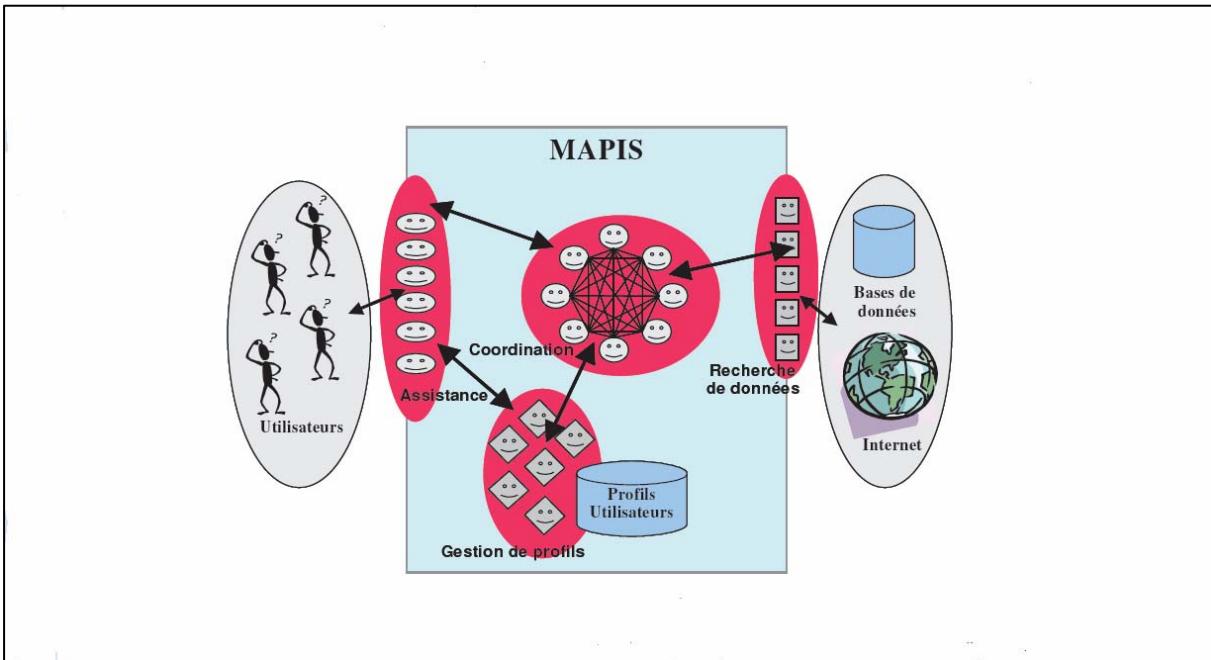


Figure 1- Architecture multi-agents du MAPIS [Petit rosé.03]

Le système MAPIS se compose de quatre agents (figure 1):

- Agent assistance : cet agent a pour rôle d'identifier les utilisateurs, comprendre leurs besoins, les assistées dans leurs activités. Les objectifs de cet agent modèle sont centrés sur l'utilisateur.
- Agents recherche d'information : ces agents ont pour rôle de récupérer les données qui sont soit centralisées, soit distribuées. Ces agents sont mobiles et capables de migrer sur un réseau informatique et de récupérer les données stockées dans des endroits différents.
- Agents gestion d'utilisateur : Ces agents sont dotés d'une base de données contenant les connaissances liées aux utilisateurs. Ils sont capables d'apprendre (selon plusieurs méthodes d'apprentissage) les besoins et les préférences de chaque utilisateur et de mettre à jour les profils utilisateurs.
- Agents coordination : le rôle de ces agents est de coordonner et planifier l'ensemble des sous-tâches afin de trouver une solution personnalisée à un utilisateur : parmi un ensemble d'offres d'itinéraires, trouvé par les agents

d'information, ces agents procèdent à un filtrage pour déterminer le meilleur itinéraire en prenant en compte le profil de l'utilisateur et les contextes liés à la demande.

PerSyst (personalization System) de [Anli.06] est bâti sur AGENPERSO qui est un système à base d'agents logiciels dont l'objectif est de fournir un système de personnalisation évolutif permettant d'intégrer facilement de nouveaux besoins en matière de personnalisation. Le système proposé par Anli [Anli.06] se compose de six agents à savoir (voir figure 2) :

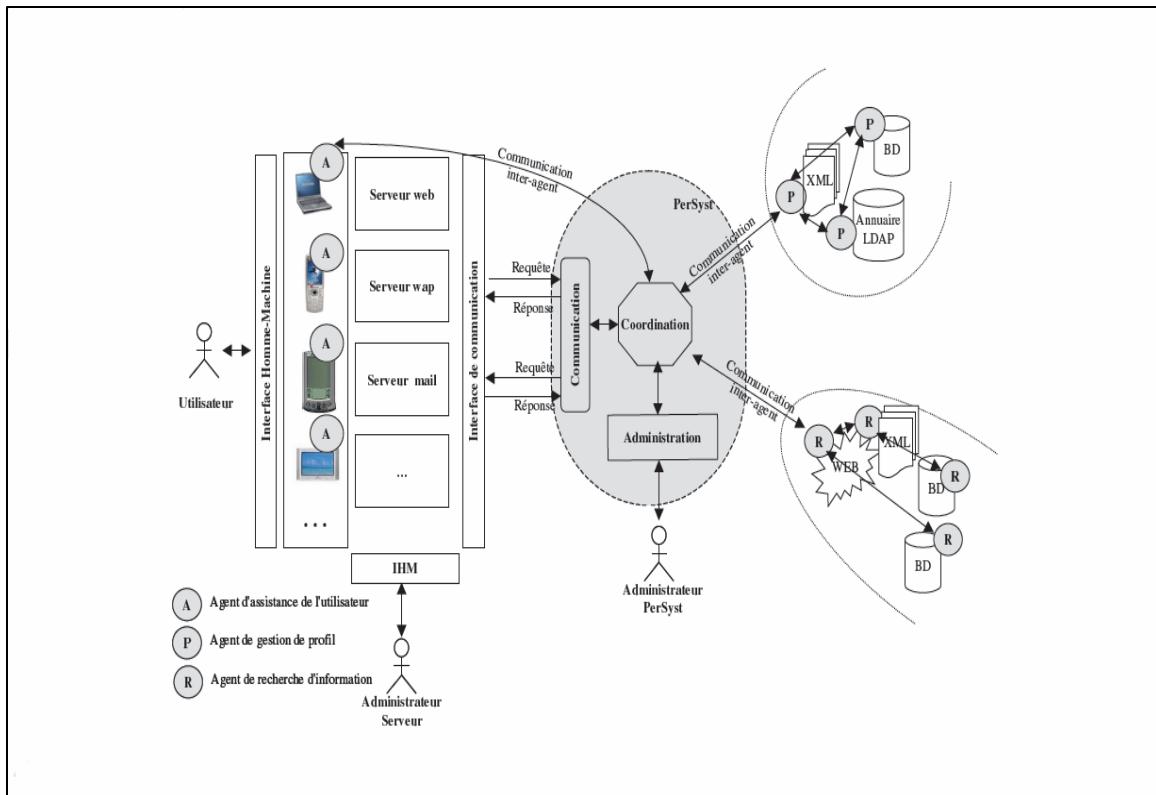


Figure 2-architecture générale de PerSyst [Anli.06]

- L'agent de communication : il assure la transmission des requêtes envoyées par les applications externes à l'agent de coordination.
- L'agent d'administration : Il traduit les requêtes de l'administrateur de PerSyst à travers une interface graphique à des requêtes compréhensibles par l'agent de coordination et inversement.
- L'agent assistant : PerSyst se sert de l'assistant pour reconnaître la plate-forme d'interaction à laquelle est connecté l'utilisateur et/ou pour choisir la plateforme sur laquelle l'information doit être acheminée.

- L’agent coordination : il assure la coordination de toutes les interactions.
- L’agent gestion de profil : Ces agents assurent la gestion des profils utilisateurs. Ils sont capables d’apprendre (selon plusieurs méthodes d’apprentissage) les besoins et les préférences de chaque utilisateur et de mettre à jour les profils utilisateurs.
- L’agent recherche d’information : Cette partie est assurée par un ou plusieurs agents. Ces agents ont pour objectif général la recherche d’information dans le web.

PerSyst a été appliqué pour la personnalisation de l’information transport et plus particulièrement les transports publics combinant différents modes de transport (itinéraire multimodaux) et différents exploitants. La personnalisation consiste à recommander un itinéraire par rapport à un ensemble d’itinéraires possibles. Ce système fournit des résultats qui sont personnalisés par rapport aux préférences de l’utilisateur. La méthode de personnalisation utilisée dans ce système se base sur le filtrage collaboratif et les réseaux Bayésiens. Le principe consiste à prédire le comportement d’un utilisateur par comparaison aux comportements d’autres utilisateurs ayant les mêmes préférences pour un contexte donné.

L’inconvénient de PerSyst de point de vue fonctionnel est qu’il ne prend en compte que les préférences de l’utilisateur et son profil par contre son adaptation au contexte environnement est limitée. De plus PerSyst n’est pas prévu pour prendre en compte des applications multiples. Il est conçu pour traiter chaque application individuellement.

Pour répondre à cette problématique de la personnalisation [Brossard .08] a développé dans PERCOMOM une approche générique de la personnalisation. La méthode PERCOMOM (**P**ERsonalization and **C**Onceptual **M**odeling **M**ethod) est une méthode utilisée pour la personnalisation de contenu des applications interactives à travers une approche dirigée par les modèles (figure 3). Elle a été appliquée dans le domaine de transport et plus spécialement dans le cadre de diffusion d’information aux usagers de transport commun dans le cadre du projet « Viatic.Mobilité » afin de rendre les applications interactives bien adaptées aux besoins et au contexte de l’utilisateur.

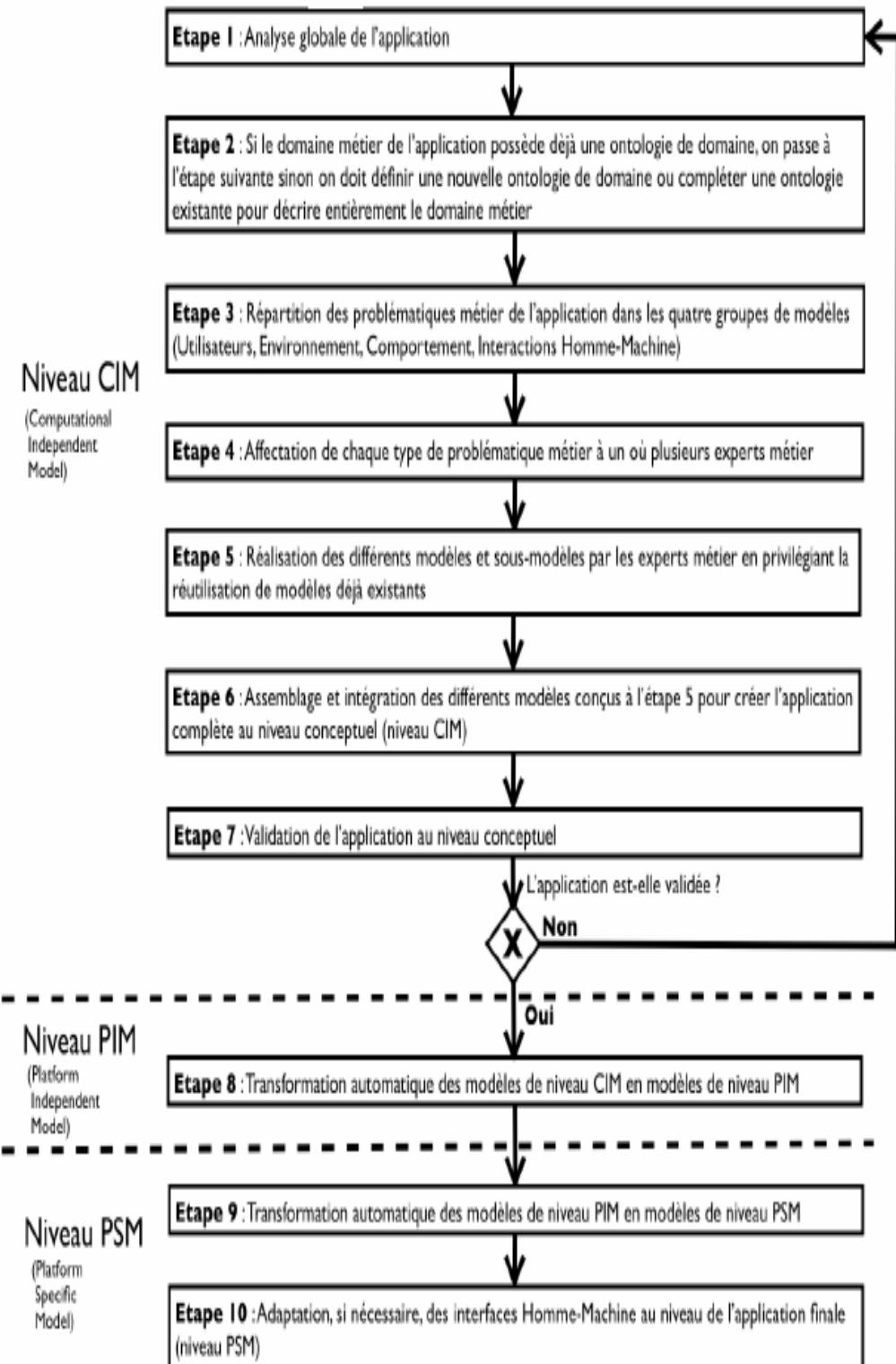


Figure 3-Les dix grandes étapes de la méthode PERCOMOM [Brossard.08]

VI. Synthèse et discussion

Ce chapitre a présenté l'état de l'art de la personnalisation dans les systèmes d'information voyageurs.

Il se dégage de ces travaux deux types de personnalisations utilisées :

- ✓ Une personnalisation par rapport au contenant.
- ✓ Une personnalisation par rapport au contenu.

Nos travaux se situent dans le contexte de personnalisation de contenu.

L'étude des systèmes d'information voyageurs nous a permis aussi de déduire que ces systèmes répondent à un ou plusieurs objectifs en terme de systèmes d'information tels que : fournir une information en temps réel, multimodale, accessible sur différents supports. Néanmoins, la personnalisation de l'information fournie reste encore limitée.

Or aujourd'hui, le défi principale des systèmes d'information voyageurs est de fournir aux utilisateurs des informations pertinentes qui conviennent mieux à ses besoins et à ses préférences .L'usager des transports ne souhaite en effet avoir à sa disposition que peu d'informations, celles qui l'intéressent d'une manière directe. Les informations offertes aux utilisateurs doivent répondre aux critères de temps, de coût et de moins de nombre de correspondances.

Nous remarquons aussi qu'aucun de ces systèmes n'offre des itinéraires selon l'activité d'attraction touristique ou de loisir préféré par les utilisateurs, ou des services répondant aux centres d'intérêts d'un utilisateur spécifique, ce qui peut présenter une piste intéressante pour améliorer la qualité de ces systèmes.

Après l'étude de ces divers systèmes de personnalisation, nous avons remarqué qu'il existe diverses méthodes de personnalisation mais à notre connaissance l'évaluation de certains systèmes est négligée.

V. Problématique

L'étude des travaux existants traitant la personnalisation de l'information et particulièrement l'information de transport multimodales nous conduit à mieux cerner certaines insuffisances dans les systèmes d'information voyageurs déjà développés.

- ✓ Nous avons remarqué aussi qu'aucun système n'intègre les critères de préférences dans le processus de recherche d'itinéraires.

- ✓ Nous avons remarqué qu'il n'existe aucune méthode de personnalisation qui intègre une étape d'évaluation du système en même temps.

Par conséquent, l'objectif de notre travail est de proposer une méthode de personnalisation de l'information selon les préférences des utilisateurs.

Nous allons présenter une méthode de personnalisation qui a pour objectif d'aider l'utilisateur à choisir son itinéraire d'une manière personnalisée selon ses besoins et préférences.

IV. Conclusion

Ce chapitre a présenté un état de l'art sur la personnalisation de l'interaction homme-machine. Des nombreux travaux sont effectués à ce sujet. La plupart de ces travaux vise à fournir des approches, des modèles et des méthodes pour la collecte d'information utilisateur, la gestion des profils utilisateurs et l'adaptation de l'interaction par rapport à ses besoins et ses préférences. Ces modèles et ces méthodes sont généralement adaptés pour répondre à une problématique précise relative à un type de personnalisation particulier. Par ailleurs, l'objectif de ces travaux vise à améliorer les services de transport commun, afin d'encourager les citoyens à les choisir en priorité par rapport aux véhicules privés.

Nous allons présenter dans ce qui suit notre système de personnalisation qui a pour objectif de fournir à l'utilisateur l'itinéraire le plus pertinent c- à dire qui correspond le plus à son profil et à ses préférences.

Le chapitre suivant a donc pour objectif de rechercher une méthode adéquate pour la personnalisation ainsi que pour le calcul d'itinéraire.

Chapitre2 : Spécification et étude des méthodes de personnalisation de calcul d'itinéraire et d'évaluation

I. Introduction

Dans le chapitre précédent, nous avons présenté les différents rôles et architectures de systèmes d'information multimodale. Ces systèmes remplissent, généralement, deux fonctions :

- Soit le calcul d'itinéraires.
- Soit la personnalisation de l'information multimodale.

Nous présenterons dans ce chapitre, les méthodes existantes qui permettent de concevoir un système d'information multimodale personnalisé. Dans la première section, nous détaillerons les approches de personnalisation utilisées. Dans la deuxième section, nous présenterons les différentes méthodes de calcul d'itinéraire utilisées dans le domaine de transport. A la fin, dans la troisième section, nous exposerons les principales méthodes d'évaluation existantes pour l'évaluation des systèmes d'information personnalisée.

Suite à l'étude de l'état de l'art présenté dans le chapitre précédent, nous avons pu montrer que les systèmes d'information multimodale actuels présentent encore certaines insuffisances.

Afin de rendre la recherche d'information plus pertinente, nous proposons d'améliorer la qualité des solutions proposées à un voyageur en tenant compte de son profil.

L'apport de notre travail consiste à proposer un système de personnalisation de l'information en se basant sur les préférences de l'utilisateur et de l'appliquer dans le domaine du transport pour la recherche d'information. Afin de mieux appréhender cette thématique, nous présentons dans la suite les principales méthodes de personnalisation, les plus connues dans la littérature.

II.2. Les méthodes de personnalisation

Les méthodes de personnalisation permettent de fournir à un utilisateur un contenu et/ou un contenant personnalisés .Dans la littérature nous distinguant deux méthodes de personnalisation :

- Les méthodes cognitives ou basé sur le contenu,
- Les méthodes sociales.

La combinaison de ces deux types de méthodes peut engendrer des méthodes qualifiées d'hybrides.

II.2.1. Les méthodes cognitives ou basé sur le contenu

La notion de méthode cognitive a été proposée par Salton & McGill (1983). Elle tire son origine de la recherche d'information traditionnelle qui utilise la description du contenu de l'information pour déterminer les profils utilisateurs correspondants [Pazzani et al, 1996]. Dans le cadre de la recherche d'information personnalisée, le profil d'utilisateur et les documents sont représentés par un vecteur de mots-clés ou de poids associés à ces mots-clés. Tous les documents sont représentés de la même manière.

La méthode consiste à calculer la distance entre le vecteur représentatif de l'utilisateur et ceux des documents. La distance la plus faible correspond au document le plus pertinent pour la requête de l'utilisateur [Korfhage.97] [Petit rosé .03].

Dans cette optique, nous trouvons plusieurs méthodes d'apprentissage automatique du profil utilisateur. Nous en citons quelques unes:

- **Technique vectorielle**

Généralement, un document ou une page web est considérée comme un ensemble de mots. Ils sont représentés par un vecteur de termes pondérés. Cette pondération est calculée par une fonction spécifique notée **TF-IDF**. La pondération est l'une des fonctions fondamentales en recherche d'information. Elle est la clé de voûte de la majorité des modèles et approches de recherche d'information proposés. Le poids d'un terme dénote son intérêt dans l'index, qui dépend du fait que le terme est important dans le document, et du fait qu'il permet de distinguer un document parmi tant autres dans la base. Le processus de représentation du document par ce vecteur est décrit dans [Memmi.00]. L'agent intelligent Web Mate utilise cette technique [Chen et al. 98]. L'algorithme d'apprentissage basé sur ces vecteurs est appelé "multi **TF-IDF** vector learning". Cet algorithme combine chaque paire de vecteur semblable. Ensuite, il trie les poids des termes dans un ordre décroissant et fournit un vecteur final.

Afin de déterminer la pertinence des nouveaux documents, ces derniers sont comparés à ce vecteur final.

L'avantage du modèle vectoriel réside particulièrement dans l'ordonnancement des documents sélectionnés selon leurs pertinences. Cependant, l'inconvénient majeur de l'approche vectorielle réside dans le fait que l'association entre les termes d'indexation n'est pas considérée. Il est impossible de représenter des phrases ou des mots multi termes. On considère effectivement que les termes sont indépendants [Zmerli.04].

- **Réseau Bayésien**

Selon Nguyen [Nguyen .05] « les réseaux bayésiens sont des graphes acycliques orientés dans lesquels les nœuds représentent des variables et les arcs représentent l'indépendance conditionnelle entre les nœuds parents et les nœuds fils. En tant que technique d'apprentissage de profil utilisateur, ils sont flexibles, puisqu'ils permettent de représenter les relations de causalité entre faits et d'émettre des prédictions sur de nombreux paramètres du système ».

D'après Naïm [Naïm et al.99] « On utilise les réseaux bayésiens pour leur capacité d'effectuer des inférences dans un contexte d'incertitude, en quelque sorte comme alternative aux systèmes experts. On les utilise aussi pour leurs algorithmes d'apprentissage, comme alternative aux autres méthodes de modélisation quantitative ».

Cependant cette méthode souffre d'un certain nombre d'inconvénient :
L'approche bayésienne est difficile à suivre d'une façon parfaitement rigoureuse pour deux raisons :

D'abord les connaissances *a priori* d'un agent sont souvent floues, mal formulées et leur traduction en un modèle probabiliste numérique est très difficile. Les hypothèses posées sont souvent fausses, il est nécessaire d'affiner les modèles après avoir vu leurs conséquences.
C'est pourquoi il est recommandé de mettre en place un cycle modélisation-inférence-évaluation permettant d'améliorer progressivement les modèles.

Il existe aussi le problème dit du "monde fermé" (*close world assumption*). Si le bon modèle n'a pas été considéré dès le départ, les inférences ne l'inventeront pas.

Le troisième grand problème de l'approche bayésienne est la difficulté calculatoire des inférences.

- **Le raisonnement à partir de cas (RàPC)**

Le raisonnement à partir de cas est une méthode qui sert à résoudre des problèmes. Ce type de raisonnement utilise des cas déjà stockés dans une base de cas. Cette méthode tire son origine des travaux de « Roger Schank » sur la mémoire dynamique et l'apprentissage, pour la

compréhension d'histoires énoncées en langage naturel [Schank.82]. Par ailleurs, ce qui motive en général l'utilisation du RàPC, c'est l'absence d'une théorie complète de la résolution de problème dans le domaine d'application, ou bien la présence d'une théorie mais elle est insuffisante pour la résolution d'un tel problème.

Le raisonnement à partir de cas est utilisé pour la résolution des divers problèmes qui appartiennent à des domaines différents tels que les services Web [lajmi et al.07], dans le domaine industriel (méthode de maintenance et diagnostic de système complexe) [Senechal et al .06].

En effet, le raisonnement par cas (RàPC) est une approche de résolution de problèmes basée sur la réutilisation par analogie d'expériences passées appelées *cas* [Lieborrtz.99].

D'après Kolodner [Kolodner.93], [Lajmi et al.07], un cas est constitué des mêmes composantes quelque soit le domaine d'application. Ces composantes sont un problème, une solution et éventuellement une évaluation.

L'ensemble des cas dont un système de RàPC, s'appelle « base de cas » ou « mémoire *de cas* ». Cette base peut contenir des expériences de problèmes résolus, ou l'on peut rechercher des expériences passées similaires au problème à résoudre. Les solutions des cas retrouvés sont utilisées comme base pour résoudre le problème courant. Un cas de cette base est appelé « *cas source* » (dont la partie problème est appelé « problème source »).

En effet, selon la résolution d'un problème [HajSaid et al.04], [Napoli.98] le raisonnement se décompose, généralement, en quatre phases principales :

- Phase de remémoration : l'objectif de cette phase est de rechercher des cas qui ont une similarité avec le problème défini ;
- phase de *réutilisation* permettant de construire une solution au problème courant en se basant sur les cas identifiés dans la phase précédente ;
- phase de *révision* de la solution qui permet de l'affiner grâce à son évaluation ;
- phase d'*apprentissage* chargée de mettre à jour les éléments du raisonnement en prenant en compte l'expérience qui vient d'être réalisée, et qui pourra ainsi être utilisée dans les raisonnements futurs.

Le principal avantage de cette méthode c'est qu'elle s'adapte au changement par le rajout des nouveaux cas. Cette méthode ne demande pas l'utilisation ni des règles heuristiques ni de modèles explicites pour la résolution d'un problème [Senechal et al .06].

L'inconvénient majeur de cette méthode est le suivant : plus la base s'accroît, et plus le temps de calcul sera long. C'est un procédé coûteux et lourd (réorganisation de la bibliothèque à chaque nouveaux cas).

- **Réseaux de neurone :**

Un réseau de neurones artificiel est un modèle de calcul dont la conception est très schématiquement inspirée du fonctionnement de vrais neurones humain [Memmi .04]. Il s'agit d'un graphe orienté. A chaque nœud est associée une fonction d'activation prédéterminée.

En effet, un neurone artificiel est composé des éléments suivants [Leroux .04] :

- Entrées : directement les entrées du système ou peuvent provenir des autres neurones
- Biais : entrée toujours à 1 qui permet d'ajuster de la flexibilité aux réseaux en permettant de varier le seuil de déclenchement du neurone par l'ajustement du poids du biais lors de l'apprentissage.
- Poids : facteur multiplicateur qui affecte l'influence de chaque entrée sur la sortie du neurone
- Noyau : intègre toutes les entrées, le biais et calcule la sortie du neurone selon une fonction d'activation qui est souvent non linéaire pour donner une plus grande flexibilité d'apprentissage.
- Sortie : directement une sortie du système ou peut être distribué vers d'autre neurones.

Le neurone calcule la somme de ses entrées puis cette valeur passe à travers la fonction d'activation pour produire sa sortie.

La fonction d'activation non linéaire permet aux réseaux de neurone artificiel de modéliser des équations dont la sortie n'est pas une combinaison linéaire des entrées. Cette caractéristique donne aux réseaux de neurone artificiel la capacité de modéliser et de résoudre des problèmes non linéaires.

Pour les réseaux de neurone on distingue deux principaux types d'architectures: réseaux à couches distinctes (sans cycles) et réseaux récurrents (avec cycles) [Lavoie .07].

Un réseau de neurones sans cycles : se compose généralement d'une succession de couches dont chacune prend ses entrées sur les sorties de la précédente. Chaque couche (i) est composée de N_i neurones, prenant leurs entrées sur les N_{i-1} neurones de la couche précédente. À chaque synapse est associé un poids synaptique, de sorte que les N_{i-1} sont multipliés par ce poids, puis additionnés par les neurones de niveau i , ce qui est équivalent à multiplier le vecteur d'entrée par une matrice de transformation. L'architecture multicouche est sans doute la plus utilisée. Nouali [Nouali.03] utilise comme technique d'apprentissage

automatique un réseau de neurone sans cycle avec trois couches dans son système conçu pour le filtrage des courriers électroniques.

✓ Les réseaux récurrents (avec cycles)

Au delà de cette structure simple, le réseau de neurones peut également contenir des boucles qui en changent radicalement les possibilités mais aussi la complexité. De la même façon que des boucles peuvent transformer une logique combinatoire en logique séquentielle, les boucles dans un réseau de neurones transforment un simple dispositif de reconnaissance d'inputs en une machine complexe capable de toutes sortes de comportements.

Pour les techniques d'apprentissage basées sur les réseaux de neurones, on distingue trois grandes classes d'apprentissages [Leroux .04] :

✓ L'apprentissage non supervisé : Lors d'un apprentissage non-supervisé, le réseau est laissé libre de converger vers n'importe quel état final lorsqu'on lui présente un motif. Ce type d'apprentissage est choisi lorsqu'il n'y a pas de connaissance à priori des sorties désirées pour des entrées données.

✓ L'apprentissage par renforcement : Dans ce cas, bien que les sorties idéales ne soient pas connues ou s'éloignent du but visé. Ainsi, les poids sont ajustés de façon plus ou moins aléatoires et la modification est conservée si l'impact est positif ou rejetée s'il est négatif.

✓ L'apprentissage supervisé : Un apprentissage est dit supervisé lorsque l'on force le réseau à converger vers un état final précis, en même temps qu'on lui présente un motif.

Cet algorithme d'apprentissage ne peut être utilisé que lorsque les combinaisons d'entrées-sorties désirées sont connues.

Une fois le type d'apprentissage est choisi, on doit déterminer si le neurone sera statistique ou dynamique.

- ✓ Dans le premier cas une phase de pré apprentissage sera effectuée.
- ✓ Dans le deuxième cas un apprentissage continu sera effectué pour permettre aux réseaux de neurone de réagir aux changements de son environnement.

Les principaux avantages de cette méthode sont : l'apprentissage avec les réseaux de neurone est performant même en présence de bruit. Les réseaux de neurone sont robustes, versatiles et peuvent s'adapter au changement de l'environnement [Senechal et al.06]. Ils sont

adaptés aussi pour l'apprentissage et la mémorisation d'un grand volume d'information [Leroux.04].

Le principal inconvénient de réseaux de neurones, la difficulté de trouver la taille optimale du réseau. Ces réseaux admettent aussi une représentation complexe [Senechal et al.06]. Le réseau reste souvent une boîte noire qui fournit une réponse quand on lui présente une donnée, mais le réseau ne fournit pas des justifications faciles à interpréter, c'est-à-dire des paramètres difficiles à interpréter « physiquement ».

Dans la littérature, on trouve d'autres méthodes cognitives qui sont utilisées selon l'objectif et le type de personnalisation souhaité. Le tableau 3 résume ces méthodes :

Approches	Méthodes de personnalisation (cognitives)	Objectifs
[Liberman et al.2001]	TF-DF (Terme Frequency - Inverse Document Frequency)	Effectuer une recommandation contextuelle en se basant sur la fréquence d'apparition du document
[Ambrosimi et al.1997]	Réseaux sémantiques pondérés	Recherche d'information documentaire, effectuer un filtrage.
[Soltysiak et al.1998]	Réseaux Bayesiens et la logique floue	Prévoir les centres d'intérêts Des utilisateurs, par l'analyse des pages Web visitées par ces derniers
[Linden et al.1997]	Les réseaux de contraintes	Décrire les préférences des utilisateurs par rapport à un ensemble des solutions possibles
[Moghrobi et al.1998]	Les réseaux de neurones	Représentation des pages Web visitées
[Coyle et al.2002]	Raisonnement à partir des cas	Domaine de trafic aérien : Ce système fournit de l'information classée selon les préférences de l'utilisateur

Tableau 3- les approches basé sur différentes méthodes de personnalisation (cognitives)

Discussion et synthèse

Nous concluons que la plupart des méthodes utilisées pour la personnalisation de contenu sont des méthodes statistiques, cependant cette approche souffre d'un certain nombre de limites :

- Cette méthode ne peut évaluer que des documents sous forme textuelle. De ce fait, les nouveaux types de documents (vidéo, photographie, son, etc.) ne peuvent pas être gérés par ces systèmes [Shardanand et al.95]. De ce fait, la difficulté d'indexer des documents, multimédia ou non, est un goulet d'étranglement pour cette approche.
- L'incapacité à traiter d'autres critères de pertinence que les critères strictement thématiques posent également problème. Le filtrage des documents basé sur le contenu ne permet pas d'intégrer d'autres facteurs de pertinence que le facteur thématique.
- Cette méthode est incapable de refléter les changements d'intérêts des utilisateurs. L'effet dit « entonnoir » restreint le champ de vision des utilisateurs. En effet, le profil évolue toujours dans le sens d'une expression du besoin de plus en plus spécifique, qui ne laisse pas de place à des documents pourtant proches mais dont la description thématique diffère fortement.

Le paradigme du filtrage collaboratif apporte précisément une réponse à ces trois limites (difficulté d'indexation, incapacité à traiter d'autres critères, effet « entonnoir »).

En réponse à la difficulté d'indexation, la sélection ne s'appuie plus sur le contenu des documents, mais sur une sorte d'indexation parallèle qui traduit les opinions que les utilisateurs ont émises sur les documents. La difficulté d'indexer les documents selon leur contenu (premier problème) est ainsi contournée.

Par conséquent, cette « indexation parallèle » présente un autre avantage, celui de refléter non seulement les goûts des utilisateurs relativement aux thèmes, mais aussi d'autres facteurs de pertinence utiles aux utilisateurs.

En effet, lorsqu'un utilisateur émet une opinion positive sur un document, il affirme non seulement que le document traite bien d'un sujet qui l'intéresse, mais aussi que ce document est de bonne qualité, et qu'il lui convient à lui personnellement (public visé).

Ainsi le problème de l'incapacité à traiter d'autres critères est également résolu.

Enfin, l'effet « entonnoir » est aussi éliminé du fait que les documents entrants ne sont pas filtrés en fonction du contenu. Pour qu'un utilisateur reçoive un document, il suffit qu'un autre utilisateur de profil proche l'ait jugé intéressant, et cela quels que soient les termes qui indexent le contenu du document. L'utilisateur peut alors ouvrir son profil sur un nouveau thème en donnant simplement un retour de pertinence positif sur ce document.

II.2.2. Les méthodes sociales ou filtrage collaboratif :

Ce type de méthode a été proposé par [Goldberg et al.92]. Les systèmes basés sur une méthode sociale, appelés aussi des systèmes de filtrage collaboratif, utilisent le jugement ou le feedback d'un ensemble d'utilisateurs concernant un ensemble d'informations pour effectuer des recommandations.

Dans ce type de filtrages, on peut utiliser la similarité entre les individus pour déterminer si une information correspond à un individu donné [Goldberg et al.92].

Cette méthode permet la prédition d'un comportement d'un utilisateur par l'analyse des comportements des autres utilisateurs [Nugyen et al.06]. Elle ne se base pas sur le profil unique de l'utilisateur mais sur l'expérience des autres utilisateurs. C'est la méthode utilisée par le célèbre site amazon (<http://www.amazon.fr>) pour la recommandation de ses produits (livres et disques) aux internautes.

Le filtrage collaboratif peut être abordé selon deux manières celle qui concerne l'utilisateur et l'autre qui concerne les ressources (informations).

Dans la littérature on distingue deux types de filtrages collaboratifs [lumineau.03] et [Berru et al.03] :

- ✓ Le filtrage collaboratif basé « mémoire ».
- ✓ Le filtrage collaboratif basé « modèle ».

II.2.2.1.1. Le filtrage collaboratif basé « mémoire ».

Le filtrage collaboratif basé sur la mémoire utilise la totalité des profils utilisateurs afin de générer une nouvelle prédition.

Le filtrage consiste à recommander à l'utilisateur courant la solution la plus appréciée par les autres utilisateurs. La recommandation est effectuée en calculant la corrélation existante entre l'utilisateur courant et les utilisateurs présentant des appétences et des intérêts communs.

Cette corrélation peut être obtenue selon différentes techniques telles que :

Le filtrage collaboratif basé sur la mémoire utilise la totalité des profils utilisateurs afin de générer une nouvelle prédition. La prédition est calculée par la formule de [Resnick et al .94] :

L'évaluation prédit sur l'item j pour l'utilisateur actif u est une somme pondérée des évaluations des autres utilisateurs :

$$P_{u,j} = \bar{e_u} + k \sum_{i=1}^n w(u,i) (e_{i,j} - \bar{e_i})$$

Où n est le nombre d'utilisateurs dans la base de données qui ont un poids non nul,
 K est un facteur de normalisation tel que la somme des valeurs absolues des poids fait 1. Le
poids $w(u,i)$ est déterminé de façon variable, selon l'algorithme.

$$\bar{e}_i = \frac{1}{|I_i|} \sum_{j \in I_i} e_{i,j}$$

Où :

\bar{e}_i : C'est l'évaluation de la moyenne pour l'utilisateur i , avec I_i est l'ensemble des items
évalués par l'utilisateur i .

- Pour l'algorithme basé sur la corrélation (projet Group Lens [Res .94]), le poids est calculé comme la corrélation entre les utilisateurs a et i , comme suit :

$$W(a,i) = \frac{\sum_j (V_{a,j} - \bar{V}_a) * (V_{i,j} - \bar{V}_i)}{\sqrt{\sum_j (V_{a,j} - \bar{V}_a)^2 * \sum_j (V_{i,j} - \bar{V}_i)^2}}$$

- La similarité des vecteurs ou le poids est calculé comme un cosinus entre les vecteurs [Ber.98].

$$W(a,i) = \sum_j \frac{V_{a,j}}{\sqrt{\sum_{k \in I_a} V_{a,k}^2}} * \frac{V_{i,j}}{\sqrt{\sum_{k \in I_i} V_{i,k}^2}}$$

Où les termes du dénominateur servent à normaliser les évaluations pour que les utilisateurs qui ont évalué plus d'items ne soient pas favorisés.

A coté de ces techniques, on trouve aussi des simples algorithmes de clusterisation qui ont pour but de réduire le nombre d'utilisateurs considérés dans le calcul de prédiction.

L'algorithme le plus connu est ce lui du plus proche voisin (K-means) [Herlocker et al. 99]. Il est le plus courant. Il consiste à examiner les k centres de cluster dans l'espace de représentation utilisateur/ressources, d'assigner temporairement chaque utilisateur au cluster le plus approprié (dont il est le plus proche du centre), puis de réitérer le processus en modifiant les centres des clusters jusqu'à obtenir un état stable. C'est un algorithme en $O(k^2n)$ où k est le nombre de clusters et n le nombre d'utilisateurs.

Ensuite, pour le calcul de la prédiction, il suffit de ne considérer que les membres du cluster de l'utilisateur courant.

[Xie et al.07] et [Papamichail et al.07] utilisent les méthodes de plus proches voisins (K-means) pour la personnalisation des données dans le domaine de commerce électronique.

[Anli .06] utilise deux méthodes de personnalisation qui se basent sur le filtrage social basé sur la mémoire. Il a utilisé comme méthode de personnalisation celle qui se base sur un raisonnement automatique. Cette méthode effectue un filtrage collaboratif en se basant sur les préférences et en analysant le comportement des utilisateurs. Le principe consiste à prédire le comportement d'un utilisateur par une comparaison des autres utilisateurs ayant les mêmes préférences pour un contexte donné. [Anli.06] a utilisé aussi une méthode statistique utilisant une approche probabiliste de Bayes.

L'avantage du filtrage collaboratif basé sur la mémoire réside dans sa simplicité de mise en œuvre et sa prise en compte de l'évolutivité de profil de l'utilisateur, cependant cette méthode souffre de deux inconvénients majeurs : (i) d'une part, la forte complexité combinatoire empêche le passage à l'échelle pour un nombre important d'utilisateurs et de ressources et (ii) d'autre part, le faible nombre de ressources communément évalué par les utilisateurs, engendre des prédictions peu pertinentes voire totalement obsolètes.

II.2.2.2. Filtrage collaboratif basé « modèle » :

Le filtrage collaboratif basé sur un modèle définit un ou plusieurs modèles qui seront instanciés pour un utilisateur donné. Pour obtenir un modèle, plusieurs méthodes sont proposées dans la littérature dont on peut citer :

II.2.2.2.1. Approche probabiliste :

Dans un contexte probabiliste, la prédition est perçue comme l'évaluation espérée que l'on calcule comme une espérance mathématique sur le profil de l'utilisateur.

- Approche par classifieur :**

L'approche par classifieur nécessite la création de classe qui partitionne l'ensemble des valeurs possible des évaluations.

Dans le cas général, deux classes sont créées : une classe «Aime» et une autre «N'aime pas» et la matrice n x m (utilisateurs/ressources) est transformée en matrice booléenne.

Dans cette approche on distingue plusieurs sous classes telle que :

- ❖ Classifieur Bayésien :**

L'utilisation d'un simple classifieur naïf de Bayes [Miyahara et al.00] se base sur l'hypothèse d'indépendance des évaluations pour une classe donnée.

Cette approche permet de déterminer la classe C la plus probable à laquelle l'utilisateur courant va appartenir.

❖ **Arbre de décision et réseau bayésiens :**

Cette approche consiste à construire un réseau où chaque nœud correspond à une ressource et l'état du nœud correspond à une valeur possible d'évaluation.

La phase d'apprentissage consiste à rechercher les dépendances entre les ressources. Dans chaque nœud, il est donc possible de construire un arbre de décision en fonction des prédecesseurs qui sont les plus à même de prédire l'état du nœud. Ces arbres de décisions sont donc une simple représentation des probabilités conditionnelles disponible au niveau du nœud et établie lors de la phase d'apprentissage.

❖ **L'approche par Clustering :**

Dans l'optique de limiter le nombre d'utilisateurs considérés dans le calcul de la prédiction.

Le clustering hiérarchisé « RecTree » [Chee et al.01] cherche à fractionner l'ensemble des utilisateurs en clique. Ces cliques sont hiérarchisées de manière à maximiser les similarités entre les membres d'une même clique et de minimiser celles entre les membres de deux cliques différentes. Le principe se base sur l'algorithme des plus proches voisins.

II.2.2.2. Le principe du clustering répété (repeated clustering) [Ungaret al. 98]

Il repose sur la généralisation des groupes d'utilisateurs et des groupes de ressources.

En effet, une première clusterisation des utilisateurs en fonction des ressources aboutit à la création de classes d'utilisateurs. Le processus de clusterisation est réitéré sur les classes issues de la première clusterisation.

Cette approche a un pouvoir de généralisation important, du fait du regroupement. Elle souffre d'un risque de sur généralisation, ce qui a pour conséquence de regrouper des utilisateurs qui ont clairement des profils très différents qui ne doivent pas s'influencer dans le calcul de prédiction.

II.2.3. Le filtrage collaboratif « hybride »:

Le principe consiste à combiner un filtrage basé sur la mémoire et un filtrage basé sur un modèle afin de limiter les inconvénients des deux méthodes et de bénéficier de leurs avantages.

L'utilisation de ces approches hybrides permet d'améliorer la pertinence des résultats de ses systèmes de filtrage en palliant certaines limites [Balabanovic et al .97].

Dans le filtrage collaboratif « hybride », on distingue plusieurs approches telles que :

1. L'approche graphique ou « horting » [Aggarwal et al.99] repose sur la construction d'un graphe orienté dans lequel les utilisateurs sont représentés par les nœuds et l'influence entre les utilisateurs est représentée par les arcs.
2. L'approche basée sur le diagnostic de la personnalité [Pennock et al.00]
3. L'approche basée la clusterisation et la réduction de dimension [Goldberg et al.00].

L'approche de filtrage collaborative hybride apporte des réponses aux problèmes rencontrés dans le filtrage basé sur la mémoire et le filtrage basé sur un modèle. C'est en cela que ces deux approches se complètent avantageusement par l'hybridation.

Nous remarquons qu'il existe une diversité des méthodes de personnalisation. [Cinquini et al.02] proposent une classification des méthodes de personnalisation selon le contexte utilisé et le type de filtrage adapté, d'une part pour répondre à la demande d'un utilisateur et, d'autre part pour le suivi de leurs centres d'intérêts.

Le tableau 4 résume les méthodes de personnalisation adaptée :

		Méthodes sociales	Méthodes cognitives
Information de contexte	Utilisateur	Filtrage collaboratif	Filtrage social

Tableau 4- Classification des méthodes de personnalisation selon le contexte utilisé et le type de filtrage adaptée de [Cinquini et al. 02]

Synthèse et discussion

Nous remarquons qu'il existe une diversité des méthodes de personnalisation. Bien que le filtrage collaboratif essaye de combler les lacunes rencontrées dans la méthode cognitive, il présente quelques limites:

Les systèmes de filtrage collaboratif souffrent aussi tous du problème de démarrage à froid. Les nouveaux utilisateurs commencent avec un profil vide et doivent le constituer à partir de zéro. Même avec un profil de démarrage, une période d'apprentissage est toujours nécessaire avant que le profil ne reflète concrètement les préférences de l'utilisateur.

Pendant cette période le système ne peut pas filtrer efficacement pour le compte de l'utilisateur.

L'évaluation des systèmes de filtrage collaboratif présente des difficultés, dont certaines s'apparentent aux difficultés d'évaluation des systèmes de recherche d'information, et d'autres sont propres au filtrage collaboratif. Ainsi la question d'évaluation des systèmes de filtrage collaboratif constitue un sujet de recherche de première importance pour le succès de ce paradigme d'accès à l'information.

III. Les méthodes de calcul d'itinéraires :

L'information multimodale est difficile à mettre en œuvre pour de plusieurs raisons : organisationnelle, économique, juridique, techniques, etc [Zidi et al.05].

De plus, les sources d'information sont nombreuses, dispersées entre les différents opérateurs de transport. Les technologies de diffusion et de présentation qui leur sont attachées sont multiples et en constante évolution.

III.1. Base de données :

Dans la littérature, on distingue plusieurs types d'algorithmes pour calculer des itinéraires. Afin de calculer le meilleur itinéraire entre deux points origine - destination, le système doit disposer d'une bonne base de données décrivant son réseau de transport. Ce dernier inclut l'ensemble des lignes, des arrêts, des stations, des modes de transport, des itinéraires et des horaires de passage. L'algorithme de calcul d'itinéraires utilise cette base de données locale pour rechercher et trouver le ou les itinéraires répondants à une requête d'un utilisateur donné. Cette utilisation du calculateur d'itinéraires favorise la coordination de l'offre des différents réseaux (correspondances, fréquences d'horaires...) pour affiner les services et augmenter le nombre de voyageurs transportés.

Généralement, cette base de données est composée de divers types d'informations qui sont collectés de diverses sources. On peut représenter ses informations distribuées par un graphe multimodal distribué, figure 4.

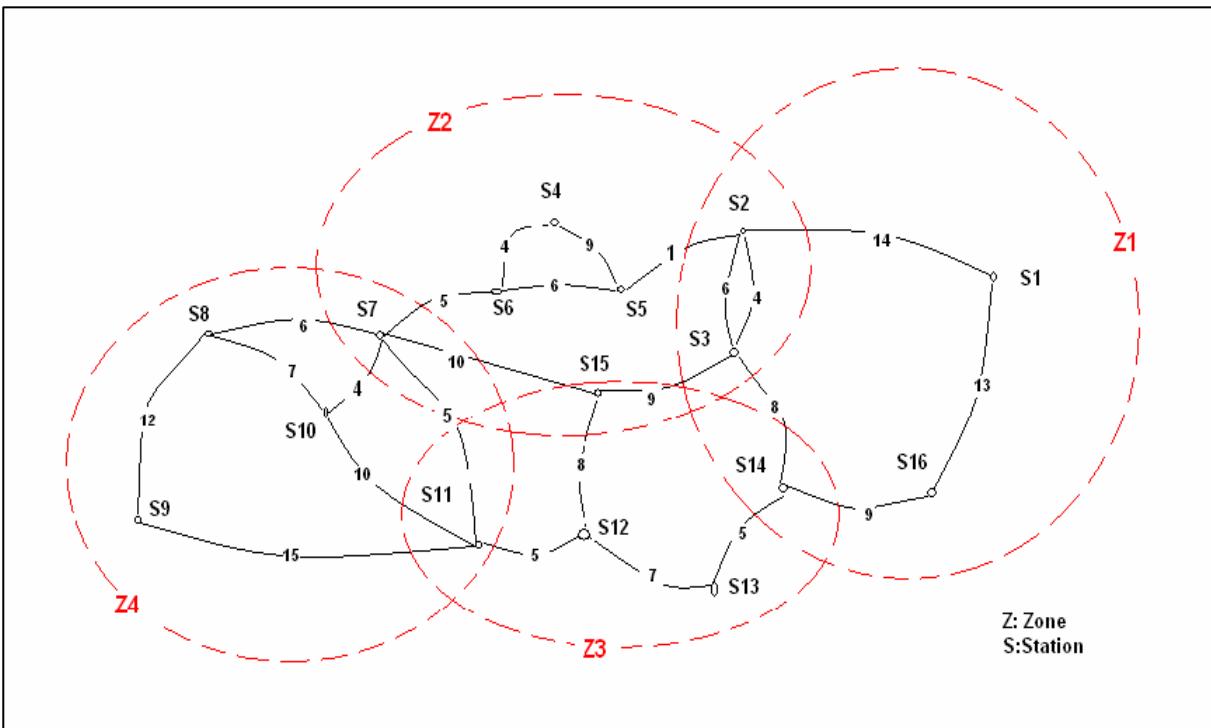


Figure 4- Un graphe multimodal et distribué

Ce graphe est divisé en plusieurs zones. Chaque source d'information est représentée par une zone. Chaque zone est composée d'un ensemble de nœuds qui représentent les stations. Le trajet parcouru par les moyens de transport entre ces deux nœuds ou stations successives est appelé arc. A ces arcs peuvent être associés des poids ; on parle alors d'arc pondérés. Il est caractérisé par un coût, une distance, un temps de parcours et un mode de transport utilisé.

III.2. Les algorithmes de calcul d'itinéraires :

Dans les systèmes d'information multimodale ou destiné au transport multimodal, nous distinguons généralement des algorithmes de calcul d'itinéraires monocritères qui dépendent d'un seul critère à optimiser mais aussi de multicritères qui dépendent de plusieurs critères à optimiser.

Dans la littérature il y a beaucoup de travaux qui ont été réalisé concernant ce domaine.

III.2.1. Algorithme de calcul d'itinéraire monocritère : (le critère de temps)

Pour calculer le **Plus Court Chemin** noté **PCC**, il faut calculer le chemin le plus rapide : si I est le chemin le plus rapide sur un graphe G , qui part d'un point de départ ou nœud de

départ A à un instant t, il doit arriver au plus tôt à une destination finale ou un nœud d'arrivée B.

Les algorithmes utilisés s'appliquent dans un graphe dynamique. La propriété de ce graphe est que le temps d'attente dans une station n'est pas fixe, il dépend du temps d'arrivée venant de la station précédente et du temps de départ allant à la station suivante. Par conséquent, la modélisation de ce graphe nécessite l'attribution de tous les horaires de départ et d'arrivée à chaque nœud et requiert aussi la connaissance à priori du poids de chaque arc. Ce poids représente le temps de parcours entre deux stations.

L'algorithme associé au plus court chemin ne prend que ces deux paramètres qui sont les horaires et le temps de parcours.

En effet, les récents travaux ont porté sur des systèmes de calcul d'itinéraires qui se basent sur un seul critère dans un réseau multimodal en réponse à une requête d'un voyageur.

L'objectif principal de travail de Kamoun [Kamoun.07] était de concevoir un système d'information voyageur coopératif (**SIMC**) à base d'architecture multi-agent, qui éviterait au voyageur de consulter plusieurs sites web différents pour planifier son voyage. Ce système permet de regrouper de manière significative les informations en interrogeant différentes sources d'information hétérogènes.

L'approche adoptée pour réaliser ce système considère que le chemin optimal est celui le plus rapide. Dans ce but, l'algorithme utilisé pour le calcul d'itinéraire entre une source et une destination est un algorithme monocritères. Kamoun [kamoun.07] optimise un seul critère qui est le critère temps.

Par exemple : si un chemin part d'une station de départ, doit arriver au plus tôt à une station d'arrivée b, le paramètre qui a été pris en considération est celui basé sur les différents horaires qui se trouvent dans chaque station. Le but est non seulement de trouver un chemin qui minimise le temps entre deux stations (ou deux nœuds) mais aussi il minimise le temps d'attente dans une station. Dans ce sens, on utilise un algorithme de plus court chemin dans un réseau dynamique.

III.2.2. Algorithme de calcul d'itinéraires multicritères :

L'objectif est d'aider l'utilisateur de transport en commun à optimiser ses déplacements selon plusieurs critères : le coût, le temps et le confort (le minimum de nombre de correspondance). Or, l'optimisation de ces critères est parfois difficile car l'optimisation de l'un va entraîner la dégradation des autres. Dans ce sens pour calculer le coût de chaque arc du graphe du réseau i [Zidi.06] utilise la fonction suivante :

$$F = \alpha.\mu(C_1) + \beta.\mu(C_2) + \delta.\mu(C_3) \quad (1)$$

Tels que C_1 , C_2 et C_3 représentent respectivement les critères d'optimisation : coût, temps de parcours et confort ; $\mu(C_i)$ est une fonction de normalisation qui permet de donner une valeur à chaque critère i dans l'intervalle $[0,1]$ et enfin α , β et δ sont les coefficients de pénalité qui sont calculés suivant le choix de l'utilisateur. En effet, ce dernier a le choix de favoriser un critère par rapport aux deux autres, ou bien de leur donner une importance égale.

Pour calculer le Plus Court Chemin note PCC et afin de trouver la solution optimale, nous pouvons appliquer une méthode déterministe (qui permet de donner la meilleure solution), comme l'algorithme de Dijkstra. Mais le problème est qu'elle est adaptée à un espace de recherche petit et non complexe lorsqu'un espace de recherche est grand, il est nécessaire d'appliquer une méthode stochastique (méthode qui permet de donner des solutions optimales) comme l'algorithme génétique. Ce dernier passe par les phases représentées ci-dessous, figure 5:

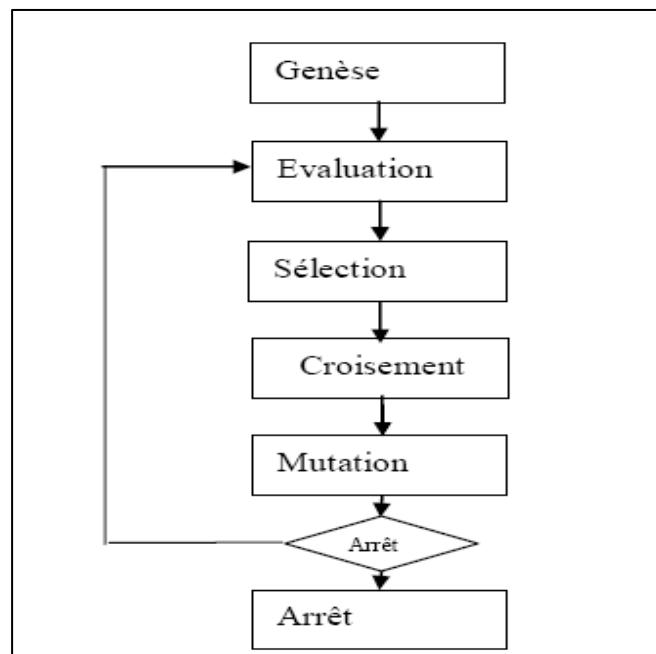


Figure 5-Les phases de l'algorithme génétique

Un itinéraire est représenté par un chromosome. Une première population (un ensemble d'itinéraires possibles) est générée dans la première phase. Chaque itinéraire est évalué dans la deuxième phase, les meilleurs itinéraires sont sélectionnés dans la phase de

'sélection'. Par la suite, chaque paire d'itinéraires est croisée afin de générer d'autres itinéraires enfants. Une mutation peut être appliquée sur des itinéraires, c'est-a-dire qu'une modification est faite sur le choix des stations des itinéraires. Ces phases sont répétées plusieurs fois jusqu'à l'obtention d'un ensemble de solutions optimisées.

En d'autres termes, les meilleurs itinéraires sont obtenus à la phase de convergence. La première population est obtenue par l'application d'un algorithme Dijkstra modifié.

Cet algorithme utilisé n'est autre que l'hybridation de l'algorithme Dijkstra modifié et de l'algorithme génétique afin d'avoir un ensemble de solutions optimisées. Parmi les travaux les plus récents on peut citer le système de calcul d'itinéraires proposé par [Zidi.06] : Le système SMAAD (**Système Multi-agents d'Aide au Déplacement**). Ces travaux sont axés sur le module de calcul d'itinéraires mais en tenant compte des différentes perturbations qui peuvent exister dans un réseau multimodal.

L'approche développée par [Zidi.06] s'appuie sur un **Système Multi Agents (SMA)** et vise à faciliter les déplacements entre une source et une destination des utilisateurs en mode normal et en mode dégradé de fonctionnement du réseau. Un déplacement en mode normal, peut se transformer en un déplacement incertain et dynamique dans un réseau perturbé. Par conséquent, l'approche préconisée était de faire appel à une méthode multicritère de recherche d'itinéraires qui consiste en une hybridation entre un algorithme de Dijkstra et un algorithme évolutionniste que nous l'avons présenté plus haut. Cette méthode prend en considération les différents critères d'un voyageur. Ces critères représentent les préférences souvent privilégiées par la majorité des voyageurs. L'étude confirme qu'ils sont essentiellement le coût, le temps de parcours et le confort (le minimum de nombre de correspondances). Ces préférences sont représentées par ordre de priorité d'un utilisateur.

Le tableau 5 illustre une comparaison entre ces deux systèmes de calcul d'itinéraires dont l'une repose sur un algorithme monocritère et l'autre sur un algorithme multicritère.

	Système (SIMC) [kamoun.07]	Système (SMAAD) [Zidi.06]
Critères	Temps	Temps, coût et le moins de

d'optimisation		correspondances
----------------	--	-----------------

Tableau 5-Comparaison entre les deux systèmes

Dans le paragraphe suivant nous allons présenter les principales méthodes d'évaluations utilisées pour évaluer les systèmes d'information personnalisée.

VI. Les méthodes d'évaluation des systèmes de personnalisations :

Dans la littérature l'évaluation des systèmes de personnalisation est presque négligée dans le domaine d'interaction Homme machine IHM.

L'évaluation d'un système de personnalisation consiste à tester et évaluer le service personnalisé obtenu. Différents types d'évaluations doivent être effectués.

VI.1.Les critères d'évaluation d'un système de personnalisation :

Avant l'évaluation d'un système d'information personnalisé (SIP), il est nécessaire de fixer des critères d'évaluation à l'avance afin de guider l'évaluateur dans l'étape d'évaluation des services de personnalisation proposé. [Anli .06] a préconisé 6 types de critères d'évaluation :

1- les évaluations fonctionnelles : il s'agit de vérifier l'adéquation des fonctionnalités du service définies dans la phase d'analyse et le service obtenu.

2- les évaluations de la personnalisation : consistent à évaluer la qualité de personnalisation réalisée. Il s'agit de vérifier si la personnalisation effectuée au niveau du service répond bien aux préférences de l'utilisateur. Il s'agit de vérifier si les informations présentées à l'utilisateur correspondent bien à celles que l'utilisateur voulait recevoir. Il s'agira donc de mesurer le degré de satisfaction de l'utilisateur par rapport à la personnalisation fournie.

3- les évaluations de performances : consistent à tester la capacité du système s'il peut fonctionner rapidement sous la pression d'une forte charge.

4- les évaluations de charges : consistent à établir le point de rupture ou la courbe de performance relative du système.

5- les évaluations techniques : Il s'agit d'évaluer la conformité du système par rapport aux normes et aux standards existants.

6-Les évaluations ergonomiques : se consacrent à la vérification de l'utilité et de l'utilisabilité [Grislin et al. 96] et [Bastien et al. 01] du service proposé.

De nombreuses méthodes dédiées aux évaluations ergonomiques sont proposées dans la littérature. Une ou plusieurs de ces méthodes devront être combinées et/ou adaptées.

Parmi ses méthodes on peut citer :

✓ **La méthode qui se base sur les grilles d'évaluation [Senach. 90] :** cette méthode se base sur le principe suivant :

Les utilisateurs répondent à certaines questions relatives à des critères ergonomiques après un test d'utilisation du système. Ces questions sont notées dans une grille.

✓ **La méthode qui se base sur les questionnaires d'utilisations [Moser et al. 71] :**

Les utilisateurs exécutent certaines tâches, et répondent à des questions liées à l'utilité et à l'utilisabilité du système pour la réalisation de ces tâches.

✓ **La méthode utilisée par [Trabelsi et al. 06] basée sur « les mouchards électroniques » :**

Cette méthode repose sur le principe suivant : Les évènements systèmes et les actions utilisateurs sont recueillis par un logiciel. Ces données sont analysées par l'ergonome pour vérifier l'utilisabilité et l'utilité du système.

✓ **La méthode d'analyse des mouvements oculaires [Abed et al. 90] :** L'activité oculaire des utilisateurs sur l'écran est enregistrée pendant l'utilisation du système. Les données recueillies sont analysées pour évaluer les caractéristiques de l'interface homme-machine.

✓ **Les méthodes d'inspection [Nielsen et al.94] :** L'utilisateur est assisté par un évaluateur pendant l'utilisation du système. L'ergonome analyse les données recueillies par l'évaluateur après le test d'utilisation du système.

✓ **Le recourt à l'expert [Hammond et al. 84] :** Le système est évalué par rapport aux recommandations et aux normes ergonomiques existantes. Ce travail est généralement effectué par un ergonome.

Une autre série des critères a été proposé par [Soui et al .07]. Ils fixent sept critères pour l'évaluation d'un système d'information personnalisé (SIP) :

- 1- **Adaptation de système aux préférences d'utilisateur :** Ce critère est associé aussi bien au contenu qu'au contenu de l'information. Les préférences de l'utilisateur se traduisent par ses goûts et ses désirs. Dans le domaine de transport, les systèmes d'information personnalisé (SIP) aident les usagers à planifier et à choisir leur itinéraires selon leurs préférences (le moins cher, le plus rapides, le moins de connexion, le plus confortables, le plus familiarisés, etc.).
- 2- **Adaptation de système aux centres d'intérêt de l'utilisateur :** Cette dimension de profil a pour objectif d'exprimer le centre d'intérêt ainsi que le domaine d'intérêt de l'utilisateur. Dans la majorité des systèmes interactifs, la dimension de centre d'intérêt est utilisée comme base de personnalisation de contenu.
- 3- **Adaptation du système à l'expérience de l'utilisateur :** La familiarité de l'utilisateur avec le système. Cette familiarité est acquise par la connaissance et à travers une longue période d'utilisation du système. Cette dimension relative à la personnalisation de contenu aussi bien que le contenu. [Bastien et al.01] considèrent que l'adaptation de système à l'expérience de l'utilisateur se traduit par une série de moyens qui permettent au système de respecter le niveau de l'expertise de l'utilisateur.
- 4- **Adaptation du système à la plateforme interactive :** adaptation de système à la plateforme (PDA, PC) avec laquelle il interagit. Ce critère est relatif au concept de plasticité utilisé dans le domaine d'interaction Homme machine (IHM). Il y a deux catégories de plateforme :
 - ✓ La catégorie Hardware : les supports avec laquelle l'utilisateur interagit le système (PC, PDA, téléphone mobile).
 - ✓ La catégorie software : qui regroupe des informations à propos de fonctionnement du système.
- 5- **Adaptation de système à la capacité physique de l'utilisateur :** le système doit être adaptable à toutes catégories d'utilisateurs (les handicapés, personnes âgées, etc.).
- 6- **Adaptation du système à son environnement d'utilisation :** le terme environnement est relatif à la paire (utilisateur, plateforme) qui peut influencer sur le processus d'utilisation. Par exemple la luminosité, le bruit, localisation géographique.
- 7- **Adaptation de système au comportement de l'utilisateur :** Ce critère est basé sur les buts que l'utilisateur désire atteindre ainsi que la procédure de navigation utilisée pour atteindre ces buts.

VI.2.Les approches d'évaluation des systèmes d'information personnalisée :

La plupart des méthodes d'évaluation nécessitent la réalisation d'un prototype afin de tester le système de personnalisation. Une approche a été proposée par [Soui et al.08]. Cette approche nécessite l'existence d'un système et l'intervention explicite de l'utilisateur.

Cette approche repose sur trois phases :

- **Phase 1** : c'est l'étape au cours de laquelle l'évaluateur choisit les tâches représentatives sur laquelle se base l'évaluation. Dans cette phase l'évaluateur prépare deux types de documents. Le premier c'est un questionnaire qui inclut des informations générales à propos de l'utilisateur. Le second est une carte d'indexation qui regroupe les différents critères d'évaluation ainsi que leurs définitions.
- **Phase 2** : c'est la phase de test pour le modèle .Elle est divisée en deux sous phases : de pré expérimentation et d'expérimentation.
- **Phase 3** : est une phase d'analyse. Dans cette phase l'évaluateur calcule un niveau de satisfaction N_s pour chaque utilisateur puis il translate le niveau de personnalisation N_p de système par la formule suivante :

$$N_p = \frac{\sum_{K=1}^n W_K \times I_a \times N_k}{\sum_{k=1}^n W_K \times I_a}$$

Avec

n : le nombre de critère avec lequel l'utilisateur est concerné.

W_k : le poids accordé selon l'importance donnée à ce critère par l'utilisateur.

I_a : l'index d'activation de critère.

1 : si l'utilisateur U est concerner par le critère K.

0 : si non

N_k : la note attribuée a l'utilisateur pour le critère K.

Cependant Anli [Anli.06] à proposer dans sa thèse une autre approche, intitulée « PerSyst », de développement de système d'information personnalisé (SIP) qui distingue le système d'information de la personnalisation elle-même et intègre une phase d'évaluation.

L'architecture de modèle PerSyst est à base d'agent logiciel. Le schéma suivant résume les phases de développement du système (voir figure :6)

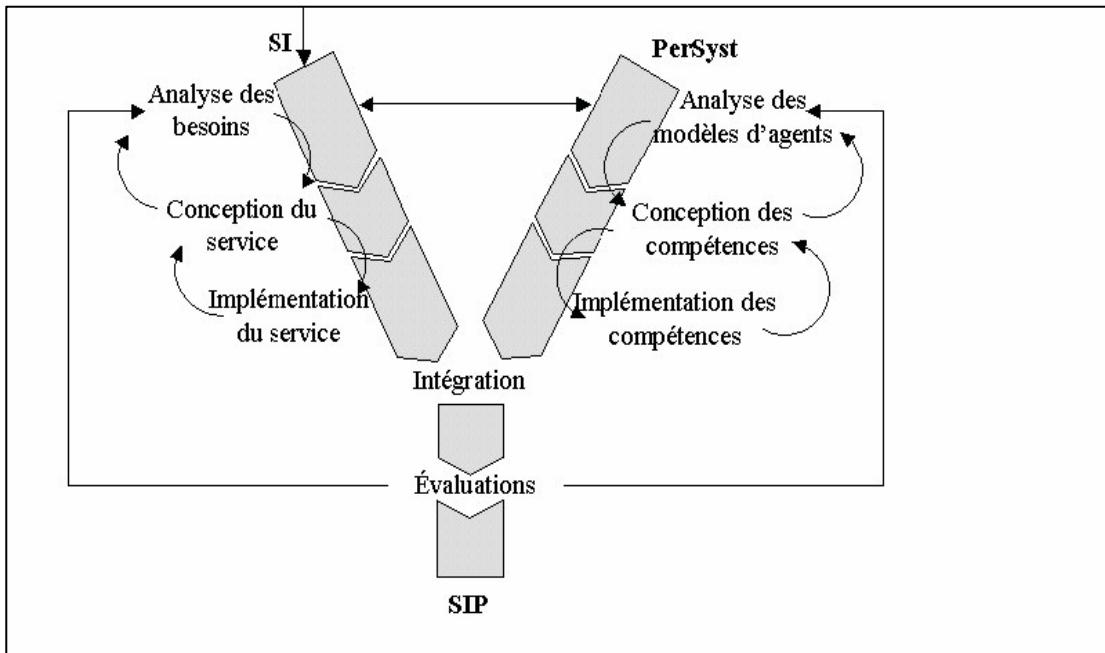


Figure 6-Méthodologie de développement d'SIP [Anli.06]

Nous remarquons que [Anli.06] a intégré une étape d'évaluation après la phase d'intégration et de réalisation du système d'information personnalisé (SIP).

Après la réalisation de (SIP) des évaluations sont effectuées. Ces évaluations peuvent être d'ordre qualitatif (touche la qualité de la personnalisation réalisée), d'ordre quantitatif (qui touche la performance globale du (SIP)), d'ordre ergonomique ou autre. Le tableau 6 regroupe les différentes méthodes d'évaluation et leurs critères :

Méthode de d'évaluation	Principe	Critères d'évaluation
Méthode d'évaluation proposée par [Anli.06]	Cette approche consiste à intégré une étape d'évaluation après la phase d'intégration et de	<ul style="list-style-type: none"> ■ Critère d'évaluation de fonctionnalité du service ■ Critère d'évaluation de la personnalisation ■ Critère d'évaluation de performances ■ Critère d'évaluation de

	réalisation du système d'information personnalisé (SIP).	charge. + Critère d'évaluation technique. + Critère d'évaluation ergonomique.
Méthode d'évaluation proposée par [Soui et al.08]	<p>Cette approche consiste à évalué un prototype. L'évaluation se déroule en trois phases :</p> <p>Phase 1 : c'est l'étape au cours de la quelle l'évaluateur choisie les tâches représentatives sur la quelle se base l'évaluation.</p> <p>Phase2 : elle est divisée en deux sous phases. La 1^{ere} de pré expérimentation et la seconde d'expérimentation.</p> <p>Phase3 :C'est une phase d'analyse.</p>	<ul style="list-style-type: none"> + Adaptation système aux préférences d'utilisateur + Adaptation de système aux centres d'intérêt de l'utilisateur. + Adaptation du système à l'expérience de l'utilisateur + Adaptation du système a la plateforme interactive + Adaptation de système à la capacité physique de l'utilisateur + Adaptation du système à son environnement d'utilisation + Adaptation de système au comportement de l'utilisateur

Tableau 6-les différentes méthodes d'évaluation et leurs critères

Synthèses et discussions

Malgré l'existence de différentes approches et méthodes pour la conception d'un système d'information personnalisé, l'évaluation de ses systèmes souffre d'une insuffisance des méthodes et des approches.

Aujourd'hui, les utilisateurs se trouvent face à des centaines d'applications interactives développées qui ne répondent pas à leurs besoins : tâches, habitudes, préférences, etc.

D'où la nécessité d'effectuer une évaluation du système afin de concevoir et réaliser des systèmes utiles (adéquation aux besoins de l'utilisateur) et utilisables (adéquation aux capacités de l'utilisateur).

Conclusion

Dans ce chapitre, nous avons présenté certaines méthodes de personnalisations.

Deux types de méthodes sont souvent utilisés: la méthode cognitive et la méthode sociale ou à filtrage collaboratif.

Dans la deuxième partie, nous avons présenté les différents algorithmes et méthodes utilisés pour le calcul d'itinéraires dans un réseau de transport multimodal.

Trouver l'information pertinente ne se limite pas à l'optimisation de l'algorithme du calcul d'itinéraires, mais aussi, pouvoir l'adapter à tous les besoins et les préférences d'un voyageur. Enfin, nous avons présenté les méthodes d'évaluation existante pour évaluer les systèmes d'information personnalisée

En effet, adapter une information à un utilisateur, c'est la rendre personnalisée. C'est pour cela que nous allons présenter dans le chapitre suivant une méthode de recherche des itinéraires basée sur les préférences de l'utilisateur.

Chapitre 3 : Recherche des itinéraires basée sur les préférences de l'utilisateur

Introduction :

Dans le chapitre précédent, nous avons vu les diverses méthodes de personnalisation ainsi que les diverses techniques de calcul d'itinéraires.

Dans ce qui suit, nous présentons deux méthodes de recherche des itinéraires basée sur les préférences de l'utilisateur.

II. Personnalisation de l'itinéraire selon les préférences de l'utilisateur :

II.1. les critères de préférences dans le domaine du transport multimodales :

La détermination des besoins des utilisateurs dans le transport multimodal a fait l'objet de plusieurs études [Lecompte et al. 00], [Mouloudi.07] et [Grotenhuis et al.07]. Ces besoins concernaient essentiellement : l'obtention d'une information en temps réel, l'accessibilité à travers différents supports et la recherche des solutions multimodales personnalisées.

Souvent, l'usager joue un rôle primordial dans la sélection et le choix d'un itinéraire. Son choix d'itinéraire ne se fait pas d'une manière arbitraire mais au contraire en se basant sur plusieurs critères de sélection.

Ces critères peuvent être d'ordre :

- économique, telle que coût (par exemple l'itinéraire le moins cher),
- de correspondance qui minimise le nombre de correspondance.
- physique ; c'est-à-dire le moins de marche à pied [Grotenhuis et al.07],
- temporelle qui fournit l'itinéraire le plus rapide.

Les critères peuvent être aussi d'ordres cognitifs [Grotenhuis et al.07] par exemple sentiment de sécurité, le moins d'effort mental concernant la recherche d'information, le moins d'effort mental concernant le changement entre les modes de transport.

Les préférences de l'utilisateur varient en fonction des catégories auxquelles l'usager appartient. Selon les études réalisées par le CERTU [Rapport.2006], il existe trois classes d'utilisateurs :

1.Les utilisateurs occasionnels : ce type d'utilisateurs se déplacent peu mais ils recherchent une information détaillée.

2.Les utilisateurs réguliers : ils sont habitués à un itinéraire. Ils restent fidèles à cet itinéraire malgré l'existence d'autres solutions. L'utilisateur régulier utilise quotidiennement soit un mode de transport soit des modes combinés. Ce type d'utilisateur a une bonne connaissance des informations liées à son déplacement.

3.Les utilisateurs mixtes : ils sont très mobiles. Ils utilisent des modes de transport combinés, ainsi que leurs propres voitures dans le cas échéant.

Les préférences de l'utilisateur varient aussi selon le type de déplacement à effectuer. Les motifs de déplacement varient d'une catégorie sociale à une autre comme le déplacement à usage professionnel ou de loisir.

II.2. Proposition d'un itinéraire selon les préférences de l'utilisateur :

Dans ce travail, nous partons de l'hypothèse qu'un utilisateur se déplace entre deux points A et B et que l'objectif est de fournir à l'utilisateur le meilleur itinéraire qui satisfait le maximum possible ses préférences (coût, temps, correspondance, distance, etc.), comme le résume le tableau suivant :

Paramètre	Notre problème
Objectif	Quel est l'itinéraire le plus adapté à l'utilisateur ?
Critères	Economique (cout) ; l'itinéraire le moins cher Temps : le plus rapide, le moins de changement, le moins compliquer... Psychologique : sécurité (bagages et personnes), confort (nombre de place assise), loisir...
solution	L'itinéraire préféré

II.2.1. Collecte d'information :

Afin d'avoir une information personnalisée et de qualité, le système d'information nécessite la collection d'informations sur ses utilisateurs afin de constituer leurs profils. Ces informations peuvent être directement fournies par lui d'une manière explicite c'est-à-dire par le biais d'un questionnaire ou d'une manière implicite. En effet, la collecte implicite d'information sur l'utilisateur, appelée aussi « **tracking** » ou « **profiling** », permet généralement de recueillir une grande quantité d'information sur l'utilisateur sans ne lui faire aucune demande explicite. Les principales sources de collecte d'information ces sont les différents sites Web visités par l'utilisateur ou des documents utilisés.

Dans notre travail, la collecte d'information (les préférences) se base sur le choix de l'utilisateur de son itinéraire parmi un ensemble proposé par le système. L'objectif est d'analyser le profil de l'utilisateur afin de mettre à jour les données pondérées de ses

préférences. Dans le domaine de transport le choix d'un itinéraire par un utilisateur se base sur un certain nombre de critères de sélection, comme nous l'avons mentionné plus haut.

Considérant une requête R, et un ensemble d'itinéraires (solutions) S_i pour cette requête.

avec $i = (1, \dots, n)$ ou n : nombre de solution possible pour la requête R.

Ensuite, nous affectons un rang pour chaque solution selon le critère désiré par l'utilisateur. Nous pouvons dire qu'un itinéraire ou (solution) est défini par un quadruplet de valeur. Ses valeurs correspondent au rang de cette solution vis à vis de ses critères.

$S_i = \{r_1, r_2, r_3, r_4\}$ Avec : r_i est les rangs de chaque critères ;

$r_i = \text{rang}(C_i)$ avec $1 < r_i < n$

Afin d'obtenir la meilleure solution (itinéraire), on considère deux valeurs r_{ki} et r_{kj} d'un critère C_k de deux solutions respectives S_i et S_j . La solution S_i est plus adaptée (ou préférée) que celle de la solution S_j si et seulement si $r_{ki} < r_{kj}$.

Ainsi, un poids est affecté à chaque préférence de l'utilisateur. Ce poids définit l'importance d'un critère vis-à-vis de l'utilisateur (Voir le tableau 7) :

	C₁ : Time	C₂ : Dist	C₃ : Cost	C₄ : Wait
itinéraire1	1	2	3	1
itinéraire2	2	1	2	4
itinéraire3	4	3	1	2
itinéraire4	3	4	4	3

Tableau 7-Rangs des critères par rapport aux solutions d'une requête de recherche d'itinéraires

II.2.2.Représentation de profil utilisateur :

Un système d'information personnalisée nécessite d'abord de disposer d'un modèle utilisateur. Ce modèle devrait contenir les informations collectées concernant l'utilisateur.

Dans le domaine de l'information multimodale, le profil de l'utilisateur est réparti selon trois catégories [Petit rosé.03] et [Anli.06] :

1. **Les données statiques** : elles représentent l'ensemble des connaissances qui décrivent directement l'utilisateur et qui peuvent contribuer à la personnalisation. Les informations

personnelles (nom, courriel, âge, sexe, etc.) et les informations de localisation «statique» (adresse de domicile, lieu de travail, etc.) et des informations générales comme la profession de l'utilisateur, ses hobbies, etc.

2. **Les données pondérées** : sont les données susceptibles d'évoluer régulièrement. Ce sont les informations représentant les préférences de l'utilisateur. Ces données pondérées constituent des critères permettant de choisir des réponses adaptées à l'utilisateur.
3. **L'historique** : représente une trace des interactions de l'utilisateur avec le système d'information personnalisé.

Cela peut servir de base de connaissance pour la mise à jour des données statiques et des données pondérées.

Cette décomposition rejoint l'approche de [Bouzagoub et al. 04] et [Kostadinov .08] qui vise lui aussi de catégoriser les informations de l'utilisateur.

Notre approche adopte à la fois une représentation **multidimensionnelle** [Kostadinov.08] [Bouzagoub et al. 04] et **ensembliste**.

Le profil de l'utilisateur est multidimensionnel car il regroupe plusieurs dimensions .En effet, le profil utilisateur est un élément clé dans le processus de recherche, la modélisation de l'utilisateur doit pouvoir capturer toutes les dimensions qui représentent l'utilisateur. Dans notre cas le profil regroupe trois dimensions :

- Les données statiques.
- Les données pondérées.
- L'historique.

Dans notre approche, on a choisi la technique d'apprentissage vectoriel, elle dépend essentiellement du choix de la représentation du profil utilisateur. Nous proposons la représentation la plus simple qui est la représentation ensembliste.

Les préférences d'un profil utilisateur sont mises sous forme d'un vecteur de termes pondérés: le poids d'un terme définit son importance par rapport à l'utilisateur. La technique d'apprentissage employée est la technique vectorielle.

Pour la modélisation de profil de l'utilisateur nous adaptons le modèle proposé par [Petit-Rosé.03] et utilisé aussi par [Anli.06].

Soit U : l'ensemble des utilisateurs du système

u : un utilisateur

Le profil de l'utilisateur représenté :

$$\mathbf{M_u} = \mathbf{S_u} \cup \mathbf{P_u} \cup \mathbf{H_u}$$

Avec

$\mathbf{M_u}$: modèle utilisateur.

$\mathbf{S_u}$: l'ensemble de données statistique.

$\mathbf{P_u}$: l'ensemble de données pondérées.

$\mathbf{H_u}$: représente l'historique de l'utilisateur.

Dans notre approche, nous nous intéressons essentiellement aux données pondérées et à l'historique. En effet, l'historique donne une trace des interactions de l'utilisateur avec le système d'information personnalisée. Il représente l'ensemble des requêtes demandées et les solutions choisies par l'utilisateur dans le passé. Cela pourrait servir comme base d'exemples pour l'apprentissage automatique des informations concernant l'utilisateur pour mettre à jour les données pondérées (les préférences). Il permet, par exemple, la détermination des préférences de l'utilisateur par rapport aux modes de transports. Par exemple, lorsqu'un utilisateur désire choisir un itinéraire parmi d'autres, il ne le fait pas au hasard. Son choix sera basé sur des critères ou préférences.

Grâce au modèle de l'utilisateur, le système dispose de connaissances concernant les utilisateurs à partir desquelles il peut émettre des hypothèses pour des préférences ou des choix d'un nouvel utilisateur. L'historique est représenté selon un ensemble de couples de couples de valeurs

$$\mathbf{H_u} = \{(\mathbf{R}_1, \mathbf{S}_{i1}), (\mathbf{R}_2, \mathbf{S}_{i2}), \dots (\mathbf{R}_i, \mathbf{S}_{ii}), \dots (\mathbf{R}_n, \mathbf{S}_{in})\} \text{ avec } i \in [1 \dots n]$$

n : nombre de requêtes qui satisfont l'utilisateur dans le système.

R_i : représente une requête utilisateur qui définit à la fois une origine et une destination d'un voyage et aussi les critères de préférences pour favoriser un itinéraire parmi plusieurs.

S_i : représente la solution en réponse à la requête R_i . Cette solution est représentée sous forme d'un itinéraire.

III. Les méthodes de personnalisation proposées :

III.1.La méthode 1 : La personnalisation basée sur le degré de pertinence :

Le rapide développement des technologies de l'information durant ces dernières années a eu pour conséquence une prolifération de sources d'informations hétérogènes. Il

devient de plus en plus difficile pour les utilisateurs de retrouver précisément ce qu'ils recherchent dans cette masse de données. En raison de cette augmentation constante du volume d'informations, nous arrivons à une situation paradoxale : jamais il n'y a eu autant d'informations disponibles, mais trouver dans cette accumulation ce que l'on recherche précisément, devient de plus en plus ardu. Le problème n'est plus tant la disponibilité de l'information mais la capacité de sélection de l'information répondant aux besoins précis d'un utilisateur.

Des travaux de personnalisation, existent dans le domaine de la recherche d'information documentaire, ayant pour objectif de délivrer à l'utilisateur une information pertinente.

En effet, les utilisateurs ont des préférences différentes, en conséquence des perceptions différentes de la notion de pertinence. Par ailleurs, plus une information est appropriée aux préférences et aux centres d'intérêts, plus elle est pertinente.

Nous trouvons dans les travaux de Tamine [Tamine et al.07] une méthode qui détermine les documents les plus pertinents par rapport à un utilisateur donné dans chaque session de recherche. Cette méthode prend en considération le profil utilisateur représenté par deux dimensions corrélées :

(i) l'historique des interactions de l'utilisateur avec le système de recherche d'information et (ii) ses centres d'intérêts.

L'historique est l'ensemble de documents déjà visité par l'utilisateur et jugée pertinents par lui lors des sessions de recherche passé.

La session de recherche S^s est représentée par une matrice Document-Terme $D^s *$

D^s l'ensemble des documents pertinents pour l'utilisateur durant une session.

Chaque ligne de la matrice S^s représente un document $d \in D^s$, chaque colonne représente un terme $t \in T^s$ où T^s est l'ensemble des termes qui indexent les documents D^s (T^s est une partie de l'ensemble des termes représentatifs des documents préalablement jugés pertinents noté $T(R_u^s)$). R_u l'ensemble des documents déjà visités et jugés pertinents par l'utilisateur lors des sessions de recherche passées.

Afin de calculer la pertinence d'un document, on calcule pour chaque terme t dans un document d à l'instant s , un coefficient dégrée de pertinence $DPT^s(t, d)$ qui traduit la pertinence relative d'un terme compte tenu des jugements de pertinence qu'il a émis et qui sont supposés êtres des indicateurs de son centre d'intérêt courant.

L'expression de ce coefficient est fondée sur l'hypothèse qu'un terme est d'autant plus important pour l'utilisateur qu'il co-occure avec les termes qui lui sont familiers en ce sens qu'ils sont présents dans des documents déjà jugés. Les dépendances entre termes associés à des documents préalablement jugés sont vues comme des règles d'association.

La pertinence d'un terme t dans un document d est calculée par le coefficient Degré de la Pertinence $DPT^S(t, d)$ qui est défini comme suit :

$$DPT^S(t, d) = \frac{W_{td}}{l(d)} \sum_{t' \neq t, t' \in T(R_u^S)} cooc(t, t') \quad (1)$$

Avec :

- W_{td} est le poids du terme t dans le document d calculé selon le schéma classique $tf * idf$ [Singhal .97] représenté comme suit :

$$W_{td} = Tf * IDf = Tf * \log\left(\frac{N}{df}\right)$$

- TF (Terme Frequency) représente le nombre d'occurrences du terme t dans le document d : un terme t est d'autant plus important qu'il est présent dans le document d .
- IDF (Inverse of Document Frequency) mesure l'importance d'un terme t dans toute la collection.

L'idée sous-jacente est que les termes qui apparaissent dans peu de documents de la collection sont plus représentatifs du contenu de ces documents que ceux qui apparaissent dans tous les documents de la collection. Cette mesure est exprimée selon la déclinaison suivante :

$$IDF = \log\left(\frac{N}{df}\right)$$

Où df est la proportion de documents contenant le terme t et N le nombre total de documents dans la collection.

- $cooc(t, t') = \frac{n}{m}$, n est le nombre de documents contenus dans l'ensemble R_u et contenant les termes t et t' , m est le nombre de documents contenus dans l'ensemble D^S contenant le terme t .

III.1.2.Degré de pertinence dans le domaine d'information voyageur :

Notre objectif est de proposer à l'utilisateur, l'itinéraire le plus préféré. Le système doit tout d'abord déterminer les meilleurs itinéraires entre une source et une destination

données pour une requête d'un utilisateur, en tenant compte de ses préférences (plus économique, plus rapide, le moins de correspondance, le mois de marche à pieds, etc).

Sur la base de l'analyse de la formule précédente de calcul de la pertinence d'un document, nous empruntons son application dans le domaine de l'information voyageur et ceci dans le but de déterminer le degré de pertinence d'un itinéraire. En effet, un document peut être assimilé à un itinéraire, et le terme à un critère ou préférence. Nous pouvons faire ultérieurement une analogie entre ces deux domaines.

Cette analogie est représentée dans le tableau suivant :

Importance du terme t dans le document d	Importance du critère i pour l'itinéraire j
Document d	Itinéraire I
Terme	Critère ou préférence : C _i
W _{dt}	W _{Ci}
TF	Poids de critère C _i
IDF	-
L(d)	Nb de critères
cooc(t, t')	Nb-It-choisie(j) / nb-Requête-Total

Tableau 8-Analogie de l'importance d'un terme t dans un document d avec l'importance d'un critère i pour un itinéraire J

La formule générale permettant de calculer le degré de pertinence d'un critère i dans un itinéraire j est:

$$DPCr(i) = \frac{Wc_i}{Nb - Critères} * \frac{Nb - It - choisie(j)}{nb - Requête - Total}$$

- ✚ **Wc_i**: est le poids de critère i dans l'itinéraire j qui est calculer selon le schéma classique tf*idf.

Pour le calcul du poids du critère i nous procédons de la manière suivante :

Nous affectons un poids e_{ij} à différentes solutions S_j sur la base du critère C_i par la formule suivante :

$$e_{ij} = \frac{1}{rang(C_i)} * \delta \quad \text{avec } \delta : \text{une constante de normalisation.}$$

- ✚ **Nb-It-choisie(j) / nb-Requête-Total** : représente l'importance d'un critère pour un utilisateur : c'est-à-dire le nombre de fois que cet utilisateur avait choisi un itinéraire

qui présente une importance pour le critère i par le nombre total des voyages du même utilisateur.

Une fois on a calculé le degré de pertinence pour chaque critère i dans un itinéraire j.

Afin de déterminer le degré de pertinence d'un itinéraire j, nous additionnons la somme de degrés de pertinence de tous les critères comme la montre la formule suivante :

$$DP_{IT(j)} = \sum_n DPCr(i)$$

Pour la normalisation DP'_{ITj} est calculé comme suit :

$$DP'_{ITj} = \frac{DP_{ITj}}{\max(DP_{ITj})}$$

Ensuite on fait un tri par ordre croissant de pertinence : l'itinéraire le plus pertinent est celui qui possède le degré de pertinence le plus élevé.

III.2.La méthode 2 : La personnalisation en se basant sur le degré la similarité entre un profil de l'utilisateur et document

Le système crée une représentation d'un document pour calculer un score de correspondance ou de similarité entre cette représentation et celle du profil utilisateur [Zmerli.04]. Ce score peut traduire aussi le degré de pertinence du système.

Afin, de mesurer le degré de ressemblance entre un document et le profil de l'utilisateur. Le document et le profil utilisateur sont représentés par le modèle vectoriel [Salton.71].

En effet, le modèle vectoriel repose sur deux hypothèses :

1 Une représentation similaire des documents et des profils utilisateurs par des vecteurs de même type.

2 Les documents les plus pertinents sont ceux qui sont les plus proches au profil utilisateur (renfermant les mêmes termes).

Dans les travaux de [Zemerli.04] chaque vecteur représentant un document est composé par un ensemble de termes. Le profil utilisateur admet aussi une représentation vectorielle formée par l'ensemble des données qui le constituent.

Une fois le document et le profil de l'utilisateur admettent à la fois la même représentation vectorielle, il est possible de comparer directement un degré de ressemblance.

Ainsi, la pertinence des documents s'en déduit suivant la règle suivante : plus un document est proche d'un profil utilisateur (plus il lui ressemble), plus il est pertinent.

Pour calculer le degré de similarité entre deux documents dans le domaine de la recherche documentaire la formule du **Cosine** est utilisée. Cette formule permet de mesurer la similarité entre deux documents. Dans notre cas, pour le filtrage collaboratif, elle est utilisée afin de mesurer le degré de similarité entre deux utilisateurs. Ceci permet de mettre à jour le profil des nouveaux utilisateurs en les comparants, aux utilisateurs antérieurs du système [Xie et al.07].

La formule du vecteur de similarité [Ber.98] :

$$w(a,i) = \sum_j \frac{e_{a,j}}{\sqrt{\sum_{k \in I_a} e_{a,k}^2}} \frac{e_{i,j}}{\sqrt{\sum_{k \in I_i} e_{i,k}^2}}$$

K : coefficient de normalisation.

$w(a,i)$: le coefficient de similarité entre l'utilisateur actif a et l'utilisateur antérieur i

$e_{a,j}$: le poids attribué par l'utilisateur a au document j.

$e_{i,j}$: le poids attribué par l'utilisateur i au document j.

$e_{a,k}$: le poids attribué par l'utilisateur a est normalisé par le coefficient de normalisation k.

$e_{i,k}$: le poids attribué par l'utilisateur i est normalisé par le coefficient de normalisation k.

III.2.1 Degré de similarité dans le domaine de l'information voyageur :

Notre objectif est de proposer à l'utilisateur l'itinéraire qui correspond le plus à ses préférences telle que :

1. plus la durée est courte et plus la solution (l'itinéraire) est préférée:

$$\forall (a,b) \in I^2, WTime(a) < WTime(b) \Rightarrow p(a) > p(b),$$

2. plus la distance est courte, plus la solution (l'itinéraire) est préférée:

$$\forall (a,b) \in I^2, WDist(a) < WDist(b) \Rightarrow p(a) > p(b),$$

3. plus le coût est faible, plus la solution (l'itinéraire) est préférée:

$$\forall (a,b) \in I^2, WCOST(a) < WCOST(b) \Rightarrow p(a) > p(b),$$

4. plus la durée d'attente est courte et plus la solution (l'itinéraire) est préférée:

$$\forall (a,b) \in I^2, WWait(a) < WWait(b) \Rightarrow p(a) > p(b)$$

Le système doit déterminer l'itinéraire le plus préférable à l'utilisateur en se basant sur les critères de préférences déjà fixés (le moins cher, le plus rapides, avec le minimum de

correspondance, le minimum de distance, le minimum de temps d'attente entre les correspondances, de marche à pieds, etc.) dans la requête lancée par l'utilisateur.

La solution délivrée par le système va être présentée sous la forme d'un itinéraire qui satisfait les préférences de l'utilisateur parmi un ensemble d'itinéraires disponibles.

Afin de faciliter l'analogie entre les deux domaines il faut que le profil de l'utilisateur et l'itinéraire admettent la même représentation vectorielle.

Un itinéraire peut être modélisé de la façon suivante :

Exemple :

$$\begin{bmatrix} t \\ d \\ m \\ p \\ c \\ a \end{bmatrix} \quad \begin{bmatrix} W_{Ct} \\ W_{Cd} \\ W_{Cm} \\ W_{Cp} \\ W_{Cc} \\ W_{Ca} \end{bmatrix}$$

Chaque itinéraire est représenté par un vecteur de terme pondéré.

t : temps

d : distance

m : moyen de transport (bus, train, voiture)

p : prix

c : correspondance

a : temps d'attente dans la correspondance

à chaque critère de préférences nous allons affecté un poids W_{Ci}

W_{Ct} : poids pour le critère de préférence temps

W_{Cd} : poids pour le critère de préférence distance

W_{Cm} : poids pour le critère de préférence moyen de transport (bus, train.. ;)

W_{Cp} : poids pour le critère de préférence prix

W_{Cc} : poids pour le critère de préférence le moins de correspondance

W_{Ca} : poids pour le critère de préférence le minimum de temps d'attente dans la correspondance.

Pour calculer le poids ou le coefficient d'importance du critère de préférence (des termes pondérés) par rapport à un itinéraire (solution) choisie, nous allons affecter des notes e_{ij} aux différents solutions S_j suivant le critère C_i par la formule suivante :

$$e_{ij} = note(C_i) = \frac{1}{rang(C_i)} * \delta \quad \text{Avec } \delta : \text{une constante de normalisation.}$$

Soit w_{ui} le poids du critère i pour l'utilisateur u . Notons e_i la note du critère i pour la solution choisie par l'utilisateur u . La nouvelle note attribuée au critère i pour l'utilisateur u sera donc définie par :

$$w_{ui} = moy(e_i)$$

$$\text{Avec moyenne } (e_i) = \frac{\sum_{i=1}^n e_i}{NB - requete - total}$$

Pour calculer le coefficient de similarité entre le profil de l'utilisateur et un itinéraire revient à comparer deux vecteurs associés au profil utilisateur et à l'itinéraire.

Il revient alors à comparer le degré de ressemblance entre l'itinéraire et le profil de l'utilisateur. Ainsi, la pertinence de choix d'itinéraire se déduit selon la règle suivante : Un itinéraire est pertinent s'il est très proche du profil de l'utilisateur.

En effet puisque l'itinéraire et le profil de l'utilisateur admettent la même représentation : modèle vectoriel. Nous pouvons calculer un coefficient de similarité en se basant sur une fonction qui permet de mesurer la colinéarité de ces deux vecteurs itinéraires profils. La formule utilisée pour calculer le degré de similarité est la formule Cosine :

$$SIM(U, I) = \sum_{i=1}^n \frac{W_{ui}}{\sqrt{\sum_{k \in la} W_{u,K}^2}} * \frac{W_{ij}}{\sqrt{\sum_{k \in li} W_{i,K}^2}}$$

U : utilisateur

I : itinéraire

n : le nombre des itinéraires possibles entre une origine et un destination.

W_{uj} : le poids accordé au critère i par l'utilisateur u .

W_{ij} : le poids accordés pour un critère i dans un itinéraire j .

$W_{U,k}$: le poids attribué par l'utilisateur u au critère i est normalisé par le coefficient de normalisation k .

$W_{i,k}$: le poids de critère i dans l'itinéraire j est normalisé par le coefficient de normalisation k .

I_a : l'ensemble des critères i évalué par l'utilisateur.

I_i : l'ensemble des critères i évalué dans un itinéraire j .

Une fois on a calculé le degré de similarité, nous classons les itinéraires dans un vecteur global selon un ordre décroissant selon ce degré. Lorsqu'on veut déterminer la pertinence des nouveaux itinéraires on les compare par rapport au vecteur global.

Conclusion :

Dans ce chapitre, nous avons présenté deux méthodes de personnalisation. La première méthode est inspirée du domaine de recherche d'information documentaire. En effet l'objectif de cette méthode est de calculer la pertinence d'un critère de préférence dans un itinéraire. L'autre méthode est dérivée aussi du domaine de recherche documentaire. L'objectif de cette méthode est de calculer le degré de similarité entre le profil de l'utilisateur et l'itinéraire afin de fournir à l'utilisateur l'itinéraire le plus pertinent et qui satisfait ses préférences. Dans le chapitre suivant nous allons développer une application dans le domaine de transport de personnes pour mettre en valeur et appliquer les deux méthodes présentées.

Chapitre4 : Application dans le domaine de transport terrestre de personnes

I. Introduction

Dans le chapitre précédent, nous avons proposé deux méthodes de personnalisation, qui permettent de fournir à l'utilisateur un itinéraire personnalisé qui répond le plus à ses besoins et ses préférences.

Dans ce qui suit, nous allons développer une application dans le domaine de transport qui se base sur les deux méthodes de personnalisation déjà présentées dans le chapitre 3. Cette application porte sur le développement d'un système d'information personnalisé dans le domaine des transports terrestres de personnes.

II. Recherche d'itinéraire personnalisé

L'objectif d'un système d'information personnalisé est de faciliter l'accès à l'information multi-modes de transport aux usagers désirant effectuer un déplacement. En effet, l'usager se trouve confronté à un ensemble de sources d'informations disparates (horaires et tarifs délivrés par les différents exploitants, cartes, sur papier, borne interactive ou par Internet) qui s'avèrent parfois difficiles à intégrer en un plan de déplacement précis et unique, adapté à ses besoins et préférences. Comme nous l'avons déjà souligné dans le chapitre 1, la personnalisation s'avère une approche pertinente pour relever ce défi.

Dans le contexte de l'information aux usagers intégrant plusieurs modes de transports et leurs connexions, notre objectif est, d'une part, d'aider l'usager dans sa démarche de recherche d'informations et, d'autre part, de lui fournir un résultat personnalisé, c'est-à-dire toute l'information nécessaire et uniquement l'information nécessaire en fonction de son destinataire [Petit-Rozé et al.04].

Dans notre travail, les services fournis par le système à développer ont été modélisés au travers des diagrammes de cas d'utilisation. La figure 7 présente les besoins fonctionnels pour le service de recherche d'itinéraire.

Les cas d'utilisation peuvent se structurer en trois catégories :

-la première catégorie concerne l'identification de l'utilisateur (cas d'utilisation « S'inscrire » et « S'authentifier »).

-la deuxième catégorie regroupe les fonctionnalités pour la recherche d'itinéraire personnalisé (cas d'utilisation « Rechercher un itinéraire », « Afficher le détail d'un itinéraire » et « Choisir un itinéraire »).

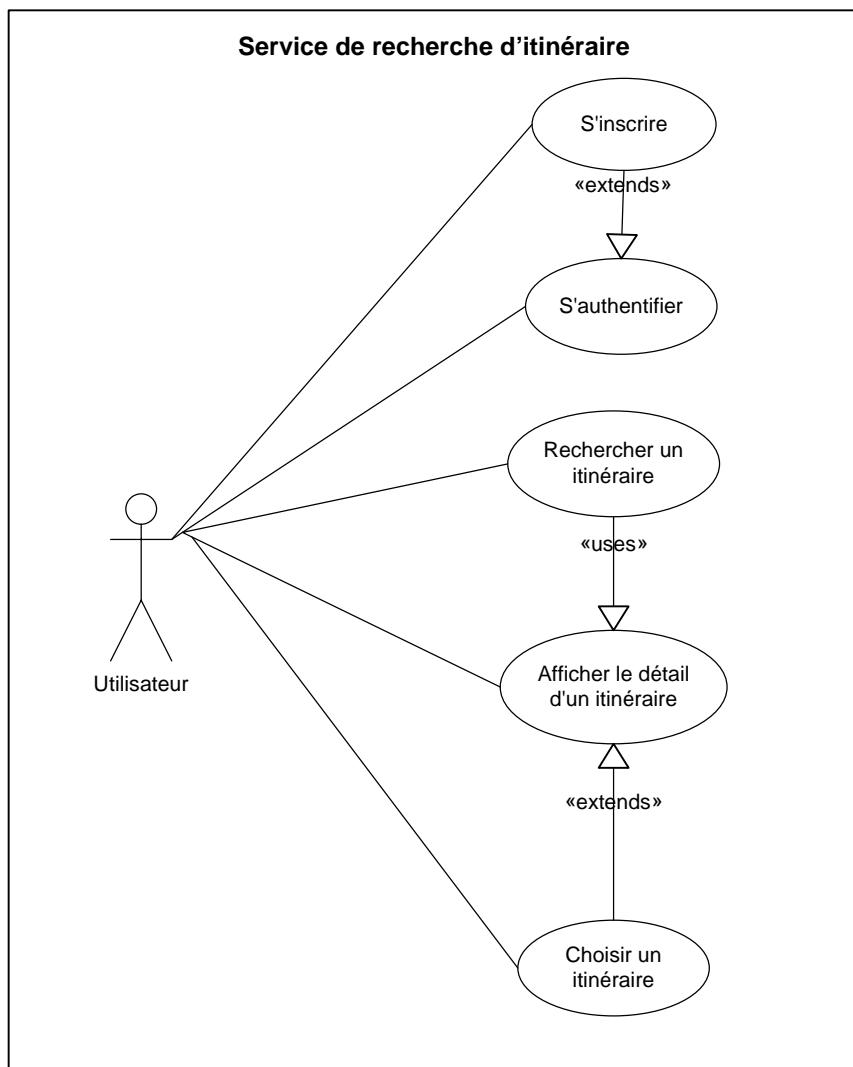


Figure 7 : Diagramme des cas d'utilisation du service de recherche d'itinéraire personnalisé

Tous d'abord, un utilisateur doit s'inscrire au pré du système. Il doit remplir un formulaire afin d'enregistrer ses coordonnées (voir figure 8). Ces dernières sont les données statiques expliquées dans le chapitre 3.

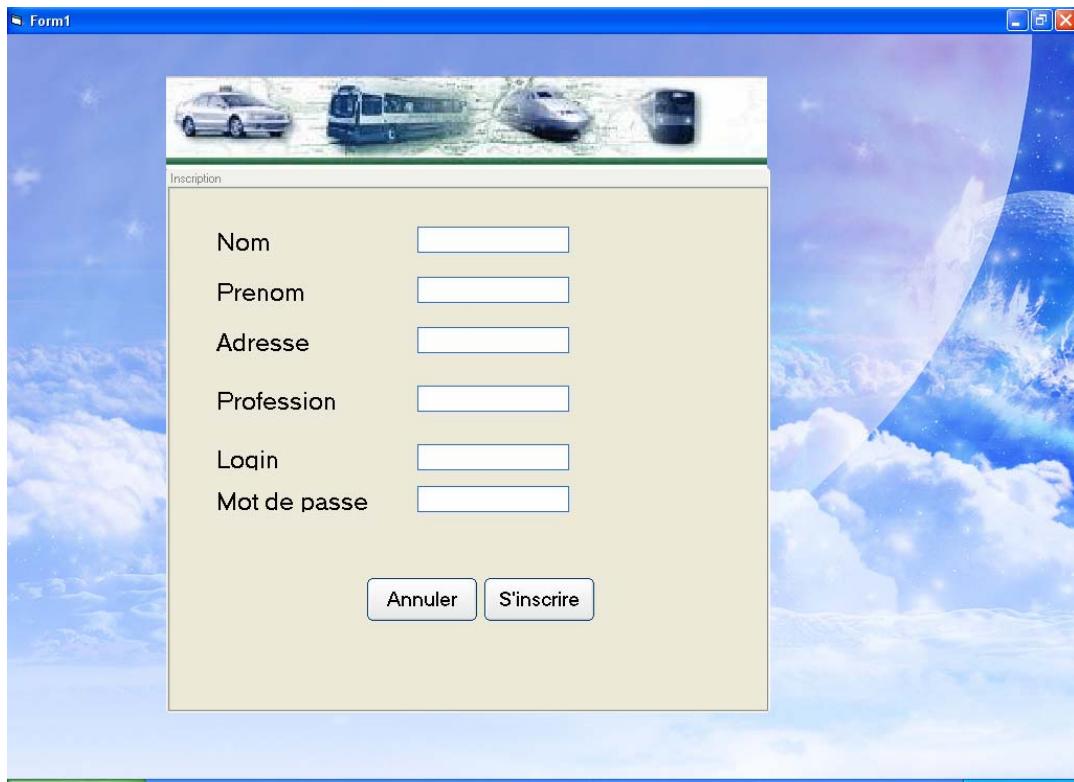


Figure 8- Inscription de l'utilisateur au système

Une fois l'utilisateur est déjà inscrit au système, à chaque session de recherche il doit s'identifier afin que le système puisse déterminer son profil (voir la figure suivante) :

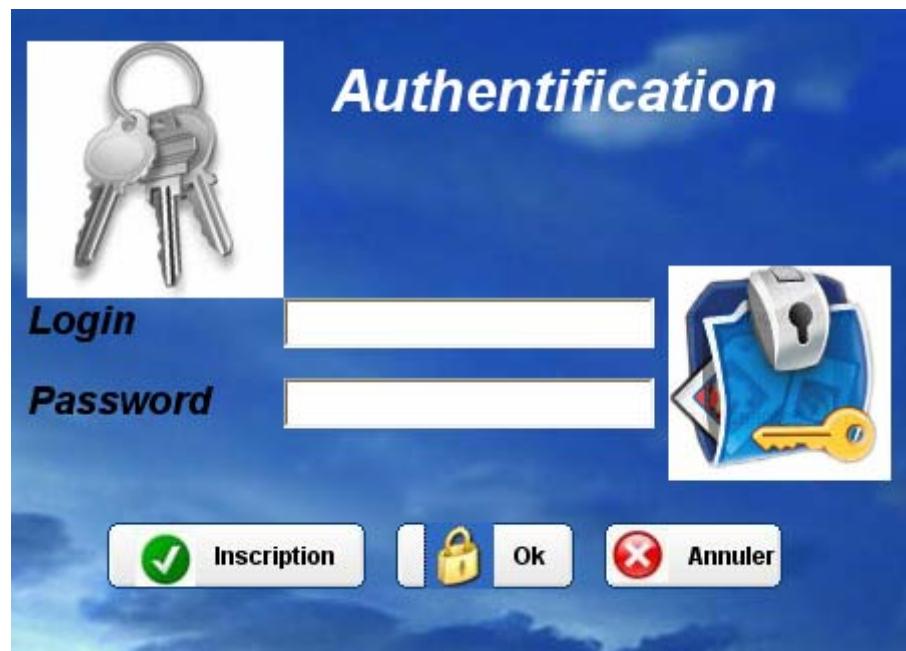


Figure 9-Identification de l'utilisateur

Recherche d'itinéraire selon préférence de l'utilisateur :

Les exemples d'illustration porteront sur le réseau transport de la Figure (ci-dessous) volontairement simplifié mais se basant sur des données réelles. Ce réseau décrit les points de connexions permettant à un usager d'effectuer un déplacement du laboratoire LAMIH de l'université de Valenciennes vers la société Archimed à Lille. Trois exploitants interviennent dans la réalisation du voyage :

- Semurval assure le déplacement de l'usager dans la ville de Valenciennes et de ses environs. Les modes de transport considérés sont le bus, la voiture et la marche à pied.

D'autres arrêts de bus auraient pu être considérés. Par exemple, à part l'arrêt « Aulnoy université », nous n'avons pas considéré les trois autres arrêts de bus qui sont situés à proximité du LAMIH. Le mode de transport « tramway » aurait pu aussi être pris en compte.

- Transpole s'occupe de la région lilloise. Les modes de transport considérés sont le bus et le métro. D'autres modes comme le vélo ou le taxi auraient pu être considérés.

- SNCF relie les deux villes de Valenciennes et de Lille par le train Ter de la région nord Pas-de-Calais. Nous avons aussi intégré la possibilité qu'un usager prenne une voiture pour aller du LAMIH à Archimed sans passer par le réseau des transports publics.

Par rapport à ce réseau, l'usager dispose de neuf itinéraires possibles pour aller du LAMIH à Archimed.

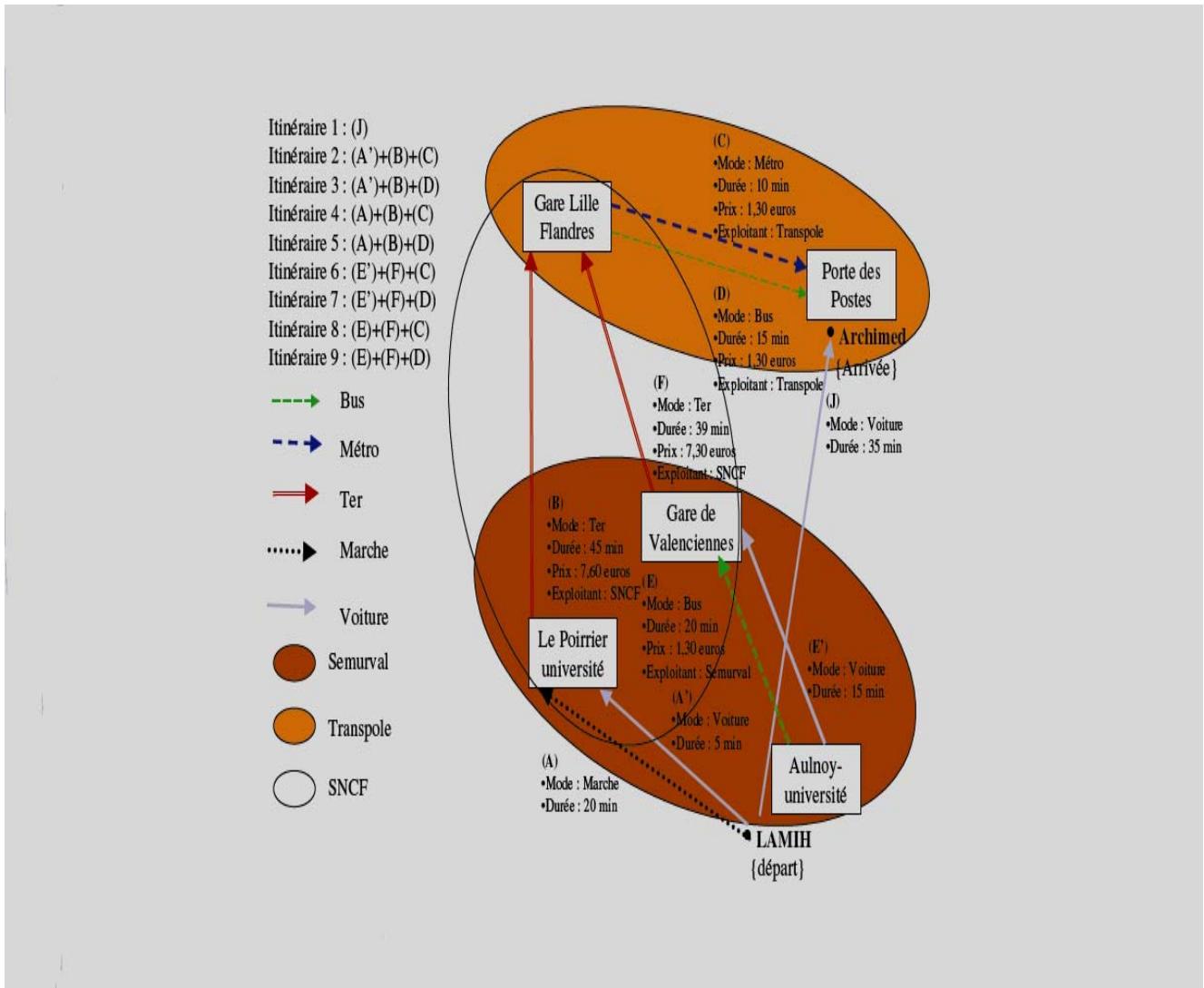


Figure 10-Choix de l'itinéraire pour un exemple de déplacement entre une origine et une destination

II.1. Application dans le domaine de transport en utilisant la première méthode :

Notre objectif est de proposer à l'utilisateur, l'itinéraire le plus préféré. Le système doit tout d'abord déterminer les meilleurs itinéraires entre une source et une destination données pour une requête d'un utilisateur, en tenant compte de ses préférences (plus économique, plus rapide, le moins de correspondances, le moins de marche à pieds).

L'utilisateur fournit le lieu de départ, le lieu d'arrivée et l'heure d'arrivée en lançant une requête R. Le système lui propose un itinéraire suivant ses préférences.

D'après la figure 10 nous trouvons 9 itinéraires possibles, qui sont :

- Itinéraire 1 : (J)
- Itinéraire 2 : (A') + (B) + (C)
- Itinéraire 3 : (A') + (B) + (D)
- Itinéraire 4 : (A) + (B) + (C)

Itinéraire 5 : (A)+(B)+(D)

Itinéraire 6 : (E')+(F)+(C)

Itinéraire 7 : (E')+(F)+(D)

Itinéraire 8 : (E)+(F)+(C)

Itinéraire 9 : (E)+(F)+(D)

Supposant que l'utilisateur à lancé 10 requête. Donc pour chaque requête l'utilisateur a choisie une solution (itinéraire) en suppose qu'il a effectué 10 voyages leur répartitions et comme suit :

Il a choisie :

2 fois l'itinéraire 1

2 fois l'itinéraire2

1 fois l'itinéraire 3

0fois l'itinéraire 4

1fois l'itinéraire 5

2fois l'itinéraire 6

1fois l'itinéraire7

1fois l'itinéraire 8

0fois l'itinéraire 9

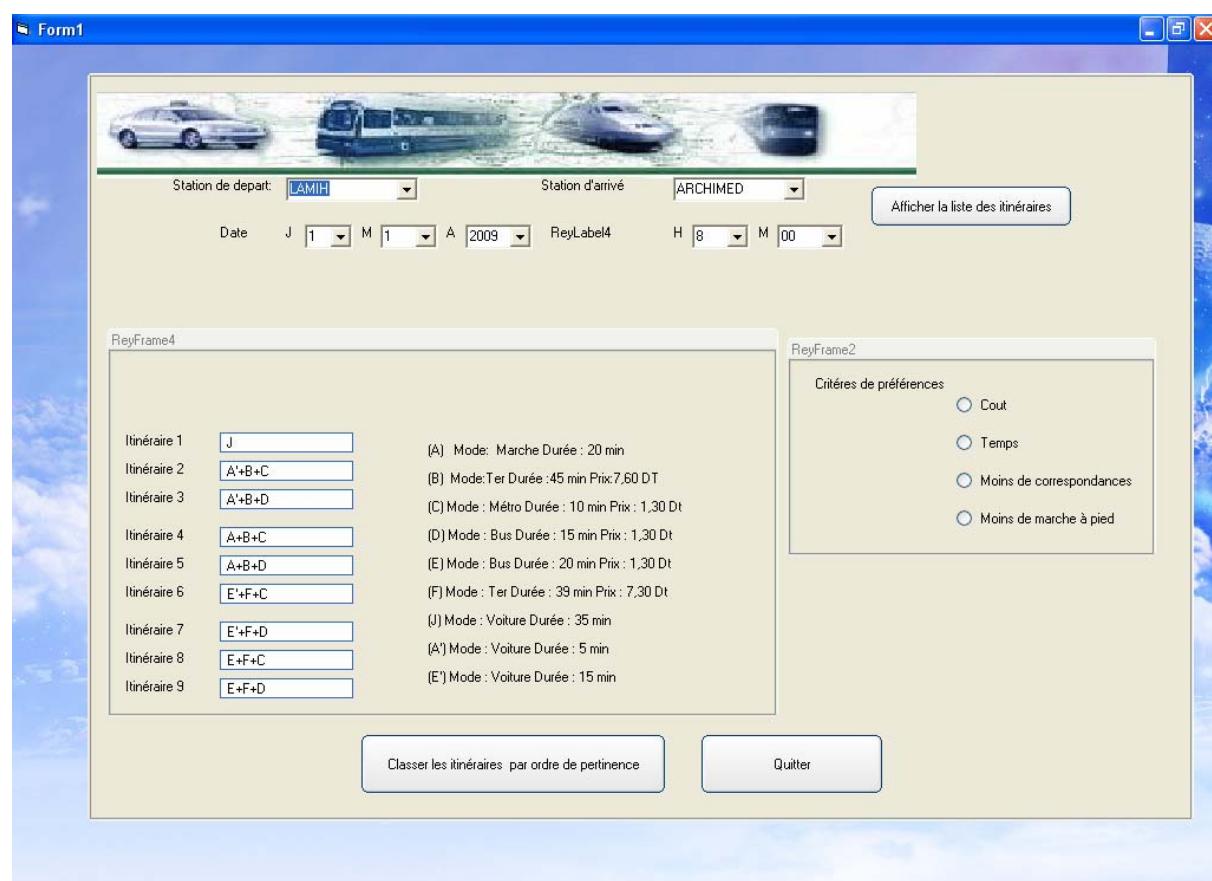


Figure 11- La liste des itinéraires possibles

Dans la première étape nous allons attribuer un rang à chaque critère. Ses valeurs correspondent au rang de cette solution vis à vis de ces critères.

$S_i = \{r_1, r_2, r_3, r_4\}$ Avec : r_i est les rangs de chaque critères ;

$r_i = \text{rang}(C_i)$ avec $1 < r_i < n$

Itinéraires	C1 : cout	C2 : temps	C3 : correspondances	C4 : moins de marche à pieds
Itinéraire1 : (J)	3	1	1	1
Itinéraire2 : (A')+(B)+(C)	1	2	2	1
Itinéraire 3 : (A')+(B)+(D)	1	4	2	1
Itinéraire 4 : (A)+(B)+(C)	1	6	2	2
Itinéraire5 : (A)+(B)+(D)	1	7	2	2
Itinéraire6 : (E')+(F)+(C)	1	3	2	1
Itinéraire7 : (E')+(F)+(D)	1	5	2	1
Itinéraire8 : (E)+(F)+(C)	2	5	2	1
Itinéraire9 : (E)+(F)+(D)	2	8	2	1

Tableau : Rangs des critères par rapport aux solutions d'une requête de recherche d'itinéraires

On associe un poids à chaque préférence de l'utilisateur. Ce poids définit l'importance d'un critère par rapport à l'utilisateur.

Nous allons affecter un poids W_{C_i} aux différentes solutions S_j suivant le critère C_i par la formule suivante :

$$W_{C_i} = \frac{1}{\text{rang}(C_i)}$$

Itinéraire (j)	Poids : C cout	Poids : C temps	Poids : C correspondance	Poids : C moins de marche à pieds
----------------	----------------	-----------------	--------------------------	-----------------------------------

Itinéraire1 : (J)	0.33	1	1	1
Itinéraire2 : (A')+(B)+(C)	1	0.5	0.5	1
Itinéraire 3 : (A')+(B)+(D)	1	0.25	0.5	1
Itinéraire 4 : (A)+(B)+(C)	1	0.166	0.5	0.5
Itinéraire5 : (A)+(B)+(D)	1	0.142	0.5	0.5
Itinéraire6 : (E')+(F)+(C)	1	0.33	0.5	1
Itinéraire7 : (E')+(F)+(D)	1	0.2	0.5	1
Itinéraire8 : (E)+(F)+(C)	0.5	0.2	0.5	1
Itinéraire9 : (E)+(F)+(D)	0.5	0.125	0.5	1

Tableau : le poids attribué aux critères de préférences

En suit nous calculons le degré de pertinence d'un critère i dans un itinéraire j selon la formule suivante :

$$DPCr(i) = \frac{Wc_i}{Nb - Critères} * \frac{Nb - It - choisie(j)}{nb - Re que te - Total}$$

Itinéraire (j)	DP Cout	DP Temps	DP Correspondance	DP moins de marche à pieds

Itinéraire1 : (J)	0.0165	0.05	0.05	0.05
Itinéraire2 : (A')+(B)+(C)	0.05	0.025	0.025	0.05
Itinéraire 3 : (A')+(B)+(D)	0.025	0.00625	0.0125	0.025
Itinéraire 4 : (A)+(B)+(C)	0	0	0	0
Itinéraire5 : (A)+(B)+(D)	0.025	0.00355	0.0125	0.0125
Itinéraire6 : (E')+(F)+(C)	0.05	0.0165	0.025	0.05
Itinéraire7 : (E')+(F)+(D)	0.025	0.005	0.0125	0.025
Itinéraire8 : (E)+(F)+(C)	0.00125	0.005	0.0125	0.025
Itinéraire9 : (E)+(F)+(D)	0	0	0	0

Tableau : Calcul de degré de pertinence de chaque critère i dans un itinéraire j

Afin de déterminer le degré de pertinence d'un itinéraire, nous additionnons la somme de degrés de pertinence des tous les critères comme la montre la formule suivante :

$$DP_{ITj} = \sum_{i=1}^n PDCr(i)$$

Itinéraire	DP _{ITj}
Itinéraire1 : (J)	0.166
Itinéraire2 : (A')+(B)+(C)	0.15
Itinéraire 3 : (A')+(B)+(D)	0.068
Itinéraire 4 : (A)+(B)+(C)	0

Itinéraire5 : (A)+(B)+(D)	0.053
Itinéraire6 : (E')+(F)+(C)	0.1415
Itinéraire7 : (E')+(F)+(D)	0.0675
Itinéraire8 : (E)+(F)+(C)	0.043
Itinéraire9 : (E)+(F)+(D)	0

Pour la normalisation DP'_{IT} est calculé comme suit :

$$DP'_{ITj} = \frac{DP_{ITj}}{\max(DP_{ITj})}$$

Ensuite, on fait un tri par ordre croissant de pertinence : l'itinéraire le plus pertinent et celui qui possède le degré de pertinence le plus élevé.

Itinéraire	DP'_{ITj}	Classement par ordre de pertinence
Itinéraire1 : (J)	1	1
Itinéraire2 : (A')+(B)+(C)	0.937	2
Itinéraire 3 : (A')+(B)+(D)	0.409	4
Itinéraire 4 : (A)+(B)+(C)	0	8
Itinéraire5 : (A)+(B)+(D)	0.319	6
Itinéraire6 : (E')+(F)+(C)	0.852	3
Itinéraire7 : (E')+(F)+(D)	0.406	5
Itinéraire8 :	0.259	7

(E)+(F)+(C)		
Itinéraire9 : (E)+(F)+(D)	0	8

Tableau : Classement des itinéraires par ordre de pertinence

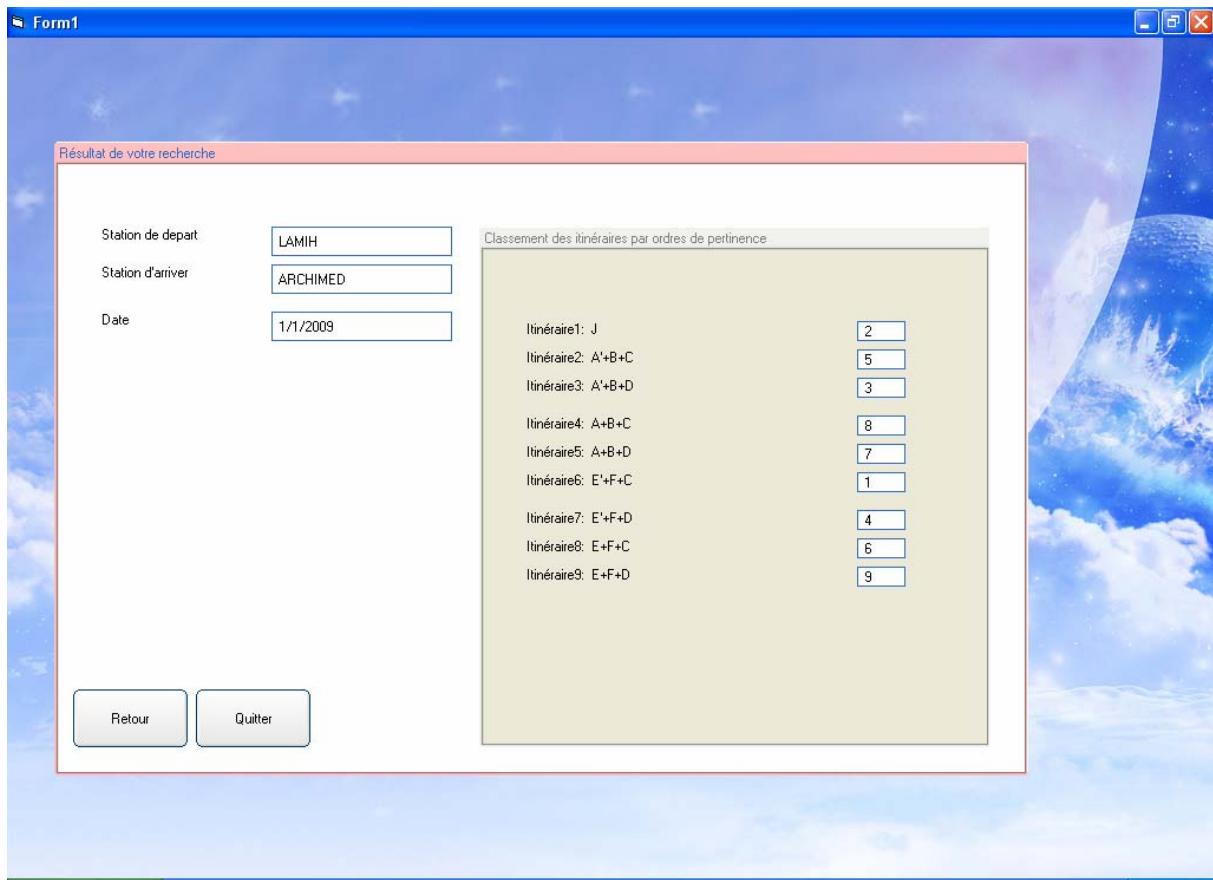


Figure 12-classement des itinéraires par ordre de pertinence

II.2.Application dans le domaine de transport en utilisant la deuxième méthode :

La deuxième méthode consiste à calculer un degré de similarité entre l’itinéraire et le profil utilisateur selon la deuxième méthode de personnalisation présenté déjà dans le chapitre3.

L’utilisateur fournit le lieu de départ, le lieu d’arrivée et l’heure d’arrivée en lançant une requête R. Le système lui propose les itinéraires les plus similaires à ses préférences (au profil utilisateur).

Pour cette application nous gardons les données :

Nous disposant de neuf itinéraires :

Itinéraire 1 : (J)
 Itinéraire 2 : (A')+(B)+(C)
 Itinéraire 3 : (A')+(B)+(D)
 Itinéraire 4 : (A)+(B)+(C)
 Itinéraire 5 : (A)+(B)+(D)
 Itinéraire 6 : (E')+(F)+(C)
 Itinéraire 7 : (E')+(F)+(D)
 Itinéraire 8 : (E)+(F)+(C)
 Itinéraire 9 : (E)+(F)+(D)

Soit R le nombre de requêtes lancées par l'utilisateur égale à 5. Chaque fois qu'il interagit avec le système, il choisit une solution (itinéraire).

Dans la première, étape nous allons attribuer un rang à chaque critère. Ses valeurs correspondent au rang de cette solution choisie par l'utilisateur vis-à-vis de ces critères.

$S_i = \{r_1, r_2, r_3, r_4\}$ Avec : r_i est les rangs de chaque critères ;

$r_i = \text{rang } (C_i)$ avec $1 < r_i < n$

Itinéraires	C1 : cout	C2 : temps	C3 : correspondances	C4 : moins de marche à pieds
Itinéraire1 : (J)	3	1	1	1
Itinéraire2 : (A')+(B)+(C)	1	2	2	1
Itinéraire 3 : (A')+(B)+(D)	1	4	2	1
Itinéraire 4 : (A)+(B)+(C)	1	6	2	2
Itinéraire5 : (A)+(B)+(D)	1	7	2	2
Itinéraire6 : (E')+(F)+(C)	1	3	2	1
Itinéraire7 : (E')+(F)+(D)	1	5	2	1
Itinéraire8 : (E)+(F)+(C)	2	5	2	1
Itinéraire9 : (E)+(F)+(D)	2	8	2	1

Tableau : Rangs des critères par rapport aux solutions d'une requête de recherche d'itinéraires

Pour calculer le poids ou le coefficient d'importance du critère de préférence (des termes pondérés) par rapport à un itinéraire (solution) choisi, nous allons affecter des notes e_{ij} aux différents solutions S_j suivant le critère C_i par la formule suivante :

$$W_{ij} = e_{ij} = note(C_i) = \frac{1}{rang(C_i)} * \delta$$

Itinéraire (j)	Poids : C cout	Poids : C temps	Poids : C correspondance	Poids : C moins de marche à pieds
Itinéraire1 : (J)	0.33 *	1	1	1
Itinéraire2 : (A')+(B)+(C)	1	0.5	0.5	1
Itinéraire 3 : (A')+(B)+(D)	1	0.25	0.5	1
Itinéraire 4 : (A)+(B)+(C)	1	0.166	0.5	0.5
Itinéraire5 : (A)+(B)+(D)	1	0.142	0.5	0.5
Itinéraire6 : (E')+(F)+(C)	1	0.33	0.5	1
Itinéraire7 : (E')+(F)+(D)	1	0.2	0.5	1
Itinéraire8 : (E)+(F)+(C)	0.5	0.2	0.5	1
Itinéraire9 : (E)+(F)+(D)	0.5	0.125	0.5	1

Tableau : le poids attribué aux critères de préférences

Nous allons calculer par la suite le poids du critère i pour l'utilisateur j

Soit W_{ui} le poids attribuer par l'utilisateur u au critère i. Notons e_i la note du critère i pour la solution choisie par l'utilisateur u. La nouvelle note attribuée au critère i pour l'utilisateur u sera donc définie par :

$$W_{ui} = moy(e_i)$$

$$\text{Avec moyenne } (e_i) = \frac{\sum_{i=1}^n e_i}{NB - requeste - total}$$

n : nombre de solution choisie

W_{ui}	Poids des critères Ci
W_{u1}	0.866
W_{u2}	0.75
W_{u3}	0.5
W_{u4}	1

Tableau : poids de critère Ci

La similarité entre l'itinéraire et le profil de l'utilisateur est calculé par la formule suivante :

$$SIM(U, I) = \sum_{i=1}^n \frac{W_{uj}}{\sqrt{\sum_{K \in I} W_{u,k}^2}} * \frac{W_{ij}}{\sqrt{\sum_{k \in I} W_{ik}^2}}$$

K : constante de normalisation

Ensuite nous classons les itinéraires par ordre de similarité croissante voir tableau suivant :

Itinéraire (j)	Degré de similarité	Classement par ordre de similarité Croissant
Itinéraire1 : (J)	0.164	2
Itinéraire2 : (A')+(B)+(C)	0.134	3
Itinéraire 3 : (A')+(B)+(D)	0.113	4
Itinéraire 4 : (A)+(B)+(C)	0.0866	8
Itinéraire5 : (A)+(B)+(D)	0.107	6
Itinéraire6 : (E')+(F)+(C)	0.264	1
Itinéraire7 : (E')+(F)+(D)	0.109	5
Itinéraire8 : (E)+(F)+(C)	0.09	7
Itinéraire9 : (E)+(F)+(D)	0.084	9

Tableau : Calcule et classement des itinéraires par degré de similarité.

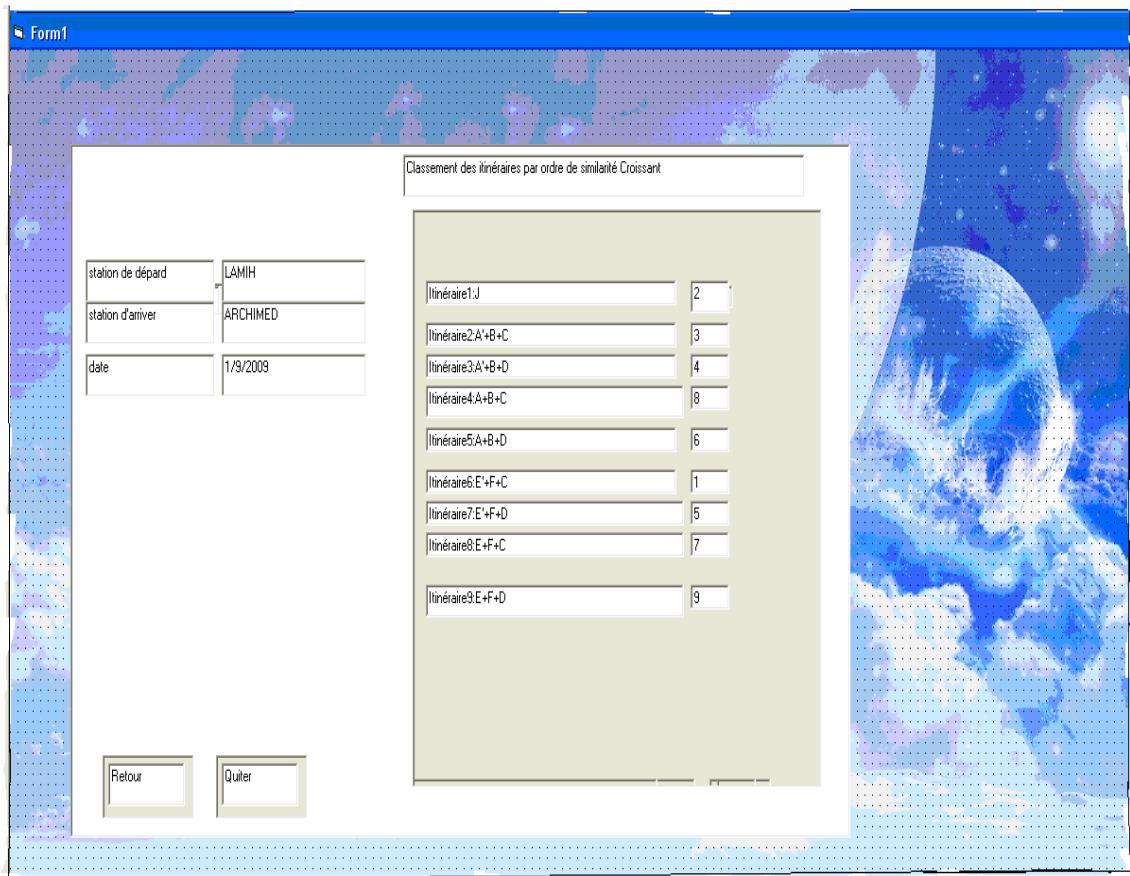


Figure : Classement des itinéraires par degré de similarité.

Comparaison des deux méthodes de personnalisation :

Itinéraires	Classement par ordre de pertinence	Classement par ordre de similarité croissant
Itinéraire1 : (J)	1	2
Itinéraire2 : (A')+(B)+(C)	2	3
Itinéraire 3 : (A')+(B)+(D)	4	4
Itinéraire 4 : (A)+(B)+(C)	8	8
Itinéraire5 :	6	6

(A)+(B)+(D)		
Itinéraire6 : (E')+(F)+(C)	3	1
Itinéraire7 : (E')+(F)+(D)	5	5
Itinéraire8 : (E)+(F)+(C)	7	7
Itinéraire9 : (E)+(F)+(D)	8	9

Tableau 9-comparaison entre les deux méthodes de personnalisation

Nous remarquons que pour les itinéraires It3, I4, It5, It7et It8, le système donne les mêmes résultats que ce soit par la méthode de mesure de degré de pertinences ou la méthode de mesure de similarité. Pour les autres itinéraires on remarque qu'il y a une différence entre les résultats.

La question qui se pose ici quelle est la méthode qui fourni la meilleure qualité en terme de personnalisation.

Conclusion :

Dans ce chapitre nous avons appliqué les deux méthodes dans le domaine de transport.

Le but de la première méthode est de fournir à l'utilisateur l'information la plus pertinente (c'est-à-dire l'itinéraire le plus pertinent). Par contre l'objectif de la deuxième méthode est de fournir à l'utilisateur l'itinéraire qui ressemble le plus à son profil (c'est-à-dire qui satisfait le plus ses besoins et préférence). Ces méthodes visent à assurer une recherche d'information plus pertinente dans le domaine de transport.

Conclusion générale

Le travail de recherche présenté dans ce mémoire, se situe dans la thématique de la personnalisation de l'information dans le domaine de transport multimodale. Ils contribuent à l'amélioration des systèmes de recherche d'information personnalisée dans le domaine de transport de personnes afin de favoriser et faciliter l'utilisation de transport en commun.

Après une introduction des principales approches et méthodes existantes pour la personnalisation, nous avons étudié les principaux termes rencontrés dans le domaine de l'information multimodale. Ainsi que les principaux axes de recherche dans le domaine de transport urbain. Par ailleurs nous nous sommes situés dans les systèmes de personnalisation de l'information voyageurs.

Cette étude nous a permis de dégager les standards et les limites de ces systèmes:

La plus part des systèmes d'information voyageur adaptent une architecture multi agents et gère des sources d'information distribué qui proviennent de plusieurs sources.

Ces systèmes sont destinés soit pour l'optimisation et le calcul des itinéraires soit pour la personnalisation de l'information.

Il n'existe aucune méthode de personnalisation qui intègre une étape d'évaluation du système en même temps.

Il y a peu de système qui intègre les critères de préférences de l'utilisateur dans le processus de recherche d'itinéraires.

En effet, dans le but d'améliorer la qualité de l'information multimodale nous avons présenté deux méthodes de personnalisation. Une méthode qui se base sur le degré de pertinence et l'autre qui se base sur le degré de similarité.

La première consiste à calculer le degré de pertinence d'un itinéraire par rapport à un ensemble d'itinéraires possibles en tenant compte des critères de préférences de l'utilisateur.

La deuxième méthode consiste à calculer un score de correspondance ou de similarité entre un itinéraire et le profil de l'utilisateur.

Ces deux méthodes contribuent à l'amélioration de processus de recherche d'itinéraire dans le domaine de transport en délivrant une information plus pertinente qui satisfait le mieux les besoins de l'utilisateur.

Mais la question qui se pose ici quelle est la méthode qui fournit la meilleure qualité en terme de personnalisation. D'où l'idée de proposer une étape d'évaluation dans un système de personnalisation qui consiste à tester et d'évaluer le service personnalisé obtenu.

Différentes perspectives sont envisagées :

- Concevoir un système qui assure une tâche de personnalisation de l'information et de l'évaluation en même temps. Afin de garantir une information personnalisée de qualité et d'éviter les charges supplémentaires pour les opérateurs qui utilisent ce genre de systèmes.
- Concevoir un système de personnalisation à base d'agent logiciel qui intègre une étape d'évaluation : ceci admet intégré des agents d'évaluation qui servent à l'évaluation de la qualité de personnalisation liée au contenu.
- L'intégration d'un outil de traçabilité au cours de la phase de conception d'un système d'information personnalisé qui a pour rôle d'effectuer une évaluation précoce du système.
- Proposer un système qui intègre à la fois le processus de calcul d'itinéraire, de personnalisation et évaluation en même temps.
- Proposer des systèmes qui proposent des itinéraires selon l'activité d'attraction touristique ou de loisir préféré par les utilisateurs, ou des services répondant aux centres d'intérêts d'un utilisateur spécifique, ce qui peut présenter une piste intéressante pour améliorer la qualité de ces systèmes.

Bibliographies

[Anli.06] Abdouroiahamane, Anli ; Thèse : « Méthodologies de développement de système d'information personnalisé : Application d'un system d'information au service des usagés des transports terrestre de personnes » ; Université de valencienne Haut-cambrésis ,2006.

[Anli et al.05] Anli A., Kolski C. et Abed M ; « Principes et architecture pour la personnalisation d'information en interaction homme-machine : Application à l'information transport ». In Proceedings of IHM 2005, International Conference Proceedings Series, ACM Press,Toulouse, pp. 123-130, Septembre 2005.

[Brossard .08] Arnaud Brossard, ; Thèse : « PERCOMOM : Une méthode de modélisation des applications interactives personnalisées appliquée à l'information voyageur dans le domaine des transports collectifs »; Université de valencienne Haut-cambrésis ,2008.

[Ciro et al.03] Ciro S., Newton V.2003; “Use Reformulated Profile in Information Filtering”, In Proceedings of the AAAI Workshop on Semantic Web Personalization, San Jose, California, 2004.

[Berrut et al. 03] Berrut C. et Denos N ; « Filtrage collaboratif. In Assistance intelligente à la recherche d'informations », E. Gaussier, M.H. Stéfanini (dir.), Hermès-Lavoisier, pp. 241-269, 2003.

[Kostadinov .08] Dimetri Kostadinov ;Thèse : « Personnalisation de l'information : une approche de gestion de profils et de reformulation de requêtes », l'université de versailles saint – quentin-en –yvelines, 2008.

[Lin et al.05] Lin C., Xue G., Zeng H., YUY ; « Using probabilistic latent semantic analysis for personalised Web search », Proceedings of the APWeb Conference, p. 707-711, 2005.

[Tamine et al.06] Lynda Tamine, Wahiba Bahsoun ; « Définition d'un profil multidimensionnel de l'utilisateur » : Vers une technique basée sur l'interaction entre dimensions Actes de CORIA,2006.

[Petit-Rosé.03] Petit-Rozé C., « Organisation Multi-Agent au service de la personnalisation de l'information : Application à un système d'information multimodale pour le transport terrestre de personnes » ; Thèse de doctorat, Université de Valenciennes et du Hainaut-Cambrésis, Décembre 2003.

[Petit-Rozé et al., 04] Petit-Rozé C., Anli A., Grislin-Le Strugeon E., Abed M., Uster G. et Kolski C., « Système d'Information Transport Personnalisée à base d'agents logiciels » ; Génie Logiciel, 70, pp. 29-38, 2004.

[Ber.98] J. Breese, D. Heckerman, and C. Kadie., “ Empirical Analysis of Predictive Algorithms for Collaborative Filtering”; Proc. of the Fourteenth Annual Conference on Uncertainty in Artificial Intelligence (UAI-98). San Francisco, CA, pp. 43-52, 1998.

[Pohl et al.97] Pohl W. and Höhle J., “Mechanisms for flexible representation and use of knowledge in user modeling shell systems”; In Jameson, A., Paris, C. and Tasso, C. (eds), User Modeling: Proceedings of the Sixth International Conference, UM97. New York : Springer Wien, pp.403-414, 1997.

[Lumineau.03] Lumineau N., « *Un tour d'horizon du filtrage collaboratif* » ; Travail réalisé dans le cadre de l'AS Personnalisation de l'Information, janvier – décembre 2003, disponible à l'adresse

www.prism.uvsq.fr/recherche/themes/sial/cnrs/Fichiers/RapportTourDHorizonDuFiltrageCollaboratif-LIP6.pdf.

[Lieberman.95] Lieberman H., “ Letizia : An Agent That Assists Web browsing”; In International Joint Conference on Artificial Intelligence (IJCAI'95 – Montreal, Canada, August 20-25). San Francisco, USA, Morgan Kaufman Publishers, 1995.

[Lieberman et al.01] Lieberman H., Fry C. and Weitzman L., “Exploring the Web with Reconnaissance Agents”; In *ACM Conference on Human-Computer Interface*. ACM Press, pp.69-75, August 2001.

[Kobsa et al.95] Kobsa A. and Pohl W., “ The User Modeling Shell System BGP-MS”; *User Modeling and User-Adapted Interaction*, Vol. 4, pp. 59-106, 1995.

[Cinquin et al.02] Cinquin L., Lalande P. A et Moreau N., « Le projet ECRM : Relation client et Internet ». Editions Eyrolle, Paris, 2002.

[Bouzeghoub et al.04] Bouzeghoub M et Kostadinov D., « *Une approche multidimensionnelle pour la personnalisation de l'information* » ; Rapport PRISM, Versailles, France, 2004.

[Sugiyama et al.04] Sugiyama K., Hatano K., Yoshikawa M., “*Adaptive web search based on user profile constructed without any effort from users* ”, In Proceeding of the 13th International WorldWide Web Conferences (WWW), New York, USA, p. 675-684, 2004.

[Hayes et al. 04] Hayes C. and Cunningham P., “Context boosting collaborative

Recommendations”; *Knowledge-Based Systems*, n° 17, pp. 131-138, 2004.

[Deschaine et al.00] Deschaine L., Brice R. and Nodine M., “ Use of InfoSleuth to CoordinateInformation Acquisition, Tracking and Analysis in Complex Applications”; In *Proceedings ofAdvanced Simulation Technologies Conference*, Washington, D.C., USA, April 2000.

[Balabanovic et al.97] Balabanovic M. and Sholham Y. Fab., “Content-based, collaborative Recommendation”; *Communications of the ACM*, vol. 40, pp. 66-72, March 1997.

[O’Hare et al. 03] O’Hare G.M.P. and O’Grady M.J., “ Gulliver’s Genie: a multi-agent system for ubiquitous and intelligent content delivery”; *Computer Communications*, 26, pp.1177-1187,2003.

[Waern et al. 98] Waern A., Averman C., Tierney M., Rudström A. and Laaksolahti J., “*ConCall: an information service for researchers based on EdInfo*”; Research report T98:04, Swedish Institute of Computer Science, October 1998.

[Asnicar et al. 97] Asnicar F.A. and Tasso C., “ IfWeb: a Prototype of User Model-Based Intelligent Agent for Document Filtering and Navigation in the World Wide Web”. In *Proceedings of the workshop « Adaptive Systems and User Modeling on the World Wide Web »*, Chia Laguna,Sardina, pp. 3-12, 2-5 June 1997.

[Keeble et al .00] Keeble R.J. and Macredie R.D., “ Assistant agents for the world wide web intelligent interface design challenges”; *Interacting with computers*, Vol. 12 No. 4, pp. 357-381, 2000.

[Oulasvirta et al. 2008] Antti Oulasvirta, [Mika Raento.](#), “ [Designing for privacy and self-presentation in social awareness, Personal and Ubiquitous Computing](#)”; v.12 n.7, p.527-542, October 2008.

[Domshlak et al. 2007] Carmel Domshlak, Thorsten Joachims., “Efficient and non-parametric reasoning over user preferences”; *User Model. User-Adapt. Interact.* p41-69, 2007.

[Mobasher.05] Mobasher B., “Web Usage mining and personalization”; In Practical Handbook Of InternetComputing. Singh Munindar P. (Ed.), 2005.

[Michlmayr et al.07] Michlmayr E., Cayzer S., “*Learning User Profiles from Tagging Data and Leveraging them for Personal(ized) Information Access*”; WWW2007, Banff, Canada, p. 2007

[Chan.00] Chan. P., “*Constructing Web User Profiles: A Non-invasive Learning Approach*”; In Web Usage Analysis and User Profiling, LNAI 1836, Springer-Verlag, p. 39-55, 2000.

[Naïm et al.99] Naïm, P & Becker, A., « *Les réseaux bayésiens. Modèles graphiques de connaissances* » ; Editions Eyrolles,1999.

[Werner et al.04] Werner P., Mushtaq N., Tolle K., Zicari R., “Building and Evaluating Non-Obvious User Profiles for Visitors of Web Site”; In Proceedings of the IEEE International Conference on E-Commerce Technology (CEC’04), p. 9-15, 2004.

[Zemirli.04] Zemirli W.Nesrine : Vers le développement d'un système de recherche d'information personnalisé intégrant le profil utilisateur. Université Paul Sabatier & Institut National Polytechnique de Toulouse ,2004.

[Goldberg et al. 92] Goldberg D., Nichols D., Oki B. and Terry D. Using collaborative filtering to weave an information tapestry. Communications of the ACM, vol. 35, pp. 61-70, 1992.

[Salton 83] Salton G. and McGill M., “ Introduction to modern information retrieval”; New York:McGraw-Hill, 1983.

[Panzani et al.96] Pazzani M., Muramatsu J., Billsus D., Syskill & Webert : Identifying interesting web sites, In Proceedings of the Thirteenth National Conference on Artificial Intelligence, Pages 54-61, 1996.

[Memmi.00] Memmi D., « Le modèle vectoriel pour le traitement des documents », 2000.

.

[Ungar et al. 98] Ungar L.H. and Foster D.P; “Clustering Methods for Collaborative Filtering”; AAAI Workshop on Recommendation Systems, M. Kaufman (Ed.) « Empirical analysis of predictive algorithms for collaborative filtering. Uncertainty in Artificial Intelligence », Proceedings of the 14th Conference, pp. 43-52, 1998.

[Nguyen.05] Nguyen T., Boucher A. « réseaux bayésiens ». Rapport d'intérêt du travail personnel encadré, 2005.

[Schank.82] Schank, R., “ Dynamic Memory: A Theory Of Reminding and Learning in Computer and People”; Cambridge university press, Cambridge,1982.

[Lajmi et al.07] Soufiane Lajmi, .Chirine Ghedira, Khaled Ghedira ., « Une méthode d'apprentissage pour la composition de services web » ; REVUE : L'objet. Volume 8 – n°2/2007

[Sénéchal et al.06] Olivier Sénéchal,.HakimSefiane., « Aide multicritère au choix d'une méthode de diagnostic pour la maintenance » ; 6ème Conférence Francophone de Modélisation et Simulation - MOSIM'06 - du 3 au 5 avril 2006 – Rabat- Maroc,2006.

[Kolodner .93] Kolodner J., « Case-Based Reasoning », San Mateo, CA : Morgan Kaufman, 1993.

[Lieborrtz .99] Lieborrtz Jay, 1999. The Hand bouk of Applied Expert Systems. Publisher: CRC Press LLC, ISBN 084933106, p. 244-250.

[HajSaid et al.04] HajSaid A, Colloc J., « Distances sémantiques pour la comparaison des connaissances objets dans le cadre du raisonnement à partir de cas » ; Mémoire de DEA Informatique Théorique et Applications, p. 25-40, 2004..

[Napoli .98] Napoli A., Groupe français de raisonnement à partir de cas ,1998.

[Memmi .04] Memmi, Daniel., « Réseaux de neurones artificiels » ; Notes de cours (Introduction à l'informatique cognitive), Programme de Doctorat en Informatique Cognitive, Université du Québec à Montréal, 2004.

[Leroux .04] Jean-Daniel Leroux., « Les réseaux de neurones artificiels » ; Concours de vulgarisation scientifique, Université de Sherbrooke, mars 2004.

[Lavoie .07] Benoit Lavoie ., « Approches symboliques et sous-symboliques en représentation de connaissances ;. Synthèse de lectures, Programme de Doctorat en Informatique Cognitive, Université du Québec à Montréal. 21p,2007.

[Nouali .03] Nouali.O ., Sélection de critères pour le filtrage automatique de messages, RECITAL 2003, Batz-sur-mer, 11-14 juin 2003.

[Moghrabi et al. 98] Moghrabi C. and Eid M.S., “ Modeling users through an expert system and a neural network”; *Computers & Industrial Engineering*, Vol. 35, No 3-4, pp. 583-586, 1998.

[Linden et al. 97] Linden G., Hanks S. and Lesh N., “ Interactive Assessment of User Preference Models : The Automated Travel Assistant”; In Jameson, A., Paris, C. and Tasso,

C. (eds), User Modeling: Proceedings of the Sixth International Conference, UM97. New York : SpringerWien, pp. 67-78, 1997.

[Lieberman et al.01] Lieberman H., Fry C. and Weitzman L., “ Exploring the Web with Reconnaissance Agents”; In ACM Conference on Human-Computer Interface. ACM Press, pp.69-75, August 2001.

[Coyle et al.02] Coyle L. and Cunningham P., “ A Case-Based Personal Travel Assistant for Elaborating User Requirements and Assessing Offers”; Proceedings of the 6th European Conference, ECCBR 2002, S. Craw, A. Preece (Eds.). LNAI Vol. 2416, pp. 505-518, Springer-Verlag, 2002.

[Ambrosini et al. 97] Ambrosini L., Cirillo V. and Micarelli A., “ A Hybrid Architecture for User-Adapted Information Filtering on the World Wide Web”; In Jameson, A., Paris, C. and Tasso,C. (Eds.), *User Modeling: Proceedings of the Sixth International Conference*, UM97, New York : Springer Wien, pp. 59-61, 1997.

[Soltysiak et al. 98] Soltysiak S. J. and Crabtree I. B.’ “ Automatic learning of user profiles – towards the personalization of agent services”; *BT Technologie*, Vol. 16 N° 3, pp. 110-117, July1998.

[Abed et al.90] Abed M. and Angue J.C., “ Using the measure of eye mouvements to modelise an operator’s activity”; *Ninth European Annual conference on “Human decision making and manual control”*, Varesse, Italy, September,1990.

[Nielsen et al. 94] Nielsen J. and Mack R. L., “ *Usability inspection methods* ”. New York : john Wiley & Sons, 1994.

[Trabelsi et al. 06] Trabelsi A., Ezzedine H. et Kolski C., « Un mouchard électronique orienté agent pour l’évaluation de systèmes interactifs de supervision » ; In *Conférence Internationale Francophone d’Automatique CIFA’2006*, Bordeaux, France, Mai ,2006.

[Senach .90] Senach B., « Evaluation ergonomique des interfaces Homme-machine : une revue de littérature » ;Rapport de recherche, INRIA, Sophia Antipolis, n° 1180, 1990.

[Moser et al. 71] Moser C. and Karlton G., “ Survey methods in social investigation ”. 2nd Edition.London : Hermann, 1971.

[Bastien et al.01] Bastien J. et Scapin D., « Evaluation des systèmes d’information et critères ergonomiques » ; In C. Kolski (Ed.), *Environnements évolués et évaluation de l’IHM. Interaction Homme-Machine pour les SI*, pp. 53-79, Paris : Editions Hermès, 2001.

[Zidi06] Zidi K., « Système Interactif d’Aide au Déplacement Multimodal (SIADM) » ; Thèse de doctorat, Université des sciences et technologies de Lille et École centrale de Lille, Décembre 2006.

[Kamoun.07] Kamoun M.« Conception d’un système d’information pour l’aide au déplacement multimodal :Une approche multi-agents pour la recherche et la composition des itinéraires en ligne.» Thèse de doctorat, Université des sciences et technologies de Lille et École centrale de Lille, Avril 2007.

[Mouloudi.07] Mouloudi A. « Intégration des besoins des utilisateurs pour la conception de systèmes d’information interactifs Application à la conception d’un système d’information voyageurs multimodal (SIVM) » ; Thèse de doctorat, Université de Technologie de Compiègne, Septembre 2007.

[Lecomte et al.00] Lecomte N. et Patesson R., «Le panel des voyageurs : une étude des activités et des besoins d’information des utilisateurs des transports publics » ; In Actes de la conférence ERGO-IHM'00. Biarritz, France, 2000.

[Chen et al.98] Chen L., Sycara K., « Web Mate: A Personal Agent for Browsing and Searching »; In proceedings of the 2nd International Conference on Autonomous Agents and Multi Agent Systems, AGENTS '98, ACM. New York, NY, USA, 1998.

[Balabanovic et al. 97] Balabanovic M. and Sholham Y. Fab ., “ Content-based, collaborative Recommendation ”. Communications of the ACM, vol. 40, pp. 66-72, March 1997.

[Tamine et al.07] Lynda Tamine-Lechani, Nesrine Zemirli, Wahiba Bahsoun., « Approche statistique pour la définition du profil d’un utilisateur de système de recherche d’information » ; Manuscrit auteur, publié dans "Information - Interaction - Intelligence (I3),pp. 5-25,2007.

[Goldberg et al.92] Goldberg D., Nichols D., Oki B. and Terry D., “ Using collaborative filtering to weave an information tapestry”; *Communications of the ACM*, vol. 35, pp. 61-70, 1992.

[Grislin et al.96] Grislin M. et Kolski C., « Evaluation des interfaces homme-machine lors dudéveloppement de système interactif » ; *Technique et Science Informatiques (TSI)*, 2, pp. 265-296, 1996.

[Korfhage .97] Korfhage R.R., “ *Information storage and retrieval* ”; Wiley Computer Publishing. ISBN 0-471-14-338-3, 1997.

[Pennock et al. 00] Pennock D., Horvitz E., Lawrence S. and Giles C., “ Collaborative Filtering by Personality Diagnosis: A Hybrid Memory- and Model-Based Approach ”; In *Proceedings of the 16th Conference on Uncertainty in Artificial Intelligence*, UAI-2000, Morgan Kaufman, San Francisco, pp. 473-480,2000.

[Aggarwal et al. 99] Aggarwal C.C., Wolf J.L., Wu K-L. and Yu P.S., “ Hortex Hatches an Egg : A New Graph- Theoretic Approach to Collaborative Filtering ”; In *Proceedings of KDD'99*, 5th

International Conference on Knowledge Discovery and Data Mining, pp. 201-212, San Diego, USA, August 1999.

[Papamichail et al.07] Georgios P. Papamichail a, Dimitrios P. Papamichail., “The k-means range algorithm for personalized data clustering in e-commerce”; European Journal of Operational Research 177 ,pp. 1400–1408,2007.

[Xie et al.07]Bo Xie ,*, Peng Han , Fan Yang , Rui-Min Shen , Hua-Jun Zeng , Zheng Chen., “DCFLA: A distributed collaborative-filtering neighbor-locating algorithm”; Information Sciences 177,pp. 1349–1363,2007.

[Grotenhuis et al.07]Jan-Willem Grotenhuis, Bart W. Wiegmansa, Piet Rietveld., “The desired quality of integrated multimodal travel information in public transport: Customer needs for time and effort savingsTransport Policy”,pp. 27–38.2007.

[Chee et al.01] Chee, J. Han, and K. Wang., “RecTree: An Efficient Collaborative Filtering Method”; Proc. 2001 Int. Conf. on Data Warehouse and Knowledge Discovery (DaWaK'01), Munich,Germany, Sept. 2001.

[Goldberg.00]Ken Goldberg, Theresa Roeder, Dhruv Huptan, and Chris Perkins. Eigentaste., “A constant time collaborative filtering algorithm”. Technical Report M00/41, IEOR and EECS Departments, UC Berkeley, Août 2000.

[Resnick et al.94]Paul Resnick, Neophytos Iacovou, Mitesh Suchak, Peter Bergstrom, John Riedl., “ GroupLens: An Open Architecture for Collaborative Filtering of Netnews”; In proc. of ACM 1994.Conference on Computer Supported Cooperative Work, Chapel Hill, Caroline du Nord,1994.

[Miyahara et al.00] K. Miyahara, M. J. Pazzani., « Collaborative filtering with the simple Bayesian classifier”; Inproc. The 6th Pacific Rim International Conference on Artificial Intelligence, Australia,2000.

[Soui et al .08] SOUI M., ABED M., “ [Evaluation of personalized interactive systems in transport applications: criteria and evaluation methods](#)”; 27th European Annual Conference on Human Decision-Making and Manual Control, EAM'08, Delft, The Netherlands, juin2008.

[Soui et al .07] SOUI M., ABED M., KOLSKI C., GHÉDIRA K., “ [Criteria devoted to evaluate personalized interactive systems](#)”; M. Abid, A. Hadj Kacem, M. Jmaiel, M. Lahiani (Ed.); *Nouvelles tendances technologiques en génie électrique & informatique, GEI'2007*, Edition CPU, Sfax, pp. 119-125, mars, ISBN 978-9973-61-631-9, 2007.

[Nugyen et al.06] Nugyen A., Denos. N., Berrut N., « Modèle des espaces de communautés orienté vers la diversité des recommandations pour les systèmes de filtrage » ; In journal Information interaction intelligence. , Vol. 6 , No. 2, pp 125-150,2006.

[Zidi et al.05] K. Zidi, S.Hammadi ., “DMAS: distributed Multi-agents System for assist users in the multimodale travals”; International Conference on Industrial engineering and System Management .IESM 2005,May 16-19, Marrakech (Morocco) , 2005..

[Shardanand et al. 95] Shardanand U. and Maes P., “ Social Information Filtering : Algorithms for Automating” “World of Mouth”; In Proceeding of the CHI-95 Conference. Denver, USA, ACM Press, May 1995.

[Hammond et al., 84] Hammond N., Hinton G., Barnard P., Maclean A., Long J. and Whiteield A.

Evaluating the interface of a document processor: a comparison of expert judgement and user observation. Proceeding of the First IFIP Conference on Human-Computer Interaction: Interact’84, London, pp. 725-729, 1984.

[Herlocker et al., 99] Herlocker J. L, Konstant J. A., Brochers A. and Riedl J., “ A algorithmic framework for perfomrming collaborative filtering”; In proc. 1999 Conf. Research and Developpement in Information retrieval, pp 230-237, Berkeley, CA, Août 1999.

[Boughanem et al.07] Mohand Boughanem, Gabriella Pasi, Henri Prade : “An Information Retrieval Driven by Ontology from Query to Document Expansion”. Proceedings of the 8th Conference on Large-Scale Semantic Access to Content (Text, Image, Video and Sound), RIAO ,2007.

[Boughanem et al.04] M. Boughanem, H. Tebri, M. Tmar., « *Filtrage d'information*. Dans : *Méthodes avancées pour la recherche d'informations* » ; Madjid Ihadjadene (Eds.), Hermès-Lavoisier, 11, Rue Lavoisier 75008 Paris, hermes science, 137-162, 2004.

[Salton.71] G. Salton., “The SMART Retrieval System: Experiments in Automatic Document Processing”; Prentice-Hall Inc, NJ,1971.
(LIN 98).

[Rapport CERTU.06] Didier Danflous., « Déploiement National des systèmes d'information multimodale » ; Rapport CERTU,2006.

(<http://www.transvilles.com>)

(<http://www.voyagessncf.com>)

(<http://www.wmdata.com>)

(<http://www.amazon.fr>)

(<http://www.transvilles.com>)

(<http://www.voyagessncf.com>)

(<http://www.wmdata.com>)

Glossaire

Atec: Association de transport en commun.

SIM: Système d'Information Multimodale.

SMA : Système Multi Agents.

SIP : Système d'Information Personnalisée.

BGP-MS: Belief,Goal and Plan Maintenance Système.

PREDIM: Plateforme de Recherché et d'Expérimentation pour le Développement de l'Information Multimodale.

PITA : Personal Intelligent Travel Assitant.

MAPIS : Multi-Agent Personalized Information System.

PerSyst : Personalization and Conceptual Modeling Method

TF-IDF: Terme Frequency-Inverse Document Frequency.

PCC: Plus Court Chemin.

SIMC: Système D'Information Multi-Agent Coopératif.

SMAAD : Système Multi-Agent d'aide au déplacement.

IHM : Interaction Homme Machine.

