

Université Mouloud Mammeri-Tizi-Ouzou
Faculté de génie électrique et informatique
Département d'informatique



Mémoire de fin d'études

En vue de l'obtention du diplôme
Master en Informatique
Option : Systèmes informatiques

Recherche d'information sur le web à base d'un profil
utilisateur

Proposé et dirigé par :

Mr DAOUI

Réalisé par :

OUBAZIZ Boudjema

Promotion 2012/2013

Listes des figures

Figure I.1. Architecture d'un système de recherche d'information.....	6
Figure I.2: Modèle de profil selon Amato	9
Figure I.3 : Méta modèle de profil.....	10
Figure I.4 : Architecture fonctionnelle d'un système d'accès personnalisé à l'information.....	15
Figure II.1. Schématisation d'une partie du graphe du Web	24
Figure II.2. Cartographie du Web	25
Figure II.3 Moteurs de Recherche.....	27
Figure II.4 Les Annuaire (Répertoires).....	28
Figure II.5 Les Meta moteurs	29
Figure II.6 Fonctionnement d'un Moteur De Recherche	30
Figure II.7 Indexation dans une page web	32
Figure II.8 résultat de l'indexation d'une page web Indexation.....	32
Figure III.1. Schéma général du système personnalisé.....	40
Figure III.2. Le prototype du fonctionnement du système.....	41
Figure III.3. Représentation multidimensionnelle d'un profil	43
Figure III.4. Sous système de gestion du profil	44
Figure III.5. Le sous système d'appariement.....	47
Figure III.6. Recherche dépendante de la localisation courante de l'utilisateur	48
Figure III.7. Extraction de la localisation courante de l'utilisateur.....	50
Figure III.8. Diagramme de cas d'utilisation "Visiteur"	59
Figure III.9. Diagramme de cas d'utilisation "Adhérent"	59
Figure III.10. Diagramme de cas d'utilisation "Administrateur"	60
Figure III.11 Diagramme de cas d'utilisation "Client"	60
Figure III.12 Diagramme de séquences du cas d'utilisation "Inscription d'un utilisateur"	61
Figure III.13 Diagramme de séquence du cas d'utilisation "recherche non personnalisée"	62
Figure III.14 Diagramme de séquence du cas d'utilisation "identification d'un Adhérent"	63
Figure III.15 Diagramme de séquence du cas d'utilisation "MAJ des préférences"	64

Figure III.16 Diagramme de séquence du cas d'utilisation “Modification des caractéristiques des hôtels	65
Figure III.17 Diagramme de classe	66
Figure IV.1. Interface principale de phpMyAdmin	72
Figure IV.2. Architecture 3 tiers.....	73
Figure IV.3. Interface d’accueil.....	74
Figure IV.4. Interface identification adhérent	74
Figure IV.5 Interface inscription abonné.....	75
Figure IV.6 Interface Abonn	76
Figure IV.7 Interface connexion administrateur.....	77
Figure IV.8 Interface d’accueil administrateur.....	77
Figure IV.9 Interface MAJ hôtels.....	78
Figure IV.10 Interface connexion client.....	79
Figure IV.11 Interface inscription client	79
Figure IV.12 Interface espace client.....	80
Figure IV.13 Organigramme de l’exécution de la recherche non personnalisée.....	81
Figure IV.14 Résultats recherche selon le contexte de la localisation	82
Figure IV.15 Caractéristiques de l’hôtel « Cotedor ».....	82
Figure IV.16 Résultats recherche sans le contexte de la localisation	83
Figure IV.17 Caractéristiques de l’hôtel « Cheraton »	83
Figure IV.18 Organigramme de l’exécution de la recherche personnalisée.....	84
Figure IV.19 Profil abonné n°1	85
Figure IV.20 Profil abonné n°2	86
Figure IV.21 Résultats recherche pour abonné n°1 selon le profil.....	87

Figure IV.22	Caractéristiques de l'hôtel « Cheraton » dans l'espace membre	88
Figure IV.23	Résultats recherche pour abonné n°2 selon le profil.....	88
Figure IV.24	Caractéristiques de l'hôtel « Atlas » dans l'espace membre.....	89
Figure IV.25	Caractéristiques de l'hôtel « Atlas » dans l'espace membre.....	90
Figure IV.26	Mise à jour du profil	91
Figure IV.27	Résultats recherche à prés mise à jour du profil	92

Liste des tableaux

Tableau I.1 Synthèse sur les domaines utilisés pour la personnalisation.....	7
Tableau I.2 Tableau des interactions de l'utilisateur	12
Tableau II.1 Comparaison entre Moteurs de recherche et Annuaire	34
Tableau II.2 : Comparaison de Syntaxes d'interrogation entre Google et Altavista	37
Tableau II.3 Comparaison de tri des résultats entre Google et Altavista	38
Tableau III.1 La structure du fichier log.....	46
Tableau III.2 Identification des cas d'utilisation.....	51

Sommaire

Introduction générale.....	1
----------------------------	---

Première partie : Etat de l'art

Chapitre I : La personnalisation de l'information.

I.1. Introduction	3
I.2. la personnalisation de l'information :	3
I.3. Les Domaines de la personnalisation :	3
I.3. 1Domaine des IHM	3
I.3. 2 Dans la communauté BD (Base de données)	4
I.3. 3 Le domaine de la recherche d'information	5
I.3. 3.1Comment fonctionne un Système de Recherche d'information(SRI).....	5
I.3. 3.2Vers une recherche d'information personnalisé.....	6
I.4. Le profil utilisateur	8
I.4. 1La notion de profil	8
I.4.2 La modélisation générique des profils.....	8
I.4.2.1 Le modèle du profil proposé par Amato.....	8
I.4.2.2 Méta modèle de profil proposé par Kostadinov et Bouzghoub.....	9
I.4.3. Le contenu du profil	10
I.4.4. Caractéristiques du profil utilisateur.....	11
I.5. Recherche d'information et le profil utilisateur.....	12
I.5.1 Exemple de profil.....	12
I.5.2 Modélisation de l'utilisateur.....	12
I.6. Les systèmes d'accès personnalisé à l'information.....	13
I.6.1. Architecture d'un système d'accès personnalisé à l'information.....	13
I.6.2. Sélection de l'information	16
I.7. Indicateurs pour la prédiction de centres d'intérêts.....	17
I.8. Conclusion.....	18

Chapitre II : Les moteurs de recherche

II.1. Introduction	19
II.2. La Cartographie du Web.....	19
II.2.1. L'adresse URL et les hyperliens	19
II.2.2 Le graphe du Web	20
II.2.3 La théorie du « Nœud papillon »	21
II.3. Les logiciels permettant la consultation des serveurs web.....	22
II.4 Les différents outils de recherche sur Internet.....	22
II.4.1. Les moteurs de recherche (<i>search engine</i>)	23
II.4.2 Les annuaires ou répertoires	23
II.4.3 Les métamoteurs	24
II.4.4 Le portail	25
II.5 Le fonctionnement d'un moteur de recherche	26
II.5.1 Le crawler	26
II.5.2. L'indexeur	27
II.5.3. L'interface Web.....	29
II.6. Les limites des moteurs de recherche	29
II.7. Avantages et inconvénients des moteurs de recherche	31
II.8. Méthodes de tri des résultats, des moteurs de recherche sur le web	32

II.8.1 Tri par pertinence.....	32
II.8.2 Tri par popularité	32
II.8.3 Tri par calcul dynamique de catégories.....	33
II.9 Cartographie des résultats	33
II.10 Comparaison entre moteurs de recherche	33
II.11 Difficultés des moteurs de recherche sur le Web.....	34
II.10. Conclusion	35

Partie II : Système proposé

Chapitre III : La conception du système

III.1. Introduction	36
III.2. Architecture générale et fonctionnement du système	36
III.3 La modélisation du profil utilisateur	38
III.3.1 Choix d'une représentation du profil.....	38
III.4 Les différents sous systèmes	40
III.4.1 sous système de gestion du profil	40
a) L'acquisition explicite du profil.....	41
b) L'acquisition implicite du profil	41
III.4.2. Le sous système de requête	42

III.4.3. Le sous système d'appariement.....	43
III.5 la géolocalisation	43
III.5.1 Objectif traité.....	43
III.5. 2 Architecture du sous-système Géo localisation	44
III.5. 3 Fonctionnement du sous –système GEO localisation	45
III.6.Modélisation des traitements	46
III.6.1 Représentation des cas d'utilisation	46
III.6.1.1 Diagramme de cas d'utilisation	54
III.6.1.2 . Diagramme de séquence	57
III.6.1.3 Diagramme de classe	61
III.6.1.4 Dictionnaire de données	62
III.6.1.5 Les modèles logiques de donnée	63
III.6.1.6 Modèle relationnel global	64
III.7 Conclusion	65
 Chapitre IV : La réalisation du système	
IV.1 Introduction	66
IV.2 Environnement de développement.....	66
IV.3. Interfaces de l'application.....	69

IV.3.1 Interface d'accueil	69
IV.3.2 Interface identification adhérent	70
IV.3.3 Interface inscription abonné	71
IV.3.4 Interface abonné	71
IV.3.5 Interface connexion Administrateur	72
IV.3.6 Interface page d'accueil Administrateur	73
IV.3.7 Interface administrateur de MAJ des hôtels	74
IV.3.8 Interface connexion client	74
IV.3.9 Interface Ajouter hôtel	75
IV.3.10 Interface espace client	76
IV.4. Mise en œuvre de la solution	76
IV.4.1 Recherche non personnalisée	76
IV.4.1 .1 Enrichir la requête d'un utilisateur selon le contexte de la localisation.....	78
IV.4.2 Recherche personnalisée	79
IV.4.2 .1 Test	81
IV.4.2 .1 Test sur la mise à jour du profil	85
IV.5 Conclusion	88
Conclusion générale	89

Annexe.....	91
Bibliographie.....	100

Les plus grandes innovations de ce siècle sont sans aucun doute l'avènement de l'informatique, du multimédia et l'essor de l'Internet. Ces innovations sont accompagnées par le développement de nouveaux supports informatiques et l'amélioration de la capacité de stockage de l'information. Depuis les années 90, c'est le service World Wide Web qui connaît le plus grand essor au niveau mondial. En effet, ce service met à la disposition de l'internaute tout type d'information organisée sous forme de documents.

Problématique

Les systèmes d'information actuels donnent l'accès à des sources de données multiples, distribuées, autonomes et potentiellement redondantes. Une des principales limites de ces systèmes est leur incapacité à discriminer les utilisateurs en fonction de leurs centres d'intérêt, de leurs préférences et de leurs contextes de requêtage, et à leur délivrer des résultats pertinents selon leurs profils respectifs. Cette limite a plusieurs conséquences pour l'utilisateur:

- Les mêmes réponses sont fournies aux mêmes requêtes quels que soient les utilisateurs qui les ont émises: les systèmes se contentant de délivrer tous les objets satisfaisant strictement les critères de la requête.
- la taille des réponses est souvent volumineuse et génère une surcharge informationnelle qui déroute ou décourage l'utilisateur dans son exploration ou sa navigation.
- la pertinence des réponses se trouve souvent réduite dans la mesure où elles ne sont pas adaptées au contexte de l'utilisateur ni à ses préférences d'ailleurs: les objets délivrés ne sont pas forcément en adéquation avec l'environnement depuis lequel interagit l'utilisateur (moment, endroit, media utilisé, etc.)
- la qualité de l'information délivrée n'est pas prise en compte; ce qui laisse souvent l'utilisateur perplexe quant à son utilité et rend difficile la prise de décision à base de cette information.

Cependant la personnalisation de l'information peut apporter des réponses satisfaisantes pour ces problèmes. Elle se présente comme une dimension qui permet la mise en œuvre de système centré utilisateur et pour cela, la personnalisation de l'information implique une modélisation des besoins de l'utilisateur sous forme d'une structure décrivant les préférences de ce dernier.

Un tel système inclut alors :

- Des modèles génériques permettant de représenter l'utilisateur, ces structures traduisent essentiellement les centres d'intérêt et le contexte de l'utilisateur.
- Des techniques qui permettent l'instanciation de ces modèles.
- Des techniques pour collecter les informations descriptives de l'utilisateur.
- Un processus d'accès à l'information intégrant les modèles liés à l'utilisateur (profil et contexte).

- Un mécanisme d'évolution de ces modèles et instances.

Dans ce cadre, on va s'intéresser aux problèmes suivants :

- Modélisation du profil (spécialisation du meta_modèle).
- Utilisation du profil.
- Evolution du profil.

Organisation du mémoire

Notre mémoire est organisé en deux parties :

- **Partie 1 : Etat de l'art :** Cette partie présente l'état de l'art sur la personnalisation de l'information ainsi que les moteurs de recherche qui existent sur le web.

Elle comporte 2 chapitres :

- ✓ Chapitre I : La personnalisation de l'information.
- ✓ Chapitre II : Les moteurs de recherche

- **Partie II : Système proposé :** Cette partie décrit notre système de recherche d'information personnalisé. Elle comprend deux chapitres :

- ✓ Chapitre III. La conception du système
- ✓ Chapitre IV. La réalisation du système

Pour clore ce mémoire, une conclusion générale est proposée afin de synthétiser notre travail et présenter d'éventuelles perspectives.

Chapitre I

La personnalisation de l'information.

I.1. Introduction

La personnalisation et la qualité de l'information constituent un enjeu majeur pour l'industrie informatique. Que ce soit dans le contexte des systèmes d'information d'entreprise, du commerce électronique, de l'accès au savoir et aux connaissances ou même des loisirs, la pertinence de l'information délivrée, son intelligibilité et son adaptation aux usages et préférences des clients constituent des facteurs clés du succès ou du rejet de ces systèmes.

A cet effet dans le présent chapitre, nous allons définir la notion de personnalisation d'information et ses domaines d'utilisations, puis nous décrivons un détail du profil utilisateur et on termine avec quelques définitions d'indicateurs participants à la mise à jour du profil.

I.2. la personnalisation de l'information

La personnalisation de l'information est une dimension qui permet la mise en œuvre d'un système centré utilisateur non dans le sens d'un utilisateur générique ⁽¹⁾ mais d'un utilisateur spécifique [1].

La personnalisation de l'information se définit, entre autres, par un ensemble de préférences individuelles représentées par des couples (attribut, valeur), par des ordonnancements de critères ou par des règles sémantiques spécifiques à chaque utilisateur ou communauté d'utilisateurs. Ces modes de spécification servent à décrire le centre d'intérêt de l'utilisateur, le niveau de qualité des données qu'il désire ou des modalités de présentation de ces données [2].

La personnalisation de l'information consiste à fournir à un utilisateur une information pertinente correspondant à ses préférences et à ses besoins.

Dans ce qui suit nous présentons quelques domaines d'application qui ont eu recours à la personnalisation :

I.3. Les Domaines de la personnalisation :

Parmi les domaines d'application qui font appel à la personnalisation on peut citer les suivants :

I.3.1 Domaine des IHM⁽²⁾ :

Dans le domaine des IHM, la notion de requête n'existe pas sous forme de langage. Les systèmes utilisent des informations sur l'utilisateur (âge, niveau d'expertise, handicaps etc.) ou sur la technologie qu'il utilise (type du media, logiciels etc.) pour lui fournir une interface d'interaction adaptée. L'ultime objectif de ces systèmes est de guider l'utilisateur dans ses recherches et de faciliter l'expression de ses besoins. Un exemple de tel système est

⁽¹⁾ Utilisateur générique : c'est l'utilisateur qui n'a pas de profil.

⁽²⁾ IHM (interface homme machine) : moyens et outils mise en œuvre à fin qu'un humain puisse contrôler et communiquer avec une machine.

‘Apt Décision’ qui représente un agent de recherche d’appartements (Shearin & Lieberman 2001). L’utilisateur soumet un certain nombre de critères de recherche (nombre de pièces, surface etc) et ensuite le système guide l’utilisateur, selon ses interactions, à travers les annonces disponibles. A chaque étape, le système analyse les actions effectuées par l’utilisateur pour lui proposer, dans la prochaine itération, des réponses conformes à ses préférences. Le contenu du profil d’un utilisateur varie selon les approches et les applications.

Dans le domaine IHM, le profil contient, par exemple, des informations permettant au système d'adapter l'affichage des résultats selon les préférences de l'utilisateur. Un exemple simple d'un tel profil, donné par Bouzeghoub & Kostadinov [3], est celui utilisé par les fournisseurs de services Web. Dans ces systèmes, le profil d’un utilisateur représente un ensemble de données personnelles (nom, prénom, langue, genre, date naissance, code postal, e-mail, profession, poste etc..., dans le cas de Yahoo par exemple) et des catégories d’intérêts qui constituent sa page d’accueil (ex. météo, football, jeux, etc.).

Le contenu de certaines catégories du centre d’intérêt peut être défini à partir des données personnelles (par exemple si l’utilisateur est né au début du mois de Juin et s’il a choisi la catégorie ‘Horoscope’, le zodiaque affiché sera Gémeaux).

La notion de profil, souvent appelé modèle d'utilisateur, se focalise plus sur le niveau d'expertise et le métier de l'utilisateur afin de déterminer le type de dialogue que le système va avoir avec lui, les métaphores graphiques les plus appropriées ainsi que les modalités de livraison des résultats qu’il attend du système d'information.

I.3.2 Dans la communauté BD (Base de données) :

Dans le domaine des BD, il n’est pas courant d’intégrer l’utilisateur dans le processus de recherche d’informations. Une requête SQL contient en général l’ensemble des critères jugés utiles à une sélection de données pertinentes. Les profils sont alors intégrés directement aux requêtes par les utilisateurs ou lors de la compilation de ces dernières ; ils sont alors pris en compte en une seule fois durant le filtrage de l'information.

Dans le domaine des BD, le profil de l’utilisateur contient des données qui expriment ses habitudes, des prédicats fréquemment utilisés dans ses requêtes ou des définitions de l’ordre de préférences de ces prédicats Koutrika & Loannidis 2004 [4]. Par exemple, le scénario de l’agence de voyage met en œuvre un agent logiciel capable de planifier un voyage complexe impliquant plusieurs lieux, plusieurs moyens de transport, l’inscription à des activités tout en résolvant des contraintes hétérogènes (dormir dans un hôtel convenable, minimiser le coût d’ensemble du voyage, diminuer le temps de correspondance, utiliser les transporteurs favoris et trouver des repas végétariens...). Imaginons que l'utilisateur d'une agence de voyage ait à résoudre le problème suivant :

Paul habite à Paris et il aime voyager pendant le week-end. Il descend d’habitude dans des hôtels au centre ville et préfère voyager en train plutôt qu’en car.

La solution proposée par l'agent doit mettre l'intérêt de l'utilisateur (Paul) au centre des recherches et exprimer par un *degré* qui est un nombre réel compris entre 0 et 1 l'intérêt de cet utilisateur. Sur chaque expression du profil, considérée comme une sous requête, est ajoutée un nombre compris entre 0 et 1 pour exprimer l'importance relative de cette expression par rapport aux autres. Ainsi la valeur 1 sur les trois premières expressions signifie que ces conditions doivent être toujours satisfaites.

Les autres expressions expriment le fait que l'utilisateur a une plus forte préférence pour les hôtels situés au centre ville (d) que pour les voyages de deux jours (e) et qu'il préfère voyager en train (f) plutôt qu'en car (g). Cet exemple exprime l'idée de personnalisation d'un voyage pour un utilisateur X, l'ensemble des préférences de ce dernier doit être considéré comme des contraintes incluses dans la requête elle-même.

Dans notre cas d'étude le domaine qui nous concerne beaucoup plus c'est la recherche d'information que nous allons introduire dans le point suivant :

I.3.3 Le domaine de la recherche d'information :

Les performances de la recherche d'information classique dépendent en grande partie de l'efficacité des processus d'indexation, d'appariement et de façon non négligeable, de la perception de l'utilisateur. Or aucun mécanisme explicite n'y est intégré pour considérer effectivement ce dernier dans le processus de sélection de l'information pertinente. C'est pourquoi les travaux ont visés la conception de système de recherche dit adaptatifs. L'objectif de ces systèmes est d'exploiter des informations extraites de l'utilisateur dans le but d'y adapter une ou plusieurs phases du processus de recherche.

Avant de présenter le fonctionnement d'un système de recherche d'information il est nécessaire de savoir qu'un système de recherche d'information SRI est un système qui gère une collection d'informations organisées sous forme d'une représentation intermédiaire reflétant aussi fidèlement que possible le contenu des documents grâce à un processus préalable d'indexation, manuelle ou automatique.

I.3.3.1. Comment fonctionne un Système de Recherche d'information(SRI)

Le fonctionnement d'un SRI peut être divisé en quatre étapes :

- **Le traitement des documents** : C'est l'étape de l'ajout des documents au système et de la construction du fichier inverse⁽³⁾.
- **Le traitement des requêtes** : Cette étape concerne surtout les systèmes statistiques et de T.L.N. (Traitement du Langage Naturel), qui doivent accomplir en aval un travail

⁽³⁾ Fichier inverse : terme utilisé en documentation pour désigner les index donnant accès aux enregistrements d'un fichier de références documentaires informatisé. Par opposition à fichier direct.

qui est partiellement accompli en amont par le chercheur en ce qui concerne les systèmes booléens: rendre les requêtes compréhensibles par la machine.

- **L'appariement des requêtes (*query matching*)** : Cette étape concerne la mise en correspondance des requêtes avec le fichier inverse et, le cas échéant, la base de connaissances.
- **La présentation des résultats** : Elle peut se faire par date, par champ ou par pertinence présumée par rapport à la requête.

La figure I.1 représente clairement l'architecture générale d'un système de recherche d'information.

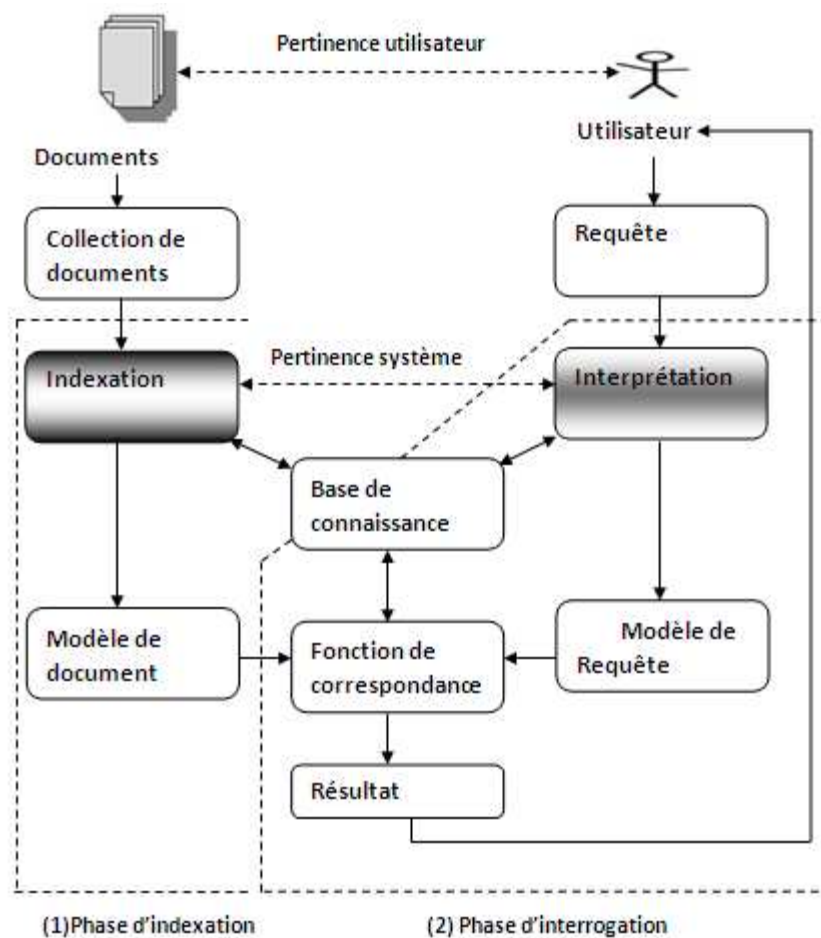


Figure I.1. Architecture d'un système de recherche d'information [5].

I.3.3.2. Vers une recherche d'information personnalisée

Le développement de système d'accès personnalisé à l'information est une solution apportée à la recherche adaptative dans un contexte de prolifération de sources d'informations volumineuses et hétérogènes.

La mise en œuvre d'un système de recherche d'information personnalisée (SRIP) s'effectue en premier par la modélisation du profil de l'utilisateur, puis son intégration dans le processus de recherche, les premières propositions dans ce sens, sont la reformulation de requête et le filtrage d'information.

➤ **Reformulation de requête**

Elle consiste à utiliser la requête initiale pour amorcer la recherche, puis à modifier celle-ci à partir des jugements de pertinence et/ou de non pertinence de l'utilisateur dans le but de répondre les termes de la requête initiale ou d'y ajouter (respectivement supprimer) d'autre termes contenus dans les documents pertinents (respectivement non pertinents). La nouvelle requête ainsi obtenue à chaque itération, permet de corriger la direction de la recherche dans le sens des documents pertinents.

➤ **Filtrage d'information**

Il consiste à comparer le contenu des documents et le profil de chacun des utilisateurs pour déterminer les documents à proposer. On distingue deux modes de filtrage qui se distinguent selon le mode d'utilisation et de construction du profil utilisateur :

✓ **le filtrage collaboratif**

Consiste à filtrer l'information destinée à un utilisateur en se fondant sur la corrélation entre son propre profil personnel et les profils des autres utilisateurs qui présentent des intérêts et préférences semblables. Le profil du groupe est construit à partir des documents consultés, jugés et annotés par les membres de groupe. Parmi les systèmes basés sur ce type de filtrage, on peut citer GroupLens et Fab présents dans [6].

✓ **le filtrage explicite**

Il est basé sur un profil utilisateur dont le contenu est décrit explicitement par l'utilisateur en spécifiant ses centres d'intérêts à l'aide d'une liste de mots clés, exemple : Les portails Web tel que francicité (*cf Partie I, Chap.02, § II.4.4.*).

Dans notre cas, nous nous sommes penchés sur la reformulation, qui est un processus évolutif, ayant pour objectif la génération de nouvelles requêtes plus adaptées que celles initialement émise par l'utilisateur.

Le tableau suivant nous résume brièvement ces domaines :

	IHM	BD	RI
Principe	Interface d'interaction adaptée	Utiliser des requêtes SQL	Utilisation des règles sémantiques
Profils utilisateur	Préférences données par l'utilisateur	Contraintes incluses dans les requêtes SQL	Les actions, les préférences et les comportements de l'utilisateur

Mode de personnalisation	Interrogation de l'utilisateur pour personnaliser l'interface	Utiliser les données qui expriment les habitudes, les prédicats fréquemment utilisés dans les requêtes de l'utilisateur	Utiliser les connaissances nécessaires à une production d'une information pertinente (couples (attribut, valeur) représentant les préférences de l'utilisateur
Objectif	Adaptation de l'interface	Trouver les critères jugés utiles à une sélection de données pertinentes.	une production d'une information pertinente adaptée à chaque utilisateur

Tableau I.1 Synthèse sur les domaines utilisés pour la personnalisation

Nous remarquons d'après l'étude précédente que le profil utilisateur est indispensable pour réaliser la personnalisation de l'information, à cet effet nous allons décrire à la profondeur la notion de profil d'utilisateur dans les paragraphes suivants.

I.4. Le profil utilisateur

La notion du profil utilisateur est apparue vers les années 80 avec les assistants et les agents d'interfaces. Cette apparition est due principalement au besoin de créer des applications personnalisées capables de s'adapter à l'utilisateur [7].

Personnaliser une application pour un utilisateur particulier nécessite de disposer d'informations sur ce dernier, ce qui permettra d'évaluer la pertinence des objets disponibles ou d'aider le système à faire des choix. Le modèle utilisateur est « une source de connaissances ou une base de données sur un utilisateur » [8].

Pour plus de précision nous allons voir la définition de profil, quelques modèles utilisés par la communauté de RI et leurs contenus.

I.4.1. La notion de profil

Un profil utilisateur est une collection d'informations sur l'utilisateur. Cette collection peut être vue comme un ensemble de caractéristiques avec des valeurs associées contenant par exemple ce que l'utilisateur préfère, ce qu'il est capable de faire...etc.

On doit également prendre en compte l'historique des actions de l'utilisateur, voir leur évolution dans le temps. [9].

Un profil utilisateur regroupe l'ensemble des connaissances nécessaires à une évaluation efficace des requêtes et à une production d'une information pertinente adaptée à chaque utilisateur [10].

I.4.2 La modélisation générique des profils

Le contenu du profil d'un utilisateur varie selon les approches et les applications, mais il manque un modèle donnant une vision globale sur tous les aspects de la prise en compte des préférences des utilisateurs. Notre but est d'avoir une vision globale sur les paramètres qui interviennent dans la personnalisation.

I.4.2.1 Le modèle du profil proposé par Amato

Amato [11] propose un modèle de contenu d'un profil utilisateur et son application pour la réalisation d'une librairie digitale (recherche et livraison personnalisé de l'information sur le web). C'est la première approche où les informations sont structurées et qui offre un modèle général. Le modèle de profil contient cinq catégories (**Figure I.2**).

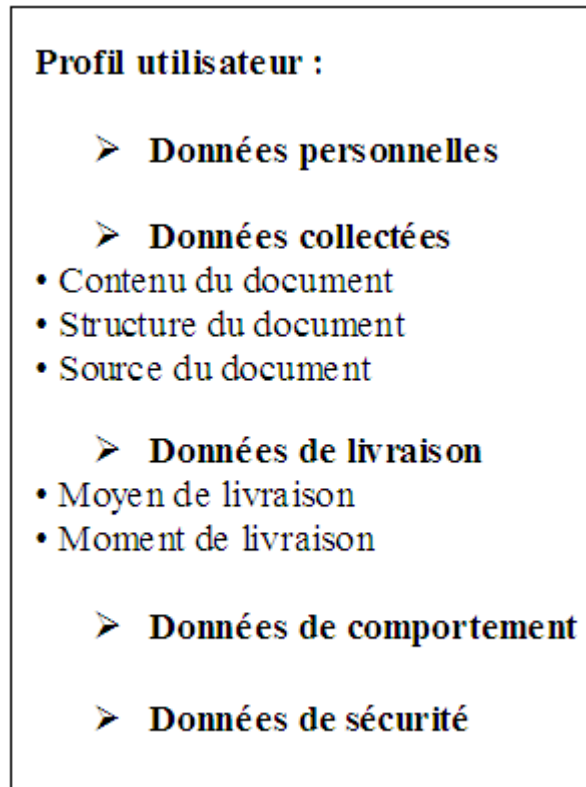


Figure I.2: Modèle de profil selon Amato [11].

I.4.2.2 Méta modèle de profil proposé par Kostadinov et Bouzghoub

Suivant la classification précédente et en appuyant sur le modèle décrit par Amato, l'objectif de Bouzghoub et Kostadinov est de proposer un ensemble de dimensions ouvertes, capables d'accueillir la plupart des informations caractérisant un profil [3]

La **Figure I.2** donne un aperçu de la structure générique d'un profil, il est composé de plusieurs dimensions. Chaque dimension est constituée d'un ensemble d'attributs éventuellement organisés en entités. Les attributs peuvent être simples ou composés. Les attributs composés sont appelés des sous-dimensions. Une sous dimension regroupe un ensemble d'attributs simples qui sont liés sémantiquement (par exemple l'adresse est composée du numéro de la rue, du nom de la rue, du code postal etc.).

Chaque attribut simple est caractérisé par son nom, le type de ses valeurs et la structure des valeurs. Le type d'un attribut peut être un des types couramment utilisés: entier, réel, chaîne de caractères, etc. La seconde caractéristique (la structure des valeurs) décrit le modèle de représentation des valeurs. Ce modèle de représentation peut être une valeur unique, un ensemble, un intervalle, un vecteur etc. À chaque attribut simple peuvent être associées une ou plusieurs valeurs. Et nous nous sommes inspirés par ce model dans notre conception

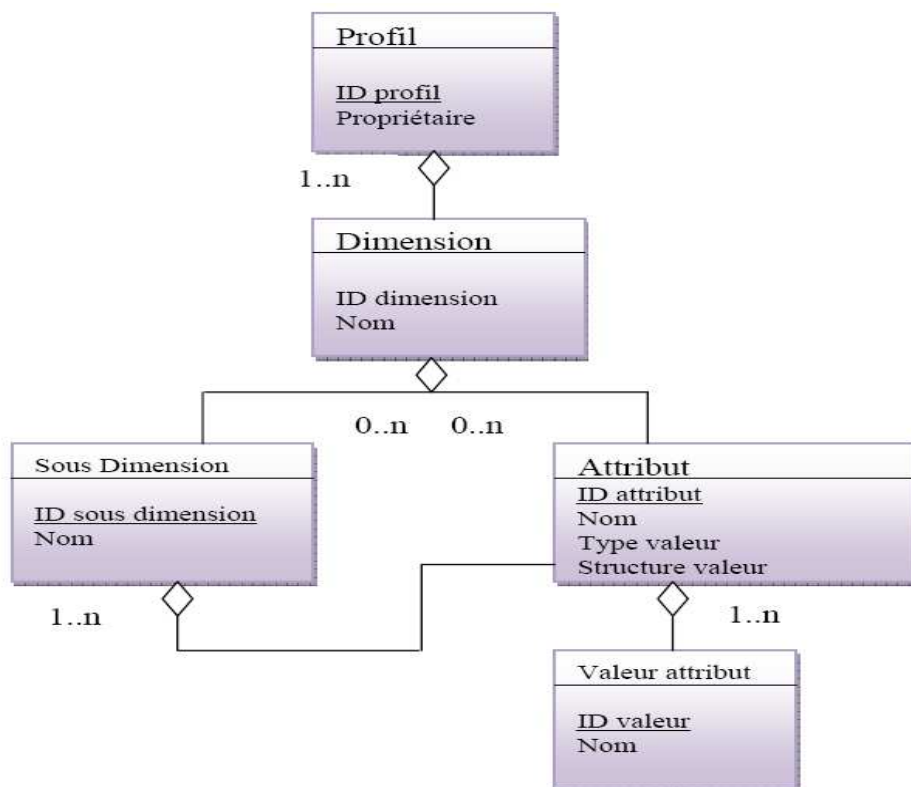


Figure I.3 : Méta modèle de profil [3]

I.4.3. Le contenu du profil

Chaque système de personnalisation propose un modèle de profil adapté au service qu'il offre. De façon formelle les données stockées dans les profils peuvent être classées selon quatre catégories :

- **Des exemples d'objets de l'espace de recherche :**

Ces modèles de profils sont utilisés en général par les systèmes de recommandation.

L'idée principale est de stocker des exemples d'objets que l'utilisateur a jugé comme pertinents ou comme inintéressants pour ensuite comparer le contenu des profils des différents utilisateurs entre eux [12,13, 14]. S'il y a des similarités entre deux profils le système va recommander à l'utilisateur des objets que d'autres clients de profils semblables ont appréciés. Comme dans certains cas les objets peuvent être volumineux (comme des livres, rapports etc.), dans ces cas il est possible de ne stocker que leurs identifiants.

➤ **Des caractéristiques extraites des objets de l'espace de recherche**

L'idée de cette approche est de trouver une forme normale permettant la représentation des profils et des objets de l'espace de recherche. Ceci est fait le plus souvent en analysant les objets pour extraire les caractéristiques pertinentes pour l'utilisateur [14, 15, 16, 17,18]. Pour la plupart des approches existantes ces caractéristiques sont des mots clés. Dans certains cas les mots clés sont combinés avec d'autres informations portant sur l'importance des mots pour l'utilisateur (fréquences d'apparition, poids, votes etc.).

Ferreira propose un système où les profils sont représentés de la même façon que les documents. Un profil (comme un document) est présenté comme un vecteur à N dimensions où les dimensions sont définies par les termes des centres d'intérêts. Les poids associés aux dimensions correspondent au degré de leur intérêt. Le but visé dans ce cas est la comparaison directe entre les profils et les éléments recherchés qui simplifie la technique de personnalisation. Seuls les documents dont la similarité avec le profil dépasse un certain seuil sont inclus dans le résultat [14].

➤ **Des attributs décrivant les préférences externes de l'utilisateur**

Les informations des deux catégories précédentes sont directement liées aux objets de l'espace de recherche et à leur contenu, mais ne portent aucune information extrinsèque sur la qualité des informations ou sur l'utilisateur. Dans cette catégorie d'informations on retrouve les préférences de l'utilisateur liées aux sources des données, la qualité des informations et le type et la structure des éléments recherchés [20, 21]. Les données personnelles font également partie de cette classe de données.

➤ **Des règles et des formules de préférences**

Dans certaines approches [22, 23, 24] le profil de l'utilisateur est remplacé par des formules de préférence qui permettent de définir un ordre entre les éléments selon la pertinence de ces éléments pour l'utilisateur. Dans d'autres systèmes comme par exemple Personnalization server [20] les préférences des utilisateurs sont capturées par des règles actives qui tentent de découvrir des généralités en regroupant les utilisateurs selon les caractéristiques comme l'âge, le sexe, les données démographiques, le nom du domaine, le type du browser, le système d'exploitation, la bande passante disponible, les activités personnelles ou la situation familiale.

I.4.4. Caractéristiques du profil utilisateur

➤ **L'acquisition explicite du profil utilisateur**

Les traits qui caractérisent le profil utilisateurs sont : ses références ou acquis (background), ses connaissances du sujet, ses objectifs (goals), ses préférences et son expérience. Peter Brusilovsky a cité deux autres caractéristiques qui sont : ses centres d'intérêts (interests) et ses traits individuels [25].

✓ **Centres d'intérêts** : l'utilisateur peut les formuler à travers l'introduction d'un ensemble de mots clés, pour décrire ses centres d'intérêts. Sur la base de ces mots clés, le système pourra lui recommander des informations qui répondent au mieux à ses attentes.

✓ **Traits individuels** : ce sont des caractéristiques stables, les traits individuels englobent les facteurs de personnalité. Ces traits ne sont pas tirés d'un simple entretien mais après des tests psychologiques.

➤ **L'acquisition implicite du profil utilisateur**

Traditionnellement la méthode de retour de pertinence exige de l'utilisateur qu'il donne explicitement son jugement à travers la spécification des mots clés ou par le marquage de documents ou encore par la réponse à des questions sur ses centres d'intérêts. Une telle méthode exige beaucoup de la part de l'utilisateur et le force à s'engager dans une activité parallèle en plus de sa recherche, ce qui risque de lui créer une surcharge cognitive. De plus, le profil d'un utilisateur doit contenir des informations précises sur ses préférences, or souvent il y a un décalage entre l'intention de l'utilisateur et ce qu'il désire réellement.

C'est la raison pour laquelle certains systèmes misent sur un processus de découverte implicite des centres d'intérêts de l'utilisateur de manière interactive et incrémentale.

Le fonctionnement de base de cette approche est réalisé par l'établissement d'un dialogue entre le système et l'utilisateur ou mieux encore en observant son comportement à travers ses différentes interactions avec le système pour récolter discrètement l'information nécessaire sur lui [26].

Dans le point suivant, nous allons donner un exemple de profil et les différents modèles de l'utilisateur dans le domaine de la recherche d'information.

I.5 Recherche d'information et le profil utilisateur

La pertinence de l'information se définit par un ensemble de critères et de préférences personnalisables spécifiques à chaque utilisateur ou communauté d'utilisateurs. Les données décrivant les utilisateurs sont souvent regroupées sous forme de profils.

I.5.1. Exemple de profil :

Dans le domaine de la RI, le profil de l'utilisateur décrit le plus souvent son centre d'intérêt et, de ce fait, est souvent confondu avec la requête de l'utilisateur. Ce profil est généralement défini à l'aide d'un vecteur de mots clés avec éventuellement un poids associé à chaque mot [14, 17, 16].

Dans le projet CASPER [13] qui présente un moteur de recherche d'emploi, le profil d'un utilisateur est défini sous la forme des statistiques des actions qu'il a effectuées sur les

offres d'emplois. L'intérêt de l'utilisateur pour une annonce est déterminé en fonction du temps qu'il a passé à la lire et du type d'action qu'il a effectuée dessus.

Prenons par exemple un utilisateur qui a lu une annonce de travail, a postulé pour une autre et a envoyé une troisième à un ami.

Annonce	Action	Nombre de clicks	Temps de lecture
Job 5	Lire	1	234
Job 56	candidater	2	186
Job 45	Envoyer à un ami	1	54

Tableau I.2 Tableau des interactions de l'utilisateur

En utilisant un ensemble de règles de décision, le système va décider quelle annonce est pertinente pour l'utilisateur. Par exemple il peut considérer que les annonces 'job56' et 'job45' sont pertinentes pour l'utilisateur en raison des actions effectuées (candidater et envoyer à un ami). Par contre il n'a fait que lire l'annonce 'job5' qui sera considérée comme inintéressante.

I.5 2. Modélisation de l'utilisateur

On cite principalement quatre types de représentation : vectorielle, sémantique, connexionniste et multidimensionnelle.

➤ **représentation vectorielle :**

Ce type de représentation s'appuie généralement sur le modèle vectorielle [27]. Le profil est représenté par un ou plusieurs vecteurs défini dans un espace de termes obtenus implicitement ou explicitement à partir de plusieurs sources d'information. Les coordonnées des vecteurs correspondent au poids des termes dans le profil. L'utilisation de plusieurs vecteurs permet de prendre en compte la diversité des domaines d'intérêts ou évolution dans le temps. Il offre l'avantage indéniable de la simplicité de mise en œuvre.

Cependant les modèles proposés ne mettent pas en évidence ni la dimension liée au temps marquant l'évolution des profils, ni à l'organisation des informations pour hiérarchiser les centres d'intérêts.

➤ **Représentation sémantique :**

La représentation sémantique met d'avantage en relief les relations de sens entre unités d'information qui représentent le profil en apportant des solutions aux problèmes de dissémination et synonymie. La direction proposée dans ce contexte, est la construction hiérarchique de concepts plutôt qu'une liste de structures indépendantes, à partir d'informations issus des fichiers *logs* la hiérarchie peut prendre compte des niveaux de

préférences de l'utilisateur et des associations latentes entre concepts et donc un raisonnement sémantique pour la dérivation du profil s'y prête aisément. Ce type de représentation utilise généralement des ontologies [28].

➤ **Représentation connexionniste :**

C'est un type de représentation basé sur l'interconnexion de nœuds représentant les termes, préférences [29] ou documents. Il offre le double avantage de la structuration et de la représentation associative permettant de considérer l'ensemble des aspects représentatifs du profil.

➤ **Représentation multidimensionnelle :**

La représentation multidimensionnelle du profil s'inscrit dans une réflexion globale sur la personnalisation de l'information, en partant du principe que la modélisation de l'utilisateur doit pouvoir capturer toutes les dimensions qui représentent l'utilisateur.

Différents travaux ont abordé cet aspect sans le couvrir dans son ensemble. Ainsi les propositions de standards P3P [30] pour la sécurisation des profils ont défini des classes distinguant les attributs démographiques des utilisateurs (identité, données personnelles), les attributs professionnels (employeur, adresse, type) et les attributs de comportement (trace de navigation). Une autre proposition faite par [20] consiste à représenter le contenu du profil utilisateur par un modèle structuré de dimensions prédéfinies.

Nous avons opté pour une représentation multidimensionnelle, le choix est justifié par le fait que le profil est caractérisé par plusieurs catégories d'informations c'est-à-dire plusieurs dimensions.

Après avoir présenté et défini la personnalisation et le profil d'utilisateur nous allons passer aux systèmes d'accès personnalisé à l'information

I.6. Les systèmes d'accès personnalisé à l'information

Un système d'accès personnalisé à l'information est un système qui intègre l'utilisateur, en tant que structure informationnelle, tout au long de la chaîne d'accès à l'information.

I.6.1. Architecture d'un système d'accès personnalisé à l'information

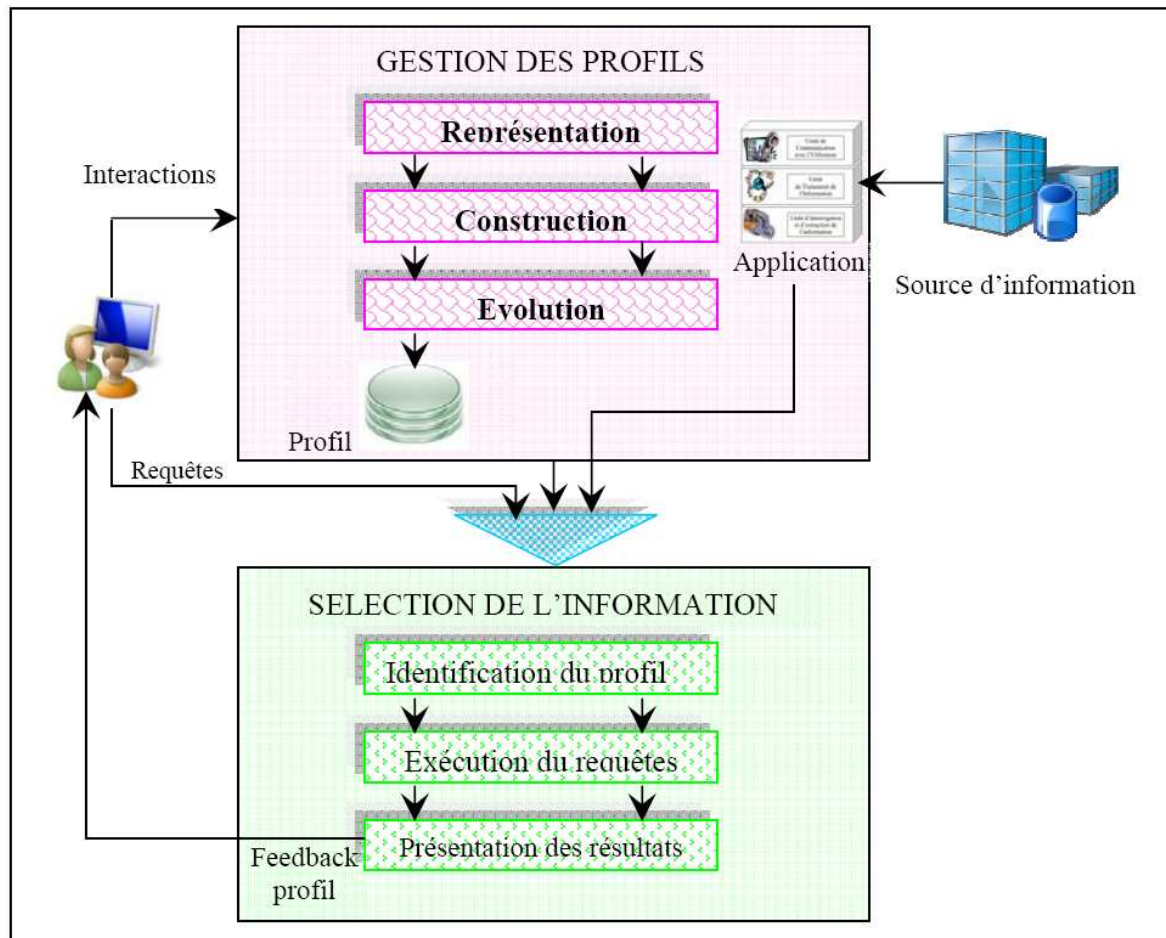


Figure I.4 : Architecture fonctionnelle d'un système d'accès personnalisé à l'information [31].

Cette architecture est centrée autour de l'utilisateur, sur la base de cette architecture ils se dégagent principalement deux fonctions fondamentales qui sont la gestion des profils et la sélection de l'information.

Les principales fonctionnalités du gestionnaire de profils sont :

➤ **Représentation du profil**

Le profil de l'utilisateur n'a pas forcément de structure explicite qui le représente. Il peut être constitué de paquets divers d'informations qui traduisent une connaissance éparse sur l'utilisateur. Dans ce sens, la représentation des profils rejoint en grande partie la représentation de l'information dans le contexte de la recherche d'information. Il n'existe pas un modèle spécifique dédié à la représentation du profil de l'utilisateur. Les modèles proposés puisent largement de ceux proposés en recherche d'information on en cite principalement quatre types de représentation : vectorielle, sémantique, connexionniste et multidimensionnelle (*cf Partie I, Chap.01, § I.5 2.*).

➤ **Construction du profil**

La construction des profils se décompose en 2 étapes :

✓ **La modélisation de profil**

La modélisation est l'une des premières fonctionnalités que doit offrir un système de personnalisation considérant le profil comme une forme de connaissance sur les habitudes et préférences de l'utilisateur. L'existence d'une méta base facilitera l'élaboration de cette connaissance en offrant une typologie de connaissance et les mécanismes de spécialisation et d'instanciation qui permettent d'adapter totalement ou partiellement la méta base aux besoins de personnalisation de l'utilisateur et/ou des applications.

✓ **Instanciation du profil**

C'est de prendre le Meta modèle, et de lui affecter des valeurs précises en fonction de la connaissance actuelle que le système a de l'utilisateur. Ces valeurs peuvent donc s'affiner au fur et à mesure que cette connaissance augmente. Ceci peut être fait par le module de génération automatique, par apprentissage ou datamining par exemple, ou manuellement par l'utilisateur via l'interface graphique [32].

➤ **Evolution du profil**

✓ **Mise à jour des profils**

Le gestionnaire de profils offre des fonctionnalités de mise à jour classique qui sont : L'insertion, la suppression et la modification de la structure et des valeurs des profils. La seule règle à respecter est que le modèle de profil générique doit inclure l'union des structures de toutes les instances de profils. Pour maintenir le profil générique le plus complet possible, chaque insertion d'une nouvelle dimension, sous dimension ou attribut dans une instance de profil se fait par le modèle générique. Par exemple si un utilisateur veut ajouter un attribut à ces données personnelles, il est obligé de créer cet attribut dans le modèle générique de profil utilisateur pour ensuite le recopier dans son profil. De cette manière la typologie des données est préservée (par exemple le nom d'un utilisateur est une chaîne de caractères dans l'ensemble des instances de profils utilisateurs) [32].

✓ **Archivage des profils**

L'archivage de profils permet de sauvegarder le contenu d'une instance de profil. Lors de cette sauvegarde, la date de création de l'archive est sauvegardée avec l'ensemble des valeurs du profil. Plusieurs archives peuvent être créées pour le même profil ce qui permet de suivre l'évolution des préférences d'un utilisateur ou d'analyser son comportement. Les archives peuvent être visualisées et il est possible de remplacer le contenu actuel du profil par une de ses archives. Ceci peut être très utile si on veut invalider une mise à jour non pertinente du profil utilisateur et revenir à une version précédente [32].

I.6.2. Sélection de l'information

Le profil d'un utilisateur est défini pour adapter le système de recherche d'information à cet utilisateur. Cette phase consiste à intégrer le profil de l'utilisateur dans le processus de recherche d'information, les informations contenues dans le profil sont exploitées pour réécrire puis exécuter la requête enfin de visualiser les résultats de la requête.

Pour pouvoir enrichir notre profil et le mettre à jours il va falloir acquérir des indicateurs pour la prédiction des centres d'intérêts et c'est le point qui sera présenter dans le paragraphe suivant.

I.7. Indicateurs pour la prédiction de centres d'intérêts

La détection implicite des préférences et des centres d'intérêts se fait à travers les comportements observables collectés par le système lors de l'interaction de l'utilisateur avec l'environnement (session de travail, consultation d'un document clé ...etc.).

Dans ce contexte plusieurs comportements peuvent êtres examinés, tels que :

- ✓ Clic de la souris ;
- ✓ Défilement en utilisant la souris ou le clavier ;
- ✓ Temps passé sur une page Web ;
- ✓ Temps de lecture d'un document ;
- ✓ Importance d'un document ;
- ✓ Impression du document ;
- ✓ Sauvegarde d'un document ;
- ✓ Copier/coller tout ou une partie d'un document ;
- ✓ Annotation de document ou de quelque uns de ses passages ;
- ✓ Evaluation qualitative d'un document
- ✓ Suivi des liens hypertextes pour arriver à une page Web ou à un document ;
- ✓ Visualisation et consultation d'un document.

Pour notre cas on a choisi la visualisation et consultation d'un document, impression du document et l'importance d'un document.

➤ Enregistrement des traces d'interaction

Les traces des interactions de l'utilisateur avec le système sont construites à partir de données qui sont produites par des spyware, ce dernier enregistre discrètement les actions de l'utilisateur durant l'utilisation des logiciels ou les traces d'interactions de celui-ci avec l'environnement de travail dans un fichier log(journal). Ces fichiers logs ont plusieurs caractéristiques qui font d'eux un moyen idéal pour le domaine de la recherche en vue de concevoir par exemple des interfaces homme/machine, adaptées. Ils permettent la collection des données sur de nombreux utilisateurs à travers le temps et à chaque fois que le logiciel est utilisé.

Pour chaque évènement un ensemble de descripteurs est sauvegardé, la cible, le contenu, le contexte de l'évènement et les valeurs descriptives des composants ayant contribués dans l'évènement, tels que le nom du bouton ayant été appuyé, son texte descriptif, etc. Dans le cadre de l'environnement d'un navigateur Web, les évènements enregistrés

seraient le mouvement de la souris, le clic sur un bouton (avec son nom « File », « Back », « Forward », « Reload »), l'usage de la barre de défilement (« Up », « Down »), le clic sur un emplacement du texte ou dans un endroit particulier de la page Web, le clavier avec un ensemble de touches utilisées, le clic sur des liens ou la fermeture du navigateur ou du document.

I.8. Conclusion

A travers ce chapitre, Nous avons constaté que la personnalisation est une approche capitale dans le développement des systèmes futurs. Nous avons défini la notion de profil utilisateur, les données contenues dans le profil qui doivent être correctes et justes car elles ont un impact direct sur les résultats qui vont être obtenus. Le fait que l'utilisateur a souvent des idées floues sur ses préférences complique l'extraction des caractéristiques pertinentes et nécessite un processus de découverte de ses préférences. Ce processus doit être fait avec l'implication minimale du côté du client et en même temps de fournir des données correctes par rapport à l'utilisateur. La construction et la mise à jour du profil sont des problèmes qui entre dans le cadre de nos objectifs.

Nous allons aborder dans le prochain chapitre les différents moteurs de recherche qui existe sur le web et leurs fonctionnements.

Chapitre II

Les moteurs de recherche

II.1. Introduction

Le développement considérable qu'a connu Internet a conduit à une croissance exponentielle du nombre de pages accessibles aux utilisateurs. De quelques centaines de milliers de pages ou documents de texte accessibles en 1993, le stock sur Internet atteint aujourd'hui quelques milliards de pages [31], avec un taux de croissance estimé à 1 million des pages supplémentaires par jour et dont la plupart sont reliées entre elles par des hyperliens.

En effet, la recherche d'information est une occupation qui est loin d'être négligeable. Plus de la moitié des internautes passent plus de 70% de leur temps de connexion à rechercher de l'information et plus de 80% ne regardent que les tous premiers résultats retournés par les outils de recherche lors d'une requête. [31],

C'est pourquoi la problématique de tout outil de recherche est de trouver un compromis entre le retour de l'information qui doit être quasi-immédiat, avec un bon rappel, et la pertinence des résultats retournés avec une bonne précision (annexe A).

Dans ce chapitre nous allons aborder les moteurs de recherche sur le web et leurs fonctionnement en commençons par : un aperçu générale sur web en suite sur les différents outils de recherche sur Internet et enfin sur le fonctionnement d'un moteur de recherche.

II.2. La Cartographie du Web

II.2.1. L'adresse URL et les hyperliens :

La structure du Web est basée sur le principe des liaisons multiples entre les pages qui le composent.

Toutes les pages sur le Web ont une adresse unique appelée URL (Unified Resource Locator). L'URL est une suite de caractères (sans espace) qui définit l'endroit (locator) où se trouve un document (ressource).

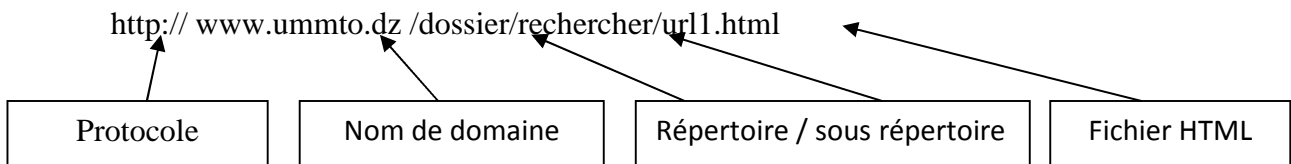
L'URL est l'adressage standard de n'importe quelle page ou document, sur n'importe quel serveur connecté sur Internet.

Il se compose :

- ✓ Du protocole de communication
- ✓ Du nom du serveur (Domain Name) où se trouve le document
- ✓ Du nom du répertoire (éventuellement) et du nom de la page

Et il s'écrit :

protocole://serveur/répertoire/page.extension



Le HTTP (Hyper Text Transfert Protocol) est le protocole de communication, entre un navigateur (le client) et un serveur connecté sur Internet, permettant le transfert des pages Web. Pages localisées grâce à l'adresse URL.

Il existe d'autres protocoles sur Internet, comme le FTP qui permet de télécharger des fichiers à distance, ou comme le Mailto pour le courrier électronique.

Le langage permettant de décrire les pages Web est le HTML (Hyper Text Markup Language). Ce langage permet de doter certains mots ou certaines images, appelés des ancres, de la fonction d'hyperlien. Cette fonction permet, par un simple clic de souris, de passer directement d'une partie d'un texte à une autre ou d'une page à une autre n'importe où sur Internet.

II.2.2 Le graphe du Web

La structure des liens logiques de toutes ces images, pages, documents d'informations sur Internet est irrégulière et a priori inconnue, d'autant qu'elle est changeante : des nouvelles pages sont créées et d'autres disparaissent en continu.

On pourrait voir cette structure comme un graphe étiqueté :

- ✓ les nœuds représentant l'URL des pages sur Internet
- ✓ les arcs représentant les hyperliens.

Et comme un graphe non connexe, c'est-à-dire que tous les nœuds ne sont pas forcément reliés entre eux.

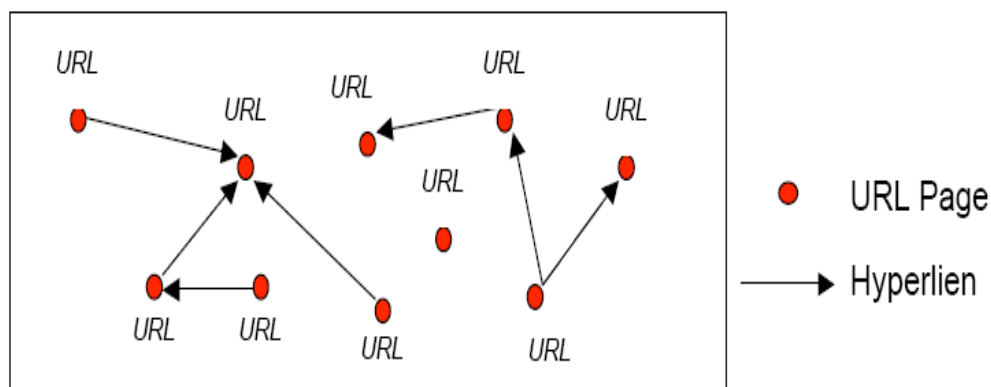


Figure II.1. Schématisation d'une partie du graphe du Web

En ce qui concerne la taille réelle du graphe, on estime :

- ✓ Qu'il contient plus de 500 milliards de nœuds (de pages),
- ✓ Que chaque nœud possède en moyenne sept liens entrant ou sortant vers d'autres pages.

Par conséquent on obtient un graphe composé de quelques 35 mille milliards de liens. Etant donné l'immensité de ce graphe, et sa progression constante tant en nombre des nœuds qu'en nombre de liens, il est souvent perçu comme étant infini et sans réelle structure.

Cependant une équipe de chercheurs d'*AltaVista*, de *Compaq* et d'*IBM* a établi une cartographie du Web. Contrairement à l'image consacrée et citée précédemment,

Internet n'aurait pas la forme « d'une toile d'araignée », mais plutôt celle d'un « nœud papillon »

II.2.3 La théorie du « Nœud papillon » :

L'étude réalisée, entre mai et octobre 1999, par des chercheurs d'*AltaVista*, de *Compaq* et d'*IBM*, sur près de 500 millions de pages Web, a permis de construire le schéma suivant décrivant la connectivité d'Internet

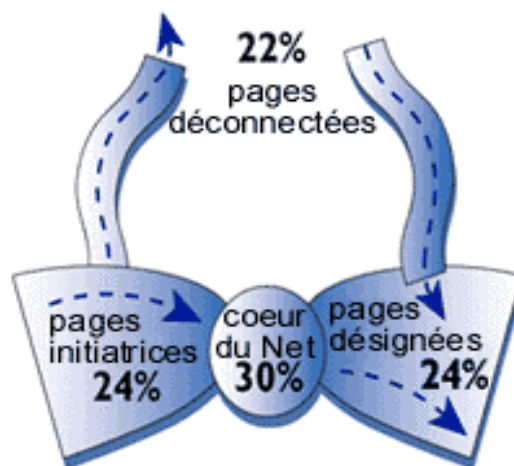


Figure II.2. Cartographie du Web

Web serait ainsi divisé en quatre grandes régions:

- ✓ Cœur du Net : 30% des pages Web constituent de véritables carrefours. De nombreux liens y mènent et ces sites renvoient sur de nombreux autres sites.
- ✓ Pages déconnectées : 11% des pages ne sont reliées à aucune autre. Pour y accéder, il faut impérativement en connaître l'adresse. 11% des pages s'échappent des pages initiatrices pour aboutir à des culs-de-sac.
- ✓ Pages initiatrices: 24% des pages Web renvoient vers des sites appartenant au cœur du Net, sans pour autant que la réciproque soit vraie. Il s'agit probablement de nouveaux sites qui n'ont pas encore été découverts, et donc pas encore reliés à ceux du cœur.

- ✓ Pages désignées: 24% de ces pages sont fréquemment désignées mais, en revanche ne renvoient sur aucun site.

Cette théorie dite du "nœud papillon" a permis la découverte de frontières rendant la navigation difficile voire, dans certains cas, impossible entre certaines régions du Web.

Cela nous donne un aperçu du monde dans lequel les outils de recherche ont dû évoluer et se doter de stratégies leur permettant de couvrir la plus grande partie du réseau.

II.3. Les logiciels permettant la consultation des serveurs web

Un navigateur (ou browser) est un logiciel qui permet de consulter le web. Ce logiciel fonctionne sur le principe du client-serveur. Le logiciel que l'on installe sur son poste de travail est le client, l'ordinateur distant est le serveur.

Les navigateurs les plus répandus aujourd'hui sont les logiciels proposés gratuitement : Internet Explorer (MICROSOFT), Firefox (MOZILLA), Netscape, Google Chrome ou bien encore Opéra (gratuit depuis le 02/10/2005).

Ils présentent l'avantage, si l'on utilise les dernières versions, d'intégrer les outils logiciels permettant de faire du ftp, de la messagerie, de créer des pages html et de visualiser des fichiers particuliers (images, image animé).

Cependant on va entamer dans ce qui suit les différents outils de recherche sur internet.

II.4 Les différents outils de recherche sur Internet

Dans cette structure en forme de « nœud papillon », il y a plusieurs milliards d'informations disponibles et dans des supports divers et variés [32] : des pages Web, de fichiers de tailles et formats différents, des images, du son, etc.

En plus les ressources sont très hétérogènes :

- ✓ sites institutionnels, universitaires, commerciaux, personnels,
- ✓ articles, romans, pages publicitaires, catalogues, et cela dans plus de 100 langues.

Le corpus de travail d'Internet est considérable, et il ne cesse pas de changer et d'augmenter en volume. Mais beaucoup d'informations disparaissent aussi ou n'ont plus de validité.

Pour s'y retrouver différents outils de recherche, couvrant des parties plus ou moins grandes d'Internet, ont été mis au point.

II.4.1. Les moteurs de recherche (*search engine*)

Un moteur de recherche (*search engine*) est un programme qui visite et indexe de manière automatique, sans intervention humaine, le texte des pages accessibles sur le Web. Il crée ainsi une base de données des mots qui seront associés aux fichiers visités.

Les moteurs de recherche mettent alors à dispositions des internautes une interface permettant d'obtenir, après une requête, la liste des pages, avec leur adresse URL, contenant les mots recherchés.

Mis à part Google, et pour n'en citer que quelques uns :

- ✓ **Altavista**, américain. Fait aussi portail.
- ✓ **AllTheWeb**, norvégien. Spécialisé aussi dans les actualités du monde.
- ✓ **Voilà**, français. Fait aussi portail.

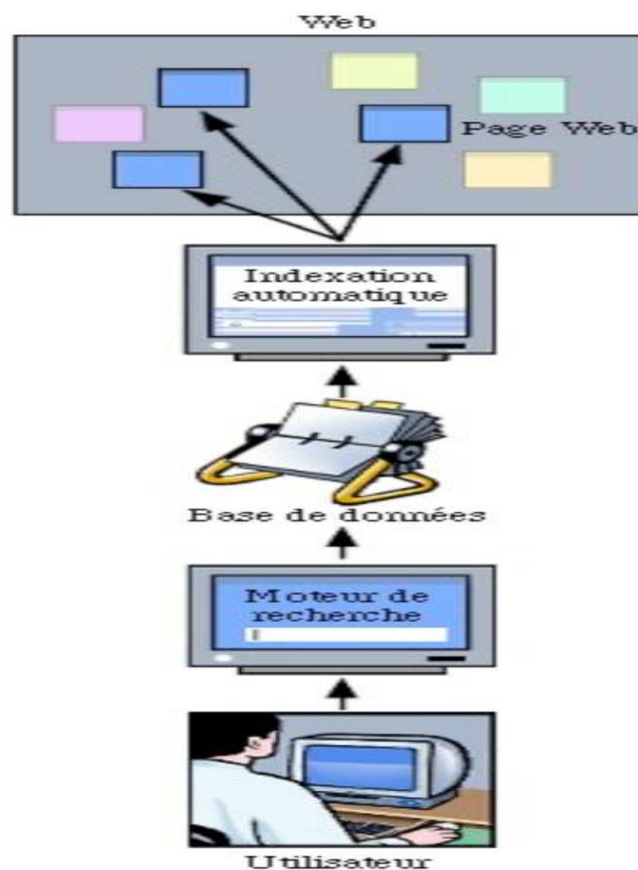


Figure II.3 Moteurs de Recherche

II.4.2 Les annuaires ou répertoires :

Les annuaires, appelés également répertoires, sont une base de données des sites jugés intéressants. Ces sites sont classés manuellement, par des documentalistes. Ils sont classés par thématique arborescente. On accède au thème recherché par une série de rubriques et sous rubriques de plus en plus précises.

La recherche sur un annuaire se fait uniquement sur les catégories, les titres et les descriptions des sites.

Les annuaires présentent souvent un nombre restreint de sites référencés.

Pour n'en citer que quelques uns :

- ✓ **Yahoo**, américain. Il fait appel au moteur Google s'il ne trouve pas de réponse.
- ✓ **Open Drectory**, américain. Des bénévoles du monde entier contribuent au classement des sites de cet annuaire.
- ✓ **Nomade**, français. Fait aussi portail.

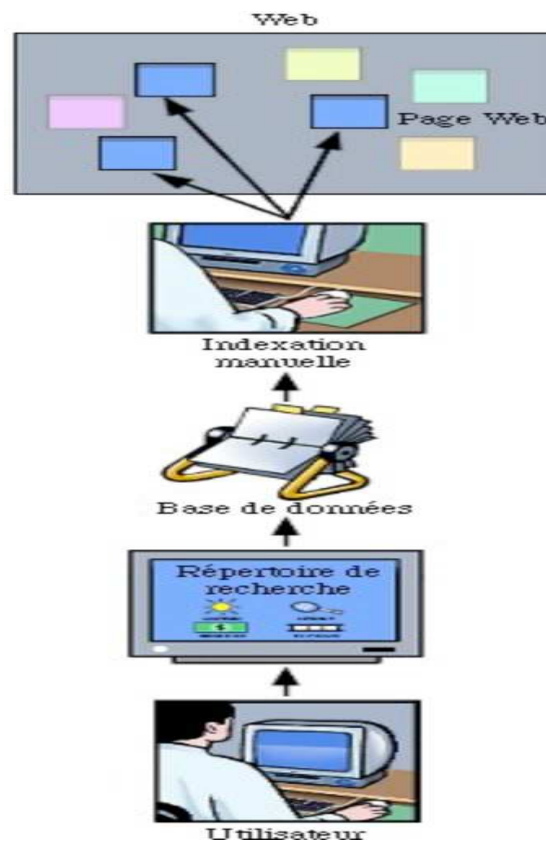


Figure II.4 Les Annuaires (Répertoires)

II.4.3 Les métamoteurs :

A partir d'une seule requête de mots clés, un méta moteur utilise plusieurs moteurs de recherche et répertoires de sites existant sur le Web et affiche les résultats fournis par les instruments de recherche qu'il a utilisés. Il propose alors les meilleurs résultats obtenus en

évitant les redondances, c'est-à-dire en choisissant les résultats communs les mieux classés et en évitant les doublons.

On trouve parmi d'autres :

- ✓ **Copernic [33]**, québécois, à installer dans l'ordinateur de l'utilisateur.
- ✓ **Kartoo[33]**, français. Il présente les résultats sous la forme d'un graphe. Il aide l'utilisateur à enrichir les requêtes par des mots ayant une relation avec les mots déjà saisis et qui sont communs aux résultats obtenus.

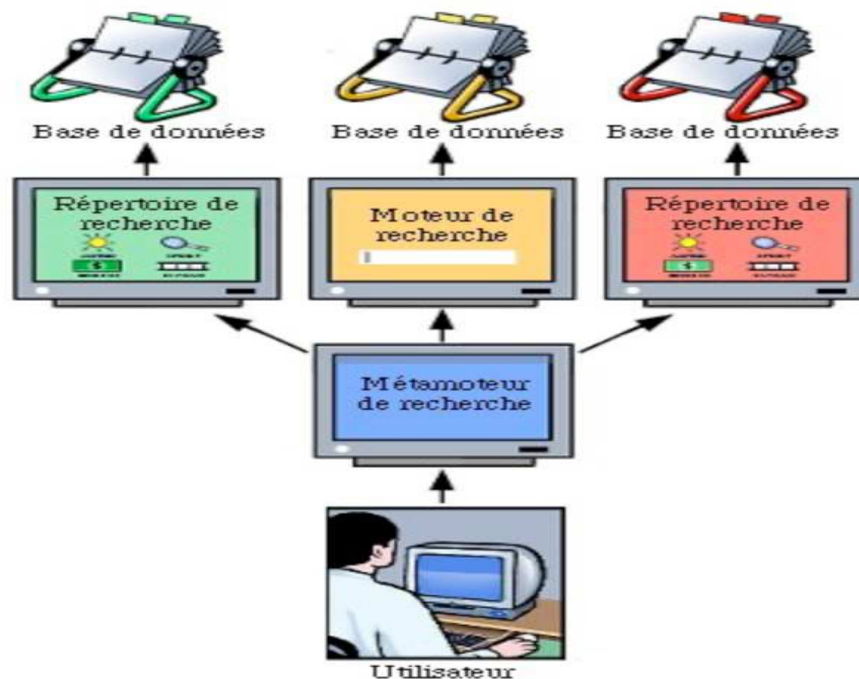


Figure II.5 Les Métamoteurs

II.4.4 Le portail :

Un portail est un outil de recherche qui offre à ses utilisateurs, en plus de la possibilité de la recherche d'une information, d'autres services existant sur Internet ; par exemple affectation d'une adresse e-mail, hébergement des pages Web, actualités, informations boursières, météo, forums, discussions en direct, espaces de commerce électronique, rencontres,etc.

L'interface de ce genre d'outil est souvent très chargée en bandeau publicitaires. Ce genre d'outil de recherche a plus tendance à rechercher les recettes publicitaires que l'efficacité de la recherche de l'information.

Exemple : **Francité [33]**, portail francophone.

Après avoir passé en revue sur les différents outils de recherche existant sur le web nous allons voir le fonctionnement d'un moteur de recherche.

II.5 Le fonctionnement d'un moteur de recherche

Google fait partie de la catégorie des moteurs de recherche. Ce qui suit sera donc consacré au fonctionnement général des moteurs de recherche [34]. Graphiquement on peut le représenter comme suit :

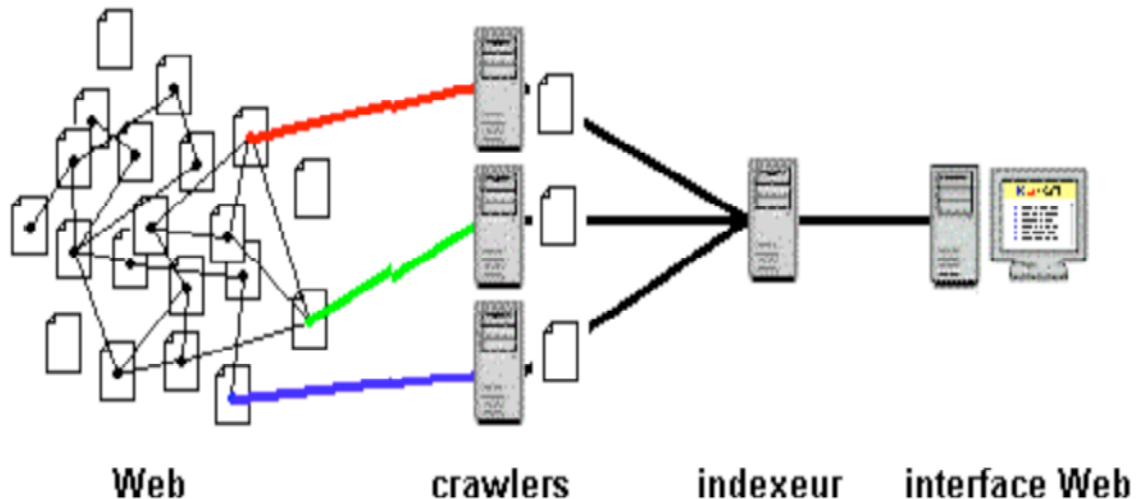


Figure II.6 Fonctionnement d'un Moteur De Recherche [34]

II.5.1 Le crawler :

Le crawler (ou spider, ou robot) est le premier maillon de la chaîne. Sa mission est « d'aspirer » les pages Web afin de constituer une base de données la plus complète possible.

Le crawler se connecte sur les pages Web suivant un plan de « route », copie leurs contenus, et les transfère à l'indexeur.

Le crawler doit travailler le plus rapidement possible et 24h/24. Actuellement la vitesse moyenne des robots est de 400 pages secondes sachant qu'un crawler travaille en parallèle avec d'autres crawlers.

Il analyse les hyperliens (les URL) contenus dans les pages qu'il copie. A chaque fois que le crawler rencontre un lien, il compare cette adresse aux adresses connues il a alors 3 possibilités :

- ✓ L'adresse est inconnue,
- ✓ L'adresse est connue mais le délai de la dernière visite sur la page est long,
- ✓ L'adresse est connue mais le délai de la dernière visite sur la page est court.

Dans les 2 premiers cas, le crawler ajoute l'adresse à son plan de « route » ; dans le dernier cas, l'adresse est ignorée.

C'est ainsi que les crawlers parcourent le Web, en sautant de page en page via les hyperliens. Mais cette méthode délaisse les sites vers lesquels aucune page ne pointe, c'est

pour cela que certains moteurs proposent aux webmasters de soumettre manuellement leurs pages aux crawlers.

II.5.2. L'indexeur

L'indexeur effectue une analyse statistique des pages aspirées par les crawlers, il ne cherche pas à comprendre le contenu des pages, il n'effectue pas d'analyse sémantique. Son rôle est de trier et classer les mots des pages

L'indexeur commence par effectuer un tri de censure, il scrute les pages à la recherche de certains mots « interdits » par le moteur. C'est ainsi que les pages à caractère raciste sont exclues de l'indexation.

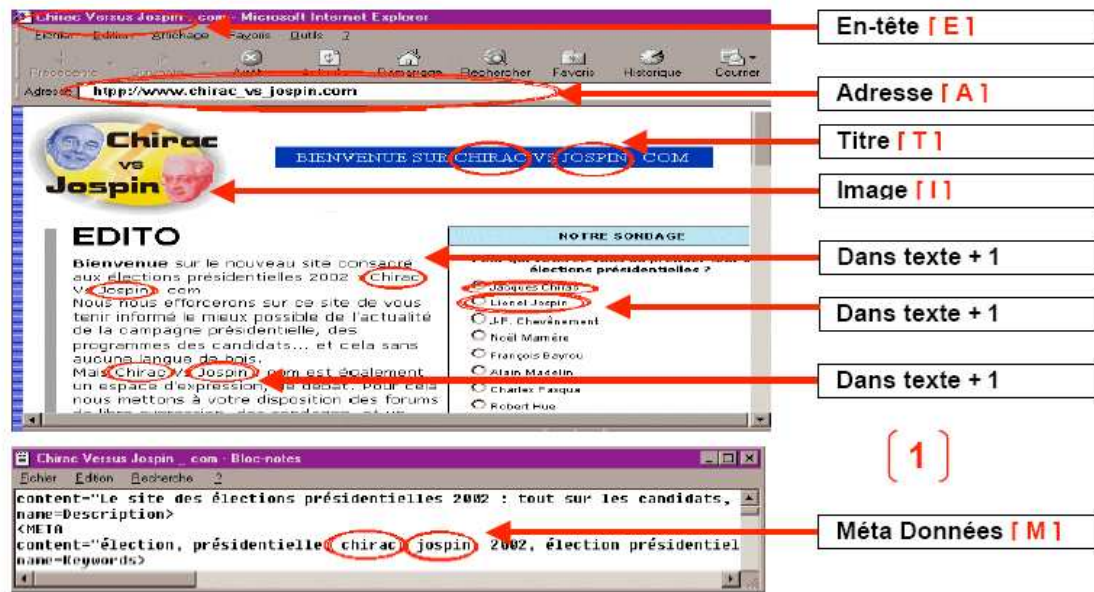
Ensuite le texte est dépouillé de sa ponctuation et de certains mot insignifiants, appelés mots vides (comme les articles, les pronoms, etc...., exemple : le – la – les – de, etc.)

Exemple simplifié d'indexation d'une page Web

L'indexeur note l'emplacement de chaque mot [voir ci-dessous (1)]. Il constitue l'**index** de la page [voir image (2) page suivante], dans lequel chaque mot est associé à un nombre d'apparition dans la page ainsi qu'à sa ou ses positions (titre, texte, lien, méta-données, etc.).Se constitue alors le « poids » du mot, son importance. (Voir **Spam indexing**⁽⁴⁾).

Une fois l'index de la page crée, l'indexeur passe à la mise à jour de l'**index inversé du mot** [voir image (3) page suivante]. Chaque élément de cet index viendra compléter la liste des pages déjà visitées contenant le même mot. Dans l'index inversé chaque mot renvoie à une liste des pages dans laquelle chaque page aura un poids donné.

⁽⁴⁾ Spam indexing : est ensemble de techniques consistant à tromper les moteurs de recherche sur la qualité d'une page ou d'un site.



Méta-données dans le code HTML de la page d'exemple.

Figure II.7 Indexation dans une page web

INDEX DE LA PAGE										INDEX INVERSE DU MOT									
URL : www.chirac_vs_jospin.com/index.htm										CHIRAC									
Occurrence mot		E	A	T	I	M	URL		E	A	T	I	M						
(2)	Jospin	20					www.chirac_vs_jospin/in...	20											
	Chirac	20					www.elections2002.fr...	13											
	Bienvenue	3					www.france_elections.fr...	9											
Une partie de l'index de la page : www.chirac vs jospin...										Une partie de l'index inversé du mot : "Chirac"									

Une partie de l'index de la page : www.chirac_vs_jospin...

Une partie de l'index inversé du mot : "Chirac"

Figure II.8 résultat de l'indexation d'une page web

Plus le ratio **mot-page-poids** sera important dans la page, mieux cette page sera classée dans l'index inversé associé au mot. Par la suite lors d'une requête faite par un internaute et contenant ce même mot, les pages apparaîtront dans cet ordre.

Pour une requête à plusieurs mots, par exemple contenant les mots *CHIRAC ET JOSPIN*, le retour se fera dans l'ordre des pages contenant le mot *CHIRAC* et parmi celles-ci, dans l'ordre des pages contenant les mots *CHIRAC* et *JOSPIN*.

A l'aide d'un programme qui s'occupe du calcul des requêtes, l'index inversé constitue le maillon entre l'indexeur et l'interface web.

II.5.3. L'interface Web

Fonctionnant sur un navigateur du type *Fozilla Firefox*, *Internet Explorer* ou autre, l'interface Web est la partie visible du moteur de recherche. C'est à son niveau que sont effectuées les requêtes des utilisateurs et l'affichage des résultats.

Le programme se trouvant du côté du moteur de recherche reçoit et compare les mots de la requête à ceux de la base de données constituée des listes des éléments **mot-page-poids** créée par l'indexeur. Après calcul, il envoie à l'interface pour l'affichage, les résultats des recherches sous la forme de classement, par ordre de pertinence, des adresses URL des pages correspondantes aux mots clés saisis.

Une étude statistique a permis de vérifier que plus de **70%** des internautes regardent seulement les tous premiers résultats retournés par un outil de recherche.

La pertinence dans les premiers résultats devient une obligation si l'on veut satisfaire l'utilisateur de l'outil. Lors de la saisie de la requête, l'interface du moteur de recherche autorise l'emploi d'opérateurs, ce qui permet de contraindre ou d'élargir le domaine de recherche.

Les trois phases, crawler, indexeur et interface Web, font partie du mode de fonctionnement commun à tous des moteurs de recherche.

A l'intérieur de chacune de ces phases, chaque moteur de recherche a sa spécificité qui le différencie des autres. et a mis au point ses propres algorithmes d'affichage afin de proposer un affichage des réponses trié. Google par exemple affiche les sites en fonction de leur notoriété (plus un site est cité, par des liens, mieux il sera placé dans la liste des réponses).

II.6. Les limites des moteurs de recherche

Ces moteurs stockent l'information trouvée à un instant T (stockage sur plusieurs Giga-octets du disque dur de l'ordinateur). Lorsque l'utilisateur lance une recherche, le moteur recherche sur son disque dur les termes correspondants puis vous proposera à l'affichage le lien vers le serveur Web qui contient l'information. Or il se peut qu'entre-temps l'information originale ait été modifiée ou ait été déplacée du serveur. Vous ne pourrez donc pas forcément retrouver l'information concernée. Google pallie ce problème en proposant les pages « en cache ».

Le créateur d'une page H T M L peut indiquer par un code spécifique qu'une page ne doit pas être indexée.

En résumé, chaque moteur a :

- ✓ Sa propre logique de parcours du réseau et de visite des sites
- ✓ Des critères spécifiques d'indexation
- ✓ Des critères d'interrogation plus ou moins puissants
- ✓ Des critères d'affichage et de tris des réponses

Dans le tableau II.1, nous illustrons les principales différences existantes entre les moteurs de recherche et les annuaires

Moteur de recherche	Annuaire
Une Recherche à travers la boîte à requêtes.	La possibilité de naviguer à travers les catégories (un annuaire classe par rubriques, par thèmes les sites qui lui sont soumis).
Un moteur référence tous les sites du web sans limite (référence tout ce qu'on lui donne).	Un annuaire n'est pas censé référencer tout le web, mais seulement prendre en compte les meilleurs sites dans les limites qu'il s'est fixé.
Indexation des pages web.	Indexation des sites web.
Référencement facile à effectuer (automatisation par des scripts) puisque seule l'URL du site est vraiment nécessaire.	Le référencier (webmaster) soumet une description et souvent, les mots clés caractérisant son site. Ces informations sont accompagnées de l'URL du site, de son mail.
L'absence de validation humaine au niveau des sites et donc, une validation à travers des filtres mis en place au niveau du programme (filtres par domaines ou par mots-clés ou par facteurs spécifiques au moteur).	Le surfeur (appelé aussi modérateur ou documentaliste, selon l'annuaire) vérifie l'intérêt du site (en le visitant) et l'accepte ou le refuse dans la base de l'annuaire en fonction de critères définis par la direction de l'annuaire.

Utilisation des logiciels qui indexent le contenu du web (robots) : indexation automatique	Indexation manuelle par le surfeur.
Le robot utilise les codes sources des pages web pour les indexer.	Le surfeur utilise les descripteurs fournis par le webmaster et, ajoute d'autres descripteurs qu'il juge intéressants.

Tableau II.1 Comparaison entre Moteurs de recherche et Annuaire**II.7. Avantages et inconvénients des moteurs de recherche****➤ Avantages**

- ✓ L'aspect automatique du système a permis à la taille de l'index d'augmenter avec la taille du Web, et cela, grâce à l'efficacité du robot.
- ✓ Une combinaison des différents opérateurs booléens et des opérateurs de proximité comme "near" (sur AltaVista) permet grandement de limiter le bruit.

➤ Inconvénients :

- ✓ Ces systèmes génèrent encore beaucoup de bruit, du fait qu'ils cherchent à indexer tout le web, ainsi l'utilisateur se trouve submergé par une masse d'information qu'il ne maîtrise pas trop.
- ✓ La difficulté de choisir les bons mots-clés, en particulier lorsque la recherche est faite vers des pages en langue étrangère
- ✓ Les moteurs de recherche évoluent dans un environnement fondamentalement hostile, dans la mesure où un grand nombre d'éditeurs de sites commerciaux ont un intérêt vital à ce que leur site soit retourné en première page des résultats, pour le plus grand nombre de requêtes possibles. On assiste donc à une véritable guerre de mesures et de contre-mesures, les premiers (entreprises et organismes) cherchent en permanence à détecter et à exploiter les faiblesses du système d'indexation et de classement du moteur, les seconds (administrateurs des moteurs de recherche) cherchent à détecter ces abus et à boucher les failles correspondantes.

II.8. Méthodes de tri des résultats, des moteurs de recherche sur le web :**II.8.1 Tri par pertinence :**

Les résultats d'une requête sont affichés selon un ordre déterminé par le calcul d'un score pour chaque réponse. La pertinence est basée sur l'un ou la combinaison des facteurs suivants, appliqués aux termes de la requête:

- ✓ **Le poids d'un mot dans un document :** est déterminé par sa place dans le document : il est maximum pour le titre et le début du texte ; à l'intérieur du texte, il est plus important si le mot est en majuscules.
- ✓ **La densité** est basée sur la fréquence d'occurrence du terme dans un document, par rapport à la taille du document. Si deux documents contiennent le même nombre d'occurrences, le document le plus petit sera donc favorisé.
- ✓ **Le poids d'un mot dans la base documentaire :** est basé sur la fréquence d'occurrence du terme dans toute la base documentaire. Les mots peu fréquents dans le corpus sont favorisés. Les mots vides de sens sont soit éliminés, soit sous-évalués.
- ✓ **La correspondance d'expression :** est basée sur la similarité entre l'expression de la requête et l'expression correspondante dans un document. Un document contenant une expression identique à celle de la requête reçoit le poids le plus élevé.
- ✓ **La relation de proximité :** est basée sur la proximité des termes de la requête entre eux dans le document. Les termes proches sont favorisés.

Cette méthode de tri est utilisée par des outils de recherche comme : Alavista, Voila et AlltheWeb.

II.8.2 Tri par popularité :

Le moteur de recherche *Google* a lancé l'algorithme de pertinence *PageRank* qui se base essentiellement sur l'indice de popularité (voir annexe B). Ce système est utilisé à la fois pour les résultats du moteur et pour le classement des sites dans son répertoire.

"Le principe de *PageRank* est simple : tout lien pointant de la page A à la page B est considéré comme un vote de la page A en faveur de la page B. Toutefois, *Google* ne limite pas son évaluation au nombre de votes (liens) reçus par la page ; il procède également à une analyse de la page qui contient le lien. Les liens présents dans des pages jugées importantes par *Google* ont plus de poids et contribuent ainsi à « élire » d'autres page".

II.8.3 Tri par calcul dynamique de catégories :

Cette méthode effectue un classement des documents trouvés dans des dossiers constitués automatiquement en fonction des réponses. Un dossier peut lui-même être constitué de sous-dossiers. Vivisimo Clustered search results⁽⁵⁾ en est un exemple.

II.9 Cartographie des résultats :

Il s'agit plus ici d'une présentation visuelle des résultats, sous forme de cartes, comme le propose Kartoo⁽⁶⁾.

En ce qui concerne le tri des résultats d'une recherche effectuée sur notre système, c'est selon la pertinence (densité, poids)

II.10 Comparaison entre moteurs de recherche :

Les moteurs malgré leur similarité dans le fonctionnement diffèrent néanmoins dans des caractéristiques telles que leurs syntaxes d'interrogation et les méthodes de tri des résultats utilisées.

Dans les tableaux 2 et 3, nous présentons à titre indicatif une synthèse comparative des deux moteurs de recherche Google et Altavista avec notre système, au travers leurs syntaxes d'interrogation et méthodes de tri respectives

Prise en compte	Google	Altavista	Notre système
Majuscules et minuscules	Sans importance	Sans importance	Avec importance
Lettres accentuées	Sans importance	Le mot est toujours recherché tel qu'il est saisi	Avec importance
Ordre des mots	Oui	Sans importance	Sans importance
Opérateur « Ou »	Oui, mais saisi en majuscules OR	Oui mais saisi en majuscules OR	Non
Opérateur « ET »	Oui AND = opérateur par défaut	Oui AND = opérateur par défaut	Non
Opérateur « SAUF »	Oui : -	Oui : - OU AND NOT	Non

⁽⁵⁾ (<http://www.vivisimo.com>)

⁽⁶⁾ (<Http://www.kartoo.com>)

Expressions :	Oui : « »	Oui : « »	Oui : « »
Troncature	Non	Oui : *	Oui : *
Synonymes : ~	Oui	Non	Oui
Recherche sur le nom de domaine du site	Oui <i>site :</i>	Oui <i>host :</i>	Oui
Recherche sur les adresse des liens	Oui <i>link :</i>	Oui <i>link :</i>	Oui
Recherche des sites similaires	Oui <i>related :</i>	Oui <i>like :</i>	Non
Recherche sur les mots	Oui <i>define :</i>	Non	Non Selon les mots clés issus de l'indexation manuelle
Recherche dans le titre	Oui <i>allintitle :</i>	Oui <i>title :</i>	Non
Recherche dans le titre et la page	Oui <i>: intitle :</i>	Non	Non
Recherche dans l'URL	Oui <i>allinurl :</i>	Oui <i>url :</i>	Non
Recherche avancée	Oui	Oui	Oui

Tableau II.2 : Comparaison de Syntaxes d'interrogation entre Google, Altavista et notre système.

➤ **Méthodes de tri des résultats :**

Google	Altavista	Notre système
Tri par indice de popularité	Tri par indice de pertinence	Tri par pertinence

Tableau II.3 Comparaison de tri des résultats entre Google, Altavista et notre système

II.11 Difficultés des moteurs de recherche sur le Web

- ✓ Les sites web sont distribués à travers la planète ;

- ✓ Le web est une gigantesque source d'information qui ne cesse d'augmenter, par conséquent la couverture de ces moteurs n'augmente pas aussi rapidement que la taille du Web ;
- ✓ Une distinction claire entre les sources dignes de foi et les autres est actuellement impossible ;
- ✓ Plusieurs sources ou collections de documents ne sont pas librement accessibles (Copyrights, collections privées).

Un ensemble de solutions a été apporté pour essayer de dépasser les problèmes cités précédemment :

- ✓ Des robots indexeurs parcourent le réseau afin de découvrir des sites web et permettre aux moteurs de recherche de les indexer ;
- ✓ Equiper les moteurs de recherche d'une capacité très importante en disque dur.
- ✓ Augmenter la fréquence de mise à jour du moteur de recherche.

II.10. Conclusion

Dans ce chapitre, nous avons vu les différents outils de recherche sur le web, leurs fonctionnement qui se base sur l'architecture client-serveur, les données qu'elle doit rechercher sont distribuées à travers la planète et leurs hétérogénéités, les difficultés qu'il y a à indexer cette gigantesque masse d'information et les quelques solutions offertes pour remédier à ces difficultés.

Nous allons dans le prochain chapitre décrire les étapes de conception de notre moteur de recherche d'information sur le web.

Chapitre III

La conception du système

III.1. Introduction :

La modélisation du profil utilisateur pour un système de recherche d'information n'est pas une tâche aisée surtout lorsque l'on sait la diversité des informations qui peuvent le caractériser, la prédominance d'une information par rapport aux autres pour la distinction des besoins de chaque utilisateur du système, sans parler de la difficulté rencontrée par les systèmes eux-mêmes à opérer une recherche personnalisée selon la particularité des besoins informationnels de chacun.

Dans ce chapitre nous présentons notre démarche à concevoir un système personnalisé pour l'interrogation d'une base de données des hôtels. A cet effet, nous commençons par décrire l'architecture générale de notre système et son fonctionnement puis modélisation du profil utilisateur où nous allons définir notre choix de représentation puis nous détaillons chaque module de notre système ensuite la modélisation des traitements avec les différents diagrammes. Notre objectif est d'arriver à un système capable de collecter le maximum d'informations concernant un utilisateur et d'enrichir son profil au fur et à mesure pour faciliter le processus de recherche, afin de satisfaire ces besoins par des documents pertinents vis-à-vis de son profil.

III.2. Architecture générale et fonctionnement du système :

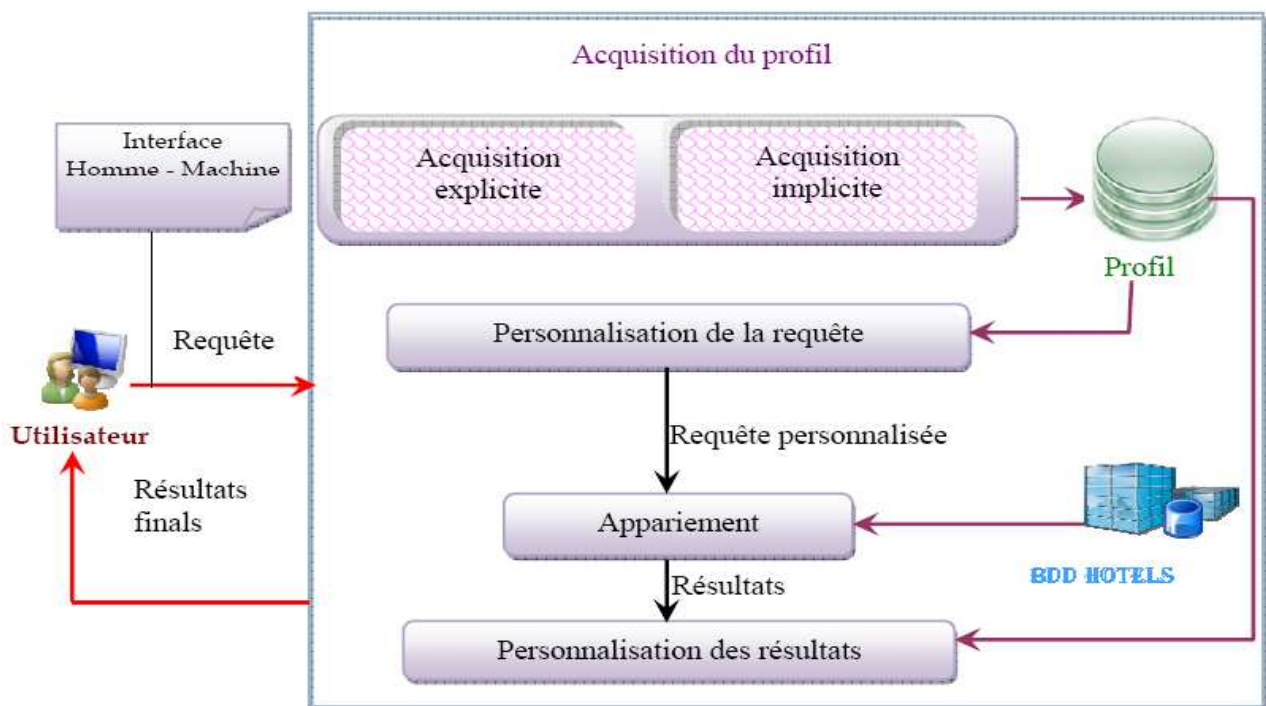


Figure III.1. Schéma général du système personnalisé.

- le prototype du fonctionnement du système :

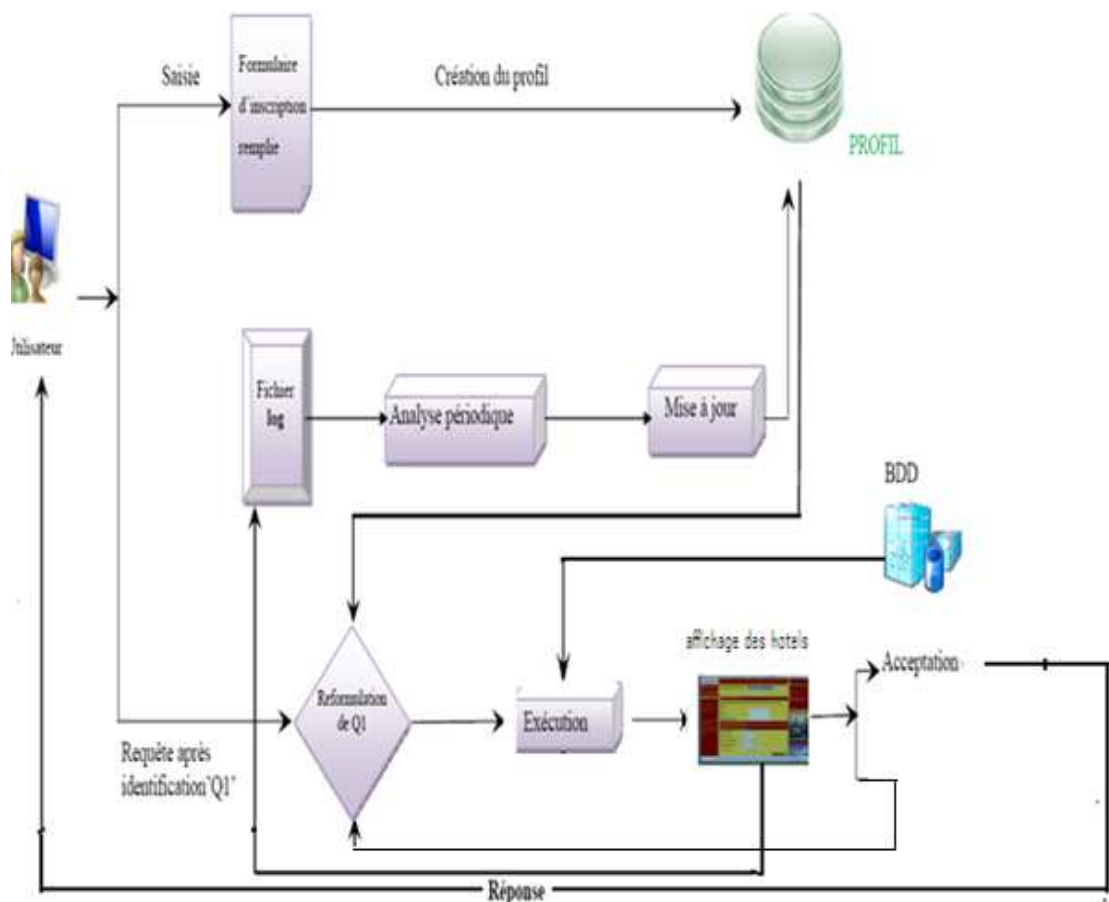


Figure III.2. Le prototype du fonctionnement du système

- Le fonctionnement du système :

C'est un système réalisé pour les services commerciaux (cas hôtels), il offre la possibilité de chercher des hôtels en introduisant la notion de personnalisation guidée par le profil utilisateur dans le contexte de la recherche d'information.

Afin de mieux répondre à ce besoin, l'utilisateur doit être identifié en s'inscrivant auprès de notre système à travers le remplissage d'un formulaire contenant des informations personnelles et des informations concernant ses préférences et ses besoins.

Un nom d'utilisateur et un mot de passe seront attribués à cet utilisateur à la fin de cette phase pour ses futurs accès au système, et son profil sera créé à partir des données fournies explicitement.

A chaque accès au système à travers le nom d'utilisateur et le mot de passe, l'utilisateur sera surveillé tout au long de son accès en enregistrant ses interactions dans un fichier nommé **log** dans la base de profil. Il doit être analysé périodiquement afin d'extraire les informations nécessaires pour la mise à jour du profil.

Lorsqu'un utilisateur pose sa requête, le profil sera utilisé dans la reformulation de cette dernière en introduisant les préférences de cet utilisateur dans sa demande. Le résultat de cette reformulation sera l'affichage des hôtels représentés par une interface. En arrivant à ce stade là, deux cas sont possibles :

- L'utilisateur est satisfait par ces résultats
- L'utilisateur n'est pas satisfait par ces résultats, c'est-à-dire que les conditions d'hôtel qui a constitué ne conviennent pas à ses besoins actuels, dans ce cas il aura la possibilité de reformuler ça requête.

III.3 La modélisation du profil utilisateur :

Il n'existe pas à l'heure actuelle de modèle global spécifique à un utilisateur [37]. La modélisation de ce dernier se base uniquement sur les informations qui le décrivent. Toute fois et à travers notre analyse en s'inspirant des travaux de bouzeghoub et kostadinov [2] qui décrivent plusieurs dimensions du profil, nous avons estimé qu'un certain nombre d'exigence auxquelles doit répondre ce dernier pour que sa description soit la plus générique que possible afin de représenter au mieux les besoins des utilisateurs .

III.3.1 Choix d'une représentation du profil :

Comme nous l'avons vu dans la partie état de l'art, le contenu du profil d'un utilisateur varie selon les approches et les applications. L'approche que nous proposons s'inscrit dans une réflexion globale sur la personnalisation de l'information en exploitant notamment les résultats acquis en RI et BD. Nous avons opté pour une représentation multidimensionnelle, le choix est justifié par le fait que le profil est caractérisé par plusieurs catégories d'informations c'est-à-dire plusieurs dimensions et ses dimensions peuvent être décomposées en plusieurs sous dimensions. Certaines de ses dimensions seront données explicitement par l'utilisateur et d'autre seront déduites implicitement par le système.

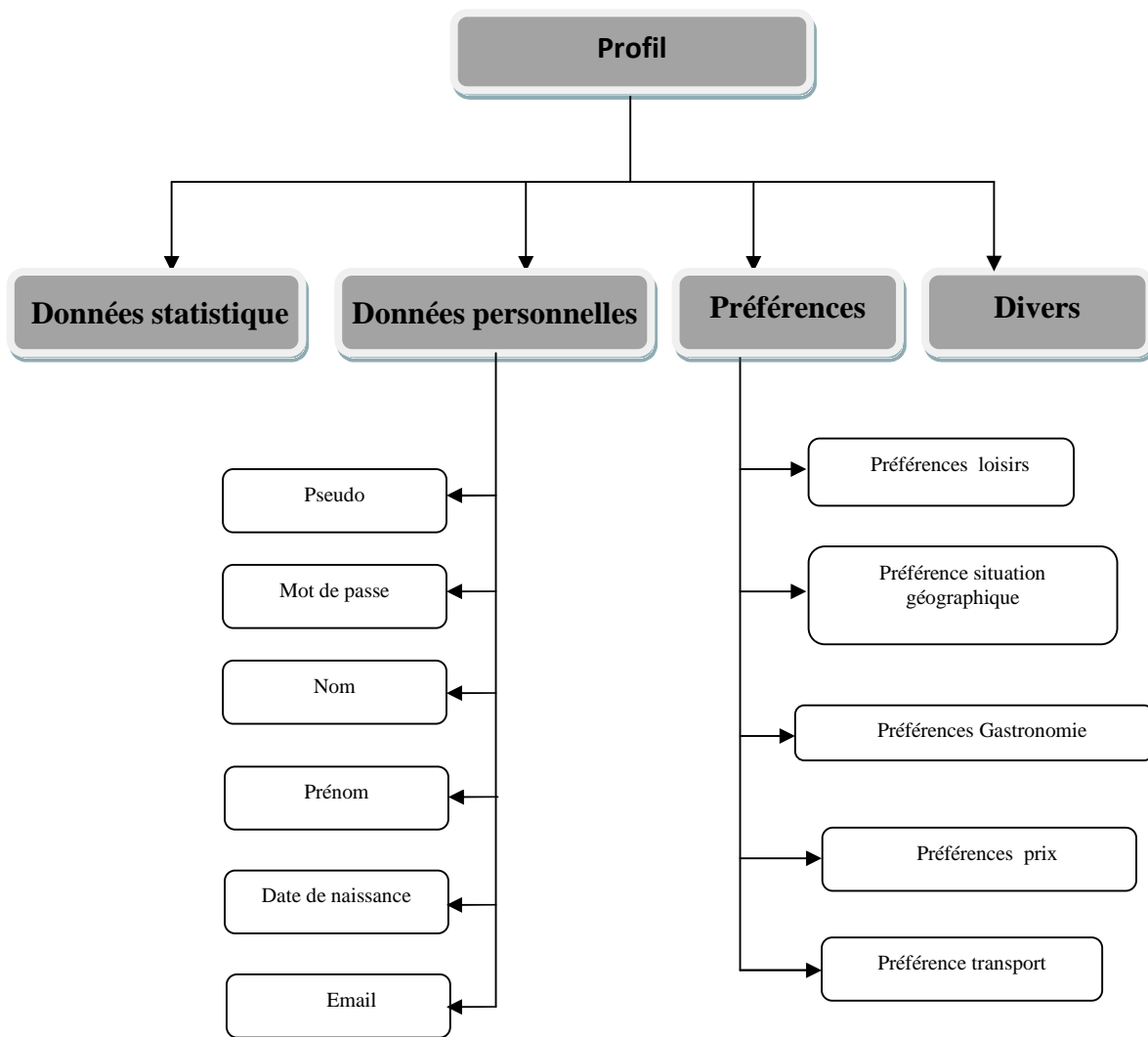


Figure III.3: représentation multidimensionnelle d'un profil

➤ **les informations personnelles :**

Les informations de cette dimension décrivent principalement :

Données personnelles : pseudo, mot de passe, nom, prénom, date de naissance, e-mail, adresse

➤ **les préférences :**

Cette dimension traduit ce que souhaite l'utilisateur obtenir, il peut souhaiter :

- Préférences loisir : salle de sport, soirée (musical/théâtrale), jeux pour enfant, salle de cinéma, shopping.
- Préférence situation géographique : ville, montagne, front de mer.etc....

- Préférences Gastronomie : française, chinoise, etc....
 - Préférences prix : intervalle de prix (prixmin, prixmax).
 - Préférence transport.
- **Les données statistiques** : cette dimension englobe les interactions de l'utilisateur avec le système et les statistiques concernant le profil lui-même, ces dernières seront utilisées pour la maintenance du profil.
- On a le nombre d'accès au système ; le degré de changement de profil c'est le nombre de fois que le profil a subi des mises à jour.
- **Divers** : l'intérêt de cette dimension est de ne pas restreindre le profil aux seules informations contenues dans les dimensions précédentes, car certaines applications demandent des informations spécifiques ne peuvent être inclus dans aucune des autres dimensions.

III.4 Les différents sous systèmes :

Nous avons décomposé notre système en trois sous système qu'ils sont :

III.4.1 sous système de gestion du profil :

Ce sous système peut être décomposé en deux modules :

- ✓ l'acquisition explicite du profil.
- ✓ l'acquisition implicite du profil.

Il se représente comme suit :

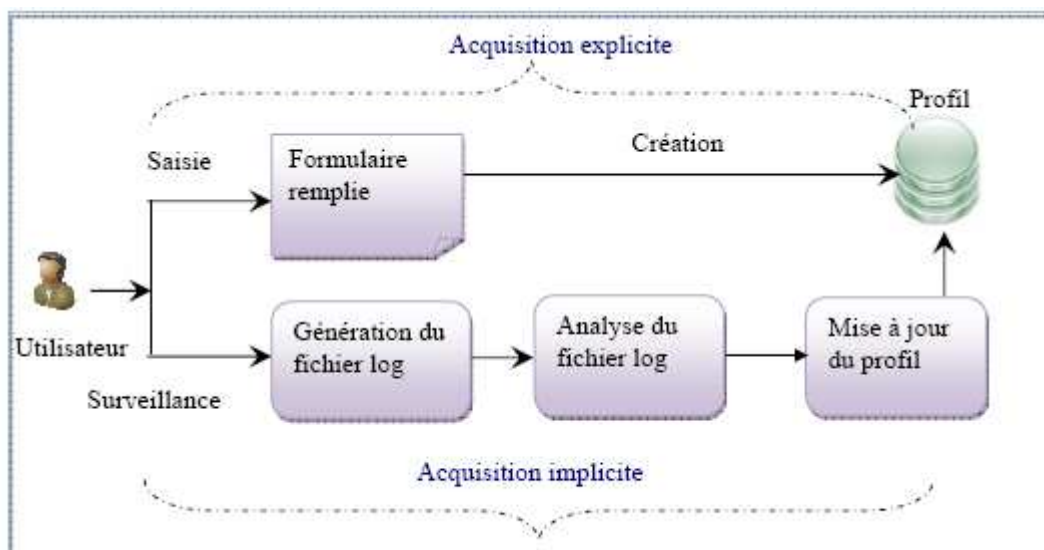


Figure III.4 Sous système de gestion du profil.

a) L'acquisition explicite du profil :

Dans son premier contact avec le système, l'utilisateur doit s'inscrire à travers le remplissage d'un formulaire où il va fournir :

- les informations personnelles concernant son pseudo, mot de passe, nom, prénom, date de naissance, e-mail. Certaines de ces informations sont obligatoires, d'autres sont facultatives.
- Les informations concernant ses besoins et ses préférences : afin de fournir à l'utilisateur des informations pertinentes adaptées à ses besoins.
- Chaque utilisateur aura la possibilité de modifier ses préférences manuellement à travers une interface de mise à jour explicite offerte par notre système.

b) L'acquisition implicite du profil

L'avantage primaire pour l'utilisation des techniques implicites est que l'utilisateur est déchargé de l'effort de mise à jour de ses préférences. Les mesures implicites sont généralement supposées être moins précises que les mesures explicites. Mais ce manque de précision est comblé par les grandes quantités de données implicites pouvant être recueillies.

De plus, ces mesures implicites peuvent être combinées avec des évaluations explicites pour obtenir une représentation plus précise des préférences de l'utilisateur. Cette phase vise à ne pas figer le profil, mais à le rendre plutôt une entité dynamique à l'intérieur du système.

Ce mode d'acquisition s'effectue grâce à un système de surveillance qui doit être déclenché dès qu'une session de recherche est ouverte, son objectif est de capturer les traces des interactions de l'utilisateur tout au long de sa session. Dans ce contexte, nous avons opté pour l'exploitation du fichier log en vue d'extraire les différentes interactions de l'utilisateur avec le système.

Deux solutions sont envisageables pour l'exploitation des fichiers **log** :

- L'utilisation des fichiers **log** générés automatiquement par des spyware, l'inconvénient de cette solution est que ces fichiers ont des formats différents et mal renseignés donc difficile d'exploiter leurs données.
- La construction de notre propre fichier **log**, c'est la solution que nous allons adopter pour notre système, elle a comme avantage d'avoir une structure connue (fixe) qu'on peut l'exploiter facilement et rapidement.

Id_log	Pseudo	Id_hot	Nom d'interaction	Date début	Date fin
L'identifiant de l'interaction	Le numéro de l'utilisateur qui fait l'interaction	L'hôtel correspond à cette interaction	Sa valeur	La date de début l'interaction	La date fin de l'interaction

Tab III.1 La structure du fichier log.

Cette table doit être analysée périodiquement afin de mettre à jour le profil de l'utilisateur.

Analyse du fichier log:

L'analyse de la table **log** vise à réaliser périodiquement une lecture statistique sur le contenu courant de cette table, en vue de mettre à jour le profil de l'utilisateur. Donc à chaque fin de période (**p**) (fixé par l'administrateur de système), le module d'analyse sera déclenché, cette procédure concerne les utilisateurs vérifiant l'une des conditions suivantes :

- La valeur **X** résultante du nombre de consultation et la durée que cet utilisateur reste sur ce lien doit être supérieur ou égale à **4** pendant cette période, tel que **X=durée+nombre de consultation** déduite après avoir effectué des statistiques sur des internautes (Forum,...).
- l'adhérent a aimé les caractéristiques de l'hôtel: si l'adhérent juge un hôtel comme pertinent alors la condition sera automatiquement vérifiée.
- l'adhérent a imprimé la page : si l'adhérent imprime les caractéristiques de l'hôtel alors il sera considéré comme pertinent donc la condition sera automatiquement vérifiée.

III.4.2. Le sous système de requête :

C'est le sous système qui s'occupe du traitement de la requête. Dans ce sous système, on a introduit un dictionnaire de synonymes de mots clés utilisés pour l'indexation des documents, pour ne pas restreindre l'utilisateur à utiliser les mêmes mots clés qui se retrouvent dans notre base de données.

Exemple : s'il tape dans sa requête **front de mer** ou bien **littoral**, les résultats seront les mêmes vu que le contexte des deux mots est le même.

III.4.3. Le sous système d'appariement

Ce sous système englobe l'exécution de la requête finale et l'affichage des résultats, plus précisément la requête finale sera exécutée en interrogeant la base de données puis le résultat va être représenté selon le choix de l'utilisateur.

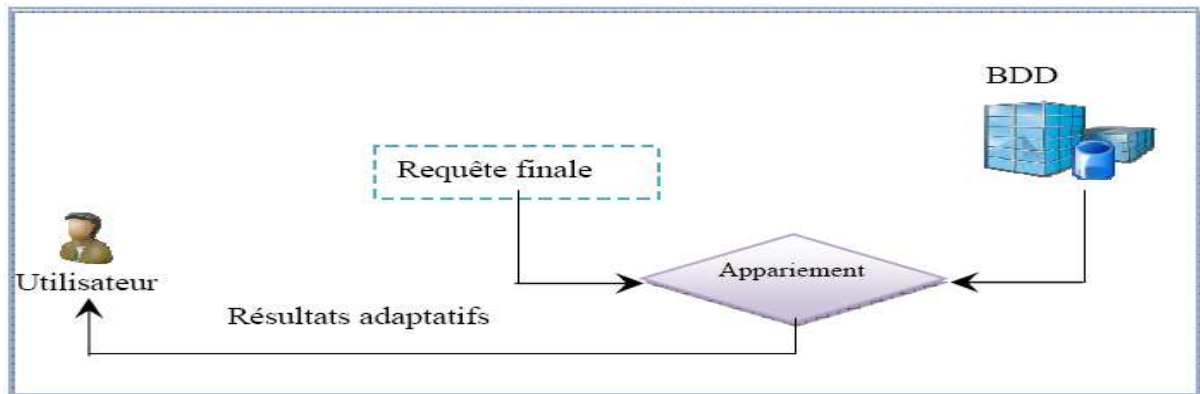


Figure III.5: Le sous système d'appariement

Pour localiser un utilisateur à un moment donné, on a pensé à introduire le procédé de la géolocalisation qui sera détaillé dans le point suivant :

III.5 la géolocalisation

La **géolocalisation** ou **géoréférencement** est un procédé permettant de positionner un objet (une personne,...) sur un plan ou une carte à l'aide de ses coordonnées géographiques déduites dans notre cas par une adresse IP.

III.5.1 Objectif traité

Près de 72% [36] des besoins informationnels des utilisateurs mobiles sont liés à des facteurs contextuels parmi eux (la localisation), dans cette partie nous étudions la mobilité de l'utilisateur, le contexte spatial sera exploité pour mieux répondre à des requêtes sensibles à la localisation (Annexe C), l'objectif étant d'incorporer la localisation de l'utilisateur à la requête et pour tenir compte des conditions de proximité.

Exemple:

Un utilisateur résidant en Algérie effectuant la requête suivante : hôtels 3 étoiles. Les résultats de recherche relatifs à un hôtel 3 étoiles sera en dinars, alors qu'un utilisateur résidant en France aura des résultats en euros.

Par analogie à l'architecture d'un système de recherche personnalisé nous avons pour :

- Les sources d'informations : la localisation géographique de l'utilisateur.
- Des stratégies de collecte de ces informations: le positionnement de l'utilisateur mobile se fera d'une manière implicite, sa localisation sera détectée de manière automatique par géo localisation par adresse IP.
- L'accès à l'information : ceci consiste à sélectionner l'information pertinente à la requête adressée au système de recherche d'informations en tenant compte de la requête et aussi de la localisation courante de l'utilisateur.

III.5. 2 Architecture du sous-système Géo localisation

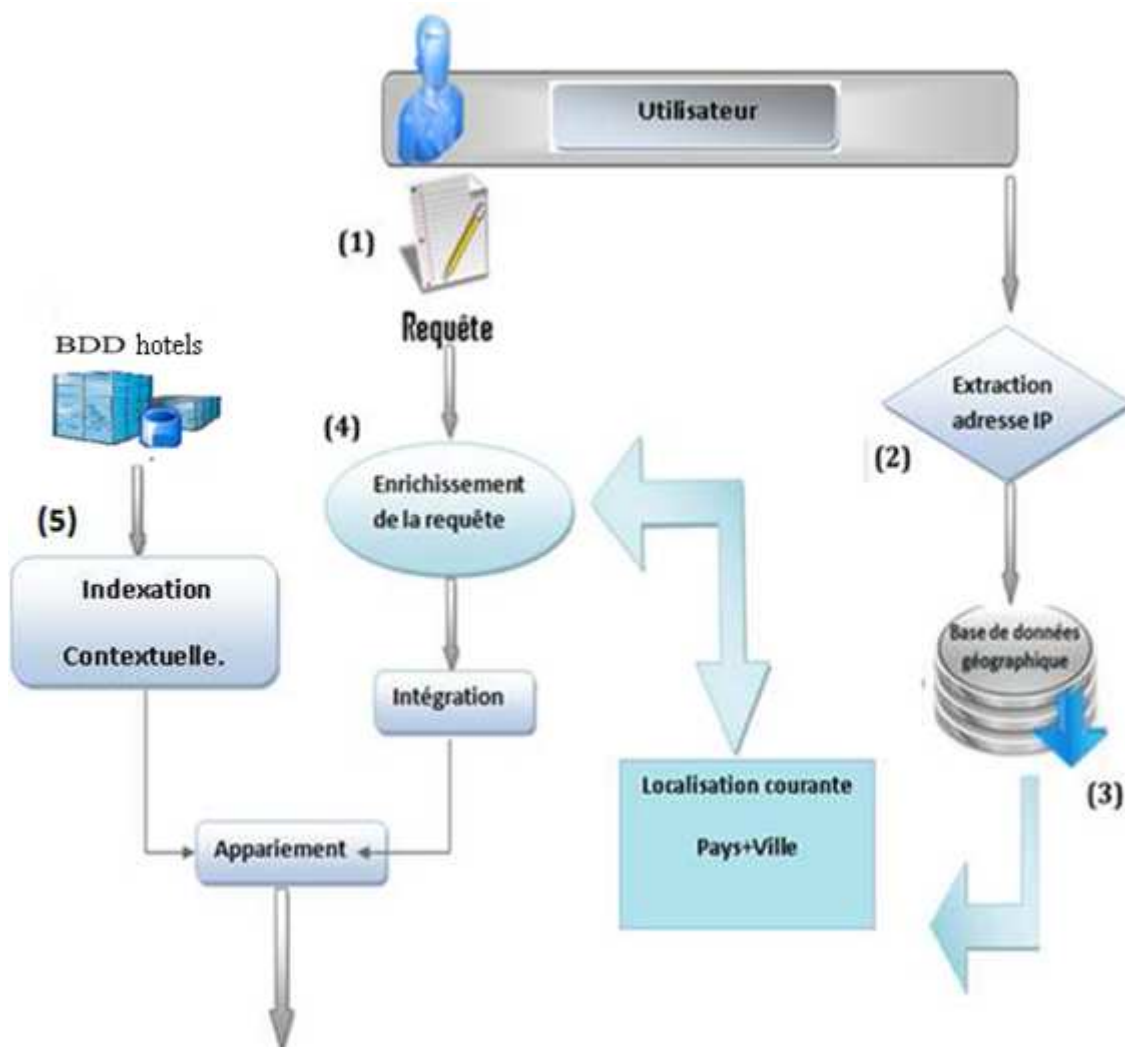


Figure III.6 Recherche dépendante de la localisation courante de l'utilisateur

III.5. 3 Fonctionnement du sous –système GEO localisation

- (1) L'utilisateur va formuler sa requête.
- (2) L'extraction de l'adresse IP de l'utilisateur est effectuée à chaque fois que l'utilisateur introduit une requête, cela permet d'offrir à l'utilisateur étant mobile et l'adresse IP dynamique des résultats toujours dépendant de sa localisation courante.
- (3) L'adresse IP courante étant détectée lors de chaque requête, nous faisons l'extraction du Pays et de la Ville courante de l'utilisateur de la base de données géographique, nous proposons de détailler la localisation de l'utilisateur (voir **Figure III.6**) au dessus.
- (4) La requête initiale est enrichie avec la localisation courante de l'utilisateur (Pays +Ville).
- (5) Cette indexation est appelée indexation contextuelle, car elle s'effectue à partir de base de donnée hôtel contenant du contexte lié à la dimension "LOCALISATION"

- La localisation de notre utilisateur se fait par : (voir **Figure III.7**).

1-Elément de localisation : C'est le positionnement par adresse IP par lequel on peut localiser notre utilisateur et ainsi déterminer sa position géographique dans le monde, on extrairait sa localisation courante représentée sous la forme du pays et de la ville.

2-Base de données géographique : L'application communique avec la base de données ceci dans le but de fournir le pays, région et la ville de l'utilisateur.

3-Utilisateur mobile : Se connecte au système et est engagé dans une situation (au travail, fait du tourisme...) et effectue une requête.

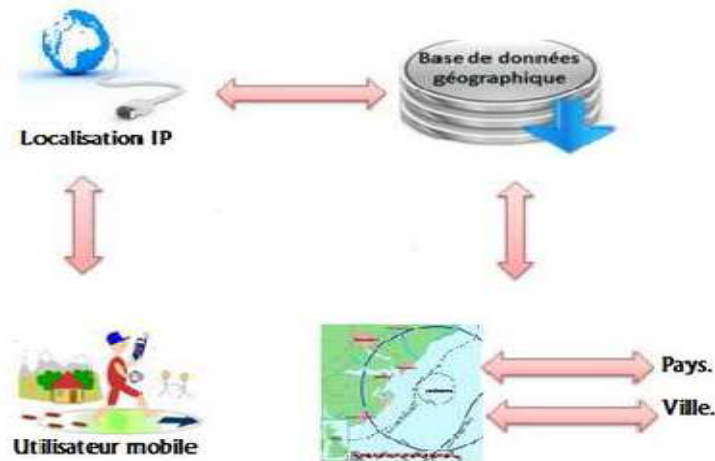


Figure III.7. Extraction de la localisation courante de l'utilisateur.

Après avoir décrit l'architecture générale de notre système et son fonctionnement puis la modélisation du profil utilisateur où nous avons défini notre choix de représentation, ensuite nous avons détaillé chaque module de notre système, maintenant nous allons passer à la modélisation des traitements avec les différents diagrammes UML.

III.6.Modélisation des traitements :

Notre choix est porté sur un ensemble de concepts proposés par le langage UML (Unified Modeling Language) devenu aujourd'hui universel, ces concepts seront un support pour l'analyse, la conception et la spécification des besoins de notre application.

De façon générale, l'UML [Annexe D] permet de modéliser :

- ✓ les utilisations de cas et de scénarios (spécifications, architecture fonctionnelle) ;
- ✓ les classes et objets (analyse technique détaillée) ;
- ✓ les composants (architecture logicielle) ;
- ✓ la distribution et déploiement (architecture technique).

Ainsi, chaque organisation est libre de choisir le processus qui lui semble le plus adapté en fonction du type d'application développées, des habitudes de ses développeurs ou encore de leur niveau de maturité objet. De fait, UML n'est pas à proprement parlé une méthode mais un support versatile adaptable à de nombreuses méthodes de développement. [39]

III.6.1 Représentation des cas d'utilisation :

➤ Identification des acteurs :

Un acteur est une entité externe (personne, machine, autre, etc.) interagissant avec le système et qui n'appartient pas à ce dernier. Les acteurs se déterminent en observant les utilisateurs directs du système, ceux qui sont responsables de son exploitation ou de sa

maintenance, ainsi que les autres systèmes qui interagissent avec le système en question. Dans notre cas on définit trois acteurs qui sont :

- ✓ L'administrateur : est la personne responsable du site web, de sa conception à sa maintenance.
- ✓ Le visiteur : la personne qui visite le site.
- ✓ L'adhérant : la personne qui possède un compte dans le site.
- ✓ Le client : la personne qui propose un hôtel à l'administrateur et qui est souvent l'administrateur de ce site.

➤ **Identification des cas d'utilisation :**

Un cas d'utilisation est utilisé pour définir le comportement d'un système ou la sémantique de toute autre entité sans révéler sa structure interne. Chaque cas d'utilisation spécifie une séquence d'actions, y compris des variantes, que l'entité réalise, en interagissant avec les acteurs de l'entité. La responsabilité d'un cas d'utilisation est de spécifier un ensemble d'instances, où une instance de cas d'utilisation représente une séquence d'actions que le système réalise et qui fournit un résultat observable par l'acteur.

➤ **Liste des cas d'utilisation :**

N°	Cas d'utilisations	Acteur
1	Inscription d'un utilisateur	Visiteur
2	Recherche non personnalisée	Visiteur
3	Authentification d'un adhérent	Adhérent
4	Recherche personnalisée	Adhérent
5	Modification des préférences	Adhérent
6	Authentification de l'administrateur	Administrateur
7	Insertion d'hôtels	Administrateur
8	Suppression d'hôtels	Administrateur
9	Modification d'hôtels	Administrateur
10	Propositions d'hôtels	client
11	Modification des caractéristiques des hôtels	client
12	Contacte administrateur	Client, adhérent, visiteur

Tab III.2 Identification des cas d'utilisation

➤ Description des cas d'utilisation

Le cas N°01:

- **Description du sommaire :**

Titre : Inscription d'un utilisateur.

But : Ce cas d'utilisation permet à un nouvel utilisateur de s'inscrire et de créer son profil sur l'annuaire.

Acteurs : Le visiteur, Son rôle est le remplissage d'un formulaire d'inscription.

- **Description des enchainements :**

Pré condition : connexion au système.

Enchainement :

1. Un visiteur accède à la page d'accueil.
2. Il demande le formulaire d'inscription.
3. Le système affiche un formulaire d'inscription.
4. Le visiteur saisit les coordonnées ainsi que ses préférences.
5. le système insère les données dans la base.
6. Le système affiche une fenêtre indiquant que son inscription est effectuée avec succès.

Liste des exceptions :

Exception 1 : le pseudo entré par l'utilisateur n'est pas disponible, c'est le cas où le visiteur saisie un pseudo qui existe déjà dans la base donc il doit le modifier.

Post condition : Néant.

Le cas N°02:

- **Description du sommaire :**

Titre : Recherche non personnalisée

But : Ce cas d'utilisation permet à l'utilisateur de visualiser les caractéristiques des hôtels.

Acteurs : Visiteur, son rôle est de lancer la requête à exécuter.

- **Description des enchainements :**

Pré condition : Connexion au système.

Enchainement :

1. l'utilisateur demande le formulaire de recherche.
2. le système lui affiche le formulaire de recherche.
3. l'utilisateur entre sa requête.
4. le système exécute la requête au niveau de la base.

5. le système affiche une fenêtre contenant les caractéristiques des hôtels résultants.

Liste des exceptions :

Exception 1 : l'hôtel demandé par l'utilisateur n'est pas disponible ; c'est le cas où le visiteur saisie des caractéristiques d'hôtels qui ne sont pas disponibles.

Post condition : Néant.

Le cas N°03:

- **Description du sommaire :**

Titre : authentification d'un adhérent.

But : ce cas d'utilisation permet à un adhérent de se connecter au système.

Acteurs : Adhérent, son rôle est de saisir son pseudo et son mot de passe.

- **Description des enchainements :**

Enchainement :

1. l'adhérent accède à la page d'accueil.
2. il demande le formulaire d'identification.
3. le système affiche un formulaire d'identification.
4. l'adhérent saisit son pseudo et son mot de passe.
5. le système lance la requête de vérification au niveau de la base.
6. le système construit la fenêtre de soumission de requête.

Liste des exceptions :

Exception 1 : pseudo inexistant, c'est le cas où l'utilisateur oublie ou saisit un pseudo qui n'existe pas dans la base, le système affiche alors une page d'erreur.

Exception 2 : mot de passe incorrect, c'est le cas où l'utilisateur oublie ou saisit un mot de passe qui ne correspond pas à son pseudo dans la base, le système affiche alors une page d'erreur.

Post condition : Néant.

Le cas N°04:

- **Description du sommaire :**

Titre : Recherche personnalisée

But : Ce cas permet à l'adhérent de visualiser les caractéristiques ainsi que le contenu de l'hôtel.

Acteurs : Adhérent, son rôle est de lancer la requête à exécuter.

- **Description des enchainements :**

Pré condition : Connexion au système et authentification à l'aide d'un pseudo et un mot de passe.

Enchaînement :

1. l'adhérant demande la page d'édition de requêtes.
2. le système lui affiche une interface d'édition de requêtes.
3. l'utilisateur saisit sa requête, puis il lance son exécution.
4. le système exécute la requête après l'avoir enrichi avec les préférences de l'adhérant.
5. le système affiche une fenêtre contenant les caractéristiques des hôtels résultants.
6. l'adhérant clique sur un lien d'hôtel choisi parmi les liens résultants.
7. le système affiche le site web choisi.

Liste des exceptions :

Exception 1 : l'hôtel demandé par l'adhérant n'est pas disponible, c'est le cas où l'adhérant saisit les caractéristiques d'hôtels qui ne sont pas disponibles.

Post condition : Néant.

Le cas N°05:

- **Description du sommaire :**

Titre : modification des préférences

But : ce cas d'utilisation permet à l'adhérant de modifier ses préférences.

Acteurs : Adhérent, son rôle est de saisir les modifications à apporter sur ses préférences.

- **Description des enchaînements :**

Pré condition : Connexion au système et authentification à l'aide d'un pseudo et un mot de passe.

Enchaînement :

1. l'adhérant demande le formulaire de mise à jour des préférences.
2. le système affiche un formulaire de mise à jour des préférences.
3. l'adhérant saisit ses nouvelles préférences.
4. le système insère les nouvelles préférences dans la base.
5. le système affiche une fenêtre indiquant que la mise à jour des préférences est effectuée avec succès.

Liste des exceptions :

Exception 1 : pseudo inexistant, c'est le cas où l'utilisateur oublie ou saisit un pseudo qui n'existe pas dans la base.

Exception 2 : mot de passe incorrect, c'est le cas où l'utilisateur oublie ou saisit un mot de passe qui ne correspond pas à son pseudo dans la base, le système affiche alors une page d'erreur.

Post condition : Néant.

Le cas N°06:

- **Description du sommaire :**

Titre : authentification de l'administrateur.

But : Ce cas d'utilisation permet à l'administrateur de se connecter au système.

Acteurs : Administrateur, son rôle est de saisir son nom d'accès et son mot de passe.

Description des enchainements :

Enchainement :

1. l'administrateur accède au formulaire d'authentification.
2. l'administrateur saisit son nom d'accès et son mot de passe.
3. le système lance la requête de vérification au niveau de la base.
4. le système affiche la fenêtre de mise à jour des hôtels.

Liste des exceptions :

Exception 1 : nom d'accès inexistant, c'est le cas où l'administrateur oublie ou saisit un nom d'accès qui n'existe pas dans la base, le système affiche alors une page d'erreur.

Exception 2 : mot de passe incorrect, c'est le cas où l'administrateur oublie ou saisit un mot de passe qui ne correspond pas à son nom d'accès dans la base, le système affiche alors une page d'erreur.

Post condition : Néant.

Le cas N°08:

- **Description du sommaire :**

Titre : Suppression d'hôtels.

But : ce cas d'utilisation permet à l'administrateur de supprimer des hôtels dans la base.

Acteurs : Administrateur, son rôle est de supprimer un hôtel en entrant son nom.

- **Description des enchainements :**

Pré condition : Connexion au système et authentification à l'aide d'un pseudo et un mot de passe.

Enchainement :

1. l'administrateur accède à la page de mise à jour des hôtels.
2. l'administrateur demande le formulaire de suppression des hôtels.

3. le système affiche le formulaire de suppression des hôtels.
4. l'administrateur soumet sa requête, en entrant le nom de l'hôtel.
5. le système exécute la requête en supprimant l'hôtel de la base de données.
6. le système affiche une fenêtre indiquant que la suppression est effectuée avec succès.

Liste des exceptions :

Exception 1 : le titre de l'hôtel n'existe pas dans la base, c'est le cas où l'administrateur saisit un nom qui n'existe pas dans la base.

Post condition : Néant.

Le cas N°09:

- **Description du sommaire :**

Titre : modification des hôtels.

But : ce cas d'utilisation permet à l'administrateur de modifier des hôtels dans la base.

Acteurs : Administrateur, son rôle est de modifier des hôtels en introduisant de nouvelles caractéristiques.

- **Description des enchainements :**

Pré condition : Connexion au système et authentification à l'aide d'un pseudo et un mot de passe.

Enchainement :

1. l'administrateur accède à la page de mise à jour des hôtels.
2. l'administrateur demande la page de modification des hôtels.
3. le système affiche le formulaire de recherche des hôtels par nom.
4. l'administrateur soumet une requête de recherche, en entrant le nom de l'hôtel.
5. le système exécute la requête en affichant le ou les hôtels correspondant à la requête.
6. l'administrateur soumet une requête de modification en sélectionnant un des hôtels résultants.
7. le système affiche un formulaire de modification.
8. l'administrateur saisit les nouvelles caractéristiques de l'hôtel.
9. le système affiche une fenêtre indiquant que la modification a été effectuée avec succès

Liste des exceptions :

Exception 1 : le titre de l'hôtel n'existe pas dans la base, c'est le cas où l'administrateur saisit un nom qui n'existe pas dans la base.

Post condition : Néant.

Le cas N°10:

- **Description du sommaire :**

Titre : Envoie hôtels.

But : ce cas d'utilisation permet au client d'envoyer à l'administrateur les caractéristiques de son propre hôtel .

Acteurs : client, son rôle est d'envoyer les caractéristiques de son hôtel.

Description des enchainements :

Pré condition : connexion au système.

Enchainement :

1. Un client accède à la page d'accueil.
2. Il demande le formulaire d'ajout d'un nouvel hôtel.
3. Le système affiche un formulaire de saisi.
4. Le client saisit ses coordonnées ainsi que les caractéristiques de son hôtel.
5. le système insère les données dans la base.
6. Le système affiche une fenêtre indiquant que son inscription est effectuée avec succès et que les informations sur son hôtel sont envoyées à l'administrateur avec succès.

Liste des exceptions :

Exception 1 : le pseudo entré par l'utilisateur n'est pas disponible, c'est le cas où le visiteur saisie un pseudo qui existe déjà dans la base donc il doit le modifier.

.Post condition : Néant.

Le cas N°11:

- **Description du sommaire :**

Titre : modification des caractéristiques des hôtels.

But : ce cas d'utilisation permet au client de modifier les caractéristiques de son hôtel.

Acteurs : Client, son rôle est de saisir les modifications à apporter à son hôtel.

- **Description des enchainements :**

Pré condition : Connexion au système et authentification à l'aide d'un pseudo et un mot de passe.

Enchainement :

1. le client demande le formulaire de mise à jour des caractéristiques de l'hôtel.
2. le système affiche un formulaire de mise à jour des caractéristiques de l'hôtel.
3. le client saisit les nouvelles caractéristiques de l'hôtel.
4. le système envoie les nouvelles caractéristiques de l'hôtel à l'administrateur.
5. le système affiche une fenêtre indiquant que les nouvelles caractéristiques de l'hôtel sont en cours de traitement par l'administrateur.

Liste des exceptions :

Exception 1 : pseudo inexistant, c'est le cas où le client oublie ou saisit un pseudo qui n'existe pas dans la base.

Exception 2 : mot de passe incorrect, c'est le cas où le client oublie ou saisit un mot de passe qui ne correspond pas à son pseudo dans la base, le système affiche alors une page d'erreur.

Post condition : Néant.

Le cas N°12:

- **Description du sommaire :**

Titre : Contacte administrateur.

But : ce cas d'utilisation permet à l'utilisateur de contacter l'administrateur, lui transmettre des idées et suggestions.

Acteurs : Visiteur, adhérent, client leurs rôles est d'envoyer leurs idées ou suggestions.

- **Description des enchainements :**

Pré condition : Connexion au système.

Enchainement :

1. l'utilisateur demande la page d'envoi de message.
2. Le système affiche un formulaire de saisie.
3. l'utilisateur rédige son message et le transmet à l'administrateur.
4. l'outil de messagerie affiche un message de confirmation d'envoi.

Liste des exceptions :

Exception 1 : l'email entré par l'utilisateur qui est erroné, c'est le cas où l'utilisateur saisie un email erroné.

Exception 2 : l'objet ou le contexte du message ne sont pas remplis, c'est le cas où l'utilisateur ne saisie pas l'objet ou le contexte du message un email.

.Post condition : Néant.

III.6.1.1 Diagramme de cas d'utilisation :

Un diagramme de cas d'utilisation est un graphe d'acteurs, un ensemble de cas d'utilisation englobés par la limite du système, des associations de communication entre les acteurs et les cas d'utilisations, et des généralisations entre cas d'utilisation. Il est destiné à représenter les besoins des utilisateurs par rapport au système.

➤ **Diagramme de cas d'utilisation "Visiteur"**

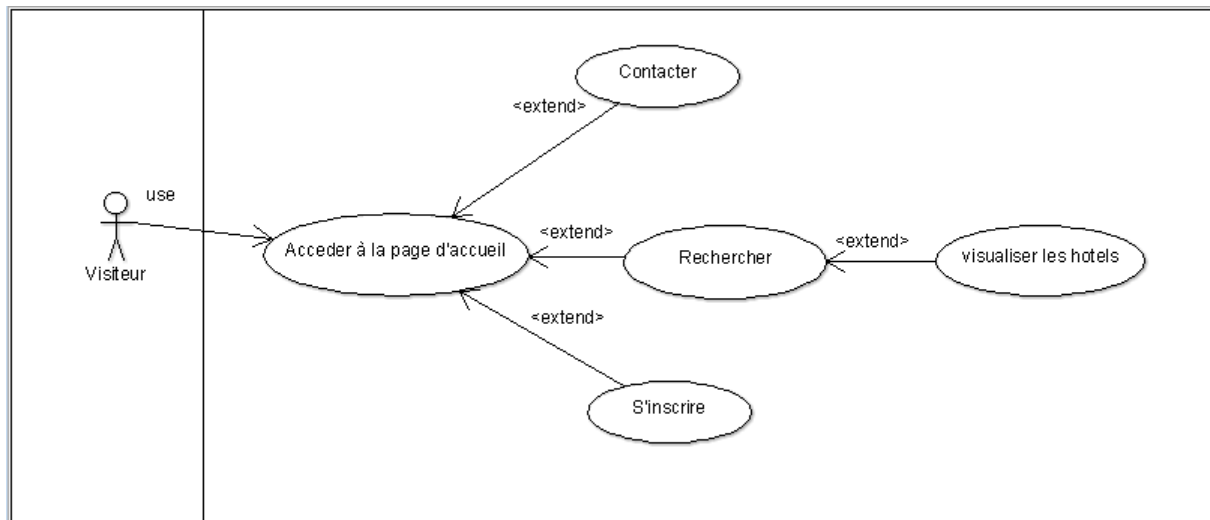


Figure III.8 Diagramme de cas d'utilisation “Visiteur”

Ce diagramme illustre les interactions du visiteur avec notre système. Celui-ci accède à la page d'accueil où il aura la possibilité soit d'accéder à l'outil de messagerie, comme il peut également s'inscrire pour devenir adhérent afin d'obtenir des résultats plus pertinents

➤ **Diagramme de cas d'utilisation “Adhérent”**

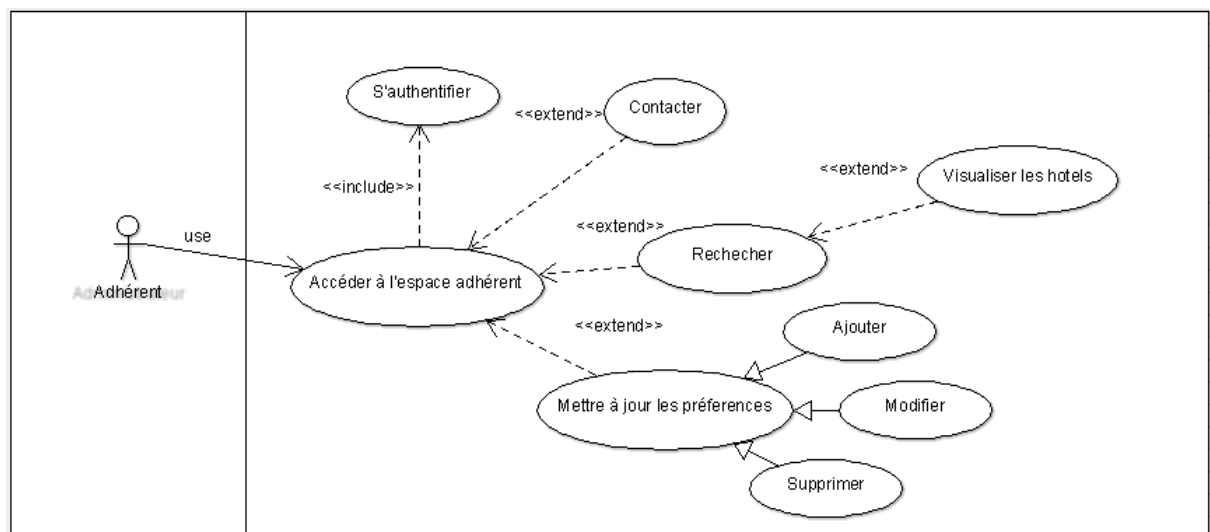


Figure III.9 Diagramme de cas d'utilisation “Adhérent”

Ce diagramme illustre les interactions de l'adhérent avec notre système, celui-ci accède à son espace après authentification où il aura la possibilité de faire une recherche personnalisée, mettre à jour ses propres préférences comme il peut accéder à la page de contact administrateur.

➤ **Diagramme de cas d'utilisation “Administrateur”**

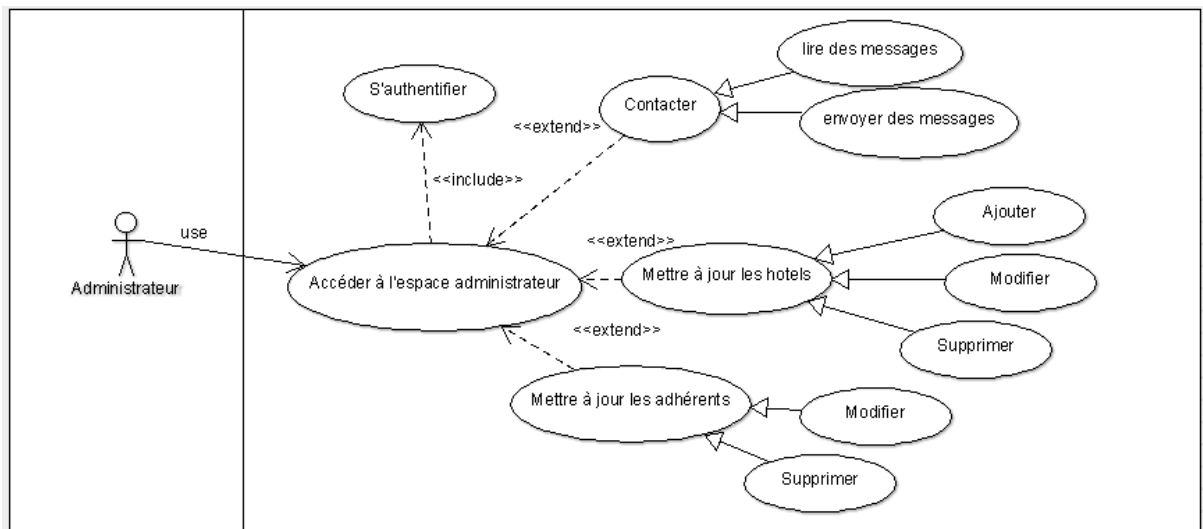


Figure III.10 Diagramme de cas d'utilisation “Administrateur”

Ce diagramme illustre l'interaction de l'administrateur avec le système, celui-ci s'authentifie et accède à son propre espace d'où il aura la possibilité d'envoyer des emails aux adhérents et aux clients, mettre à jour les hôtels de la base, accepter ou refuser les nouveaux hôtels, mettre à jour les comptes adhérents ceci dit supprimer l'adhérent, récupérer mot de passe adhérent, etc.....

➤ **Diagramme de cas d'utilisation “Client”**

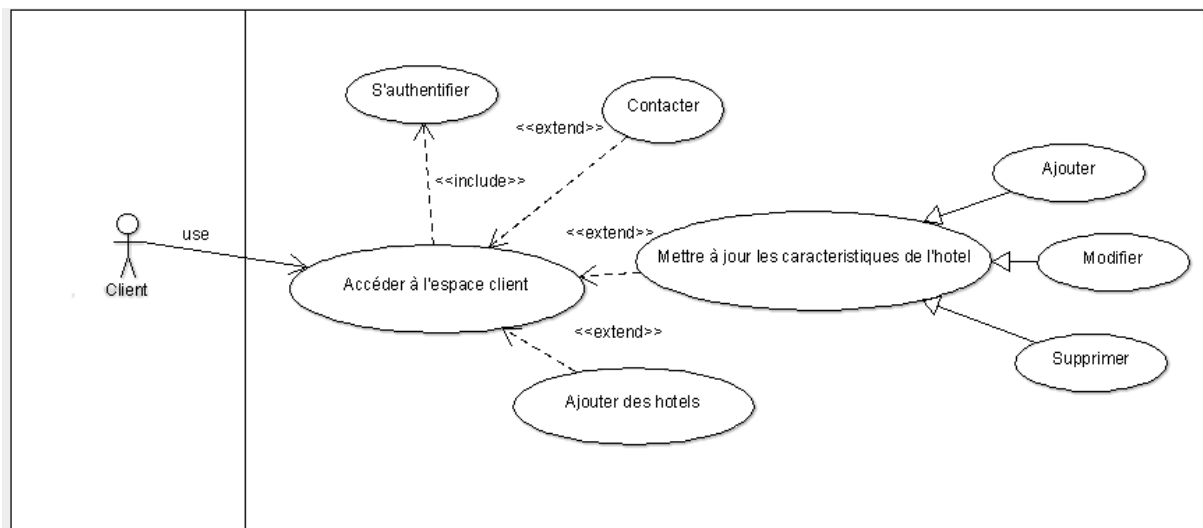


Figure III.11 Diagramme de cas d'utilisation “Client”

Ce diagramme illustre les interactions du client avec notre système, celui-ci accède à son espace après authentification où il aura la possibilité de modifier les caractéristiques de ses hôtels, ajouter d'autres hôtels, il peut aussi accéder à la page de contact administrateur, et modifier son mot de passe s'il en a envie.

III.6.1.2 . Diagramme de séquence :

Le diagramme de séquence permet de décrire les scénarios de chaque cas d'utilisation en mettant l'accent sur la chronologie des opérations en interaction avec les objets. Un diagramme de séquence montre une interaction présentée en séquence dans le temps. En particulier, il montre aussi les objets qui participent à l'interaction par leur "ligne de vie" et les messages qu'ils échangent présenté en séquence dans le temps. [62]

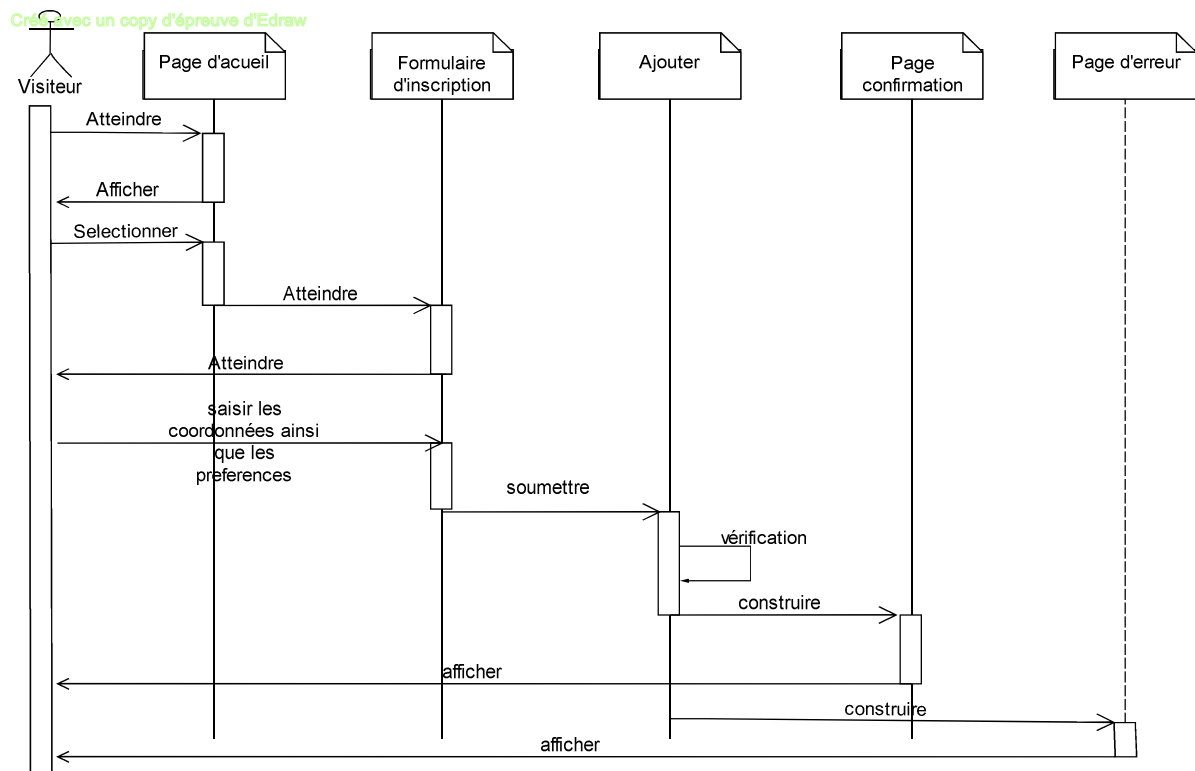


Figure III.12 Diagramme de séquences du cas d'utilisation “Inscription d’un visiteur ”

Ce diagramme illustre l'interaction entre le système et le visiteur lors de son inscription selon l'ordre chronologique, dans ce cas ici présent, le visiteur accède à la page d'accueil du système puis demande le formulaire d'inscription, le système lui renvoi ce formulaire par la suite le visiteur le remplit et soumet sa requête qui sera vérifiée par le système sur son existence dans la base de données, si c'est le cas, il lui affiche une page de confirmation pour qu'il accède à son espace, sinon une page d'erreur lui sera affichée.

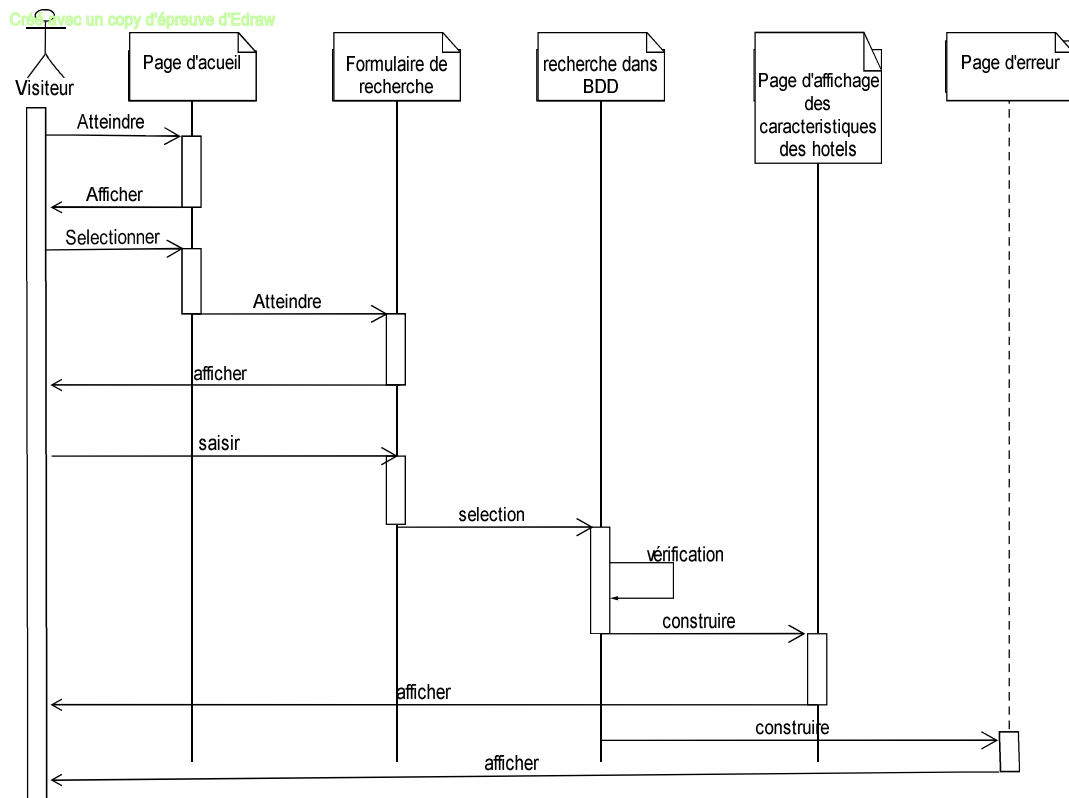


Figure III.13 Diagramme de séquence du cas d'utilisation “recherche non personnalisée”

Ce diagramme illustre l’interaction entre le système et le visiteur et ceci lors de sa recherche, ce dernier accède à la page d’accueil puis atteint le formulaire de recherche, le système lui envoie sa demande puis le visiteur soumet une requête, le système lui renvoie des résultats s’ils existent dans la base sinon le message « il n’y a aucun résultat » s’affichera.

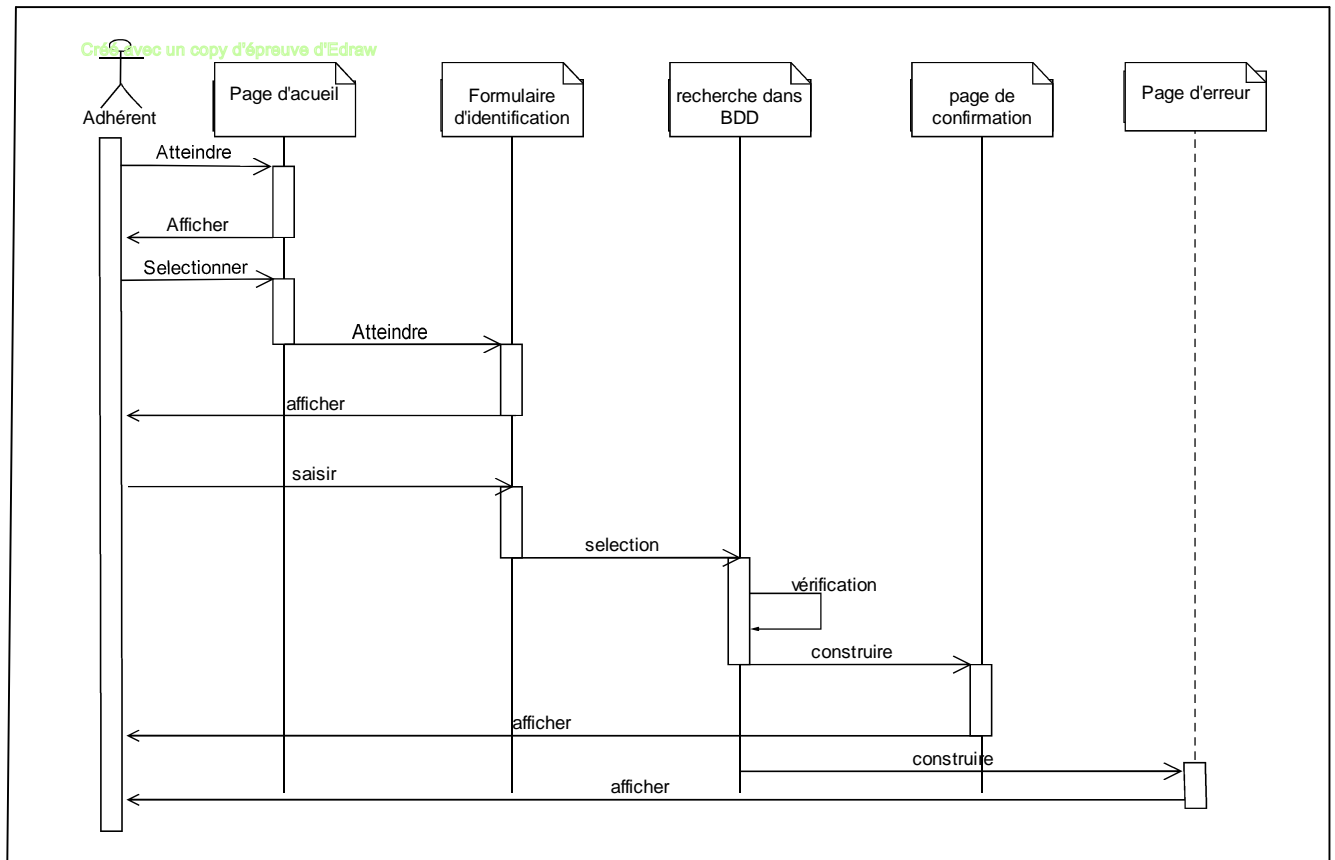


Figure III.14 Diagramme de s  quence du cas d'utilisation " identification d'un Adh  rent"

Ce diagramme illustre l'interaction entre le syst  me et le visiteur lors de son identification, ce dernier acc  de a la page d'accueil puis atteint le formulaire pour s'identifier, le syst  me lui envoi sa demande puis l'adh  rant soumet sa requ  te, le syst  me lui renvoi    son tour une r  ponse, si elle existe dans la base il acc  de    son espace sinon une erreur s'affiche.

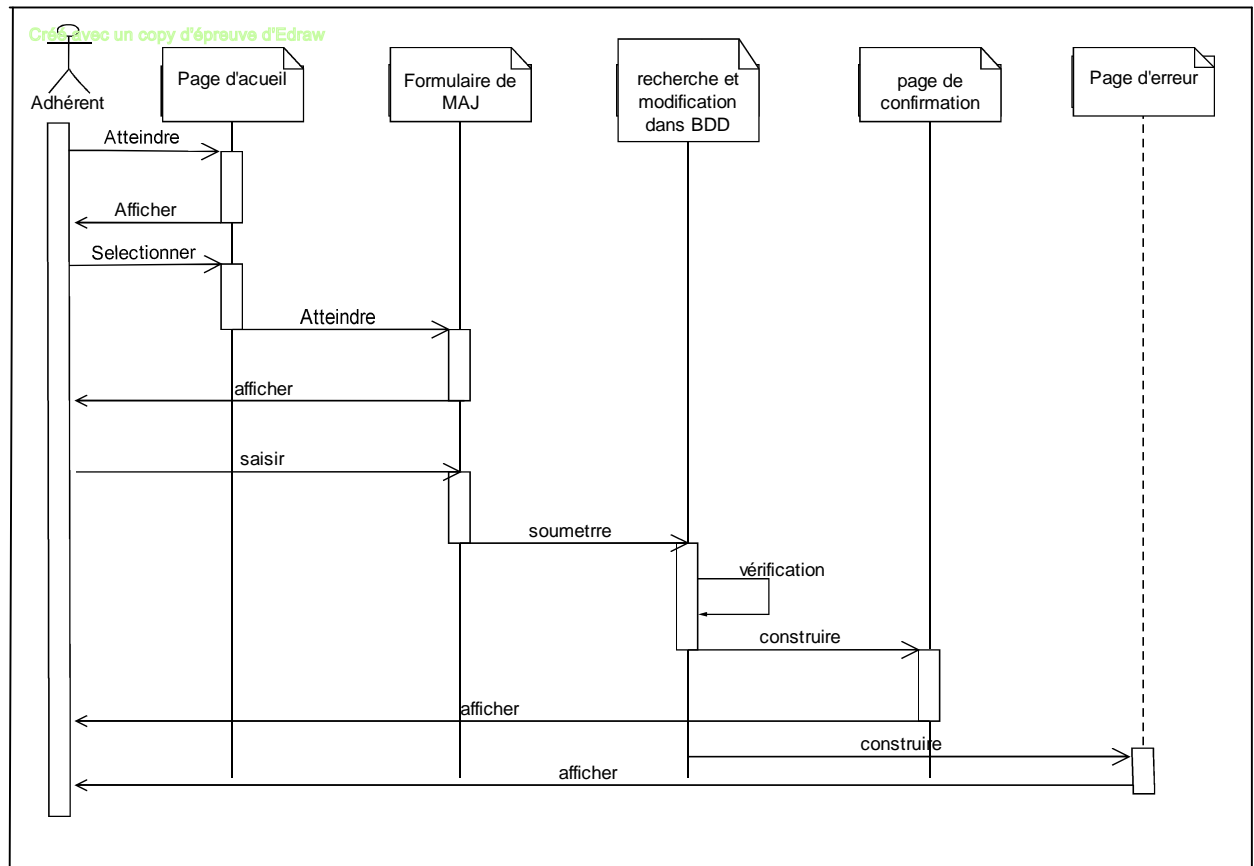


Figure III.15 Diagramme de s  quence du cas d'utilisation “MAJ des pr  f  rences”

Ce diagramme illustre l’interaction entre le syst  me et l’adh  rent lors de mise    jour des pr  f  rences de ce dernier qui acc  de    son espace puis atteint le formulaire de MAJ, le syst  me lui envoie sa demande puis l’adh  rent modifie ses pr  f  rences que le syst  me sauvegarde dans la base.

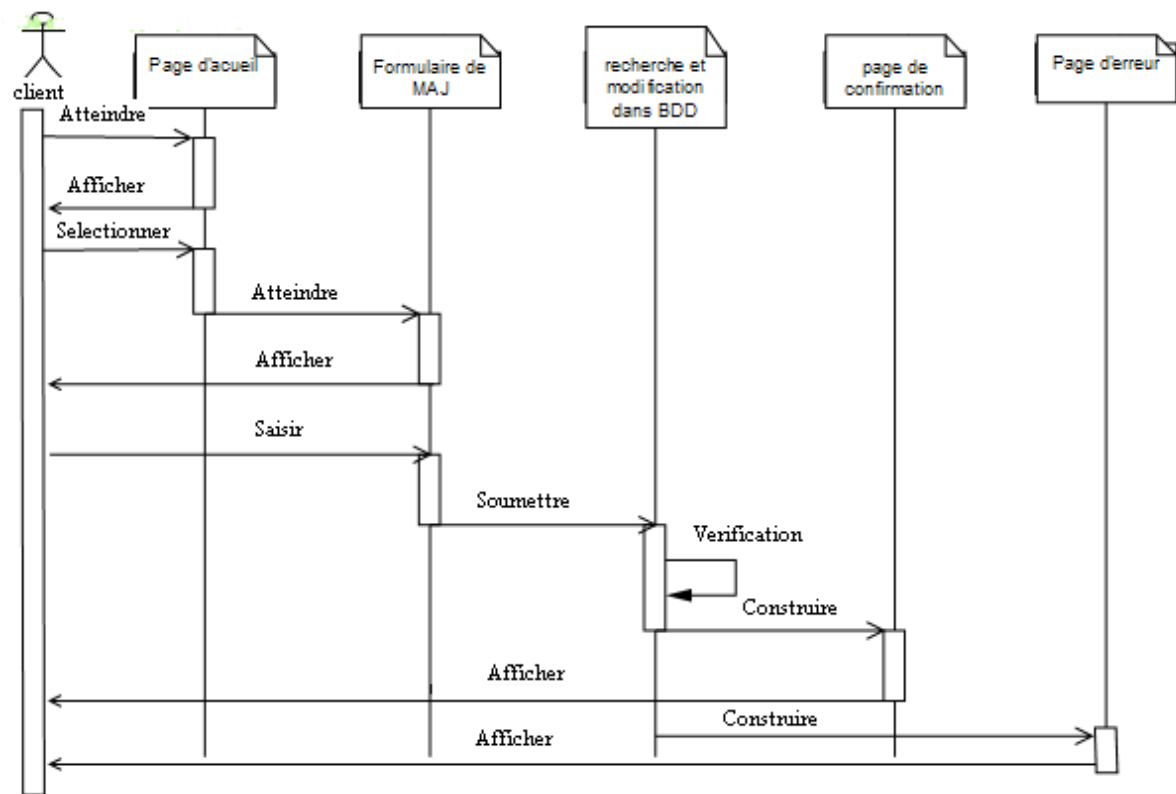


Figure III.16 Diagramme de séquence du cas d'utilisation ‘Modification des caractéristiques des hôtels.’

Ce diagramme illustre l’interaction entre le système et le client lors de mise à jour des caractéristiques de l’hôtel de ce dernier qui accède à son espace puis atteint le formulaire de MAJ, le système lui envoie sa demande puis le client modifie les caractéristiques de l’hôtel que le système transmettra à l’administrateur pour les traiter (accepter ou refuser).

III.6.1.3 Diagramme de classe :

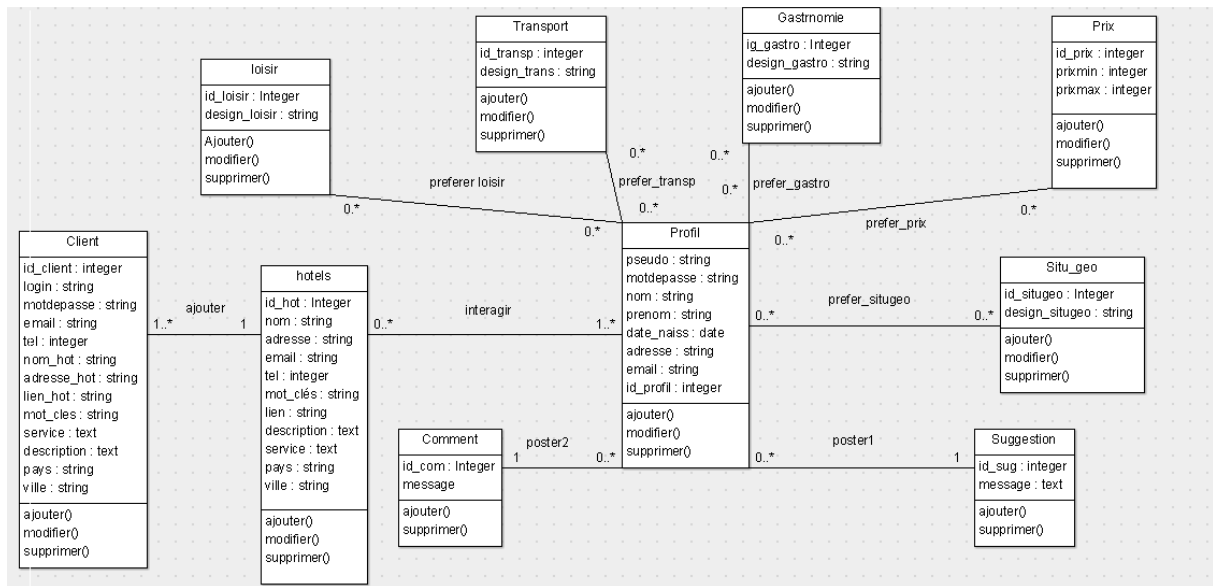


Figure III.17 Diagramme de classe

III.6.1.4 Dictionnaire de données :

Table	Identificateur	Attribut	Désignation	Type	Taille
Profil	Id_profil	Id_profil	Identificateur du profil	entier	11
		Pseudo	Pseudo de l'adhérant	Alpha-num	25
		Motdepasse	Mot de passe de l'adhérant	Alpha-	
		Nom	nom de l'adhérant	num	25
		Prénom	prénom de l'adhérant	Alpha-	25
		Date_naiss	date de naissance //	num	
		Adresse	adresse //		50
		Email	email //	date	35
Loisir	Id_loisir	Id_loisir	Identificateur du loisir	entier	35
		Design loisir	La désignation de loisir	Alpha-num	40
gastron	Id_gastro	Id_gastro	Identificateur gastronomie	entier	25
		Design_gastro	La désignation	Alpha-	30

			gastronomie	num	
Prix	Id_prix	Id_prix	Identificateur prix	entier	11
		Min	L'intervalle du prix		11
		max	Le prix minimum	entier	
			Le prix maximum	entier	11
Situatio geograp	Id_stgeo	Id_stgeo	Identificateur situation	entier	11
			Géographique		
		Design_stgeo	Désignation situation géographique	Alpha-num	30
transpo	Id_trans	Id_trans	Identificateur transport	entier	11
		Design_trans	Désignation transport	Alpha-num	30
comme	Id_com	Id_com	Identificateur de	entier	11
			commentaire		
		message	éditer un message	Text	50
suggest	Id_sug	Id_sug	Identificateur suggestion	entier	11
		message	Faire une suggestion	Text	50
Hotel	Id_hotel	Id_hotel	Identificateur hôtel	entier	11
		Nom	Nom du l'hôtel	Alpha-num	20
		Adresse	adresse du l'hôtel	Alpha-num	50
		Mots_cle	Les mots clés du l'hôtel	Alpha-num	80
		Liens	lien du l'hôtel	Alpha-num	50
Client	Id_client	Id_client	Identificateur client	entier	11
		login	Login client	Alpha-num	25

Tab III.3 Dictionnaire de données.

III.6.1.5 Les modèles logiques de donnée :

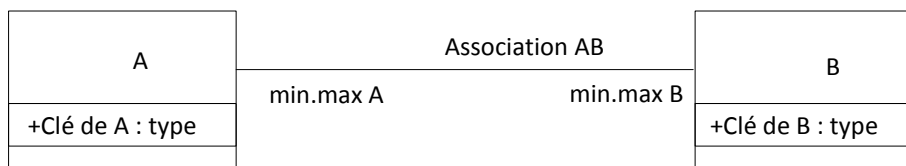
Afin d'assurer la cohérence de la base de données lors du passage du modèle conceptuel (le digramme de classes) au modèle logique, un certain nombre de règles doivent être respectées. Ces règles sont les suivantes :

A- Les classes : La règle est assez simple, « chaque classe devient une relation ».

Chaque attribut de la classe devient un attribut de la relation. Concernant la clé, soit il existe une clé pour la classe qui deviendra la clé de la relation, ou bien, on ajoute une clé artificielle à la relation.

B- Traduire les associations : L'objectif est de mémoriser les liens entre les objets des classes reliées par l'association. La solution à cette mémorisation dépend de la cardinalité de l'association, plus exactement de ses maximums.

Soient les deux classes A et B suivantes, qui sont reliées par une association AB :



Le tableau suivant examine les cas possibles concernant le maximum des cardinalités de A et de B :

MaxA \ MaxB	1	>1
1	Ajouter la clé de B dans la relation de A comme attribut ou Ajouter la clé de A dans la relation de B comme attribut.	Créer une relation AB ayant comme attribut la clé de A et la clé de B.
>1	Ajouter la clé de A dans la relation de B comme attribut.	Ajouter la clé de B dans la relation de A comme attribut.

Tab III.4 Tableau de cardinalités.

En se basant sur les règles citées précédemment pour le passage du modèle conceptuel, nous obtenons le modèle relationnel que nous allons définir dans le point suivant

III.6.1.6 Modèle relationnel global :

A partir de la description conceptuelle que nous avons effectuée, on peut réaliser le modèle relationnel suivant :

Profil (**pseudo**, mot de passe, nom, prénom, âge, adresse, email) ;

Situa_géo (**id_géo**, design_géo) ;

Gastronomies (**id_gastro**, design_gastro) ;

Loisirs (**id_loisir**, design_loisir) ;

Prix (**id_prix**, interval, min,max) ;

Transport (**id_transp**, design_transp) ;

Prefer_géo(**pseudo**,**id_geo**,valeur_geo) ;

Prefer_gastro (**pseudo**,**id_gastro**,valeur_gastro) ;

Prefer_prix(**pseudo**, **id_prix**, valeur_prix) ;

Prefer_transport (**pseudo**,**id_transport**,valeur_transport) ;

Prefer_loisir(**pseudo**,**id_loisir**,valeur_loisir) ;

Hotel(**id_hot**,nom,adresse,email,tel,mots_clés,lien,description,services,pays,ville,login,**id_geo**
id_prix,login) ;

log (**id_log**,nom_d'interaction,date, temps, **pseudo**, **id_hot**) ;

Client(**id_client**,login,motdepasse,email,tel,nom_hot, adresse_hot,lien_hot,mots_cles,services
Description, pays, ville)

III.7 Conclusion :

Dans ce chapitre nous avons essayé de montrer l'approche utilisée pour concevoir notre système personnalisé pour l'interrogation d'une base de données concernant les hôtels comme cas d'étude, auquel nous avons intégré le profil utilisateur et la possibilité de l'enrichir pour permettre de servir chaque adhérent à son juste besoin.

Dans le chapitre suivant, nous allons présenter l'implémentation du système et la mise en œuvre de la démarche proposée dans cette étude conceptuelle.

Chapitre IV

La réalisation du système

IV.1 Introduction

Après avoir proposé notre solution puis élaborer sa conception, nous allons passer à son implémentation. Pour ce faire, nous allons d'abord décrire l'environnement de développement et les outils utilisés. Ensuite, nous allons citer les principales classes développées et décrire les différentes fenêtres de l'application. Enfin, nous allons présenter quelques tests effectués, sur notre application.

IV.2 Environnement de développement

Langage Java: Nous avons choisi pour l'implémentation de notre système le langage JAVA. Ce dernier a été mis au point par Sun Microsystems en 1991. C'est un langage orienté objet créé à base du langage c++. Il est multi plate forme grâce à sa machine virtuelle appelée Java Virtuelle Machine(JVM). Nous avons aussi choisi le langage JAVA pour l'ensemble d'outils et API open source qu'il offre.

JSP⁽⁵⁾ : Pour générer nos pages web et y intégrer du contenu dynamique nous avons utilisé JavaServer Pages ou JSP qui est une technique basée sur Java, cette technique permet au code Java et à certaines actions prédéfinies d'être ajoutés dans un contenu statique.

Servlet : Une **servlet** est une classe Java qui permet de créer dynamiquement des données au sein d'un serveur HTTP. Ces données sont le plus généralement présentées au format HTML, mais elles peuvent également l'être au format XML ou tout autre format destiné aux navigateurs web. Les servlets utilisent l'API **Java Servlet** (package javax.servlet).

Une servlet s'exécute dynamiquement sur le serveur web et permet l'extension des fonctions de ce dernier, typiquement : accès à des bases de données, transactions d'e-commerce, etc. Une servlet peut être chargée automatiquement lors du démarrage du serveur web ou lors de la première requête du client. Une fois chargées, les servlets restent actives dans l'attente d'autres requêtes du client.

L'utilisation de servlets se fait par le biais d'un conteneur de servlets (framework) côté serveur. Celui-ci constitue l'environnement d'exécution de la servlet et lui permet de persister entre les requêtes des clients. L'API définit les relations entre le conteneur et la servlet. Le conteneur reçoit la requête du client, et sélectionne la servlet qui aura à la traiter. Le conteneur fournit également tout un ensemble de services standards pour simplifier la gestion des requêtes et des sessions.

⁽⁵⁾ http://fr.wikipedia.org/wiki/JavaServer_Pages

Une ou plusieurs servlets constituent les *applications web* Java, leur principe de fonctionnement (architecture, fonctionnalités, configuration, déploiement) est décrit suivant une spécification officielle, menée par Sun Microsystems et à laquelle toute personne peut contribuer par le biais des Java Specification Requests (Java Community Process). La version actuelle des spécifications servlet est la **3.0**.

Apache Tomcat 6: est un serveur Web qui gère les servlets et les JSP. Comme Tomcat inclut un serveur HTTP interne, il est aussi considéré comme un serveur HTTP. Tomcat est un outil open source qui a été écrit en langage Java, il peut donc s'exécuter via la JVM (machine virtuelle java) sur n'importe quel système d'exploitation la supportant.

Environnement de développement intégré (IDE) : Nous avons utilisé ECLIPSE HELIOS comme environnement de développement intégré (IDE). Il à été développé par IBM et considéré comme un IDE multi langages.

PhpMyadmin : L'outil PhpMyadmin est développé en PHP (ensemble de script PHP), il offre une interface graphique pour l'administration des bases de données MySQL via un navigateur Web. Ses fonctions principales sont :

- Création de nouvelles bases de données ;
- Création / suppression / modification des tables ;
- Edition, ajout et suppression de champs ;
- Exécution de commandes SQL et de requêtes ;
- Gestion des privilèges des utilisateurs.

La figure suivante montre l'interface principale de phpMyAdmin :

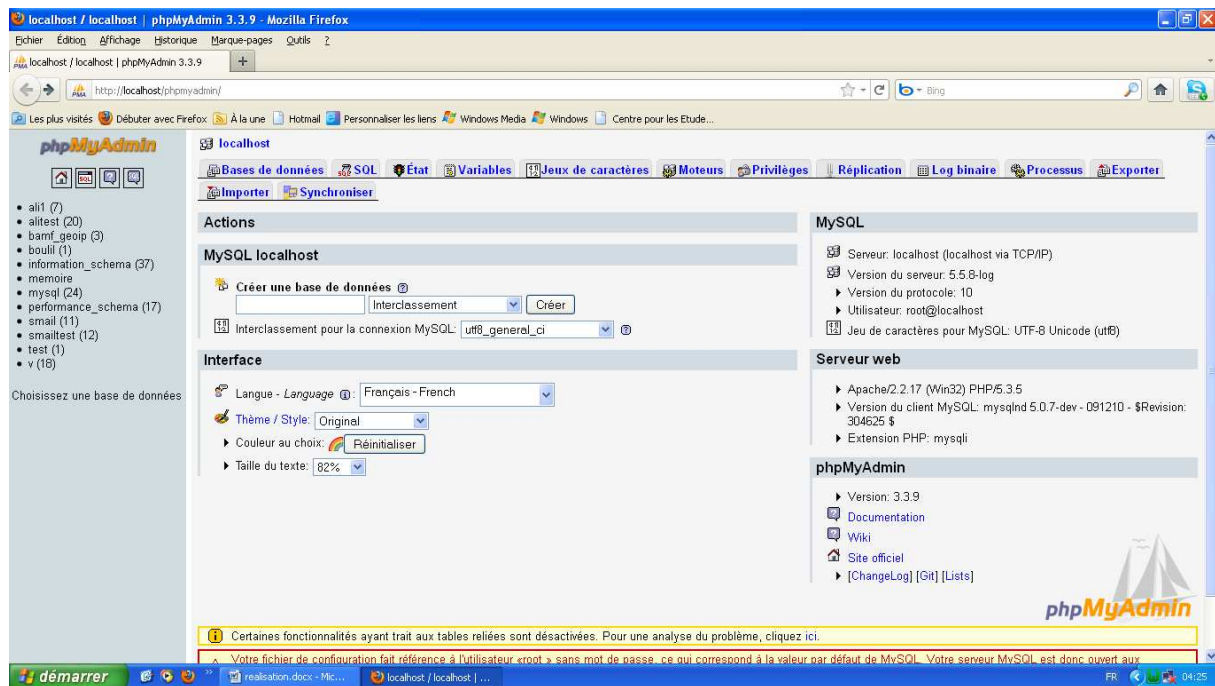


Figure IV.1 Interface principale de phpMyAdmin.

Notre système est doté d'une architecture à 3 niveaux (3 tiers) où il existe un niveau intermédiaire, c'est-à-dire que l'on a généralement une architecture partagée entre :

- Couche présentation
- Couche métier
- Couche accès aux données

Cette architecture peut être définie comme étant un modèle logique d'architecture applicative qui vise à séparer très nettement trois couches logicielles au sein d'une même application ou système, à modéliser et présenter cette application comme un empilement de trois strates dont le rôle est le suivant :

- La présentation des données : correspondant à l'affichage, la restitution sur le poste de travail, le dialogue avec l'utilisateur
- Le traitement métier des données : correspondant à la mise en œuvre de l'ensemble des règles de gestion et de la logique applicative
- L'accès aux données persistantes (persistance en anglais) : correspondant aux données qui sont destinées à être conservées sur la durée, voire de manière définitive

Dans cette approche, les couches communiquent entre elles au travers d'un « modèle d'échange », et chacune d'entre elles propose un ensemble de services rendus. Les services d'une couche sont mis à disposition de la couche supérieure. On s'interdit par conséquent qu'une couche invoque les services d'une couche plus basse que la couche immédiatement inférieure ou plus haute que la couche immédiatement supérieure (chaque niveau ne communique qu'avec ses voisins immédiats).

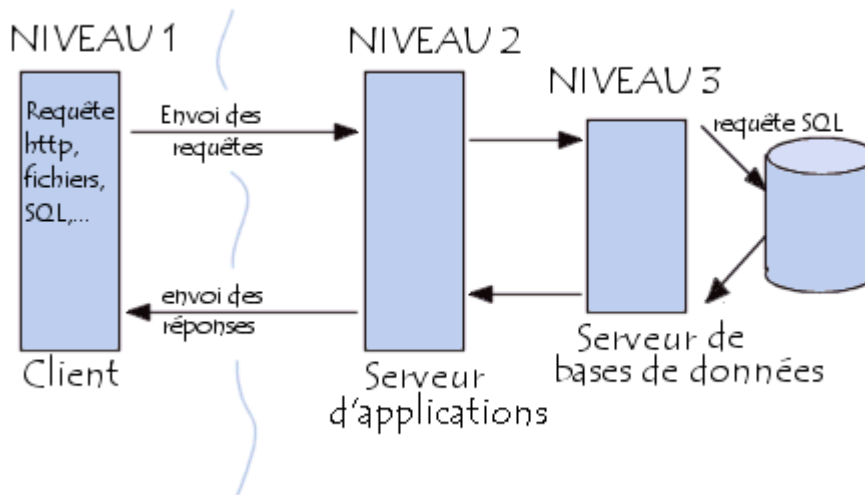


Figure IV.2 Architecture 3 tiers.

IV.3. Interfaces de l'application

Afin d'exploiter au mieux notre système de recherche d'information, nous avons mis en place une interface homme machine simple dont nous décrivons les principales fenêtres ci-dessous.

IV.3.1 Interface d'accueil

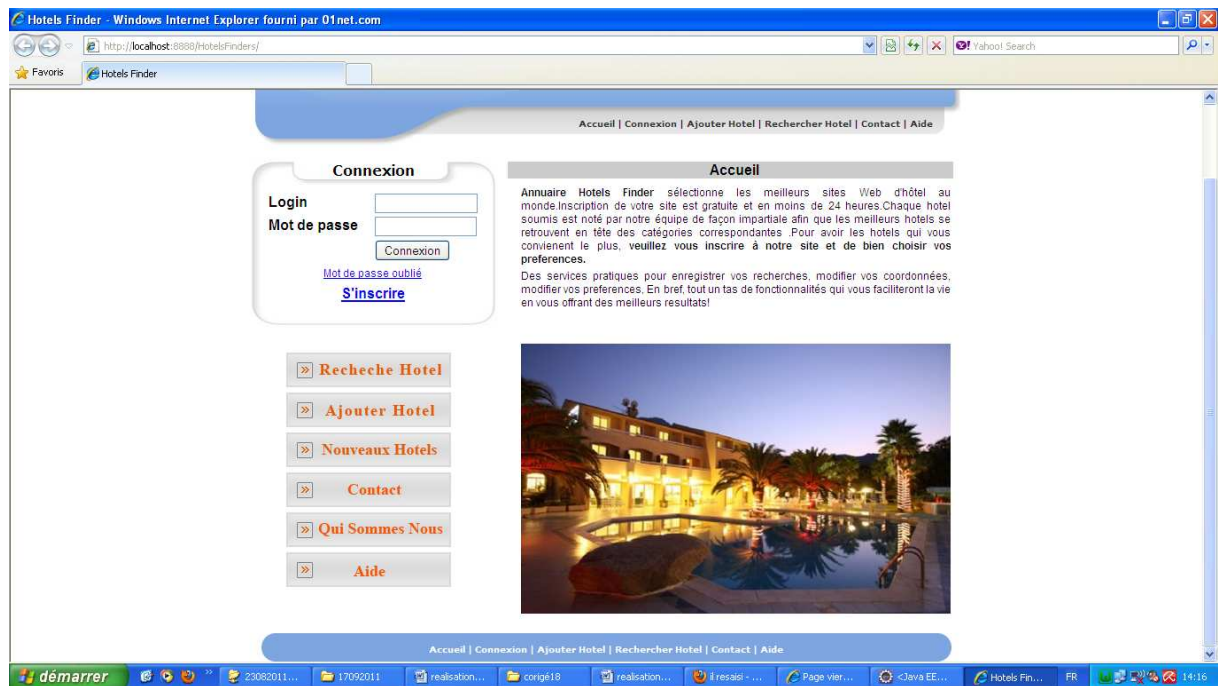


Figure IV.3 Interface d'accueil

Le visiteur de notre site web peut chercher un hôtel, s'inscrire pour devenir membre, ajouter un hôtel, contacter l'administrateur.

IV.3.2 Interface identification adhérent :

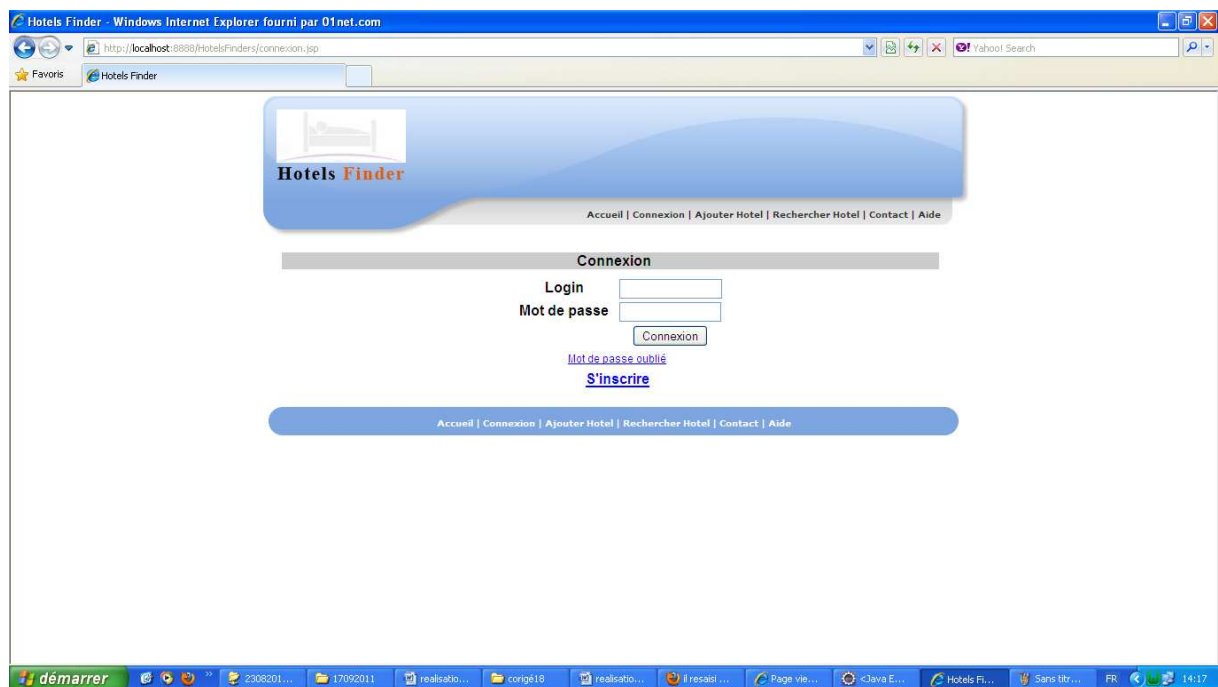


Figure IV.4 Interface identification adhérent

En cliquant sur le bouton **connexion** après avoir introduit le pseudo et mot de passe l'utilisateur accède a son espace.si il a oublié son mot de passe, il aura la possibilité de le récupérer en contactant l'administrateur en cliquant sur le lien « mot de passe oublié ».

IV.3.3 Interface inscription abonné

The screenshot shows a web browser window titled 'Hotels Finder - Mozilla Firefox'. The address bar displays 'http://localhost:8888/HotelsFinders/inscire.jsp'. The page content includes a navigation bar with links: 'Accueil | Connexion | Ajouter Hotel | Rechercher Hotel | Contact | Aide'. Below this is a header for 'Hotels Finder'. The main form is titled 'S'inscrire' and is divided into two sections: 'Données personnelles' and 'Préférences'. The 'Données personnelles' section contains the following fields: 'Pseudo', 'Mot de passe', 'Retaper Mot de passe', 'Nom', 'Prenom', 'Date de naissance' (with dropdowns for 'jj', 'mm', and 'aa'), 'Adresse', and 'E-mail'. The 'Préférences' section contains: 'Loisirs', 'Situation géographique', 'Transport', 'Prix' (with 'Max' and 'Min' input fields), 'Monnaies' (with a dropdown), 'Gastronomie', and 'Autre'. A 'Valider' button is located at the bottom of the form. The Windows taskbar at the bottom shows various open applications and the system clock.

Figure IV.5 Interface inscription abonné

En cliquant sur le bouton valider[4] après avoir saisi les informations le visiteur sera ajouté dans la base de donnée et devient adhérent ou il pourra accéder a son espace.

IV.3.4 Interface abonné



Figure IV.6 Interface Abonné

En cliquant sur le bouton **rechercher** et après avoir tapé les mots clé l'adhérant va avoir les résultats selon ses préférences sauvegardées par le système et en cliquant sur le bouton **Modifier profil**, l'adhérant peut modifier ses préférences et il peut aussi contacter l'administrateur en cliquant sur le bouton **Contact**.

IV.3.5 Interface connexion Administrateur

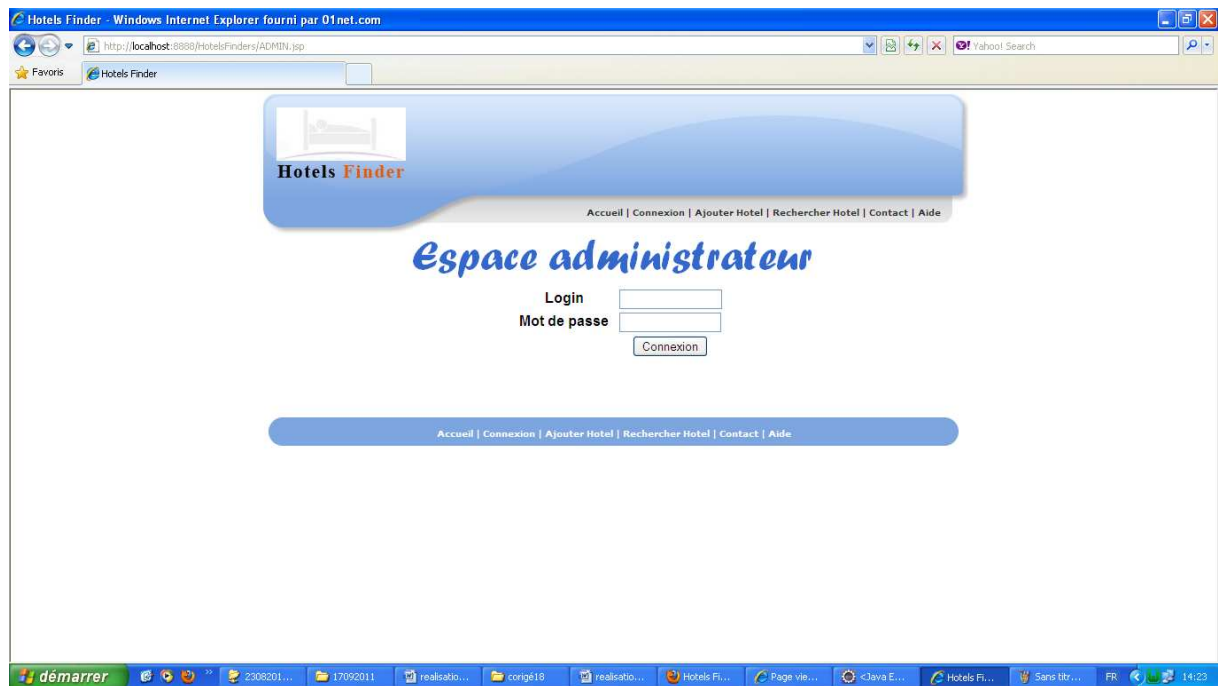


Figure IV.7 Interface connexion administrateur

En cliquant sur le bouton **connexion** après avoir introduit le pseudo et mot de passe l'administrateur va accéder a son espace.

IV.3.6 Interface page d'accueil Administrateur

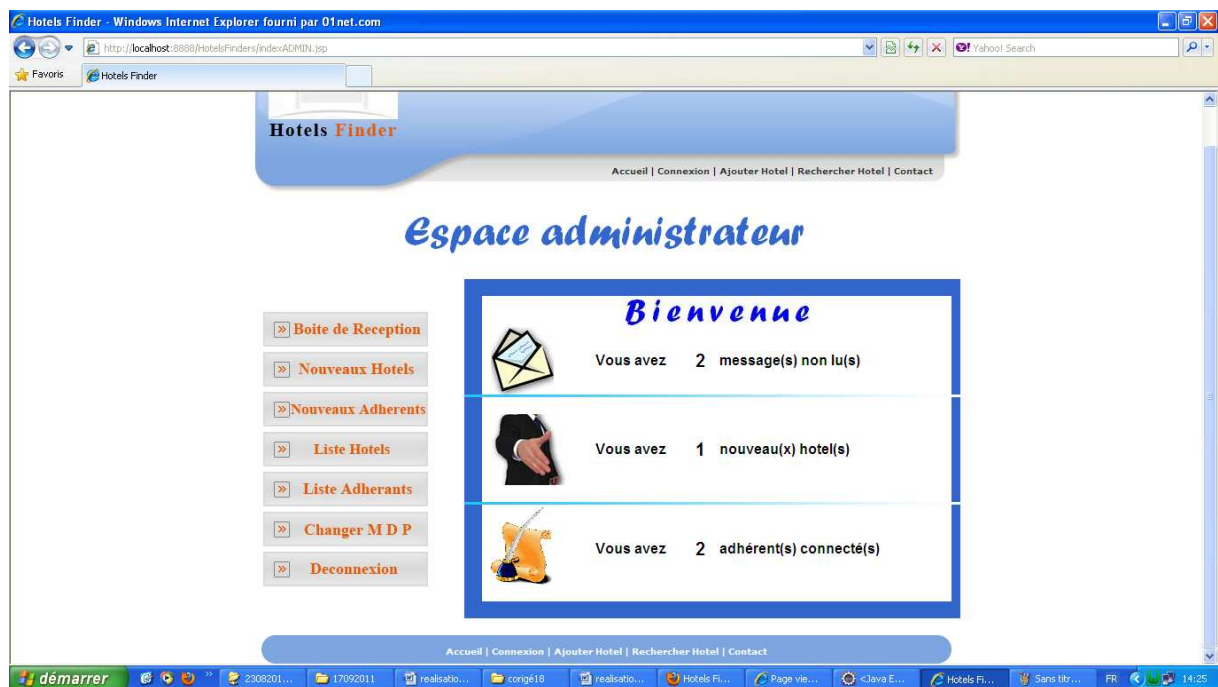


Figure IV.8 Interface d'accueil administrateur

En cliquant sur le bouton **Boite de Réception**, l'administrateur peut accéder à sa boîte de réception de message. En cliquant sur le bouton **Nouveaux Hôtels**, l'administrateur peut visualiser les nouveaux hôtels et les traiter. Il peut accéder à la liste des hôtels, liste adhérents en cliquant respectivement sur le bouton **Liste Hôtels**, **Liste Adhérents**.

IV.3.7 Interface administrateur de MAJ des hôtels



Figure IV.9 Interface MAJ hôtels

L'administrateur peut modifier, supprimer l'hôtel en cliquant respectivement sur le lien **Modifier**, **Supprimer** en bas de page.

IV.3.8 Interface connexion client :

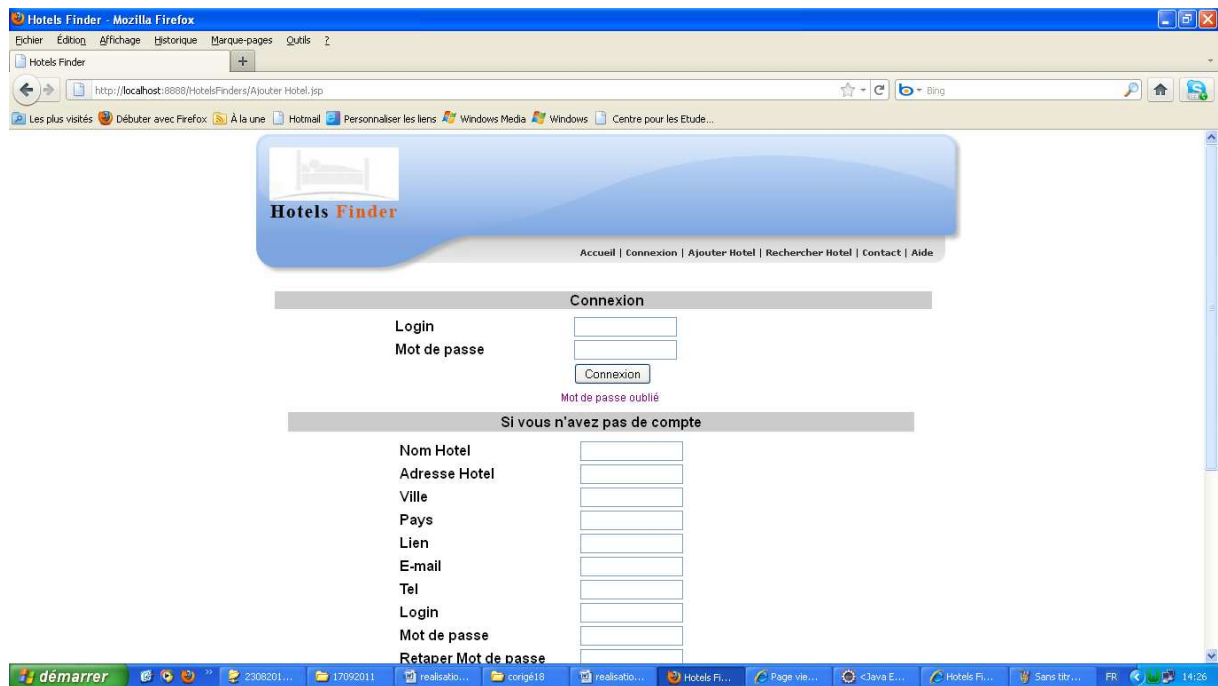


Figure IV.10 Interface connexion client

En cliquant sur le bouton **connexion** après avoir introduit le pseudo et mot de passe le client accède à son espace. Si il a oublié son mot de passe, il aura la possibilité de le récupérer en contactant l'administrateur en cliquant sur le lien « mot de passe oublié ».

IV.3.9 Interface Ajouter hôtel :

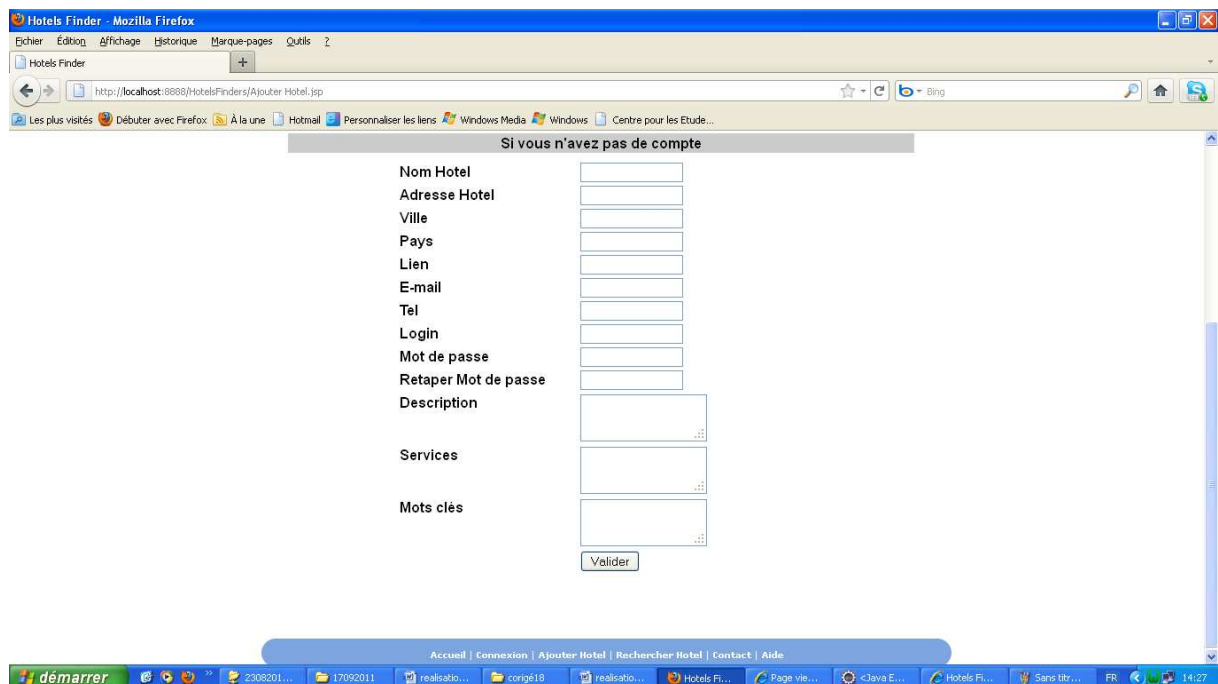


Figure IV.11 Interface inscription client

En cliquant sur le bouton valider après avoir saisi les informations personnelles et informations sur de l'hôtel, le client sera ajouté dans la base de donnée et devient client ou il pourra accéder à son espace par contre l'hôtel ne sera pas accepté qu'après traitement par l'administrateur.

IV.3.10 Interface espace client :

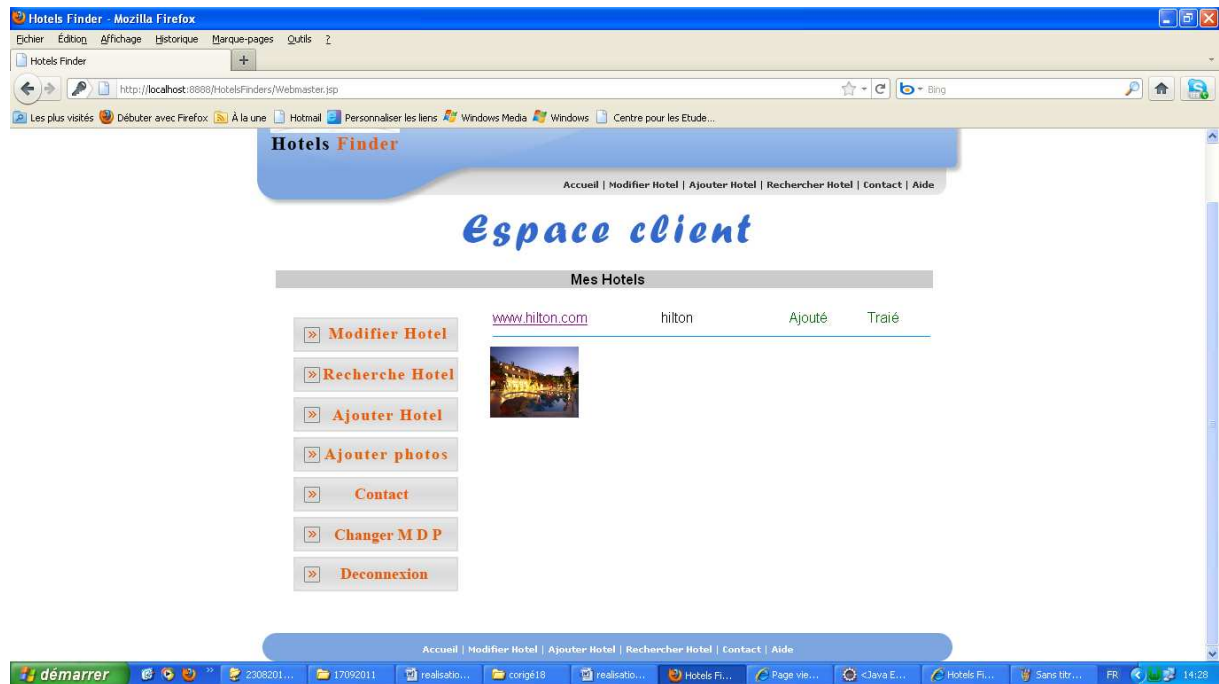


Figure IV.12 Interface espace client

Le client peut modifier les caractéristiques de son hôtel par le bouton **Modifier Hôtel**, il peut ajouter d'autre hôtels par le bouton **Ajouter Hôtel**, il peut ajouter des photos à son hôtel par le bouton **Ajouter Photos**, il peut aussi contacter l'administrateur par le bouton **Contact**.

IV.4. Mise en œuvre de la solution :

IV.4.1 Recherche non personnalisée :

Si un utilisateur désire effectuer une recherche sur les hôtels disponibles dans la base, il devra après avoir saisi l'url du site, cliquer sur le bouton recherche hôtel se trouvant sur la page d'accueil, un formulaire de recherche lui sera retourné, lui permettant ainsi de faire des recherches selon les mots clés qu'il a saisi. Ces recherches qu'on a appelé anonymes car on ne connaît pas leur émetteur vu qu'il n'est pas inscrit, sont bien évidemment dépourvu de personnalisation. Dans cette recherche, on a proposé d'intégrer le procédé de la géolocalisation dans le processus de recherche pour aider l'utilisateur à avoir des résultats pertinents selon le contexte de la localisation.

Organigramme de fonctionnement : la recherche non personnalisée se fait selon l'organigramme suivant :

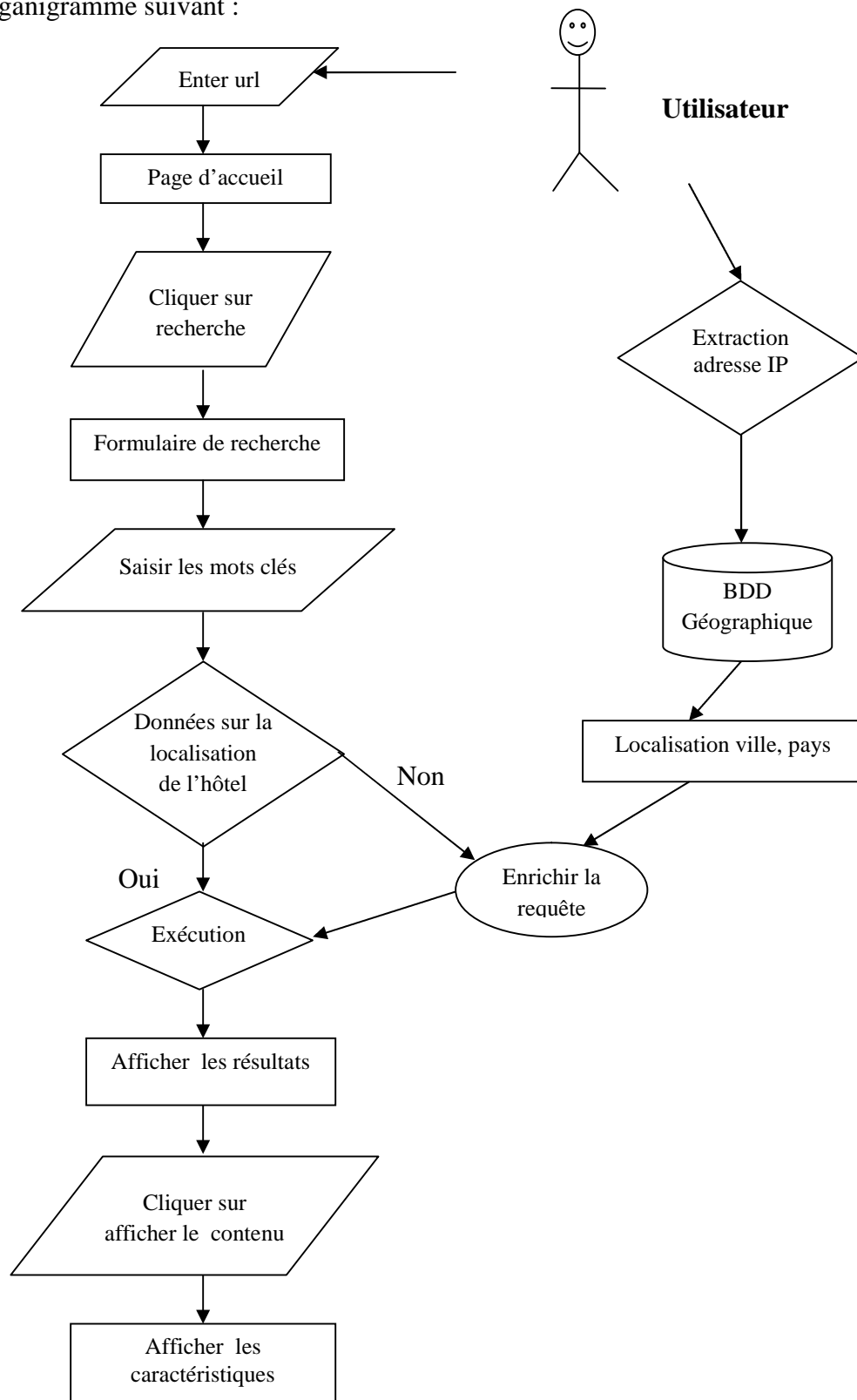


Figure IV.13 Organigramme de l'exécution de la recherche non personnalisée

IV.4.1 .1 Enrichir la requête d'un utilisateur selon le contexte de la localisation :

Test :

Un utilisateur qui ne spécifie pas la localisation de l'hôtel dans sa requête, notre système lui proposera les hôtels qui lui sont proches.

Exemple : un utilisateur qui se situe à « Oran », s'il saisi la requête « 3etoile » alors il aura les résultats suivants :

Resultats Recherche		
Mots clés	<input type="text" value="3etoile"/>	<input type="button" value="Rechercher"/>
www.Cotedor.com	Cotedor	Afficher
www.Cherton.com	Cherton	Afficher
www.Atlas.com	Atlas	Afficher
www.djurjura.com	djurjura	Afficher

Figure IV.14 Résultats recherche selon le contexte de la localisation

-Afficher les caractéristiques de l'hôtel qui est affiché le premier dans la liste des résultats :

Nom Hotel:	Cotedor
Adresse Hotel:	oran
Ville Hotel:	oran
Pays Hotel:	algerie
Lien:	www.Cotedor.com
E-mail:	Cotedor@yahoo.com
Tel:	 0021341000000 

Figure IV.15 Caractéristiques de l'hôtel « Cotedor »

Remarque : On remarque que l'hôtel « Cotedor » correspond mieux à la localisation du visiteur.

Par contre sans le procédé de la localisation, les résultats seront ainsi

Resultats Recherche		
Mots clés	<input type="text" value="3etoile"/>	<input type="button" value="Rechercher"/>
www.Cheraton.com	Cheraton	Afficher
www.Cotedor.com	Cotedor	Afficher
www.Atlas.com	Atlas	Afficher
www.djurjura.com	djurdjura	Afficher

Figure IV.16 Résultats recherche sans le contexte de la localisation

-Afficher les caractéristiques de l'hôtel qui est affiché le premier dans la liste des résultats :



Nom Hotel:	Cheraton
Adresse Hotel:	alger
Ville Hotel:	alger
Pays Hotel:	algerie
Lien:	www.Cheraton.com
E-mail:	Cheraton@gmail.com
Tel:	 0021321000000 
Description:	Cheraton est un hotel restaurant 3etoile qui se situe a la ville d alger.

Figure IV.17 Caractéristiques de l'hôtel « Cheraton »

Remarque : On remarque que l'hôtel «Cheraton» ne correspond pas à la localisation du visiteur.

IV.4.2 Recherche personnalisée :

Pour pouvoir procéder à une recherche personnalisée, l'abonné devra d'abord passer par l'identification présenté précédemment, ensuite il pourra saisir sa requête sous forme de mots clés ,le système récupérera les mots clés saisis par l'abonné, auxquels il ajoutera les préférences récupérées du profil de l'abonné ,puis affichera les résultats .

Organigramme de fonctionnement : la recherche personnalisée se fait selon l'organigramme suivant :

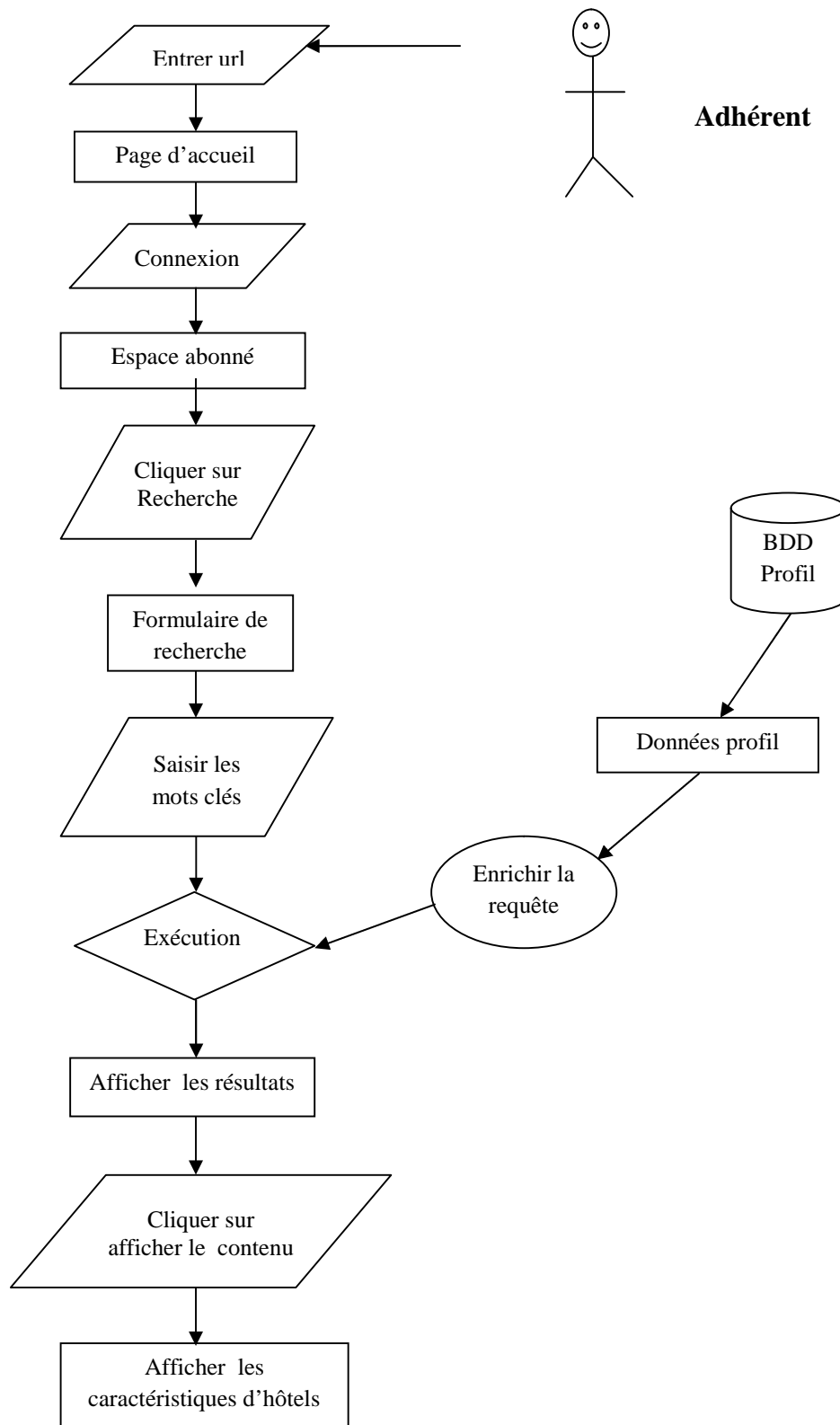


Figure IV.18 Organigramme de l'exécution de la recherche personnalisée.

IV.4.2 .1 Test :

On prend par exemple deux abonnés avec deux profils différents :

Profil abonné n°1 :

Données personnelles :	
Nom :	abonné1
Prenom :	
Date de naissance :	1970-01-01
Adresse :	alger
E-mail :	abonné1@yahoo.com
Preferences :	
Loisirs :	sona
Situation géographique :	ville
Transport :	
Prix Min :	50
Prix Max :	80
Monnaies :	dinarALG
Gastronomie :	orientale
Autre :	

Figure IV.19 Profil abonné n°1.

Profil abonné n°2 :

Données personnelles :

Nom : abonné2
Prenom :
Date de naissance : 1986-01-01
Adresse : tizi ouzou
E-mail : abonné2@hotmail.com

Preferences :

Loisirs : natation
Situation géographique : montagne
Transport :
Prix Min : 70
Prix Max : 120
Monnaies : dinarALG
Gastronomie : française
Autre :

Figure IV.20 Profil abonné n°2.

On suppose que les deux abonnés saisissent la requête « piscine, 3etoile » :

Résultats pour l'abonné n°1 :

Mots clés

www.Cheraton.com	Cheraton	Afficher
www.Hilton.com	Hilton	Afficher
www.Cotedor.com	Cotedor	Afficher
www.Atlas.com	Atlas	Afficher
www.djurjura.com	djurdjura	Afficher
www.hilton.com	hilton	Afficher

Figure IV.21 Résultats recherche pour abonné n°1 selon le profil.

Afficher les caractéristiques de l'hôtel qui est affiché le premier dans la liste des résultats :

Nom Hotel: Cheraton
Adresse Hotel: alger
Ville Hotel: alger
Pays Hotel: algerie
Lien: www.Cheraton.com
E-mail: Cheraton@gmail.com
Tel:  0021321000000 
Description: Cheraton est un hotel restaurant 3etoile qui se situe a la ville d alger.
Services: parking,piscine,sona,restaurant.
Mots clés: alger,piscine,3etoile,sona,ville,cheraton



[j'aime](#) [imprimer](#)

Figure IV.22 Caractéristiques de l'hôtel « Cheraton » dans l'espace membre.

Remarque : On remarque que l'hôtel « Cherraton » correspond mieux à son profil.

Résultats pour l'abonné n°2 :

Mots clés

www.Atlas.com	Atlas	Afficher
www.djurjura.com	djurdjura	Afficher
www.Hilton.com	Hilton	Afficher
www.Cheraton.com	Cheraton	Afficher
www.Cotedor.com	Cotedor	Afficher
www.hilton.com	hilton	Afficher

Figure IV.23 Résultats recherche pour abonné n°2 selon le profil.

Afficher les caractéristiques de l'hôtel qui est affiché le premier dans la liste des résultats :

Nom Hotel:	Atlas
Adresse Hotel:	tizi ouzou
Ville Hotel:	tizi ouzou
Pays Hotel:	algerie
Lien:	www.Atlas.com
E-mail:	Atlas@yahoo.fr
Tel:	 0021326555555 
Description:	Atlas est un hotel 3etoile qui se situe en plein coeur de Djurjura
Services:	parking,sona,animaux admis,restaurant.
Mots clés:	atlas,tikjda,tizi ouzou,3etoile,montagne,piscine

[j'aime](#)[imprimer](#)

Figure IV.24 Caractéristiques de l'hôtel « Atlas » dans l'espace membre.

On remarque que les caractéristiques de l'hôtel « Atlas » correspond mieux au profil.

IV.4.2 .1 Test sur la mise à jour du profil :

Si l'abonné n°1 consulte l'hôtel « Atlas », voir figure IV.24 au-dessus

Nom Hotel: Atlas
Adresse Hotel: tizi ouzou
Ville Hotel: tizi ouzou
Pays Hotel: algerie
Lien: www.Atlas.com
E-mail: Atlas@yahoo.fr
Tel:  0021326555555 
Description: Atlas est un hotel 3etoile qui se situe en plein coeur de Djurjura
Services: parking,sona,animaux admis,restaurant.
Mots clés: atlas,tikjda,tizi ouzou,3etoile,montagne,piscine



(1)  [j'aime](#) [imprimer](#)

Figure IV.25 Caractéristiques de l'hôtel « Atlas » dans l'espace membre.

S'il considère cet hôtel comme pertinent en cliquant par exemple sur « j'aime » voir (1) dans la figure IV.25, alors son profil sera mis à jour.

Données personnelles :

Nom : abonné1
Prenom :
Date de naissance : 1970-01-01
Adresse : alger
E-mail : abonné1@yahoo.com

Preferences :

Loisirs : sona
Situation géographique : ville
Transport :
Prix Min : 50
Prix Max : 80
Monnaies : dinarALG
Gastronomie : orientale
Autre : atlas,tikjda,tizi
ouzou,3etoile,montagne,piscine

Figure IV.26 Mise à jour du profil.

Remarque : On remarque que les mots clés de l'hôtel Atlas sont insérer dans champ autre du profil de l'abonné n°1 (voir figure IV.26) d'où le profil a été mis à jours.

Maintenant s'il ressaisi la requête précédente « piscine, 3etoile », voici les résultats :

Mots clés	<input type="text" value="piscine,3etoile"/>	<input type="button" value="Rechercher"/>
www.Atlas.com	Atlas	Afficher
www.Cheraton.com	Cheraton	Afficher
www.Cotedor.com	Cotedor	Afficher
www.djurjura.com	djurdjura	Afficher
www.Hilton.com	Hilton	Afficher
www.hilton.com	hilton	Afficher

Figure IV.27 Résultats recherche à prés mise à jour du profil.

On remarque que le classement des résultats est différent par rapport à celui d'au-dessus malgré que la requête est la même et cela revient à la mise à jour du profil après l'interaction de l'abonné n°1 avec le système.

IV.5 Conclusion

Dans ce chapitre, nous avons présenté les outils de développement puis quelques interfaces de notre système ensuite on a présenté les deux organigrammes de la recherche avec des tests concrets dans le but de montrer l'impact de l'intégration du profil utilisateur ainsi que la géolocalisation dans la pertinence des résultats dans la recherche d'information.

Conclusion Générale

Notre travail s'inscrit dans le domaine de la personnalisation de l'information dans le contexte de la recherche d'information, où nous avons présenté une interface personnalisée pour un annuaire de recherche d'hôtels prenant en compte le profil de l'utilisateur afin de lui offrir des informations pertinentes respectant ses besoins, car aujourd'hui, l'information est considérée comme le produit le plus stratégique : connaître la bonne information, au bon moment, juste au moment où elle est disponible, est le but à tous.

Les principaux problèmes auxquels ce mémoire a apporté des contributions sont :

- (i) La modélisation et la gestion des connaissances qui décrivent l'utilisateur et
- (ii) L'exploitation du profil dans la reformulation des requêtes et l'affichage des résultats de manière adaptatif.

Afin d'aboutir à ces objectifs, nous avons au cours de l'étude bibliographique passé en revue les notions théoriques entourant le sujet, à commencer par donner des notions de base sur la personnalisation et le profil utilisateur, puis les principaux domaines de la personnalisation (RI, IHM, BDD) où nous avons bien détaillé celle concernant la recherche d'information dans laquelle nous présentons une modélisation générique de profil, ce dernier se présente sous la forme de méta modèle qui permet de classer l'ensemble des informations décrivant l'utilisateur, cette modélisation nécessite une identification et une modélisation conceptuelle des différents attributs du modèle, elle a l'avantage d'une implémentation rapide mais reste limitée aux seuls types d'informations prévues.

Le travail réalisé dans ce mémoire a permis la modélisation et la construction dynamique du profil utilisateur, notre contribution se situe à plusieurs niveaux à savoir: D'une part la modélisation des dimensions composant le profil, nous avons estimé que pour offrir à l'utilisateur un espace de recherche personnalisé, notre profil devait être caractérisé par plusieurs dimensions pour contenir la description la plus exhaustive que possible des préférences de l'utilisateur, et d'autre part, la modélisation multidimensionnelle nous a permis de savoir ce que veut un utilisateur de notre système.

En ce qui concerne l'acquisition du profil utilisateur, nous avons estimé que l'acquisition explicite ne pouvait pas suffire à elle seule pour couvrir l'éventail des besoins de l'utilisateur.

C'est la raison pour laquelle nous avons adoptée une démarche hybride combinant l'acquisition explicite et implicite des besoins de l'utilisateur. Ceci est rendu possible à travers la

Conclusion Générale

mise en place d'un système de surveillance pour capturer les principales interactions de l'utilisateur. Ces dernières sont enregistrées dans le fichier **log** qui doit être analysé périodiquement afin de mettre à jour le profil de l'utilisateur. Le but est non seulement de pallier le problème de surcharge cognitive mais aussi et surtout de prédire d'autres préférences non exprimés au départ par lui.

Après avoir modélisé le profil de l'utilisateur, nous l'avons utilisé dans la reformulation des requêtes, afin de générer une requête personnalisée.

Par ailleurs, le contexte spatial a été exploité pour mieux répondre à des requêtes sensibles à la localisation, nous avons incorporé la localisation du client à la requête pour tenir compte des conditions de proximité

Comme la problématique abordée dans notre projet est récente et fait encore l'objet de recherches actives, nous avons été confrontées à quelques difficultés au niveau de l'implémentation de notre prototype, et surtout dans le processus d'analyse du fichier log, Mais les objectifs de notre étude ont été atteints, il y'a toujours place pour son amélioration.

Sur le plan pratique, nous avons réalisé une grande partie du prototype que nous avons conçu à savoir, l'acquisition du profil utilisateur de deux manières différentes explicite et implicite, la reformulation de la requête de l'utilisateur par la méthode d'enrichissement afin de générer la requête personnalisée et enfin l'exécution de cette dernière et l'affichage des résultats de manière adaptatif.

Perspectives

Pour la personnalisation de l'information dans le contexte de la recherche d'information, il est claire que le profil utilisateur est un domaine de recherche très vaste et couvre plusieurs aspects qu'il serait intéressant d'approfondir, tels que :

- Evaluation des performances : faire des tests pour savoir l'impact de la structure du profil sur les performances de système.
- La prise en compte de la notion des préférences composées, c'est-à-dire personnaliser une requête en prenant en compte un groupe de profil utilisateur au lieu d'un seul.
- Faire le perfectionnement des interfaces de notre application.

Annexe A

Le rappel : Le rappel est défini par le nombre de documents pertinents retrouvés au regard du nombre de documents pertinents que possède la base de données. Cela signifie que lorsque l'utilisateur interroge la base il souhaite voir apparaître tous les documents qui pourraient répondre à son besoin d'information. Si cette adéquation entre le questionnement de l'utilisateur et le nombre de documents présentés est importante alors le taux de rappel est élevé. A l'inverse si le système possède de nombreux documents intéressants mais que ceux-ci n'apparaissent pas on parle de silence. Le silence s'oppose au rappel.

$$Rappel_i = \frac{\text{documents correctement attribués à la classe } i}{\text{nombre de documents appartenant à la classe } i}$$

La précision : La précision est le nombre de documents pertinents retrouvés rapporté au nombre de documents total proposé par le moteur de recherche pour une requête donnée. Le principe est le suivant: quand un utilisateur interroge une base de données, il souhaite que les documents proposées en réponse à son interrogation correspondent à son attente. Tous les documents retournés superflus ou non pertinents constituent du bruit. La précision s'oppose à ce bruit documentaire. Si elle est élevée, cela signifie que peu de documents inutiles sont proposés par le système et que ce dernier peut être considéré comme "précis". On calcule la précision avec la formule suivante:

$$Précision_i = \frac{\text{documents correctement attribués à la classe } i}{\text{nombre de documents attribués à la classe } i}$$

Annexe B

Le **PageRank** ou *PR* est l'algorithme d'analyse des liens concourant au système de classement des pages Web utilisé par le moteur de recherche Google pour déterminer l'ordre dans les résultats de recherche qu'il fournit. De nos jours le PageRank ne fournit qu'un indice parmi tant d'autres dans l'algorithme qui permet de classer les pages du Web dans les résultats de recherche de Google. Ce système a été inventé par Larry Page, cofondateur de Google. Ce mot est une marque déposée.

Le principe de base est d'attribuer à chaque page une valeur (ou score) proportionnelle au nombre de fois que passerait par cette page un utilisateur parcourant le graphe du Web en

cliquant aléatoirement, sur un des liens apparaissant sur chaque page. Ainsi, une page a un *PageRank* d'autant plus important qu'est grande la somme des *PageRanks* des pages qui pointent vers elle (elle comprise, s'il y a des liens internes).

Plus formellement, le déplacement de l'utilisateur est une marche aléatoire sur le graphe du Web, c'est-à-dire le graphe orienté dont les sommets représentent les pages du Web et les arcs les hyperliens. En supposant que l'utilisateur choisisse chaque lien indépendamment des pages précédemment visitées (le réalisme d'une telle hypothèse pouvant être discuté), il s'agit d'un processus de Markov. Le *PageRank* est alors simplement la probabilité stationnaire d'une chaîne de Markov, c'est-à-dire un vecteur de Perron-Frobenius de la matrice d'adjacence du graphe du Web. La taille (gigantesque) de ce graphe et son évolution dynamique (modifications de pages et hyperliens, connexion ou déconnexion de serveurs Web...) rendent cependant impossible un calcul direct de ce vecteur propre : des algorithmes d'approximation sont utilisés.

De nombreuses corrections et améliorations ont été apportées à cet algorithme, certaines étant décrites dans le brevet déposé le 17 avril 2007, d'autres ne restant connues que de Google. En particulier, il est important de garantir que des modifications trop locales du graphe du Web n'entraînent pas d'augmentation disproportionnée du *PageRank* de certaines pages, ceci afin d'éviter que des utilisateurs (par exemple des sites commerciaux) ne "boostent" artificiellement leur *PageRank*. Par exemple, dans l'algorithme de base décrit ci-dessus, ajouter de nombreux liens internes sur une page Web (ce qui est très simple à faire pour un particulier) permet d'augmenter son *PageRank* (cette stratégie ne marche pas avec le *PageRank* actuel de Google).

Les internautes peuvent obtenir une approximation du classement de chaque page en consultant la zone *PageRank* de la Google Toolbar, qui indique sa valeur sur une échelle de 0 à 10 (Échelle logarithmique). Il existe aussi de nombreux outils pour l'obtenir sans afficher la toolbar. L'affichage de ce "Pagerank" a été abandonné par Google en 2009 il est remplacé par le "TrustRank".

Annexe C

Méthodes de positionnement:

1-Positionnement par GPS:

Le système GPS permet d'avoir la position tridimensionnelle (latitude, longitude et altitude) d'un utilisateur, de manière continue et instantanée, en tout endroit sur Terre, un code pseudo-aléatoire est envoyé par chaque satellite GPS. Le récepteur GPS reçoit ce signal et il l'utilise pour calculer la distance qui le sépare de ces satellites.

La connaissance du positionnement du satellite s'effectue grâce à une table que chaque récepteur GPS possède, cette table contient la position des différents satellites.

2-Positionnement par GSM :

1)- L'identification par cellule :

Cette méthode permet de localiser un mobile grâce à l'adresse de la BTS sur laquelle il est connecté, on appelle cette identification Cell ID (identification par cellule). Ses avantages sont : sa rapidité, simplicité et son coût réduit. Néanmoins, sa précision du positionnement est très faible (100 à 800 m en ville et jusqu'à 10 km en campagne).

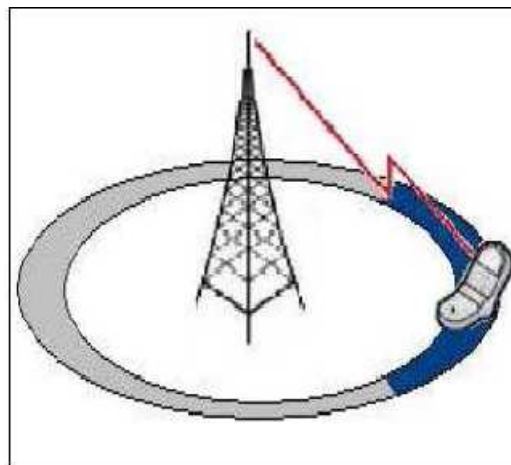


Figure Identification par cellule

2)- L'identification par triangulation:

Qui est une autre technique de positionnement par GSM, le principe de cette méthode est qu'au lieu d'avoir une station BTS, on calcule la distance qui sépare le mobile de trois stations BTS.

Ses avantages sont :

Elle permet de positionner un mobile d'une manière plus précise (elle est de 125 mètres en milieu urbain et de 4 kilomètres en milieu rural) néanmoins le temps de localisation est un peu plus long qu'avec les technologies Cell ID, environ 5 secondes.

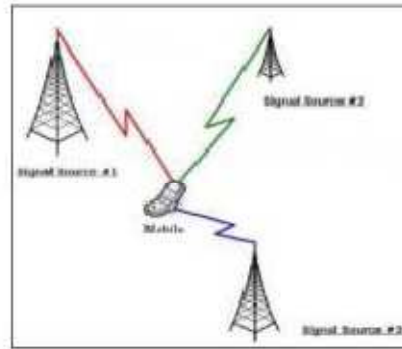


Figure: L'identification par triangulation

3)- L'identification par temps:

La troisième méthode de localisation d'un mobile se base sur le temps écoulé entre l'émission et la réception du signal entre le mobile et la station BTS. Après l'analyse du temps écoulé, il sera utilisé pour déterminer la position du mobile.

Annexe D

UML (Unified Modeling Language)

1. Introduction [37]

UML (Unified Modeling Language) est un langage standard conçu pour l'écriture de plans d'élaboration de logiciels. Il peut être utilisé pour visualiser, spécifier, construire et documenter les artefacts d'un système à fortes composantes logicielles.

UML est adapté à la modélisation de systèmes, depuis les systèmes informatiques d'entreprise jusqu'aux applications distribuées basées sur le Web, c'est un langage très expressif qui couvre toutes les perspectives nécessaires au développement et au déploiement de tels systèmes. Pour apprendre à s'en servir efficacement, il faut d'abord s'appuyer sur une représentation conceptuelle de ce langage, ce qui nécessite l'assimilation de trois éléments fondamentaux : les briques de base d'UML, les règles qui déterminent la manière de les assembler et quelques mécanismes généraux qui s'appliquent à ce langage.

2. Définition [37]

UML est un langage de formalisme. Il constitue une méthode de développement logiciel. Il est indépendant des processus, même si, idéalement, il doit être utilisé dans le cadre d'un processus fondé sur les cas d'utilisation, centré sur l'architecture, itératif et incrémental.

UML s'articule autour de neuf diagrammes différents, chacun d'eux étant dédié à la représentation des concepts particuliers d'un système logiciel. Par ailleurs, UML modélise le

système suivant deux modes de représentation : l'un concerne la structure du système pris " au repos " (diagramme de cas d'utilisation, de classe...etc.) l'autre concerne sa dynamique de fonctionnement (diagramme de séquence, d'états transitions...etc.). Les deux représentations sont nécessaires et complémentaires pour schématiser la façon dont est composé le système et comment ses composantes fonctionnent entre elles.

3. Le modèle conceptuel d'UML :

Le modèle conceptuel d'UML comprend les notions de base génériques du langage. Il définit trois sortes de briques de base :

- Les éléments, qui sont les abstractions essentielles à un modèle.
- Les relations, qui constituent des liens entre ces éléments.
- Les diagrammes, qui regroupent des éléments et des liens au sein de divers ensembles.

Dans l'étude conceptuelle de notre plate-forme, nous n'avons utilisé que 3 diagrammes, à savoir le diagramme de cas d'utilisation, le diagramme de classe et le diagramme de séquence. Nous allons décrire ces diagrammes ainsi que les éléments et les relations les constituant dans ce qui suit.

4. Les Diagrammes d'UML

Un diagramme est la représentation graphique d'un ensemble d'éléments qui constituent un système. Il se présente sous la forme d'un graphe connexe où les sommets correspondent aux éléments et les arcs aux relations. Les diagrammes servent à visualiser un système sous différents perspectives, ce sont des " vues " qui permettent d'isoler certaines parties des modèles pour les rendre plus aisément compréhensibles.

4.1. Diagramme de cas d'utilisations

Ce diagramme, décrit les fonctions du système selon le point de vue de ses utilisateurs futurs, il permet de :

- Structurer les besoins des utilisateurs et les objectif correspondants d'un système.
- Centrer l'expression des exigences du système sur ses utilisateurs : ils partent du principe que les objectifs du système sont tous motivés.
- Identifier les utilisateurs du système (acteurs) et leur interaction avec le système, classer les acteurs et structurer les objectifs du système.

Ce type de diagrammes intervient tout au long du cycle de développement, il permet de définir les limites du système.

Eléments de Base des Cas d'Utilisation

- **Acteur** : entité externe qui agit sur le système (opérateur, autre système...).

Un acteur peut consulter ou modifier l'état du système. En réponse à l'action de cet acteur, le système fournit un service qui correspond à son besoin.

➤ **Les relations :** Il existe trois types de relations entre cas d'utilisation

1- La relation de généralisation : le cas d'utilisation enfant est une spécialisation du cas d'utilisation parent.

2- La relation d'inclusion (uses): le cas d'utilisation source comprend également le comportement de son cas d'utilisation destination. Cette relation a un caractère obligatoire (à la différence de la généralisation) et permet ainsi de décomposer des comportements partageables entre plusieurs cas d'utilisation différents.

3- La relation d'extension (extend): le cas d'utilisation source ajoute son comportement au cas d'utilisation destination. L'extension peut être soumise à condition. Cette relation permet de modéliser des variantes de comportement d'un cas d'utilisation.

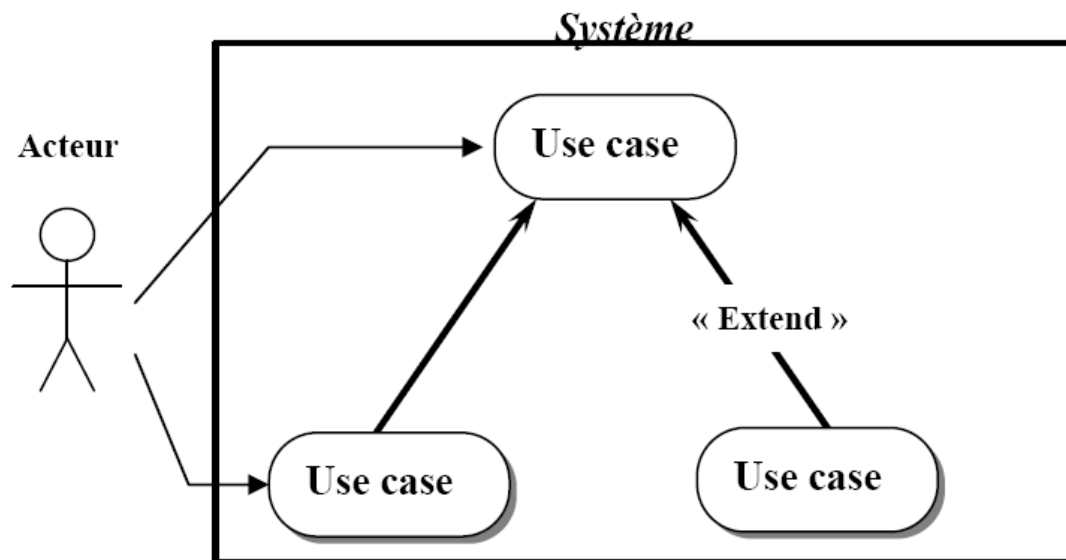


Figure 1: Représentation graphique d'un cas d'utilisation [38].

4.2. Diagramme de classe

Structure des données du système défini comme un ensemble de relations entre classes. Un diagramme de classes est une collection d'éléments de modélisation statiques (classes, paquets...), qui montre la structure d'un modèle. Ce diagramme fait abstraction des aspects dynamiques et temporels.

Pour un modèle complexe, plusieurs diagrammes de classes complémentaires doivent être construits.

Eléments de Base

- **Classe** : C'est une description d'un ensemble d'objets qui partage les mêmes attributs, opérations, méthodes, relations et contraintes. Une instance d'une classe est appelée objet.

Elle est représentée par un rectangle, qui, en règle générale, contient son nom, ses attributs et ses opérations.

- **Packages** : ils représentent les parties organisationnelles des modèles UML. Ce sont les boites dans lesquelles un modèle peut être décomposé. Un paquetage regroupe des éléments de modélisation, selon des critères purement logiques. Il permet d'encapsuler des éléments de modélisation (il possède une interface) et de structurer un système en catégories (vue logique) et sous-systèmes (vue des composants). Il sert de "brique" de base dans la construction d'une architecture. Un paquetage représente le bon niveau de granularité pour la réutilisation et est aussi un espace de nom. A la différence des composants (qui existent lors de la phase d'exécution), un paquetage est purement conceptuel (il existe seulement lors de la phase de développement). Il est représenté par un dossier étiqueté, contenant son nom et parfois son contenu.
- **Association** : Une *association* est une relation structurelle bidirectionnelle qui décrit un ensemble de liens entre différents éléments (la nature des liens est ajoutée comme décoration). Elle est représentée par une ligne qui peut être fléchée ; elle comprend parfois une étiquette et souvent d'autres décorations, comme la multiplicité et les noms de rôles.
- **Classe association** : Une association peut être représentée par une classe appelée classe associative ou classe association. Utile par exemple, lorsque l'association a des attributs ou bien qu'on souhaite lui attacher des opérations. La notation UML utilise une ligne pointillée pour attacher une classe à une association. La classe association contiennent des attributs sans participer à des relations avec d'autres classes.

Le schéma ci-dessous représente un exemple des notions que nous venons de définir.

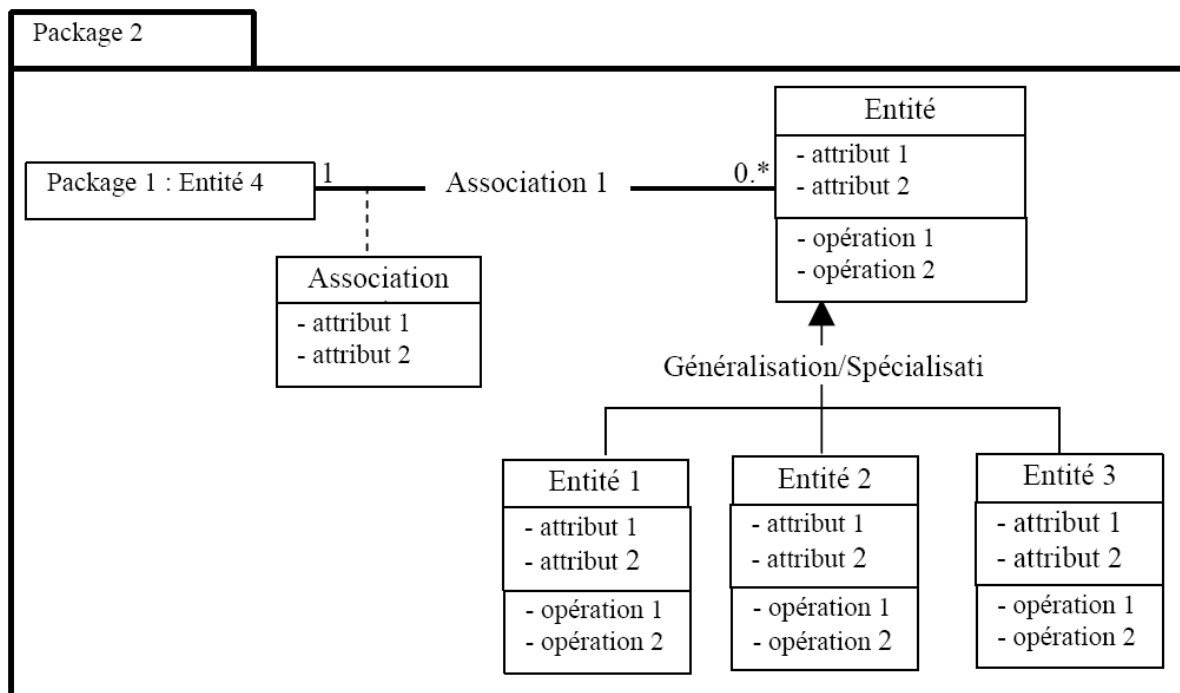


Figure 2 : Représentation d'un diagramme de classe [38].

4.4. Diagramme de séquence

Il représente des interactions temporelles entre objets dans la réalisation d'une interaction homme-système. Il montre des interactions entre objets selon un point de vue temporel. Ce type de diagrammes sert à modéliser les aspects dynamiques des systèmes temps réels et des scénarios complexes mettant en œuvre peu d'objets. L'accent est mis sur la chronologie des envois de messages. La représentation se concentre sur l'expression des interactions et non pas sur l'état ou le contexte des objets. Ce diagramme est usuellement utilisé pour illustrer les diagrammes de cas d'utilisation.

Dans un diagramme de séquence, les objets sont associés à une ligne de vie. La dimension verticale de celle-ci représente l'écoulement du temps (du haut vers le bas). Notons que la disposition des objets sur l'axe horizontal n'est pas importante dans ce type de diagrammes. Les messages sont représentés par des flèches et leur ordre est donné par leurs positions sur la ligne de vie. Le concept de message unifie toutes les formes de communication entre objets (appel de procédure, événement discret, signal entre flots d'exécution ou interruption matérielle).

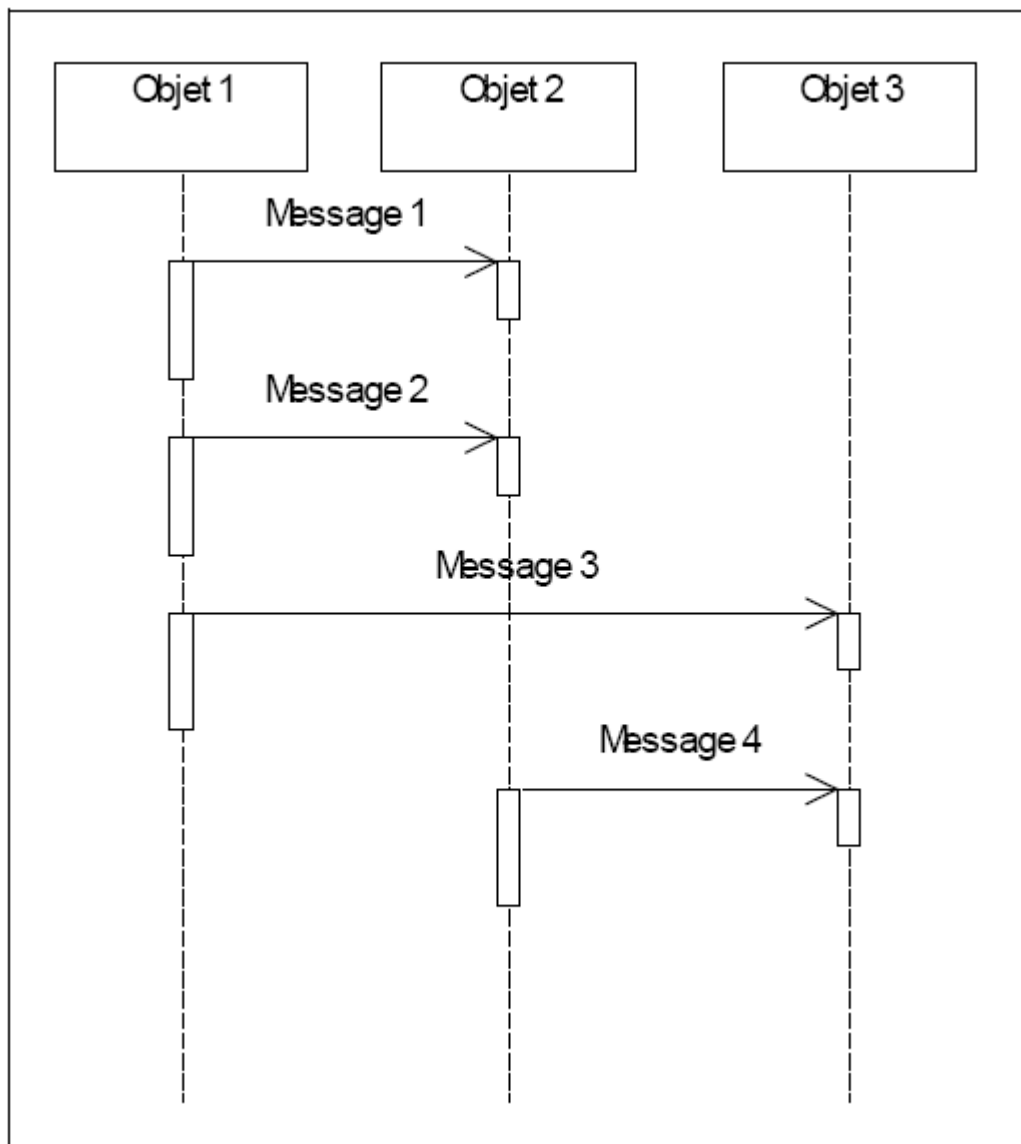


Figure 3 : Représentation d'un diagramme de séquence

Bibliographie

- [1] : « Lynda Tamine-Lechani, Nesrine Zemirli, Wahiba Bahsoun ». Approche statistique pour la définition du profil d'un utilisateur de système de recherche d'information
- [2] : « Mokrane Bouzeghoub, Dimitre Kostadinov INRIA Rocquencourt et laboratoire PRiSM, université de Versailles 45, avenue des Etats-Unis, 78035 Versailles » une approche multidimensionnelle pour la personnalisation de l'information « Mokrane Bouzeghoub, Dimitre Kostadinov INRIA Rocquencourt et laboratoire PRiSM, université de Versailles 45, avenue des Etats-Unis, 78035 Versailles »
- [3] : Bouzeghoub M., Kostadinov D., « Personnalisation de l'information : aperçu de l'état de l'art et définition d'un modèle flexible de profils », *CORIA 05, Grenoble*, p. 201-218, 2005.
- [4] : Koutrika G., Ioannidis Y., « Personalized Queries under a Generalized Preference Model », *ICDE 05, Tokyo, Japan*, p. 841-852, 2005.
- [5] : Jean Martine « Un modèle vectoriel relationnel de recherche d'information adapté aux images », 2004.
- [6] : A. Pretschner, S. Gauch, Personalization on the Web, Technical report ITTC-FY2000-TR-13591-01, Information and telecommunication technology center, Department of electrical engineering and computer science, University of Kansas, December 1999.
- [7] : Kobsa A. « Generic User Modeling Systems », *Journal On User Modeling and User-Adapted Interaction*, vol. 11, 2001, P49-63.
- [8] : McTear M. F. 1993. User modelling for adaptive computer systems : à survey of recent developments. *Artificial Intelligence Review*, 7 : P.157-184, 1993.
- [9] : http://maia.loria.fr/lexique/profil_utilisateur.html.
- [10] : Mokrane Bouzeghoub, Dimitre Kostadinov. Personnalisation de l'information : aperçu de l'état de l'art et définition d'un modèle flexible de profils. Laboratoire PRiSM ; Université de Versailles 45, avenue des Etats-Unis, 78035 Versailles {Prénom.Nom@prism.uvsq.fr}.
- [11] : Giuseppe Amato, Umberto Straccia “User Profile Modeling and Applications to Digital Libraries”, Proc. 3rd European Conf. Research and Advanced Technology for Digital Libraries, ECDL, 1999.
- [12] : Angela Carrillo-Ramos, Marlène Villanova-Oliver, Jérôme Gensel et Hervé Martin. Gestion des préférences utilisateurs pour les Systèmes d'Information ubiquitaires *Laboratoire*

LSR-IMAG, équipe SIGMA. 681, rue de la Passerelle 38402, Saint Martin d'Hères France {carrillo, villanov, gense, martin@imag.fr}.

- [13]: K. Bradley, Rafter, B. Smyth Case-Based User Profiling for Content Personalisation, , Lecture Notes in Computer Science, 2000
- [14]: B. Mobasher, H. Dai, T. Luo, M. Nakagawa, Discovery and Evaluation of Aggregate Usage Profiles for Web Personalization, Data Mining and Knowledge Discovery, 6(1):61-82, 2002.
- [15]: MySDI: A Generic Architecture to Develop SDI Personalised services, J. Ferreira, A. Silva, ICEIS (1) 2001: 262-270.
- [16]: A. Pretschner, S. Gauch, Ontology Based Personalized Search, roc 1th IEEE Intl. On Tools with Artificial Intelligence, pp. 391-398, Chicago, November 1999.
- [17]: W. B. Croft, St. Cronen-Townsend, V. Lavrenko, Relevance Feedback and Personalization: A Language Modeling Perspective, Computer Science, DELOS Workshop, 2001.
- [18]: Automatic learning of user profiles-towards the personalisation of gent services, S J Soltysiak, I B Crabtree, BT Technol J. Vol 16 No 3, July 1998, pp 110-117.
- [19]: S. Shearin, H. Lieberman, Intelligent Profiling by Example, MIT Media Lab, Cambridge, 2001.
- [20]: S. Y. Jung, J-H. Hong, T-S. Kim A Formal Model for User Preference, IEEE International Conference on Data Mining, 2002.
- [21]: A Review and Analysis of Commercial User Modeling Servers for Personalization on the World Wide Web, J. Fink, A. Kobsa, User Modeling and User-Adapted Interaction 10: pp. 209-249, 2000.
- [22]: Querying with Intrinsic Preferences, Jan Chomicki, EDBT, 2002
- [23]: Which Are My Preferred Items?, R. Torlone, P. Ciaccia, Workshop on Recommendation and Personalization in eCommerce (RPEC 2002), Malaga, Spain, 2002.
- [24]: Preference: Putting More Knowledge into Queries, M. Lacroix, P. Lavency, Proceeding of the 13th VLDB Conference, Brighton, 1987.
- [25]: Brusilovsky P. 1996. Methods and Techniques of Adaptive Hypermedia. In Journal of user modelling and user-adapted interaction. 6, 2-3, P. 87-129, 1996.
- [26]: D. KELLY, J. TEEVAN, « Implicit Feedback for Inferring User Preference: A Bibliography », SCILS, Rutgers University, New Brunswick et CSAIL, Massachusetts

Bibliographie et Webliographie

Institute of Technology Cambridge, USA, diane@scils.rutgers.edu, teevan@csail.mit.edu, 2002.

[27]: G. Salton, The SMART retrieval system: experiments in automatic document processing, prentice Hall Inc, NJ. 1971.

[28] : A. Pretschner, S. Gauch, Personalization on the Web, Technical report ITTC-FY2000-TR-13591-01, Information and telecommunication technology center, Department of electrical engineering and computer science, University of Kansas, December 1999.

[29]: [JH 1993] Jennings, A, Higuchi, H, A user model neural network for a personal news service. User modelling and user adapted interaction, 3(1), 1:25, 1993.

[30]: The Skyline Operator, S. Borzsonyi, D. Kossmann, K. Stocker, IEEE Conf. On data Engineering, 2001.

[31]: Personnalisation de l'information : une approche de gestion de profils et de reformulation de requêtes.

[32] : Les mécanismes de recherche d'informations utilisés par GOOGLE

[33] : www.abondance.fr

[34] : <http://www.lesannuaires.com/annuaire-meta-moteur.html>

[35] : <http://www.francite.com/>

[36] : www.ccf.d.crosemont.qc.ca/cours

[37]: Sohn, Griswold, Hollan, Diary « *Study of Mobile Information Needs* », Florence, Italy, Avril 2008.

[38]: P.A. Muller, " La genèse d'UML", 2004

[39]: ZAKARIA C., et STAMBOUL M.Y, Thèse de fin d'études en vue d'obtention du diplôme d'état en informatique. Option : système d'information Thème : « Conception et réalisation d'un S.I pour la gestion et le suivi de cours médiatisés »Proposé par : L'Office National de l'Enseignement et de la Formation à Distance (O.N.E.F.D ex C.N.E.G).Année 2001/2002