

MH 004-70-1

# UNIVERSITE DE SAAD DAHLED DE BLIDA



Faculté des Sciences  
Département d'Informatique

Mémoire Présenté par :

Djerrah Leila

Et

Khali Rômaïssa

En vue d'obtenir le diplôme de Master

Domaine : Informatique

Filière : Informatique

Spécialité : Ingénierie des logiciels

## Définition d'un model flexible de profil pour un OLAP personnalisé

Soutenue le : Septembre 2011, devant le jury composé de :

MA-004-70-1

Mme. Ouahmane

Président

Melle. Ali Mazighi

Examinateur

Melle. Elmes

Examinateur

Mme REZOUG.N

Promotrice

Blida, 2010/2011

# Remerciements

*Nous remercions d'abord Dieu le tout puissant de nous avoir donné le courage et la volonté d'achever ce travail et sans lequel il n'aurait jamais été accompli.*

*Nous tenons à remercier et à témoigner toute notre reconnaissance à tous le personnel du vice réctorat en l'occurrence Mr CHERGUI, Mme LEKHAL, Mr DJELOUAH, Melle DJALLATOU pour leurs patiences, compréhensions durant l'élaboration de ce mémoire.*

*Nos remerciements les plus sincères, accompagnés de toute notre gratitude vont tout d'abord à notre promotrice Mme REZOUG pour nous avoir proposé ce sujet, pour les conseils qu'elle n'a cessé de nous prodiguer et surtout pour la confiance qu'elle nous a accordé pour la réalisation de ce projet.*

*Nous tenons à remercier tout particulièrement Mr Rabah Abdelkarim et Mr Ali Racem Zakaria qui nous ont apportées un grand soutien moral par leurs amitiés et leurs conseils, trouvent dans ce message l'expression de notre gratitude et notre profonde reconnaissance.*

*Nous remercions également toute personne ayant contribué à notre éducation et notre formation.*

*Enfin, nos remerciements vont à toute personne ayant contribué, de près ou de loin, à l'aboutissement de ce travail.*

# Dédicaces

*Je dédie ce modeste travail*

*À mes parents que je ne remercierai jamais assez pour toute l'aide  
qu'ils m'ont prodiguée,*

*À ma mère pour son amour inconditionnel et sa présence à mes côtés  
dans les moments difficiles*

*À mon père pour son soutien, ses encouragements et ses précieux  
conseils durant toute ma vie.*

*À mes sœurs*

*Naila et Ilham synonyme de sagesse et gentillesse pour qui ma  
réussite est très importante.*

*Assala et Niama synonyme de générosité qui m'a appris l'amour et le  
partage dans la vie.*

*À mon petit frère Aymen (elmouchaghîb) synonyme de réconfort et  
d'écoute.*

*À mes chers grands parents, mes tantes, mes oncles, cousines et cousins  
qui je sais que ma réussite est très importante pour eux.*

*À mon binôme synonyme de patience avec qui j'ai partagé des  
moments les difficultés relatives et de joie au suivi de notre formation.*

*À tous mes enseignants pour le savoir et les connaissances qu'ils m'ont  
inculqué.*

*À mes meilleurs amis ( Assia, Aicha et Krimo) synonyme de  
protection et fidélité et qui m'ont encouragée aux moments difficiles  
que Dieu les accueille dans son vaste paradis.*

*À mes collègues de travail synonyme de compréhension et d'amitié.*

*Romaissa*

# Dédicaces

*Je dédie ce modeste travail*

*A ma chère mère, que j'aime énormément et dont je vois la parfaite et unique maman au mode qui est toujours prête à se sacrifier pour le bonheur de ces enfants.*

*A mon cher père, que je vois le père dévoué à sa famille, qui est toujours présents dans les meilleurs et les mauvais moments.*

*A mes petits frères Adel et Nabil, pour leurs complicités et leurs taquineries.*

*A mes grands parents, pour leurs amours et leurs sollicitudes à mon égard me marqueront à jamais.*

*A mes tantes et mes oncles, qui je sais que ma réussite est très importante pour eux.*

*A mon binôme, avec qui j'ai partagé la joie et les difficultés relatives au suivi de la formation, pour sa motivation et ses judicieuses propositions.*

*A mon ami d'enfance, pour sa fidélité, sa patience, sa compréhension, sa gentillesse et ses encouragements aux moments difficiles que dieu le bénisse.*

*A tous mes amis, avec qui j'ai partagé les meilleurs moments de ma vie.*

*Leila qui vous aime*

## \*ملخص\*

شهدنا في السنوات الأخيرة تضخما في نسبة المعلومات نظرا لنمو القدرة على الحوسبة والتخزين، وفي هذا السياق من الضروري تطوير أدوات المعلوماتية اللازمة لتسهيل البحث والاستخلاص السريع للمعلومات الموافقة للبحث. التخصيص هو أحد هذه الأدوات التي تلبي احتياجات كل مستخدم.

تحتوي هذه الرسالة على قسمين، يختص القسم الأول بدراسة تخزين البيانات بانتظام، تخصيص وتعريف نموذج الجانب المستخدم. يركز الجزء الثاني على تصميم وتنفيذ نظام التخصيص على أساس إعادة صياغة الأسئلة الموجهة من قبل التشكيل الجانب للمستخدم.

يهدف تخصيص المعلومات إلى توفير المعلومات المتعلقة بالمستخدم، وعادة هذا ما يتم مع اعتبار مجموعة من المعارف، وغالبا ما يشار إليه بواسطة مفاهيم الجانب والسياق واصفا المستخدمين، تفضيلاتهم والبيئة المتواجدين فيها. يسمى نموذج المستخدم الذي تمثل فيه كل هذه المعلومات بالجانب.

يتمثل العمل الأول للرسالة المطروحة في اقتراح نموذج مفتوح للجانب القادر على اكتساب كل المعلومات التي تصف المستخدم. أما المهمة الثانية فهي دراسة منهج إعادة صياغة الاستعلام استنادا على تقنية تخصيص الاستعلام لتخصيص المعلومات.

## RESUME

Ces dernières années, on assiste à une forte inflation de volume d'informations en raison de l'accroissement des capacités de calcul et de stockage, dans ce contexte de surcharge d'informations, il y a lieu de développer des outils informatiques pour faciliter la recherche et l'extraction rapide de l'information pertinente. Un de ces outils est la personnalisation qui consiste à répondre au mieux aux attentes de chaque utilisateur.

Ce mémoire contient deux parties, La première est orientée vers l'étude de l'état de l'art sur les entrepôts de données, la personnalisation et la définition d'un modèle de profil utilisateur. La seconde partie se focalise sur la conception et la réalisation d'un système de personnalisation basé sur la reformulation des requêtes guidée par le profil utilisateur.

La personnalisation de l'information a pour objectif de fournir l'information pertinente à l'utilisateur, cela se fait généralement par la prise en compte d'un ensemble de connaissances, désignées souvent par les notions de profil et de contexte, qui décrivent les utilisateurs, leurs préférences et l'environnement dans lequel ils se trouvent. L'ensemble de ces informations est représenté dans un modèle d'utilisateur appelé souvent profil.

Le premier travail de ce mémoire est de proposer un modèle ouvert de profil capable d'acquérir toutes les informations décrivant l'utilisateur. La seconde contribution est l'étude de l'approche de reformulation des requêtes basées sur la technique d'enrichissement des requêtes pour la personnalisation de l'information.

## *ABSTRACT*

In recent years, there has been a high inflation volume of information because of the increase in computing ability and storage. In this context, if the information overloaded, it should be necessary to develop computer tools to facilitate research and rapid extraction of relevant information. One such tool is the customization that is to best meet the needs of each user.

This thesis contains two parts. The first is oriented towards the study of the state of data warehousing, customization and the definition of a model of user profile. The second part focuses on the design and implementation of a customization system based on the reformulation of queries guided by the user profile.

Customizing the information intended to provide relevant information to the user; this is usually done by taking into account a set of knowledge, often referred to by the notion of profile and context, describing the users, their preferences and the environment in which they are located. All of this information is represented in a user model which is often called profile.

The first task of this work is to provide an open model profile capable of acquiring all the information describing the user. The second contribution is the study of the approach of query reformulation based on query enrichment technique customizing information.

**Table des matières**

**INTRODUCTION GENERALE**

I. Introduction.....	2
II. Problématiques.....	4
III. Objectifs.....	6
Organisation du mémoire.....	7

**PARTIE I: ETAT DE L'ART**

**CHAPITRE I: *Entrepôts de données et les systèmes OLAP***

I. Introduction.....	10
II. Définitions et généralités sur les entrepôts de données .....	10
II.1 Entrepôt de données (Data warehouse).....	10
II.2 Les caractéristiques des données stockées dans un entrepôt de données ...	11
II.3 Objectif d'un entrepôt de données .....	11
II.4 L'architecture d'un entrepôt de données .....	13
II.4.1 Le système source.....	13
II.4.2 La zone de préparation des données.....	14
II.4.3 Le serveur de présentation.....	14
II.4.4 L'entrepôt de données.....	14
II.4.5 Le magasin de données.....	14
II.4.6 Portail de restitution.....	14
II.4.7 Métadonnées.....	15
II.5 Le fonctionnement du système datawarehouse.....	16
II.5.1 L'acquisition des données.....	16
II.5.2 Le stockage des données dans le datawarehouse.....	16
II.5.3 L'alimentation des datamarts (les magasins de données).....	17
II.5.4 L'exploitation de l'information : la restitution des données.....	17
II.6 Modélisation multidimensionnelle.....	18
II.6.1 Concepts de modélisation.....	18
II.6.1.1 Concept de fait.....	18
II.6.1.2 Concept de dimension.....	19
II.6.2 Schémas de modélisation multidimensionnelle.....	20

<b>II.6.2.1</b> Modèles en étoile.....	20
<b>II.6.2.2</b> Modèles en flocon en neige.....	21
<b>II.6.2.3</b> Modèles en constellation.....	22
<b>III</b> Les outils d'analyse de données .....	22
<b>III.1</b> OLAP (On Line Analytical Processing).....	23
<b>III. 1. 1.</b> Le concept OLAP .....	23
<b>III. 1. 2.</b> Les 12 règles OLAP .....	23
<b>III. 1. 3.</b> Les différents outils OLAP.....	26
<b>III.1. 3. 1</b> Les outils MOLAP (Multidimensional OLAP).....	26
<b>III. 1. 3.2</b> Les outils ROLAP (Relational OLAP).....	27
<b>III. 1. 3.3</b> Les outils HOLAP (Hybride OLAP).....	28
<b>III.1.4.</b> Opérations sur les données d'OLAP.....	28
<b>III.1.4.1</b> Opérations OLAP liées à la structure.....	28
<b>III.1.4.2</b> Les opérations OLAP liées à la granularité.....	32
<b>IV.</b> Les langages d'interrogation des entrepôts de données.....	33
<b>IV.1</b> Les langages de données .....	33
<b>IV.1.1</b> Types de langages de données .....	33
<b>IV.1.1.1</b> Le langage de définition de données (LDD) .....	33
<b>IV.1.1.2</b> Le langage de contrôle des données (LCD) .....	34
<b>IV.1.1.3</b> Langage de manipulation des données (LMD) .....	34
<b>V. CONCLUSION</b> .....	35
 <b>CHAPITRE II: Accès personnalisé à l'information</b>	
<b>I. Introduction</b> .....	37
<b>II.</b> La personnalisation de l'information.....	37
<b>II.1</b> Définition.....	37
<b>II.2</b> Les Domaines d'utilisation de la personnalisation de l'information.....	38
<b>II.2.1</b> Dans le domaine de l'interaction homme-machine (IHM) .....	38
<b>II.2.2</b> Dans le domaine de la recherche d'information (RI) .....	39
<b>II.2.3</b> Dans le domaine des bases de données (BD) .....	39
<b>II.2.4</b> La personnalisation dans les entrepôts de données .....	40
<b>III.</b> Architecture d'un système personnalisé de base de données.....	41
<b>IV.</b> Profil utilisateur.....	42

IV.1 Définition.....	42
IV.2 Le contenu du profil.....	42
IV.3 Exemples de profil .....	43
V. La notion de préférence .....	46
V.1. L'expression des préférences .....	46
V.1.1 L'approche quantitative .....	46
V.1.2 L'approche qualitative .....	47
VI. Approches de représentation du profil utilisateur.....	47
VI.1 Représentation ensembliste.....	48
VI.2 Représentation sémantique .....	48
VI.3 Représentation conceptuelle.....	48
VI.4 Représentation multidimensionnelle.....	49
VII. Approches de construction du profil utilisateur.....	49
VII.1 Acquisition des données utilisateurs.....	49
VII.1.1 L'acquisition explicite.....	50
VII.1.2 L'acquisition implicite.....	50
VII.2 Construction du profil.....	51
VII.2.1 La modélisation de profil .....	51
VII.2.2 Instanciation du profil .....	51
VIII. Travaux de personnalisation dans les entrepôts de données .....	51
VIII.1 Concepts et définition .....	52
VIII.1.1 Référence d'une cellule .....	52
VIII.1.2 Structure.....	52
VIII.1.3 Grille d'analyse .....	52
VIII.1.4 Visualisation .....	52
VIII.1.5 La contrainte de visualisation.....	52
VIII.2 Travaux de BELLATRECHE et AL 2005.....	53
VIII.3 Travaux de BELLATRECHE et AL 2006 .....	53
VIII.4 Travaux de MOULOUDI 2007.....	53
VIII.5 Travaux de RAVAT et AL, 2008.....	54
VIII.6 Travaux de MEZIANI, 2009.....	54
IX. CONCLUSION.....	55

**PARTIE II: CONCEPTION ET REALISATION**

**CHAPITRE I: CONCEPTION**

<b>I. Introduction</b> .....	58
<b>II</b> Méthodologie de conception.....	58
<b>III</b> Modélisation du profil utilisateur.....	58
<b>III.1</b> Choix d'une représentation du profil.....	59
<b>III.2</b> Les différentes dimensions, sous dimensions et attributs du profil.....	61
<b>III.3</b> Le méta - description du modèle de profil.....	62
<b>III.3.1</b> Identification des classes.....	62
<b>III.3.2</b> Liste des associations .....	63
<b>IV</b> Architecture logicielle de l'approche de la personnalisation .....	65
<b>IV.1</b> Description des différents sous-systèmes.....	66
<b>IV.1.1</b> Sous-système de la gestion des profils.....	66
<b>IV.1.2</b> Sous système d'Appariement des profils.....	68
<b>IV.1.3</b> Sous système de la mise en œuvre de la personnalisation.....	73
<b>IV.1.3.1</b> L'enrichissement des requêtes.....	73
<b>IV.1.3.1.1</b> Sélection des prédicats pertinents.....	74
<b>IV.1.3.1.2</b> Intégration des préférences.....	74
<b>V.</b> Présentation de l'entrepôt a utilisé .....	78
<b>VI. CONCLUSION</b> .....	79
 <b>CHAPITRE II: REALISATION</b>	
<b>I. Introduction</b> .....	81
<b>II. Architecture matérielle</b> .....	81
<b>II.1</b> Schéma « un tiers ».....	81
<b>II.2</b> Schéma « 2 tiers ».....	81
<b>III. Environnement de développement</b> .....	82
<b>III.1</b> Description du système de gestion de données.....	82
<b>III.2</b> Choix du langage de programmation.....	83
<b>IV. Interface de test</b> .....	85

<b>IV.1</b> Interface d'authentification.....	85
<b>IV.2</b> Interface de création d'un nouveau compte .....	86
<b>IV.3</b> Interface d'accueil.....	87
<b>IV.4</b> Interface de gestion des préférences .....	88
<b>IV.5</b> Interface de d'appariement.....	89
<b>IV.6</b> Interface d'affichage de résultats.....	90
<b>V. Conclusion</b> .....	91
<b>CONCLUSION &amp; PERSPECTIVES</b> .....	93
<b>BIBLIOGRAPHIE</b> .....	96

*Liste des figures*

*Chapitre : Entrepôts de données et OLAP*

<i>Numéro</i>	<i>Désignation</i>	<i>page</i>
<b>Figure n° II.1</b>	Architecture d'un Data Warehouse	15
<b>Figure n° II.2</b>	Le processus d'acquisition des données	16
<b>Figure n° II.3</b>	L'alimentation des datamarts	17
<b>Figure n° II.4</b>	La restitution des données	17
<b>Figure n° II.5</b>	Exemple de table de faits	18
<b>Figure n° II.6</b>	Exemple de tables de dimensions	19
<b>Figure n° II.7</b>	Schéma en étoile	20
<b>Figure n° II.8</b>	Schéma en Flocon	21
<b>Figure n° II.9</b>	Schéma en constellation de faits	22
<b>Figure n° III.1</b>	Processus d'extraction des connaissances	22
<b>Figure n° III.2</b>	Architecture d'un produit MOLAP	26
<b>Figure n° III.3</b>	Architecture d'un produit ROLAP	27
<b>Figure n° III.4</b>	Architecture d'un produit HOLAP	28
<b>Figure n° III.5</b>	Opération Dice	29
<b>Figure n° III.6</b>	Opération Rotate	29
<b>Figure n° III.7</b>	Opération Switch	30
<b>Figure n° III.8</b>	Opération Slice	30
<b>Figure n° III.9</b>	Opération Split	31
<b>Figure n° III.10</b>	Opération Nest	31
<b>Figure n° III.11</b>	Opération Push	32
<b>Figure n° III.12</b>	Application du forage vers le bas et vers le haut sur le cube "Ventes"	33

*Chapitre : Accès personnalisé à l'information*

<i>Numéro</i>	<i>Désignation</i>	<i>page</i>
<b>Figure n° III.1</b>	Architecture d'un système de base de données personnalisé	41
<b>Figure n° IV.1</b>	Représentation graphique du profil utilisateur	45

**Chapitre : Conception**

<i>Numéro</i>	<i>Désignation</i>	<i>page</i>
Figure n° III.1	représentation multidimensionnelle et hiérarchique d'un profil	60
Figure n° III.2	Diagramme de classe pour la base de profil	64
Figure n° IV.1	schéma général du système de la personnalisation.	65
Figure n° IV.2	Acquisition explicite du profil utilisateur	66
Figure n° IV.3	Sous-système de la gestion du profil utilisateur	67
Figure n° IV.4	Exemple de profils utilisateurs	69
Figure n° IV.5	Algorithme de Matching	73
Figure n° IV.6	Algorithme d'enrichissement des requêtes.	75
Figure n° IV.7	Sous système de traitement de la requête.	77
Figure n° IV.8	Entrepôt de données à utiliser	78

**Chapitre : Réalisation**

<i>Numéro</i>	<i>Désignation</i>	<i>page</i>
Figure n° II.1	Schéma « 2 tiers » (Client/Serveur)	82
Figure n° III.1	le SGBD SQL Server 2005	82
Figure n° III.2	la plateforme JAVA 2	84
Figure n° III.3	Environnement de développement eclipse 3.2	84
Figure n° IV.1	Interface d'authentification	85
Figure n° IV.2	Introduction de données personnelles	86
Figure n° IV.3	Interface de gestion des préférences	87
Figure n° IV.4	Interface de gestion des préférences	88
Figure n° IV.5	Interface de d'appariement	89
Figure n° IV.6	Interface d'affichage de résultats	90
Figure n° IV.7	Interface d'affichage de résultats d'un profil similaire	91

*Liste des tableaux*

<i>Numéro</i>	<i>Désignation</i>	<i>Page</i>
<b>Tableau n° II.1</b>	comparaison entre les caractéristiques des processus OLTP et OLAP	13
<b>Tableau n° IV.1</b>	Exemple d'un profil utilisateur	44
<b>Tableau n° II.1</b>	les différentes exigences pour la description d'un profil.	59

*Liste des abréviations*

<i>Abréviation</i>	<i>Désignation</i>
<b>BDD</b>	<i>Base de données</i>
<b>HOLAP</b>	<i>Hybrid On-Line Analysis Processing.</i>
<b>ROLAP</b>	<i>Relational On-Line Analysis Processing.</i>
<b>MOLAP</b>	<i>Multidimensional On-Line Analysis Processing.</i>
<b>MDX</b>	<i>Multidimensional Data Expression</i>
<b>OLAP</b>	<i>On-Line Analysis Processing.</i>
<b>OLTP</b>	<i>On-Line Transactional Processing.</i>
<b>SGBD</b>	<i>Système de Gestion de Bases de Données.</i>
<b>SQL</b>	<i>Structured Query Language</i>
<b>SID</b>	<i>Système d'Information Decisional.</i>

INTRODUCTION

GÉNÉRALE

## **I. Introduction**

Les systèmes de recherche d'information sont des outils qui ont permis, jusqu'à aujourd'hui d'améliorer la qualité des services d'accès à l'information. Au fur et à mesure que le volume de données s'accroît et les informations se diversifient, ces systèmes (moteur web, SGBD....) délivrent des résultats massifs, générant ainsi une surcharge informationnelle en réponse aux requêtes des utilisateurs dans laquelle il est difficile de distinguer l'information pertinente d'une information secondaire. Donc le problème n'est pas dans la disponibilité de l'information mais dans sa pertinence relative aux besoins précis de l'utilisateur, c'est pourquoi les travaux s'orientent actuellement vers la révision de cycle de vie d'une requête dans la perspective d'intégrer l'utilisateur comme composante du modèle globale de recherche et ce dans le but de lui délivrer une information pertinente adaptée à ses besoins et ses préférences précis.

Ces travaux s'inscrivent dans le cadre de la personnalisation de l'information qui est vue comme l'une des solutions pouvant maintenir le web comme une ressource d'information viable.

La personnalisation de l'information constitue un enjeu majeur pour l'industrie informatique, son but est de faciliter l'expression du besoin de l'utilisateur et de lui permettre d'obtenir des informations pertinentes lors de ses interactions avec un système d'information.

La pertinence de l'information délivrée et son adaptation aux préférences des utilisateurs sont des facteurs clés du succès ou du rejet d'un tel système. Pour rendre la personnalisation effective, on a besoin de collecter et sauvegarder l'ensemble des critères et des préférences personnalisables spécifiques à chaque utilisateur. Ces données sont souvent regroupées sous forme de profil. Le contenu de ce dernier peut varier selon les approches et les applications. Le profil de l'utilisateur peut être construit explicitement ou implicitement.

Dans l'approche explicite, l'utilisateur fournit lui-même les informations le concernant (données personnelles, préférences, etc.). Dans l'approche implicite, les informations du profil sont acquises par des techniques d'apprentissage ou de datamining exploitant les historiques des actions et des choix passés des utilisateurs.

Du point de vue recherche, la personnalisation de l'information a été abordée principalement par la communauté Recherche d'Information (RI), la communauté Bases de Données (BD) et la communauté interface homme machine (IHM). Bien que les deux premiers domaines sont très proches dans leur objectif final (accès efficace à l'information) et qui ont le même but (personnalisation des requêtes), les deux domaines de recherche se distinguent généralement par la nature de l'information traitée (documents textuels pour la RI et tables structurées pour les BDs) et le mode d'accès à cette information (accès par mots-clés plus ou moins complexes et pas à pas pour la RI face à un accès par expressions logiques et de façon globale pour les BDs).

Notre contribution se situe dans le cadre de la personnalisation de l'information dans le domaine des Bases de Données multidimensionnelles autrement dit entrepôt de données.

## II. Problématiques

Les systèmes d'information actuels donnent l'accès à des sources de données multiples, distribuées 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 requête, 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.
- La taille des réponses est souvent volumineuse et génère une surcharge informationnelle qui dérouté ou décourage l'utilisateur dans son exploration ou sa navigation.
- La qualité de l'information délivrée n'est pas prise en compte; ce qui rend difficile la prise de décision à base de cette information.
- Il est difficile de distinguer l'information pertinente d'une information secondaire.
- La pertinence des résultats souvent se réduit dans la mesure où ils ne sont pas adaptés ni aux contextes ni aux préférences utilisateurs.
- L'organisation, la réutilisation et le partages des analyses ne sont pas prises en compte.

Pour cela, faire appel aux outils de personnalisation de l'information peut apporter des réponses satisfaisantes à ces problèmes.

Personnaliser une information ou une application à un utilisateur particulier nécessite donc de disposer d'informations sur ce dernier permettant d'aider le système à faire des choix. Ces données persistantes caractérisant un utilisateur ou un groupe d'utilisateurs sont souvent organisées et structurées sous forme d'un modèle, appelé souvent, le modèle utilisateur (ou le profil utilisateur).

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

- Modélisation du profil utilisateur.
- Utilisation du profil dans la reformulation des requêtes.

### ***II.1. Modélisation du profil***

La problématique consiste à identifier les informations qui permettent de caractériser le mieux un utilisateur pour un type d'application particulier.

La problématique de recherche dans cette perspective consiste à répondre aux questions suivantes :

- Quelles dimensions sont pertinentes et pour quel type d'application ?
- Quels attributs de dimension caractérisent le mieux les besoins d'un utilisateur ?

### ***II.2. Utilisation du profil***

Les questions a posées dans cette partie sont :

- Quelles parties de profil interviennent dans l'enrichissement des requêtes utilisateurs et comment se fait cet enrichissement ?
- Quelles parties de profil influent sur l'exécution et l'optimisation des requêtes ?

### **III. Objectifs**

Dans le cadre d'un système dédié exclusivement à l'accès à l'information, l'objectif de la personnalisation est d'intégrer l'utilisateur dans tout le processus de recherche afin de lui délivrer une information pertinente en fonction de ses caractéristiques spécifiques. Ainsi, toute information sur l'utilisateur, comme ses préférences, ses centres d'intérêts, ses besoins en information et son environnement de recherche sont de ce fait supposés pertinents et exploitables par le système de personnalisation. L'ensemble de ces informations va correspondre à ce que l'on nomme le **contexte de l'utilisateur** ou dans un cadre plus spécifique **profil utilisateur**.

L'idée principale de l'accès personnalisée aux données est de fournir à l'utilisateur des informations pertinentes adaptées à ses besoins et ses centres d'intérêts. En d'autres termes, deux utilisateurs différents qui expriment la même requête, auront des résultats différents. Donc il est important de trouver une nouvelle solution plus adaptée qui aidera l'utilisateur dans son accès à l'information.

L'objectif de notre travail est de concevoir et réaliser un système de personnalisation d'accès aux entrepôts de données centré utilisateur et de :

- Modéliser et gérer les préférences utilisateur (profil) : création d'une structure du profil utilisateur.
- Interroger l'entrepôt de données d'une manière personnalisée et cela par la prise en compte des préférences d'un utilisateur dans la reformulation de ses requêtes ensuite les exécuter sur l'entrepôt de données.
- Offrir à l'utilisateur un espace d'analyse personnalisé (propre à lui) qui va lui faciliter ses analyses.
- Fournir à l'utilisateur une information pertinente en fonction de ses caractéristiques spécifiques en écartant les données inutiles.

Pour cela nous nous proposons de classifier les différentes connaissances constituant un profil et d'établir un cadre de référence pour définir et manipuler des profils. Ce cadre de référence se présente sous forme d'un modèle générique décrivant la typologie des informations nécessaire à la définition d'un profil dans un système d'accès personnalisé aux entrepôts de données.

Ce mémoire est composé des parties suivantes dont le contenu est exposé brièvement si dessous :

*La Partie I : Etat de l'art, est constituée de deux chapitres qui sont :*

**Le Chapitre I : Les entrepôts de données et OLAP.**

Ce chapitre présente différents aspects relatifs à la modélisation, à la conception des entrepôts de données et à leur fonctionnement. Nous parlerons aussi des systèmes OLAP.

**Le Chapitre II : Accès personnalisé à l'information.**

Ce deuxième chapitre est dédié à l'accès personnalisé à l'information. Nous parlerons de la personnalisation à travers différents domaines qui sont la recherche d'informations, l'interaction homme-machine et les bases de données. Nous évoquons également les notions de profil utilisateur et des préférences utilisateur. Nous aborderons le domaine de la personnalisation dans les entrepôts de données et nous présenterons enfin, les travaux émergents dans ce contexte.

*La Partie II : Conception et Réalisation*

**Le Chapitre I : Conception**

L'objectif principal de ce chapitre est de définir l'architecture de notre prototype. Pour cela, nous identifierons et nous décrirons les différents sous systèmes de notre prototype.

**Le Chapitre II : Réalisation**

Cette section décrit l'environnement et les outils de réalisation de notre prototype, et cela en présentant les choix effectués pour sa réalisation. Enfin nous terminerons par une conclusion générale sur le travail réalisé et les différentes perspectives proposées.

# *PARTIE*

## **I**

### **. ETAT DE L'ART**

- ✓ Entrepôt de données et OLAP
- ✓ Accès personnalisé à l'information

# Entrepôts de données et OLAP

## **I. Introduction**

Dans ce premier chapitre, nous présenterons les concepts généraux qui caractérisent les entrepôts de données, leurs objectifs, leurs modélisations...etc., ensuite nous présenterons les systèmes OLAP (On Line Analytical Process), qui permettent l'interrogation efficace des entrepôts de données selon une vue multidimensionnelle de celle-ci, et nous terminerons par la présentation des langages d'interrogation des entrepôts de données en évoquons les différents types de langages de données.

## **II. Définitions et généralités sur les entrepôts de données**

### **II.1 Entrepôt de données (Data warehouse) [BOUZIDI et SADKI, 10]**

Le terme « entrepôt de données » est certainement l'un des plus répandus dans le monde de l'informatique en général et dans l'informatique décisionnelle en particulier, du moins, en fin de la décennie 90 et début de 21ème siècle.

Un entrepôt de données « Data Warehouse » est défini de plusieurs façons comme suit :

- **Un ED est une structure informatique dans laquelle est centralisé un volume important de données consolidées à partir des différentes sources de renseignements d'une entreprise (notamment les BDs internes) et qui est conçue de manière que les personnes intéressées aient accès rapidement à l'information stratégique dont elles ont besoin.**
- « Fait une référence à une collection des technologies d'aide à la décision permettant à des (managers, dirigeants et analystes) de prendre des décisions pertinentes et rapides».
- « Un ED est une collection de données **thématiques, intégrées, non volatiles et historiées, organisées pour la prise de décision.** »

## **II.2. Caractéristiques des données stockées dans un entrepôt de données**

### ➤ **Orientées sujets**

Les données de l'entrepôt sont structurées par thèmes (sujets). L'intérêt de cette organisation est de disposer de l'ensemble des informations utiles sur un sujet le plus souvent transversal aux structures fonctionnelles et organisationnelles de l'entreprise.

### ➤ **Intégrées**

Les données alimentant l'entrepôt de données proviennent de multiples sources hétérogènes. Ces données doivent être converties, reformatées et nettoyées, de façon à avoir une seule vision globale dans l'entrepôt de données.

### ➤ **Historiées**

Dans un entrepôt de données, la donnée ne doit jamais être mise à jour en mode « Annule et remplace ». Une empreinte de temps doit être associée à la donnée (dater) afin d'être capable d'identifier une valeur particulière dans le temps.

### ➤ **Non volatiles**

Une donnée entrée dans l'entrepôt n'a pas vocation à être supprimée.

## **II.3. Objectif d'un entrepôt de données [BOUZIDI et SADKI, 10] [BOUKHALFA, 02]**

L'entrepôt de données joue un rôle stratégique dans la vie d'une entreprise. Il stocke des données pertinentes aux besoins de prise de décision en provenance des systèmes opérationnels de l'entreprise et d'autres sources externes. A la différence d'une base de données classique supportant des requêtes transactionnelles de type OLTP (On-line Transaction Processing), un entrepôt de données est conçu pour supporter des requêtes de type OLAP (On-line Analytical Processing).

Les sources de données d'une entreprise sont des bases de données de production hétérogènes gérées par les applications OLTP qui sont adaptées pour supporter un processus d'analyse pour l'aide à la prise de décision. Par contre les entrepôts de données nous permettent de valoriser et d'analyser ces données unifiées. Ils servent aussi à :

- Intégrer des sources de données, donc résoudre le problème d'hétérogénéité des différentes sources.
- Supporter un processus d'analyse en ligne, exploiter l'entrepôt de données.
- Construire de l'information utile pour l'aide à la prise de décision.

Les bases de données sont utilisées dans les entreprises pour gérer les importants volumes d'informations contenus dans leurs systèmes opérationnels. Ces données sont gérées selon des processus transactionnels en ligne (OLTP : "*On-line Transactional Processing*"). L'exploitation de l'information contenue dans ces systèmes opérationnels est devenue une préoccupation essentielle pour les dirigeants des entreprises qui désirent améliorer leur prise de décision par une meilleure connaissance de leurs propres activités, de celle de la concurrence, des employés, des clients et des fournisseurs. Les entreprises sont donc à la recherche des systèmes supportant efficacement les applications d'aide à la décision. Ces applications décisionnelles utilisent des processus d'analyse en ligne des données (OLAP : "*On-line Analytical Processing*") Ces processus répondent aux besoins spécifiques des analyses d'information.

Le Tableau suivant compare les caractéristiques des processus OLTP et OLAP

	Bases de production	Entrepôts de données
<b>Données</b>	Exhaustives, détaillées	Résumées, agrégées
	Courantes	Statiques
	Mises à jour	Recalculées
	Dynamiques	Statiques
	Orientées applications	Orientées sujets d'analyse
	De l'ordre des giga-octets	De l'ordre des téraoctets
<b>Utilisateurs</b>	Agents opérationnels	Décideurs
	Nombreux	Peu
	Concurrents	Non concurrents
	Mises à jour et interrogations	Interrogations
	Requêtes Prédéfinies	Requêtes imprévisibles
	Réponses immédiates	Réponses moins rapides
	Accès à peu d'informations	Accès à de nombreuses informations

**Tableau n° II.1 : Comparaison entre les caractéristiques des processus OLTP et OLAP**

#### **II.4. L'architecture d'un entrepôt de données**

On distingue sept parties constituant l'architecture de l'entrepôt de données :

##### **II.4.1. Le système source**

« *Système opérationnel d'enregistrement, dont la fonction consiste à capturer les transactions liées à l'activité de l'entreprise* » [Kim, 97]. Il peut être représenté par les systèmes de productions ou bien de données externes.

#### II.4.2. La zone de préparation des données

« Ensemble de processus qui nettoient, transforment, combinent, archivent, suppriment les doublons, c'est-à-dire prépare les données sources en vue de leur intégration puis de leur exploitation au sein de l'entrepôt de données » [Kim, 97].

Cette zone comprend tout ce qui se trouve entre les systèmes sources et le serveur de présentation. La frontière qui détermine la zone de préparation des données est que la zone de préparation ne doit en aucun cas être accessible à l'utilisateur final par requête ou par un ou quelques autres services de présentation.

#### II.4.3. Le serveur de présentation

« Machine cible sur laquelle l'entrepôt de données est stocké et organisé pour répondre en accès direct aux requêtes émises par des utilisateurs » [Kim, 97].

Une fois les données chargées dans le Data warehouse, les utilisateurs peuvent y accéder pour exécuter leurs requêtes, analyser et visualiser l'information... etc.

#### II.4.4. L'entrepôt de données

« L'entrepôt de données correspond à la source de données interrogeable de l'entreprise » [Kim, 97]. Il est alimenté par la zone de préparation des données, Il contient toutes les données de l'entreprise.

#### II.4.5. Le magasin de données

« Le magasin de données ou Data Mart est défini comme un sous-ensemble logique d'un entrepôt de données » [Kim, 97]. On considère souvent le magasin de données comme la réduction de l'entrepôt de données à un seul processus ou à un groupe de processus ciblant un groupe métier spécifique.

#### II.4.6. Portail de restitution

C'est la part publique de l'entrepôt de données [Kim, 97]. Il représente ce que voient les utilisateurs, les outils avec lesquels ils travaillent. Les services offerts par le portail de restitution sont les services d'accès aux données, les applications de modélisation et de Data Mining. Les services d'accès aux données comprennent : la navigation dans l'entrepôt, dans les métas données, la surveillance de l'activité et la gestion des requêtes.

Les applications de modélisation offrent différents types d'analyse basées sur des modèles tels que modèles financiers, systèmes d'évaluation des clients, optimisation des processus et prévision, ainsi que les activités centrales du Data Mining telles que la catégorisation, la classification, l'estimation et la prédiction et finalement le regroupement.

#### II.4.7. Métadonnées

Ce sont toutes les informations de l'environnement de l'entrepôt de données qui ne constituent pas les données proprement dites. Ce sont les « données sur les données ».

La figure suivante représente l'architecture du Data Warehouse :

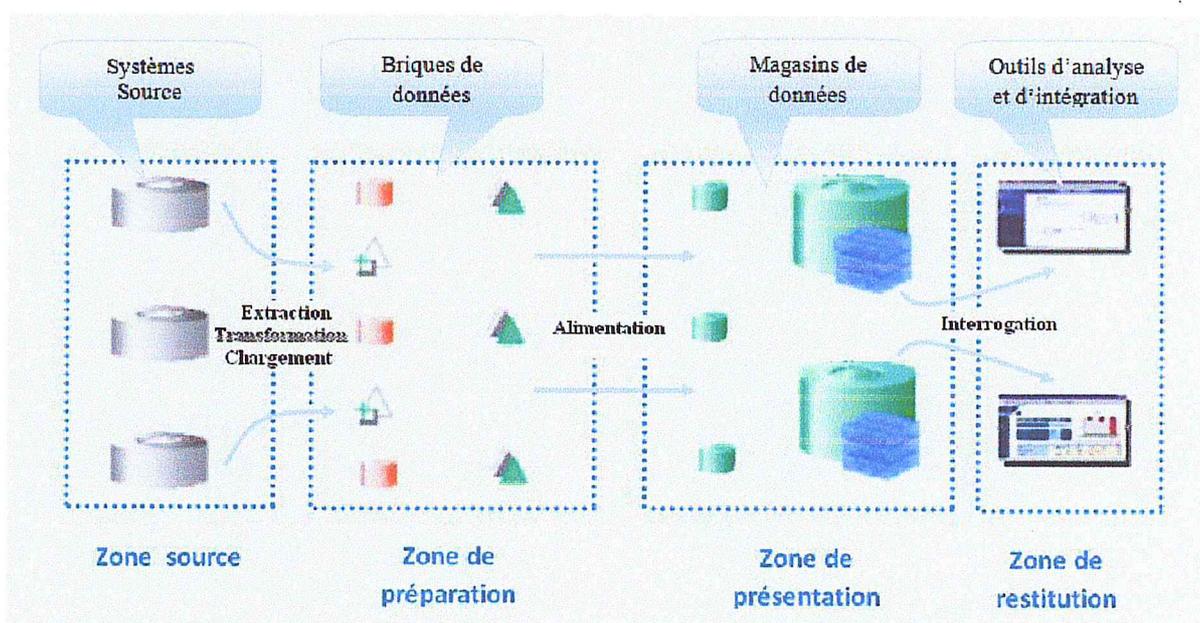


Figure n°II.1: Architecture d'un Data Warehous

## II.5. Le fonctionnement du système datawarehouse [HACIANE, 07]

Un système datawarehouse fonctionne selon les étapes suivantes :

### II.5.1 L'acquisition des données

Cette étape constitue la frontière entre le système décisionnel et les systèmes opérationnels. Elle consiste à extraire les données utiles des systèmes opérationnels qui dans de nombreux cas sont hétérogènes, diffusés et complexes.

La problématique de l'alimentation d'un datawarehouse se résout par la mise en place d'un processus en cinq phases : *découvrir, extraire, transformer, transporter et charger*.

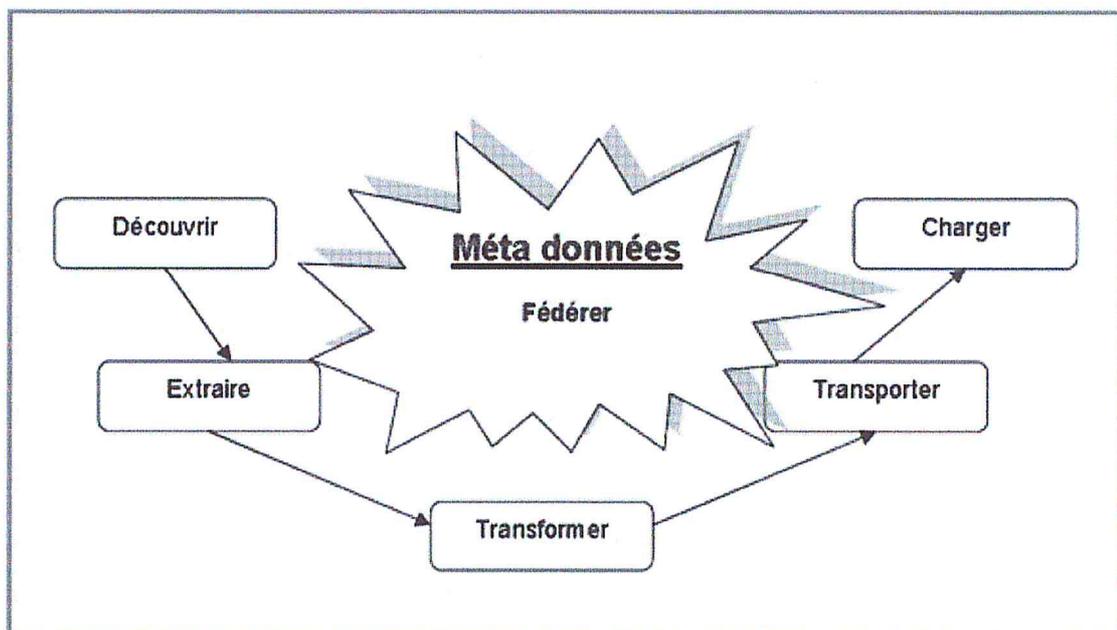


Figure n°II.2: Le processus d'acquisition des données

### II.5.2. Le stockage des données dans le datawarehouse (l'entrepôt de données)

C'est le point central de stockage de toutes les données de l'entreprise concernées par le système décisionnel. Les données du datawarehouse sont, rappelons-le, orientées sujet, intégrées, non volatile et historiées pour le support du processus d'aide à la décision.

### II.5.3 L'alimentation des datamarts (les magasins de données)

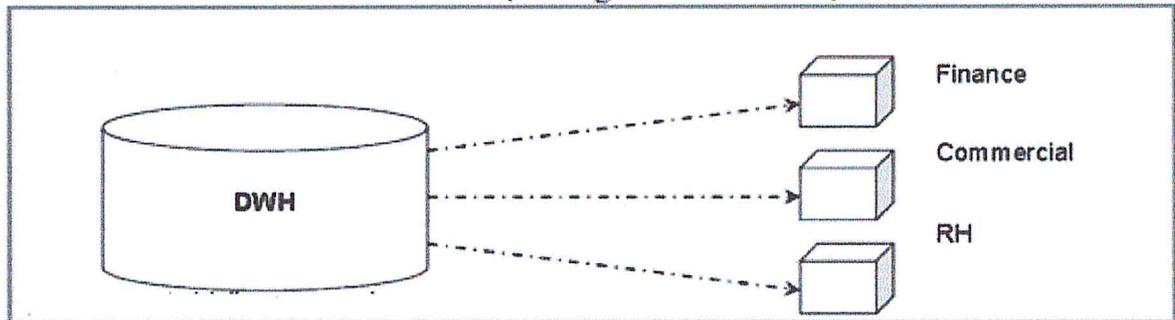


Figure n°II.3 : L'alimentation des datamarts

- Le datawarehouse est le sas central de contrôle, garant de la qualité et de l'intégrité de l'information. Son principal objectif est d'optimiser l'approvisionnement des magasins de données.
- Chaque magasin de données est conçu pour répondre à un enjeu métier. Les données y sont structurées en fonction de la problématique traitée. Le stockage de données y est généralement multidimensionnel.

### II.5.4. L'exploitation de l'information : la restitution des données

C'est le bout de la chaîne. Il s'agit du point d'utilisation du système par les utilisateurs. La satisfaction de ceux-ci dépend de la capacité des outils de restitution à répondre à leur besoin en information et aide à la décision. Les types de restitution possible sont :

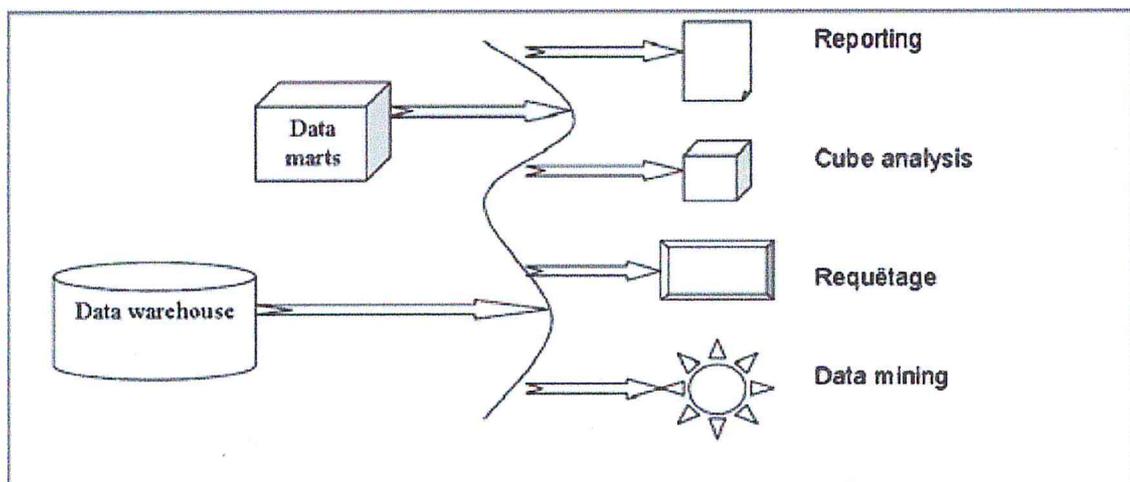


Figure n° II.4: La restitution des données

## II.6 Modélisation multidimensionnelle

La **modélisation multidimensionnelle** consiste à considérer un sujet analysé comme un point dans un espace à plusieurs dimensions. Les données sont organisées de manière à mettre en évidence le sujet analysé et les différentes perspectives de l'analyse.

### II.6.1 Concepts de modélisation

Conceptuellement, cette modélisation multidimensionnelle a donné naissance aux concepts de fait et de dimension.

#### II.6.1.1 Concept de fait

Une table de fait est une table qui contient les données observables (les faits) que l'on possède sur un sujet et que l'on veut étudier, selon divers axes d'analyse (les dimensions).

Une table de faits contient les valeurs numériques de ce qu'on désire mesurer. Il peut s'agir du chiffre d'affaire (montant en argent des ventes), du nombre d'unités vendues d'un produit, etc.

Elle contient aussi les clés associées aux dimensions. Il s'agit de clés étrangères vers les dimensions.

Une table de fait contient plus d'enregistrements qu'une table de dimension.

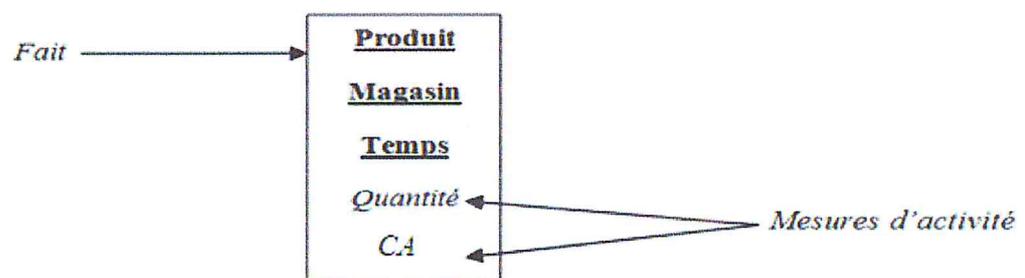


Figure n° II.5: Exemple de table de faits

### II.6.1.2 Concept de dimension

Une dimension est une table qui représente un axe d'analyse selon lequel on veut étudier des données observables (les faits) qui, soumises à une analyse multidimensionnelle, donnent aux utilisateurs des renseignements nécessaires à la prise de décision.

On appelle donc « dimension » un axe d'analyse. Il peut s'agir des clients ou des produits d'une entreprise, d'une période de temps, des activités menées au sein d'une société, etc.

Une table de dimension contient le détail sur les faits, elle contient les informations descriptives des valeurs numériques de la table de faits

Une table de dimension contient en général beaucoup moins d'enregistrements qu'une table de faits

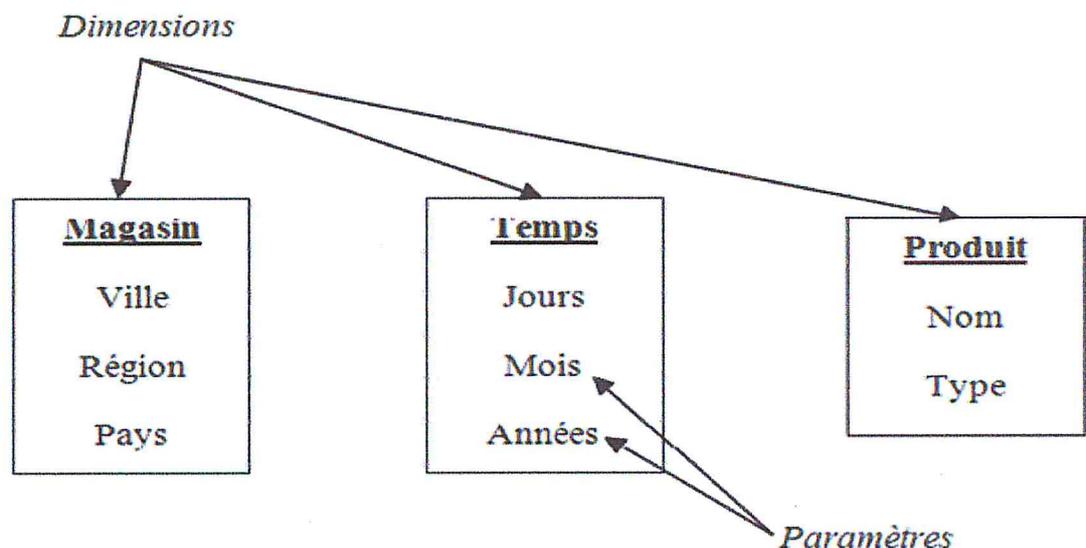


Figure n°II.6: Exemple de tables de dimensions

## II.6.2 Schémas de modélisation multidimensionnelle

### II.6.2.1 Modèles en étoile

Dans ce modèle, au centre de l'étoile se trouve la table de faits. L'identifiant de cette table est une clé multiple composée de la concaténation des clés de chacune des dimensions d'analyse. Autour de la table de faits on trouve tous les paramètres qui caractérisent les dimensions d'analyse. Ces caractéristiques sont regroupées dans des tables de dimension.

#### Avantages :

Les plus grands avantages de ce type de modèle sont la lisibilité et la performance.

- La lisibilité : La finalité de ce modèle est très évidente et définit clairement les indicateurs d'analyse (Facilité de navigation)
- La performance : les chemins d'accès à la base de données sont prévisibles (nombre de jointures limité).

#### Inconvénients :

- Toutes les dimensions ne concernent pas les mesures
- Redondances dans les dimensions

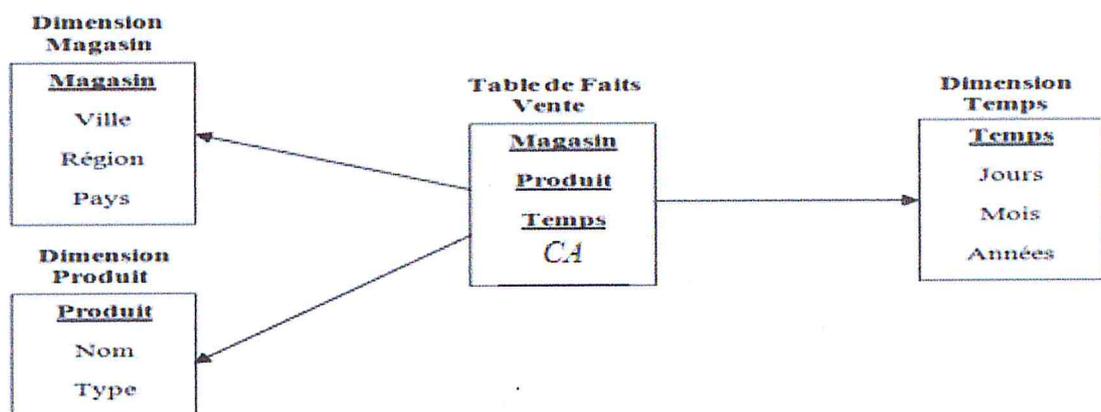


Figure n°II.7: Schéma en étoile

### II.6.2.2 Modèles en flocon en neige

La modélisation en flocon est une modélisation en étoile pour laquelle on éclate les tables de dimensions en sous tables selon la hiérarchie de cette dimension. L'avantage de cette modélisation est de formaliser une hiérarchie au sein d'une dimension. Par contre la modélisation en flocon induit une dénormalisation des dimensions générant une plus grande complexité en terme de lisibilité et de gestion.

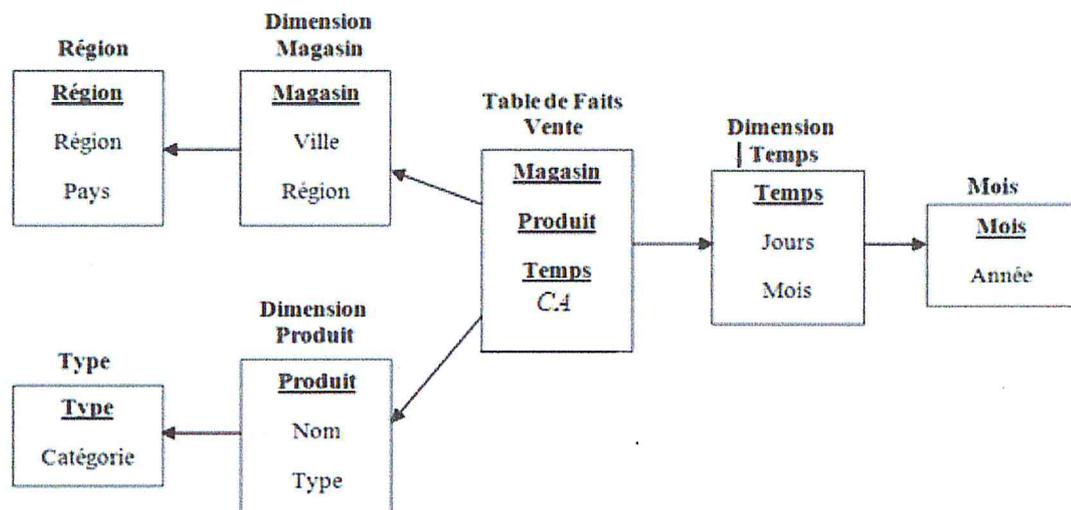


Figure n°II.8: Schéma en Flocon

#### Avantages :

- réduction du volume,
- Permettre des analyses par pallier (drill down/Roll up) sur la dimension hiérarchisée.

#### Inconvénients :

- Navigation difficile ;
- Nombreuses jointures

### II.6.2.3 Modèles en constellation

Il s'agit de fusionner plusieurs modèles en étoile qui utilisent des dimensions communes. Un modèle en constellation comprend donc plusieurs faits et de dimensions communes ou non.

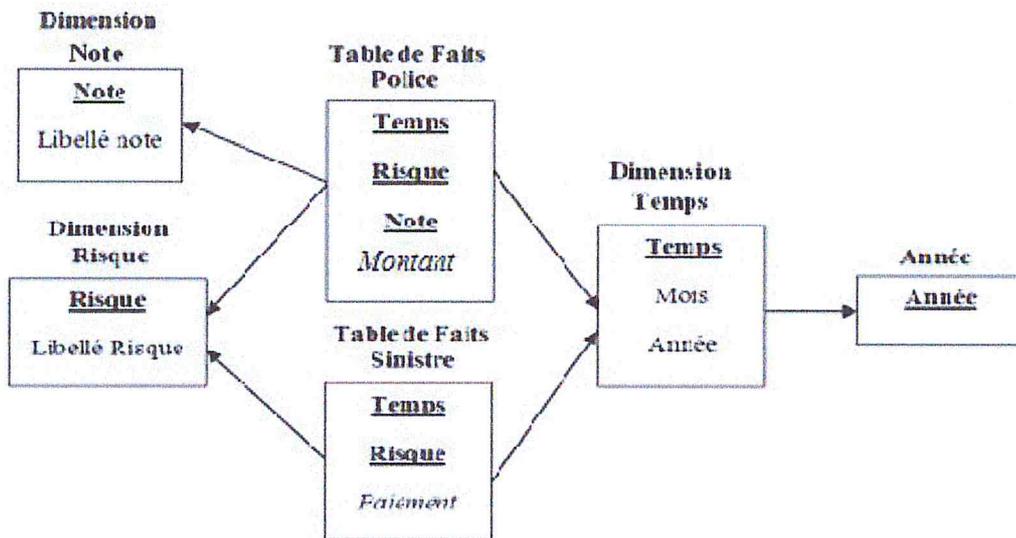


Figure n°II.9 : Schéma en constellation de faits

### III. Les outils d'analyse de données

L'objectif principal des entrepôts de données est d'emmagasiner les données accumulées par les systèmes opérationnels après une série de traitements, pour permettre d'appliquer des outils de découverte de connaissance comme l'OLAP et les techniques de Data Mining, afin d'aider à prendre des bonnes décisions aux bons moments. Et le schéma suivant montre les processus d'extraction de connaissances.

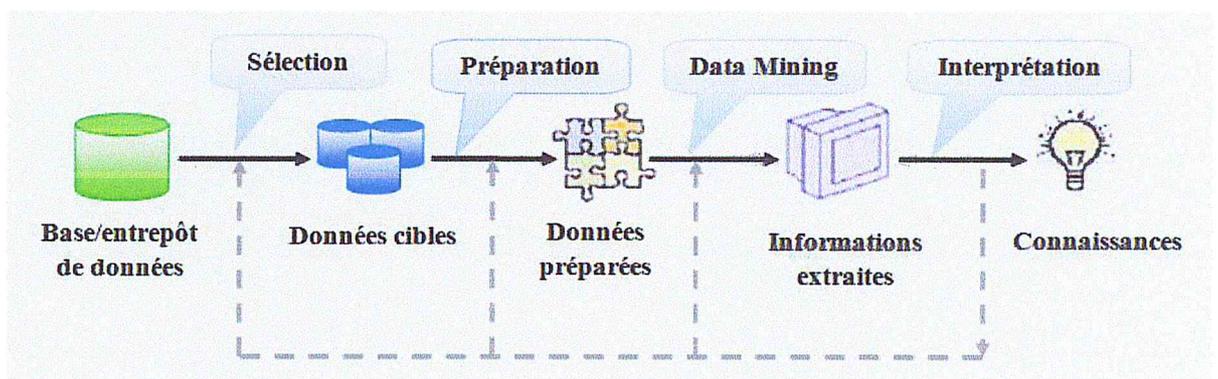


Figure n° III.1: Processus d'extraction de connaissances

### III.1. OLAP (On Line Analytical Processing)

#### III. 1. 1. Le concept OLAP

« OLAP est une activité globale de requêtage et de présentation des données textuelles et numériques contenues dans l'entrepôt de données ; style d'interrogation et de présentation spécifiquement dimensionnelle ». La technologie OLAP est non relationnelle et presque toujours basée sur un cube de données multidimensionnel. Les bases de données OLAP sont également connues sous le nom de bases de données multidimensionnelles.

Les bases de données multidimensionnelles se présentent sous forme d'un cube. Chaque cube représente une dimension d'analyse. Il est difficile de représenter sur le papier plus de trois dimensions, mais un cube d'analyse peut parfaitement comporter dix ou vingt dimensions d'analyse simultanées.

Un cahier des charges comprenant douze règles que doivent satisfaire les systèmes décisionnels. Ces règles sont détaillées dans ce qui suit.

#### III. 1. 2. Les 12 règles OLAP

Afin de formaliser le concept OLAP, fin 1993, E. Codd publie un article intitulé « Providing OLAP to User Analysts » aux Etats-Unis, dans lequel il définit douze règles que tout système de pilotage multidimensionnel devrait respecter. Ces règles sont les suivantes :

➤ **Vue multidimensionnelle**

La base s'appuie sur un hyper cube (cube à n dimensions). L'administrateur définit une fois pour toutes les dimensions qui représentent une façon de trier l'information et regroupent une liste de membres du même type (temps, produit, région, ...).

L'analyse pourra ainsi être affinée dans le détail. L'utilisateur choisit deux ou trois critères à visualiser sous forme de tableau ou de cube. Il peut également faire pivoter les axes d'analyse pour projeter les informations sous un angle différent. Ce critère est certainement le critère-clé du concept OLAP car il reflète la dimensionnalité de l'entreprise telle que la perçoivent ses membres qui ne sont autres que les utilisateurs du système.

➤ **Transparence du serveur OLAP à différents types de logiciels**

Permet d'implanter le système OLAP sans affecter les fonctionnalités du système central. Ainsi, l'utilisateur doit pouvoir utiliser ses progiciels habituels (tableur, reporting, interface graphique, ...) sans percevoir la présence d'un outil OLAP. L'utilisateur ne doit pas se rendre compte de la provenance des données si celles-ci proviennent de sources hétérogènes.

➤ **Accessibilité à de nombreuses sources de données**

Les outils OLAP ont leur propre schéma logique de stockage de données physiques mais doivent accéder aux données et réaliser n'importe quelle conversion pour présenter une vue simple et cohérente des données.

Ils doivent savoir d'où proviennent les données. En fait, par cette règle, le Dr Codd a essentiellement décrit les outils OLAP comme middleware, se plaçant entre des sources de données hétérogènes et une application OLAP.

➤ **Performance du système de Reporting**

L'augmentation du nombre de dimensions ou du volume de la base de données ne doit pas entraîner de dégradation visible par l'utilisateur.

➤ **Architecture Client/serveur**

La plupart des données pour OLAP sont stockées sur des gros systèmes et sont accessibles via des PC. Il est donc nécessaire que les produits OLAP soient capables de travailler dans un environnement Client/serveur.

➤ **Dimensions Génériques**

Toutes les dimensions doivent être équivalentes en structure et en calcul. Il ne doit exister qu'une seule structure logique pour toutes les dimensions. Toute fonction qui s'applique à une dimension doit être aussi capable de s'appliquer à une autre dimension.

➤ **Gestion dynamique des matrices creuses**

Le schéma physique des outils OLAP doit s'adapter entièrement au modèle d'analyse spécifique créé pour optimiser la gestion des matrices creuses. En effet, dans une analyse à la fois sur les produits et les régions, tous les produits ne sont pas vendus dans toutes les régions.

➤ **Support multiutilisateurs**

Support des accès concurrents (récupération, mise à jour, ...), garantie de l'intégrité et de la sécurité afin que plusieurs utilisateurs puissent accéder au même modèle d'analyse ou encore créer des modèles d'analyse provenant des mêmes données de l'entreprise.

➤ **Calculs à travers les dimensions**

Les opérations doivent pouvoir s'effectuer sur toutes les dimensions et ne doivent pas faire intervenir l'utilisateur pour définir un calcul hiérarchique.

➤ **Manipulation intuitive des données**

L'utilisateur dispose d'une ergonomie de consultation, toute manipulation doit être accomplie via une action directe sur les cellules du modèle sans utiliser des menus ou des chemins multiples à travers l'interface utilisateur, on parle ici de navigation.

➤ **Souplesse et facilité de constitution des rapports**

L'analyse et la présentation des données sont plus simples lorsque les lignes, colonnes et cellules de données qui doivent être comparées, sont organisées de façon logique, par des regroupements correspondant à la vision de l'entreprise. C'est pour cela que l'élaboration de rapports doit être souple et conviviale.

➤ **Nombre illimité de niveaux d'agrégation et de dimensions**

Tout outil OLAP doit gérer au moins 5 à 10 dimensions. Le nombre des niveaux d'agrégation est illimité.

### III. 1. 3 Les différents outils OLAP

Plusieurs versions d'OLAP s'affrontent actuellement. Les outils MOLAP « Multidimensionnel OLAP » d'une part qui s'appuient sur une base données multidimensionnelles. Les outils ROLAP « Relationnel OLAP » d'autre part, qui représente leur équivalent sur une base de données relationnelle.

#### III.1.3.1 Les outils MOLAP (Multidimensional OLAP)

Les systèmes multidimensionnels OLAP utilisent une base de données multidimensionnelle pour stocker les données de l'entrepôt et les applications analytiques sont construites directement sur elle. Dans cette architecture, le système de base de données multidimensionnel sert tant au niveau de stockage qu'au niveau de gestion des données.

La figure suivante montre une architecture pour les systèmes MOLAP.

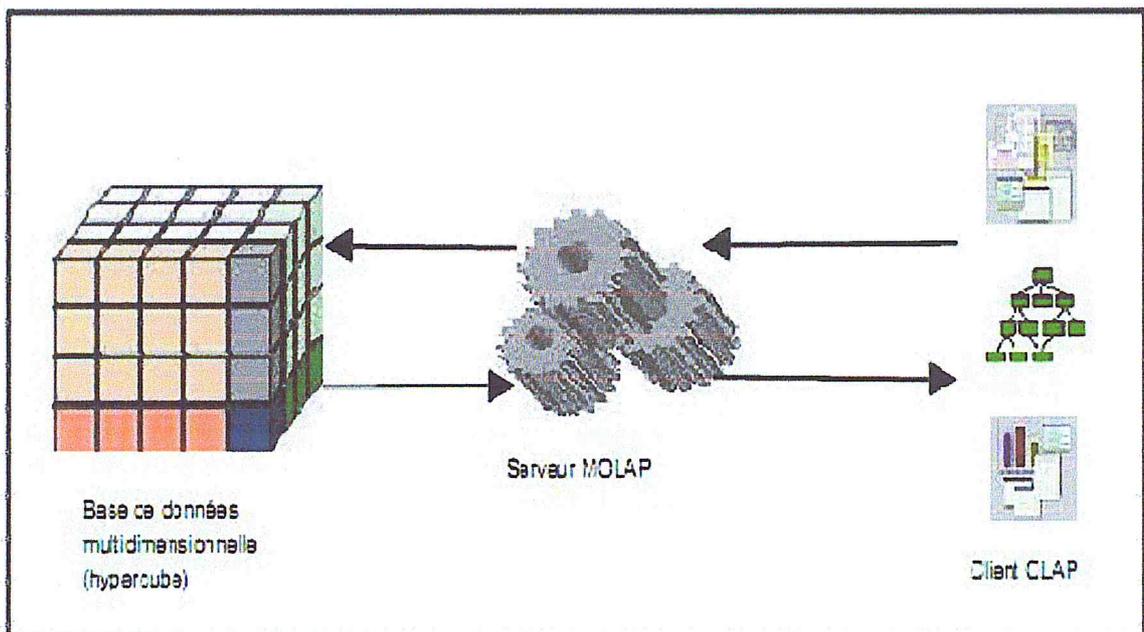


Figure n°III.2: Architecture d'un produit MOLAP

### III. 1. 3.2 Les outils ROLAP (Relational OLAP)

« Ensemble d'interfaces utilisateur et d'applications qui donnent une vision dimensionnelle des bases de données relationnelles » [Kim, 05].

Le système ROLAP est capable de simuler le comportement d'un SGBD multidimensionnel en exploitant un SGBD classique. On entend par cela qu'il conserve la table de la donnée de base sur le serveur OLAP d'origine (ou dans un entrepôt de données) mais utilise un jeu indépendant de tables relationnelles pour stocker les références dimensionnelles et les données agrégées, ce qui dispense de l'acquisition d'un SGBD multidimensionnel. De ce fait, lorsque l'utilisateur interroge la base de données relationnelle, un cube multidimensionnel est fourni en réponse.

L'architecture des systèmes ROLAP est montrée dans la figure suivante :

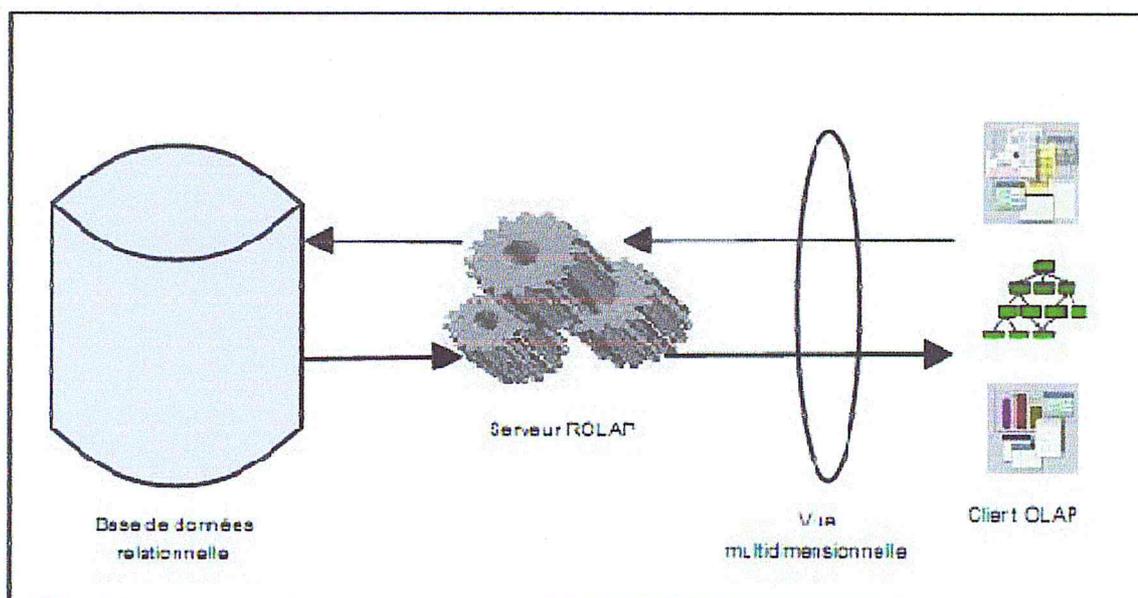


Figure n°III.3: Architecture d'un produit ROLAP

### III.1.3.3 Les outils HOLAP (Hybride OLAP)

Un système HOLAP est un système qui supporte et intègre un stockage des données multidimensionnelles et relationnelles d'une manière équivalente pour profiter des caractéristiques de correspondance et des techniques d'optimisation.

La figure suivante montre une architecture des systèmes HOLAP :

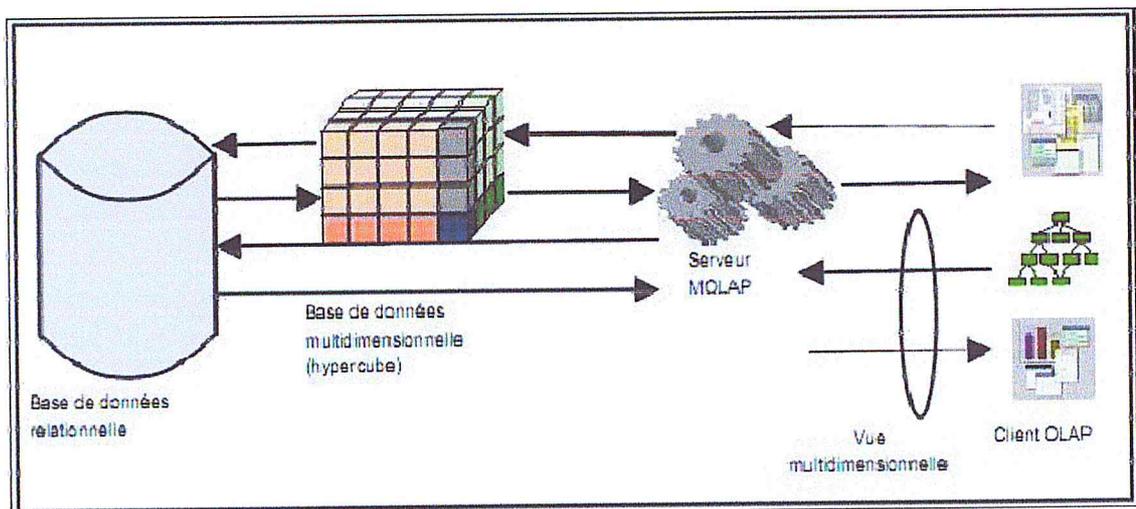


Figure n°III.4: Architecture d'un produit HOLAP

### III.1.4. Opérations sur les données d'OLAP

#### III.1.4.1. Opérations OLAP liées à la structure

Ce sont les opérations qui agissent sur la structure multidimensionnelle, qui visent à changer le point de vue des données.

- **Extraction d'un bloc de données (Dice) :** Elle extrait un « sous-cube » du cube principal. L'opération est illustrée dans la figure suivante :

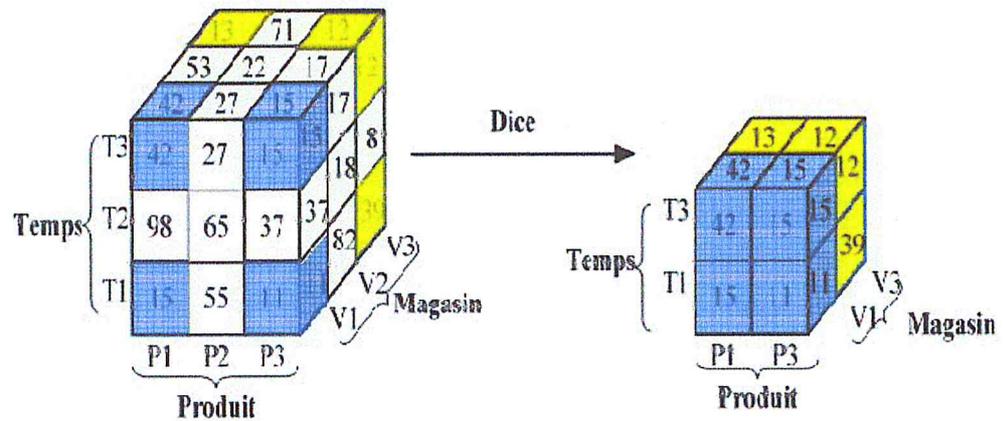


Figure n°III.5 : Opération Dice

- **Pivot (Rotate)** : Consiste à faire effectuer à un cube une rotation autour d'un de ces axes passant par le centre de deux faces opposées, de manière à présenter un ensemble différent. Un exemple explicatif est représenté dans la figure suivante :

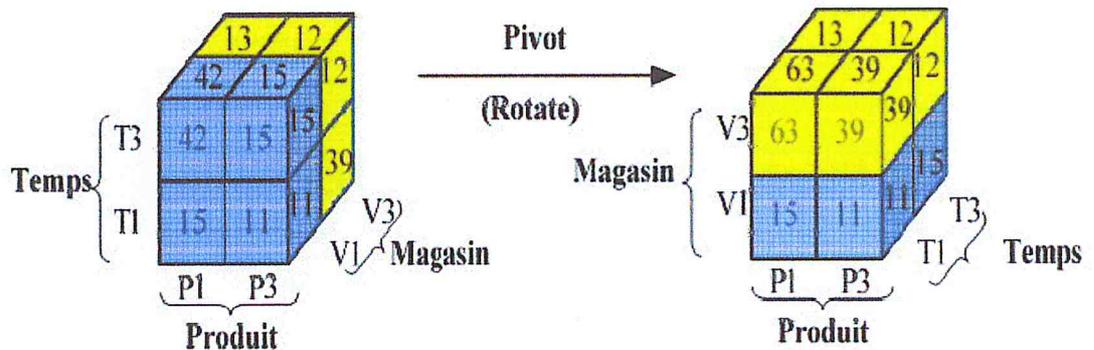


Figure n°III.6 : Opération Rotate

- **Permutation (Switch)** : Consiste à inter-changer la position des membres d'une dimension. Cette opération est illustrée dans la figure suivante :

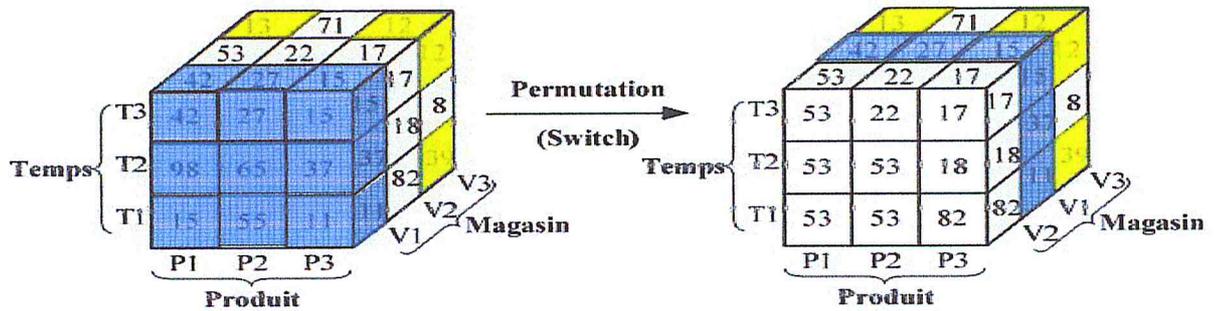


Figure n°III.7: Opération Switch

- **Extraction d'une Tranche du cube (Slice) :** Consiste à sélectionner une dimension, il s'agit de couper une tranche du cube afin d'observer les données de la dimension. Cette opération est illustrée dans la figure suivante :

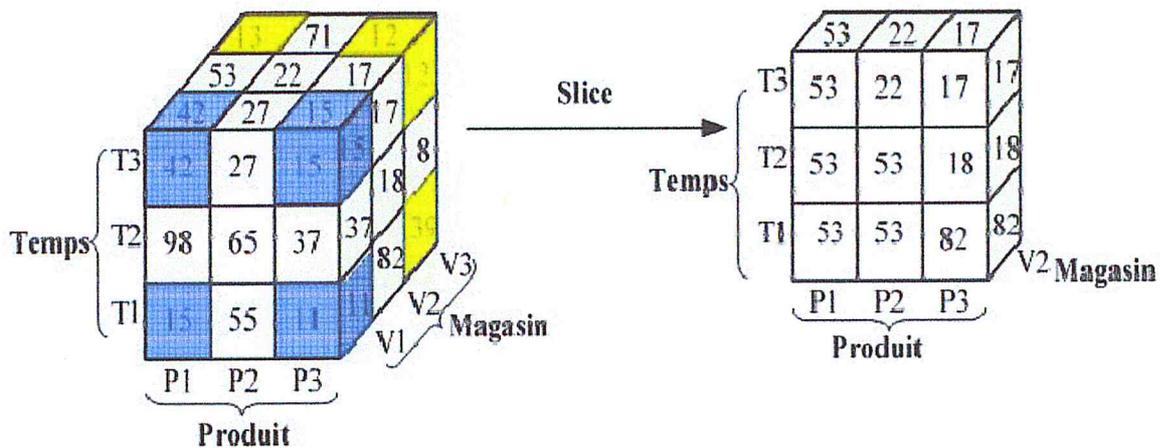


Figure n°III.8 : Opération Slice

- **Division (Split) :** Consiste à présenter chaque tranche du cube et passer d'une présentation tridimensionnelle d'un cube à sa présentation sous forme d'un ensemble de tables. Sa généralisation permet de découper un hyper cube de dimension 4 en cubes

**Exemple :** Si nous avons un hyper cube de quatre (04) dimensions qui sont :

1. La dimension Temps qui contient trois membres T1, T2 et T3.
2. La dimension Produit qui contient trois membres P1, P2 et P3.
3. La dimension Magasin qui contient trois membres V1, V2 et V3
4. La dimension Client qui contient deux membres C1 et C2.

Après avoir effectué l'opération Split sur l'hyper cube, le résultat obtenu est présenté dans la figure suivante :

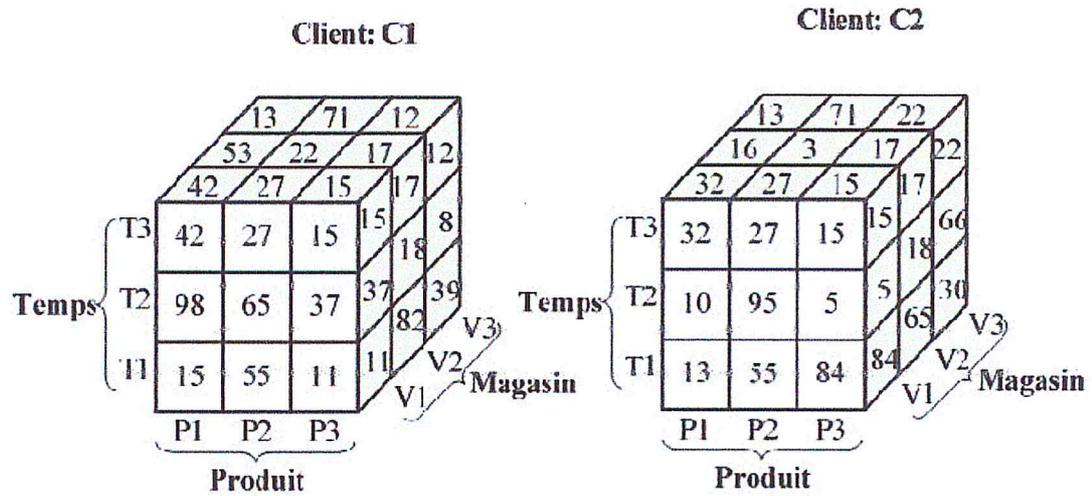


Figure n°III.9: Opération Split

➤ **Emboîtement (Nest)**

Permet d'imbriquer des membres à partir du cube. L'intérêt de cette opération est qu'elle permet de grouper sur une même représentation bidimensionnelle toutes les informations (mesures et membres) d'un cube, quel que soit le nombre de ses dimensions. Un exemple de l'opération d'emboîtement est illustré dans la figure ci-dessous

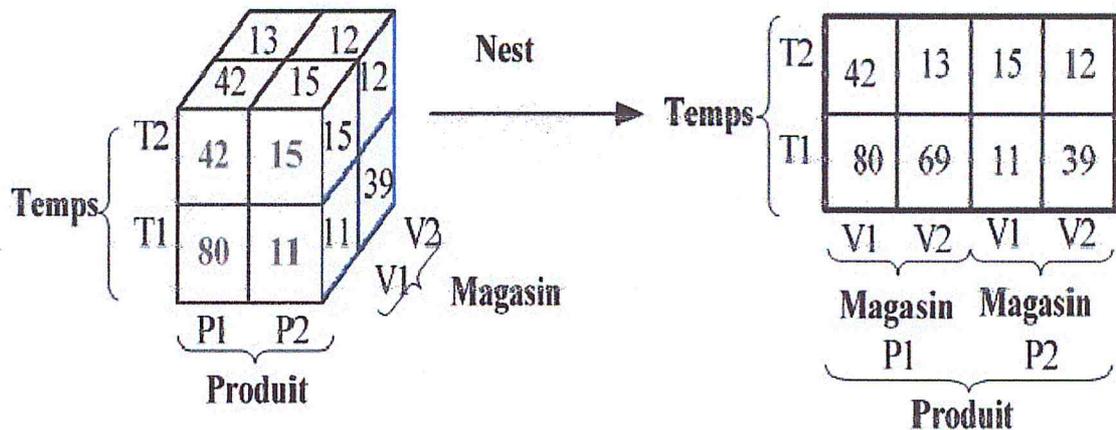


Figure n°III.10 : Opération Nest

- **Enfoncement (Push)** : Permet de combiner les membres d'une dimension aux mesures du cube, identiquement équivalant de faire passer des membres comme contenu de cellules. La figure suivante illustre l'enfoncement.

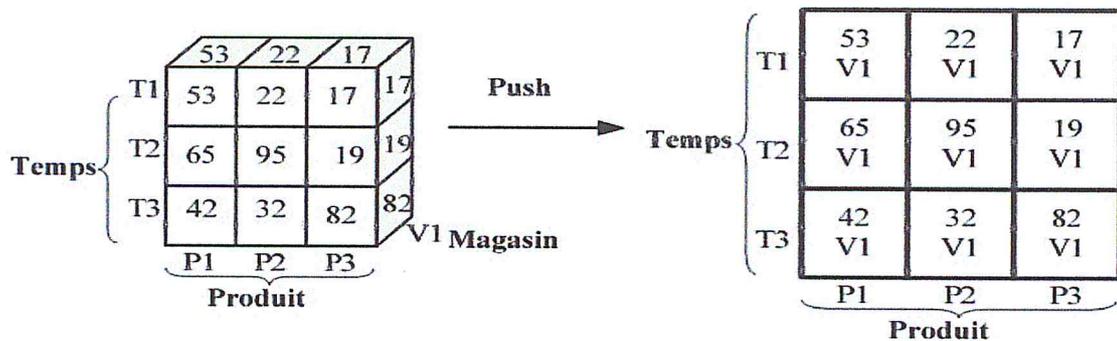


Figure n°III.11 : Opération Push

### III.1.4.2. Les opérations OLAP liées à la granularité

Les opérations agissant sur la granularité d'observation des données, caractérisent la hiérarchie de navigation entre les différents niveaux. Elles correspondent aux opérations suivantes :

- **Forage vers le haut (Drill-up ou Roll-up ou Scale-up)** : Consiste à représenter les données du cube à un niveau de granularité supérieur conformément à la hiérarchie définie sur la dimension. Une fonction d'agrégation est utilisée (somme, moyenne, etc.) en paramètre à l'opération, indique comment sont calculés les valeurs du niveau supérieur à partir de celles du niveau inférieur.
- **Forage vers le bas (Drill-down ou Roll-down ou Scale-down)** : Elle fait l'inverse de la précédente (Roll-up), elle consiste à représenter les données du cube à un niveau de granularité de niveau inférieur, donc sous une forme plus détaillée (descendre dans la hiérarchie d'une dimension). Cette opération est illustrée dans la figure suivante :

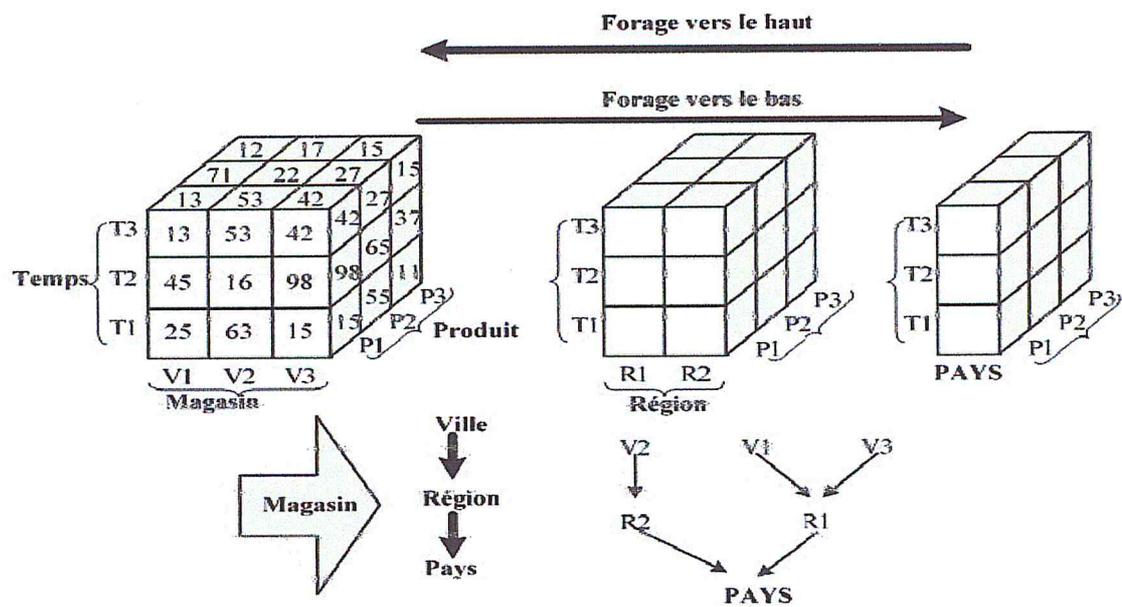


Figure n°III.12: Application du forage vers le bas et vers le haut sur le cube "Ventes"

## IV. Les langages d'interrogation des entrepôts de données

### IV.1 Les langages de données

La représentation multidimensionnelle des données a incité l'émergence de nouveaux langages de données permettant de définir, et de manipuler les schémas des entrepôts de données d'une manière assimilable par la machine. Ces langages de données sont constitués d'un ensemble d'opérations qui peuvent être classées selon le type de langage de données.

#### IV.1.1 Types de langages de données

Dans [MEZIANI et TEBANI, 09], les langages de données dédiés aux entrepôts de données sont classés en trois types et qui se présentent comme suit :

##### IV.1.1.1 Le langage de définition de données (LDD)

Ce langage permet de définir les opérateurs nécessaires à la définition d'un schéma multidimensionnel comme la création des cubes (définition des faits), dimensions et hiérarchies ainsi que la mise à jour (la modification et / ou la suppression) de ses éléments.

#### **IV.1.1.2 Le langage de contrôle des données (LCD)**

Ce langage de gérer la sécurité de l'entrepôt de données, il définit les opérateurs nécessaires à l'affectation et à la suppression des droits des utilisateurs (les permissions) sur les éléments du schéma multidimensionnel.

#### **IV.1.1.3 Langage de manipulation des données (LMD)**

Ce langage adresse des requêtes à l'entrepôt de données et définit les opérateurs nécessaires à la consultation des données.

#### ***Exemple :***

Dans les bases de données, SQL (Structured Query Language) est un langage de données composé de ces trois langages vus précédemment.

- **Définition de données en SQL** : Création d'une table

```
CREATE TABLE Client (NumCl: integer, Nom: string, Ville: string).
```

- **Contrôle de données en SQL** : Gérer les permissions

```
GRANT ALL PRIVILEGES ON Client FOR 'Utilisateur'.
```

- **Manipulation des données en SQL** : Consulter les données

```
SELECT Nom  
FROM Client  
WHERE Ville = 'V1'
```

## **V. Conclusion**

Dans ce chapitre, nous avons évoqué les principaux concepts liés à la conception et à l'exploitation des entrepôts de données, où nous avons vu les caractéristiques des données de l'entrepôt et l'architecture associée.

Nous avons vu, également, les différents modèles OLAP et les opérations de manipulation de ces derniers qui nous permettent l'interrogation efficace des entrepôts de données.

Nous avons constaté, également, que l'entrepôt de données permet d'offrir des contextes d'analyses aux utilisateurs pour les aider dans leur prise de décision. Il s'avère alors que les utilisateurs peuvent avoir besoin de contextes d'analyses spécifiques, répondant à des besoins particuliers voire individuels.

L'émergence de nouveaux besoins d'analyse individuels fait alors apparaître la nécessité d'une personnalisation des analyses, qui placerait l'utilisateur au cœur du processus décisionnel. C'est ce que nous présenterons dans le chapitre suivant.

# Accès personnalisé à l'information

## **I. Introduction**

La personnalisation de l'information constitue 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.

Actuellement, les chercheurs travaillent sur la manière de faciliter l'expression du besoin de l'utilisateur et de lui permettre d'obtenir des informations pertinentes lors de ses accès aux systèmes en modélisant le **profil utilisateur**. Ils travaillent aussi à l'adaptation des systèmes de base de données aux différents utilisateurs qui n'ont pas une idée précise sur l'information qu'ils recherchent c'est qu'ils appellent **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.

Nous nous proposons, dans ce chapitre, d'introduire le cadre théorique dans lequel s'inscrit notre travail, en définissant tout d'abord la personnalisation de l'information ainsi ces domaines d'utilisations, et en détaillant, par la suite, le modèle de profil, la notion de préférence, l'approche de représentation d'un profil, l'approche de construction d'un profil, et enfin nous terminerons par la représentation de quelques travaux sur le profil.

## **II. La personnalisation de l'information**

### **II.1 Définition**

- 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 mais d'un utilisateur spécifique [TAMINE ET ALL ,06].
- 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 [BOUZEGHOUB ET KOSTADINOV, 05].

- La personnalisation est définie comme un processus qui change la fonctionnalité, l'interface, la teneur en information, ou l'aspect d'un système pour augmenter sa pertinence personnelle en fonction des caractéristiques sociodémographiques déclarées de l'utilisateur (sexe, âge, lieu de résidence, etc.) et/ou de son comportement observé. Elle est considérée comme une des portées du processus de l'adaptation. Elle vise à répondre de façon adaptée aux besoins et aux caractéristiques de chaque utilisateur. Il s'agit de gérer la connaissance relative à l'utilisateur et exploite cette connaissance pour décider ce qui doit être présent à l'utilisateur. [AI MAKSSOUD, 08]

## **II.2 Les Domaines d'utilisation de la personnalisation de l'information**

Il est difficile d'identifier la personnalisation de l'information comme un domaine de recherche en soi puisqu'elle intervient dans de nombreuses technologies. Aussi les recherches en personnalisation se trouvent-elles éclatées dans de nombreux articles et de nombreux projets, à priori centrés sur des thématiques de langage de requêtes, d'optimisation de requêtes, de gestion mémoire... etc. Dans cette section, nous traitons des travaux consacrés à la personnalisation dans des domaines aussi variés que l'interaction homme-machine (IHM), les bases de données (BD), la recherche d'informations (RI), mais également des travaux plus récents dans le contexte des entrepôts de données.

### **II.2.1 Dans le domaine de l'interaction homme-machine (IHM)**

Dans le domaine des IHM, la personnalisation se focalise principalement 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.

Dans ce domaine, le profil va contenir des informations qui vont permettre au système d'adapter l'affichage des résultats selon les préférences de l'utilisateur. C'est le cas de l'environnement Yahoo ! Qui recueille dans le profil un certain nombre d'informations personnelles et adapte la page d'accueil en fonction des centres d'intérêt de l'internaute.

### **II.2.2 Dans le domaine de la recherche d'information (RI)**

Le profil utilisateur peut être représenté de différentes manières dont nous évoquons ici quelques exemples.

- Dans certain cas, le profil utilisateur peut être confondu (unir en mêlant) avec la requête elle-même de l'utilisateur. Dans ce cas, le profil utilisateur est alors défini par un vecteur de mots-clés, avec éventuellement un poids associé à chaque mot-clé. [PRETSCHNER et GAUCH, 99] [FEREIRA et SILVA, 01].
- Un profil utilisateur peut également contenir les statistiques d'actions avec le système (nombre de clicks, temps de lecture, etc.), ceci permet par la suite d'inférer (déduire) sur les préférences en connaissant davantage son comportement. [BRADLEY et AL, 00].
- Une autre alternative consiste à stocker dans le profil utilisateur des fonctions d'utilités sur un domaine d'intérêt, qui permettent d'exprimer l'importance relative à des sujets de ce domaine, les uns par rapport aux autres. [CHERNIACK et AL, 03].

### **II.2.3 Dans le domaine des bases de données (BD)**

Dans le domaine des BD, l'utilisateur ne fait pas partie du processus de recherche d'informations, la requête contient en général l'ensemble des critères considérés nécessaires à produire des données pertinentes, et le profil utilisateur peut contenir par exemple les habitudes d'interrogation de celui-ci, en l'occurrence les prédicats souvent utilisés dans ces requêtes ou des ordres dans ces prédicats, ils sont alors intégrés directement aux requêtes par les utilisateurs ou lors de la compilation de ces dernières; donc ils sont pris en compte en une seule fois durant l'exécution de la requête. [KOUTRIKA ET IOANNIDIS, 04]

## **II.2.4 La personnalisation dans les entrepôts de données**

Si la personnalisation n'est pas une idée nouvelle dans les domaines précédemment évoqués, elle constitue un axe de recherche émergent dans le domaine des entrepôts de données. L'intérêt de cet axe de recherche peut être motivé à la fois vis-à-vis de la volumétrie des données connue pour être importante dans les entrepôts de données et du rôle central que joue l'utilisateur dans le processus décisionnel.

En effet, il est en interaction directe avec le système au niveau de l'analyse des données, en particulier dans le contexte de la navigation.

Ils existent aussi d'autres exemples des domaines d'utilisation de la personnalisation telles que :

### ➤ **Systeme d'informations mobiles**

Les services accessibles via les téléphones mobiles et les systèmes embarqués requièrent d'un côté une personnalisation leur permettant de limiter l'effort et le temps passé à la recherche de l'information pertinente, et une adaptabilité aux contraintes physiques ou techniques (écran de petite taille, clavier absent ou réduit, bande passante limitée, etc.).

Les vendeurs de service présentent souvent ces techniques comme des moyens d'aide à la décision pour l'utilisateur.

### ➤ **Accès aux bibliothèques électroniques**

La personnalisation s'applique à différents niveaux :

- limiter l'accès aux seuls documents aux quels l'abonné a souscrit.
- guider la navigation de l'utilisateur au sein de ces documents selon sa requête du moment et recommander les nouveautés en fonction du profil.

### ➤ **Configuration des logiciels (réseau, composantes)**

La technologie informatique elle-même fait appel de plus en plus aux techniques de la personnalisation telle que la configuration de systèmes d'exploitation, de protocole réseaux ou de services web en fonction des besoins des utilisateurs.

### III. Architecture d'un système personnalisé de base de données

L'architecture générale d'un système de base de données personnalisé est basé essentiellement sur :

- **la création de profil** : Qui s'occupe de la représentation des caractéristiques d'utilisateur sous forme d'un **modèle de profil** et constitue le noyau central de la personnalisation.
- **La personnalisation de la requête** : Qui s'occupe de l'intégration des informations du profil à la requête entrante afin de sélectionner le contenu ;
- **La personnalisation de présentation** : Qui s'occupe de la présentation de résultat d'une manière adaptative selon les préférences de l'utilisateur.

La figure suivante illustre cette architecture :

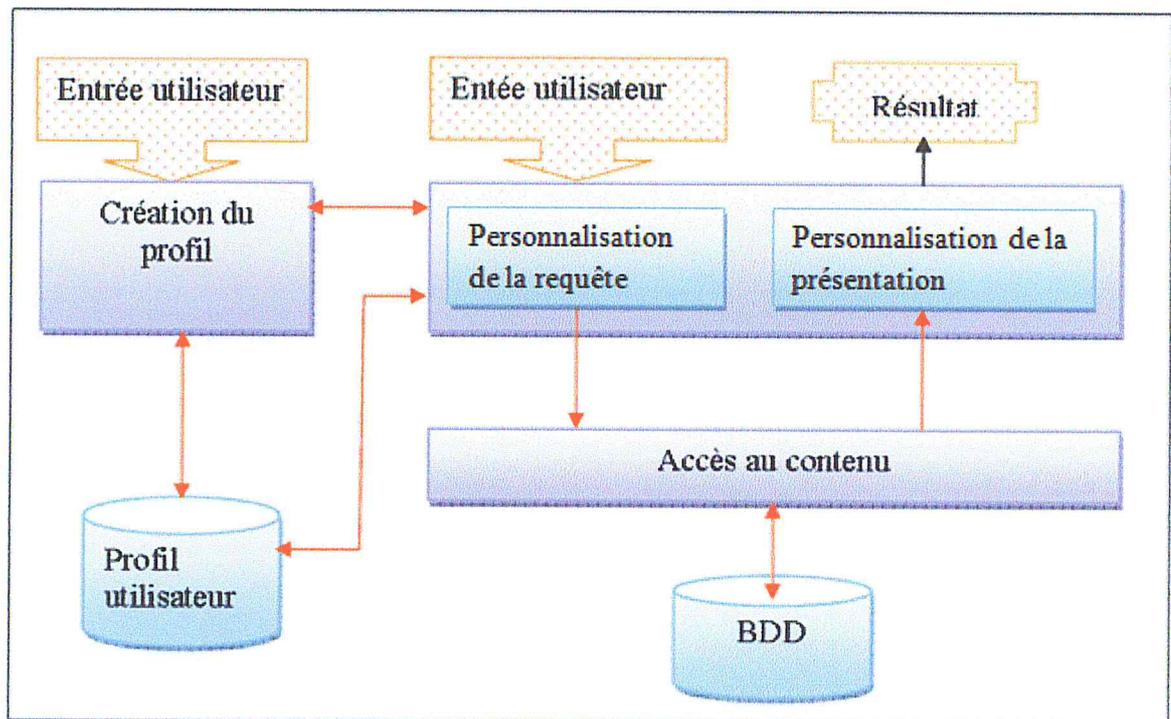


Figure n° III.1: Architecture d'un système de base de données personnalisé

[DJEMAI et GHOUALI, 08]

## **IV. Profil utilisateur**

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

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 ».

### **IV.1 Définition**

La notion de profil utilisateur apparaît comme étant à la base de la personnalisation, mais elle est loin d'être définie de façon standard malgré les Plusieurs approches développées.

- Dans [MEZIANI et TEBANI, 09]: Un profil utilisateur est une collection d'information sur un utilisateur, cette collection peut être vue comme un ensemble de caractéristiques avec des valeurs associées contenant par exemple les préférences d'utilisateurs.
  
- Et dans [ZEMIRLI, 08]: Un profil utilisateur toute structure qui permet de modéliser et de stocker les données caractérisant l'utilisateur. Ces données représentent les centres d'intérêts, les préférences et les besoins en informations de l'utilisateur ou un groupe d'utilisateurs.

### **IV.2 Le contenu du profil**

Dans cette section, nous allons voir les informations que les systèmes stockent le plus souvent sous forme de profils afin de personnaliser leurs services. De façon formelle, le profil est une mise en équation des préférences d'un utilisateur. Il doit contenir toutes les informations nécessaires pour le fonctionnement des systèmes qui l'utilisent.

Chaque système de personnalisation propose un modèle de profil adapté au service qu'il offre. La classification, l'organisation et la structuration des données de profils est un élément clé si on veut avoir une vision globale de la personnalisation.

- Dans [FAVRE, 07] un profil utilisateur peut être représenté simplement par un ensemble de caractéristiques permettant au système d'identifier et modéliser l'utilisateur tout en respectant sa vie privée. Cependant, même si cette vision fait consensus, la structure et le contenu des différents profils pouvant être identifiés dans la littérature sont très hétérogènes d'une application à l'autre. En effet, même si plusieurs applications visent à remplir la même tâche, on peut souligner le fait qu'il existe presque autant de types de profils utilisateurs qu'il y a d'applications.
- Un profil utilisateur ne se limite pas aux préférences sur le contenu des données mais contient également [KOSTADINOV, 08]:
  - Des informations personnelles, que l'on appelle aussi démographiques, telles que le nom, le prénom, l'âge, etc.
  - Des préférences sur le contenu de la réponse à la requête, des préférences sur la présentation de la réponse, et / ou des conditions d'exploitation.

Ces informations peuvent être complétées pour adapter encore mieux le processus de la personnalisation de l'information.

### **IV.3 Exemples 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.

Dans le projet CASPER [DJEMAI et GHOUALI, 08] 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.

Annonces	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 n° IV.1 : Exemple d'un profil 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.

- Un autre exemple de profil est celui qui est représenté sous forme d'un graphe. Ici le profil de l'utilisateur est composé d'un ensemble des prédicats pondérés. Le poids du prédicat exprime son intérêt relatif pour l'utilisateur. Il est spécifié par un réel compris entre 0 et 1. Les poids relatifs des prédicats sont représentés sur les arcs correspondants.

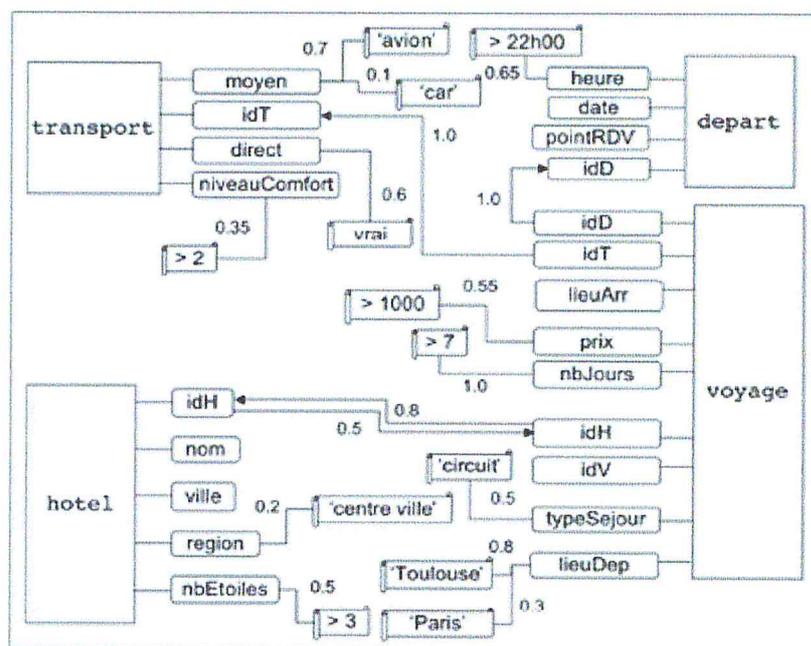


Figure n° IV.1 : Représentation graphique du profil utilisateur

Dans l'exemple 1 ci-dessous, le profil décrit un utilisateur qui aime voyager en avion pour des séjours de plus de 7 jours, préfère les vols directs dont le niveau de confort est supérieur à 2, descend dans les hôtels de plus de 3 étoiles et n'aime pas les circuits touristiques. A chaque prédicat décrivant un élément du centre d'intérêt est associé un poids qui exprime son importance relative par rapport aux autres éléments de ce centre d'intérêt.

**Exemple 1** : Profil utilisateur

```
P1 = {
  VOYAGE.idT=TRANSPORT.idT 1.0 (a)
  VOYAGE.idH=HOTEL.idH 1.0 (b)
  VOYAGE.nb_re_jours>7 1.0 (c)
  TRANSPORT.moyen='avion' 0.7 (d)
  TRANSPORT.type_trajet='direct' 0.6 (e)
  HOTEL.nb_re_etoiles>3 0.5 (f)
  VOYAGE.type_formule<>'circuit' 0.4 (g)
  TRANSPORT.comfort>2 0.4 (h)
}
```

## V. La notion de préférence

Une préférence utilisateur est un ensemble de descriptions englobant: ce qu'un utilisateur envisage d'accomplir dans le système (les activités), le type et l'ordre des résultats de ces activités (le contenu), et la manière dont il aimerait que l'information soit affichée [DJEMAI et GHOUALI, 08].

De cette définition nous distinguons trois types de préférences :

- **Les Préférences d'Activité** : Concernent les activités qu'un utilisateur souhaite et peut accomplir dans le système.
- **Les Préférences de Résultat** : Concernent le contenu et le format préféré des résultats des fonctionnalités (par exemple, de la vidéo, du texte, ou des images).
- **Les Préférences d'Affichage** : Concernent la manière dont l'utilisateur souhaite que l'information soit affichée.

Un utilisateur peut avoir des préférences particulières telles que la langue, la méthode de travail seul ou en groupe, les types d'informations recherchées, les auteurs ou équipe de recherche auxquels il s'intéresse, etc.

### V.1. L'expression des préférences [MEZIANI et TEBANI, 09]

**V.1.1 L'approche quantitative** : Dans cette approche la comparaison 'mieux que' est exprimée en assignant des scores numériques définissent des degrés d'intérêt d'un utilisateur à ces objets.

*Exemple 1:*

Soit  $P$  la fonction attribuant des scores numériques aux auteurs Victor hugo et William schekespir. On dit qu'un utilisateur préfère **Victor** mieux que la **Willaim** si  $P(\text{Victor}) > P(\text{Willaim})$ .

**V.1.2 L'approche qualitative :** Dans cette approche la préférence est spécifiée directement par des relations binaires de préférence. Par exemple, les formules logiques et les constructeurs de préférences spécifiques, cette préférence est comme un ordre partiel sur le domaine des valeurs associées à l'ensemble des attributs : c'est une relation binaire (notée  $>p$ ) entre les tuples d'une même relation. On dira que le tuple **X** est 'mieux que' le tuple **Y** dans le sens d'une préférence 'P' Si  $X >P Y$ .

*Exemple 2 : Soit la table ci-dessous décrivant une instance de la relation Livre. Le schéma de cette relation est Livre (titre, genre, Prix).*

Livre	titre	genre	prix	
	Les hauts des hurles vents	roman	100	<b>X</b>
	Guerre et paix	roman	50	<b>Y</b>

On considère dans cet exemple qu'un utilisateur exprime les préférences suivantes sur les tuples de la relation 'Livre' :

*P1 préférer un tuple de Livre à un autre si et seulement si leur genre est identique et le prix dans le premier tuple est inférieur.*

Sur cet exemple, si on considère la relation  $>p1$  obtenue à partir de l'expression de préférence P1, on a : le tuple **Y** est préféré au tuple **X** «  $Y >p X$  »

## VI. Approches de représentation du profil utilisateur

La représentation de l'utilisateur à travers la notion de profil permet de cibler ses besoins spécifiques dans le but d'améliorer les performances de recherche. Le profil de l'utilisateur, constitué de paquets divers d'informations le caractérisant, traduit une connaissance éparse sur l'utilisateur. Un modèle de représentation permet d'organiser ces éléments afin de faciliter leur exploitation dans le processus d'accès à l'information. On distingue quatre principales approches de représentation : ensembliste, connexionniste, conceptuelle et multidimensionnelle.

## **VI.1 Représentation ensembliste**

L'approche ensembliste consiste à représenter le profil de l'utilisateur par des paquets de termes pondérés. Ces paquets de termes, traduisant les centres d'intérêts de l'utilisateur, peuvent être regroupés différemment selon l'approche suivie pour considérer le profil de l'utilisateur.

On distingue dans la littérature trois grandes approches de représentation du profil utilisateur basées sur ce modèle :

- Par une liste de mots clés, où chaque mot correspond à un centre d'intérêt spécifique .
- Par un vecteur de termes pondérés pour chaque centre d'intérêt.
- Par un ensemble de vecteurs de termes pondérés (ou non) indépendants, pour prendre en compte des centres d'intérêt multiples où chaque vecteur correspond à un domaine d'intérêt.

Le poids associé à chaque terme permet de représenter son degré d'importance dans le profil de l'utilisateur. [ZEMIRLI, 08] [AI MAKSSOUD, 08] [DJEMAI et GHOUALI, 08]

## **VI.2 Représentation sémantique**

Elle utilise essentiellement les ontologies pour mettre en relief les relations sémantiques qui relient les unités informationnelles qui composent le profil utilisateur. [BOULKRINAT, 07]

## **VI.3 Représentation conceptuelle**

Cette représentation est essentiellement basée sur l'utilisation d'ontologies ou des ou des réseaux probabilistes. Dans ce type d'approche les liens entre les concepts sont explicitement induits de l'ontologie et le profil résultant inclura des relations informationnelles plus diverses et spécifiques. La représentation conceptuelle est semblable à la représentation sémantique, dans le sens, où elle représente les centres d'intérêts de l'utilisateur par un réseau de nœuds conceptuels.

Cependant, dans l'approche conceptuelle, les nœuds correspondent à des domaines abstraits représentant les centres d'intérêts de l'utilisateur, contrairement à l'approche sémantique où les centres d'intérêts sont représentés par un mot spécifique ou ensemble de mots relatifs. [ZEMIRLI, 08]

#### **VI.4 Représentation multidimensionnelle**

Les utilisateurs sont divers et complexes, ils sont caractérisés par des modèles différents et font partie d'une communauté. Ils effectuent des tâches multiples ayant des buts différents. Ils ont également des activités simultanées de recherche. Une autre représentation possible du profil est la représentation multidimensionnelle. Cette représentation 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. [ZEMIRLI, 08] [DJEMAI et GHOUALI, 08]

### **VII. Approches de construction du profil utilisateur**

La construction du profil traduit un processus qui permet d'instancier sa représentation. L'approche de construction dépend fortement de la représentation choisie pour le profil utilisateur: les techniques utilisées par les systèmes différents selon qu'ils représentent le profil par un (des) vecteur(s) de termes ou par des classes (hiérarchiques ou pas).

Cependant la démarche de construction commune à tous les systèmes est la suivante : on commence par collecter des informations sur l'utilisateur à partir de sources d'informations diverses, puis on applique des techniques et des algorithmes pour apprendre à partir de ces informations le profil de l'utilisateur. La construction du profil s'effectue donc en deux étapes :(1) l'acquisition et la collecte des données utilisateur ; (2) puis la construction proprement dite du profil.

#### **VII.1 Acquisition des données utilisateurs**

Cette phase consiste à collecter les informations pertinentes pour instancier le profil de l'utilisateur. Le processus d'acquisition des données de l'utilisateur peut collecter ces informations soit directement à partir de la machine de l'utilisateur (côté client) ou à partir de l'application (côté serveur). Ce processus d'acquisition peut être explicite et/ou implicite :

En effet, toute interaction de l'utilisateur avec le système est considérée comme une estimation de son jugement d'intérêts, basé sur la durée de l'interaction avec le système.

## **VII.2 Construction du profil**

La construction des profils se décompose en 2 étapes : (1) La modélisation de profil et (2) l'instanciation de profil.

### **VII.2.1 La modélisation de profil**

La modélisation est l'une des premières fonctionnalités que doit offrir un système de personnalisation considérant le profil utilisateur comme une forme de connaissance, son but est de sélectionner les informations les plus appropriées qui traduisent les intérêts de celui-ci. Cette modélisation consiste à désigner une structure (Méta model) pour stocker toutes les informations qui caractérisent l'utilisateur et qui décrivent principalement ses centres d'intérêts en plus d'autres informations relatives à ses préférences, les objectifs de sa recherche, ses traits individuels et son expérience, etc. [BOULKRINAT, 07]

### **VII.2.2 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. [DJEMAI et GHOUALI, 08]

## **VIII. Travaux de personnalisation dans les entrepôts de données**

Lors de notre étude bibliographique, nous avons pu examiner de nombreux travaux traitant de la personnalisation. Même si l'objectif est le même et qui est celui d'adapter la réponse à la requête utilisateur en faisant intervenir ses préférences, les méthodes utilisées n'abordent pas nécessairement ce problème de la même manière.

Avant de présenter ces travaux nous allons définir quelques concepts essentiels pour mieux se rendre compte.

## **VIII.1 Concepts et définition**

### **VIII.1.1 Référence d'une cellule**

Une cellule ou fait de  $c$  est un tuple de l'instance de la table des faits  $F$ , et sa référence est un tuple et elle est définie à partir de la table des faits.

### **VIII.1.2 Structure**

Définit le placement des dimensions sur les différents axes et l'ordre d'apparition des membres de chaque dimension sur un axe donné. Et décrit avec précision la table croisée utilisée pour afficher une instance donnée d'un cube.

Une structure d'affichage spécifie le moyen d'organiser un ensemble de références dans un tableau croisé, notamment en précisant comment placer les dimensions du cube analysé sur les axes d'analyse. Étant donné qu'un utilisateur peut préférer voir un même ensemble de références en utilisant une structure donnée plutôt qu'une autre.

### **VIII.1.3 Grille d'analyse**

Une grille d'analyse est le résultat (ou visualisation) de l'exécution d'une requête sous la forme d'une grille de cellules. Elle est constituée seulement d'un ensemble de références de structures. Veut dire qu'elle ne contient pas de faits. Et pour chaque grille possède une structure et un ensemble de références propre à elle.

### **VIII.1.4 Visualisation**

Une visualisation permet de présenter les données d'un cube à un utilisateur sous forme d'une table croisée. Elle contient les données du cube et est caractérisée par une structure qui permet de préciser la forme de présentation des données du cube à l'affichage.

### **VIII.1.5 La contrainte de visualisation**

Elle indique si un cube avec sa structure peut être visualisé.

### **VIII.2 Travaux de BELLATRECHE et AL 2005**

Dans [BELLATRECHE et AL, 2005] la personnalisation consiste à transformer la requête utilisateur avant l'accès au cube de données, l'utilisateur est interrogé afin de donner ses préférences utilisateurs qui classent les membres, les dimensions et la contrainte de visualisation qui contrôle les résultats sont stockées dans le profil utilisateur. Et une définition du profil utilisateur dans le contexte OLAP est proposée.

### **VIII.3 Travaux de BELLATRECHE et AL 2006**

Dans [BELLATRECHE et AL, 2006] il s'agit de combiner la requête originale de l'utilisateur ( $q$ ) avec le profil utilisateur ( $\Gamma$ ) qui contient ses préférences et vérifiant les contraintes de visualisation pour avoir une requête ( $q^*$ ) personnalisée et prête à être exécutée sur le cube de données.

### **VIII.4 Travaux de MOULOUDI 2007**

Dans [MOULOUDI, 2007] l'approche de personnalisation des requêtes OLAP peut être effectuée à l'aide de méthodes différentes et qui se présentent ainsi :

- dans la première, il s'agit de personnaliser la requête de l'utilisateur qui veut dire transformer la requête de l'utilisateur avant l'accès aux cubes de données,
- et l'autre consiste à personnaliser la réponse de la requête utilisateur qui veut dire transformer le résultat de la requête utilisateur après l'accès au cube de données.

L'auteur a montré que sous certaines conditions que les deux types de personnalisation aboutissent aux mêmes résultats. Mais lors de l'implémentation, l'auteur s'est limité à la première seulement.

Dans cette approche, la personnalisation de la visualisation d'une requête OLAP exprimée en langage MDX passe par la personnalisation d'une grille d'analyse associée à cette visualisation.

### **VIII.5 Travaux de [RAVAT et AL, 08]**

Dans [RAVAT et AL, 08] l'acquisition des données modélisant le profil utilisateur est faite d'une manière implicite en les récoltant de l'expertise de l'utilisateur décideur (commentaires, discussions, prises de décision...).

Pour mieux représenter les besoins de l'utilisateur en matière de données analysées il a proposé une définition d'un modèle de préférence, cette proposition repose sur une approche qualitative représentant les préférences de l'utilisateur par une relation d'ordre exprimée sur les données.

### **VIII.6 Travaux de [MEZIANI, 09]**

Concernant la modélisation du profil utilisateur [MEZIANI, 09] a utilisé l'acquisition explicite parce que il a vu qu'elle est suffisante pour couvrir toutes les préférences exprimée par l'utilisateur. Ensuite, pour l'approche de personnalisation il a adopté une démarche hybride combinant celles inspirées de la recherche d'information [BELLETRECHE et AL, 2005], [BELLETRECHE et AL, 2006], [MOULOUDI et AL, 2005] qui permettent la transformation des requêtes utilisateur avec la prise en compte de ses préférences et celle de réutilisation de ces requêtes [PITOURA et STEFANIDIS, 2002].

Concernant la réalisation, il a réalisé un prototype qui permet l'acquisition des préférences utilisateur d'une manière explicites, la personnalisation de la requête MDX générée par utilisateur qui consiste à la transformation, l'exécution de celle –là, la présentation de ses résultats et sa réutilisation.

## **IX. CONCLUSION**

A travers ce chapitre nous avons constaté que l'objectif de la personnalisation est d'éviter à l'utilisateur d'écrire à chaque fois la partie commune à ses requêtes, l'élaboration d'un modèle de profil générique est le premier pas vers la construction de systèmes natifs de personnalisation capables de rechercher et extraire des informations de qualité selon les préférences d'un utilisateur. C'est pour cette raison que les chercheurs intègrent le profil de l'utilisateur en le modélisant sous forme d'un modèle de préférence. Ce modèle représente le noyau de la personnalisation qui consiste à modéliser le profil de l'utilisateur.

La partie de l'état de l'art étant terminée, nous entamerons dans le chapitre suivant la conception de notre prototype en mettant en application les notions décrites précédemment.

# *PARTIE*

## **II**

### **. Conception et Réalisation**

✓ Conception

✓ Réalisation

# Conception

## **I. Introduction**

La modélisation du profil utilisateur pour un système de personnalisation 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 un résultat personnalisé selon la particularité des besoins informationnels de chacun.

Dans ce contexte, se situe le but de notre projet qui consiste à introduire la personnalisation dans le domaine des entrepôts de données, dont la modélisation du profil utilisateur est le composant principal dans ce type de personnalisation.

## **II. Méthodologie de conception**

Pour parvenir à la mise en œuvre de notre système, nous avons estimé que notre conception devrait répondre à un double souci :

- Cibler les paramètres qui caractérisent au mieux un utilisateur en quête d'informations pour les modéliser dans son profil ;
- Concevoir un système qui gère le profil à la bonne prise en charge dans le processus d'analyse.

## **III. Modélisation du profil utilisateur**

Il n'existe pas à l'heure actuel de modèle global spécifique à un utilisateur. 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 qui décrivent plusieurs dimensions du profil, Nous avons estimé qu'un certain nombre d'exigence aux quelles doit répondre ce dernier pour que sa description soit la plus générique que possible afin de représenter au mieux les besoin des utilisateurs.

Ces exigences sont traduites à travers la réponse aux questions suivantes :

N°	Question	Description
1	Qu'est-il ?	Informations personnelles.
2	Il veut obtenir des informations, comment ?	Préférences
3	Il veut ces informations, pour quel but ?	Objectifs visés à travers la recherche ou l'usage à faire des informations
4	Est-ce que je désire que les autres sachent ce que je recherche comme information ou non ?	Sécurité et confidentialité.
5	Quelle est la qualité espérée Concernant les services offerts?	Annotation de système.
6	Divers.	toute information utile pour consolider la description du profil.

**Tableau n° II.1 : les différentes exigences pour la description d'un profil.**

### III.1 Choix d'une représentation du profil

La plupart des systèmes connus actuellement se basent sur une représentation vectorielle contenant les mots clés pondérés selon leur importance dans la collection ou bien selon l'intérêt que leur accorde l'utilisateur. La similarité entre deux profils, est estimée par le calcul de la distance qui sépare leurs vecteurs correspondants, en utilisant un des algorithmes tel que celui des k proches voisins, le modèle vectoriel, etc.

La représentation vectorielle du profil présente cependant plusieurs limites, parmi les quelles :

- Elle ne prend pas en considération l'ambiguïté des mots-clés qui peuvent avoir plusieurs sens et qu'un concept peut être décrit par plusieurs mots ;
- Les mots ne décrivent pas correctement le contexte.
- Les mots-clés manquants dans les vecteurs profils doivent être complétés, avant de pouvoir les comparer de manière correcte.

En égard ces limites et contraintes, nous avons opté pour une représentation **multidimensionnelle/hierarchique**. Ce choix est justifié par le fait que le profil est caractérisé par plusieurs catégories d'informations où chaque catégorie ou dimension peut être décrite au besoin de manière hiérarchique. Dans ce contexte, les centres d'intérêts sont la partie du profil qui sera soumise à une telle description.

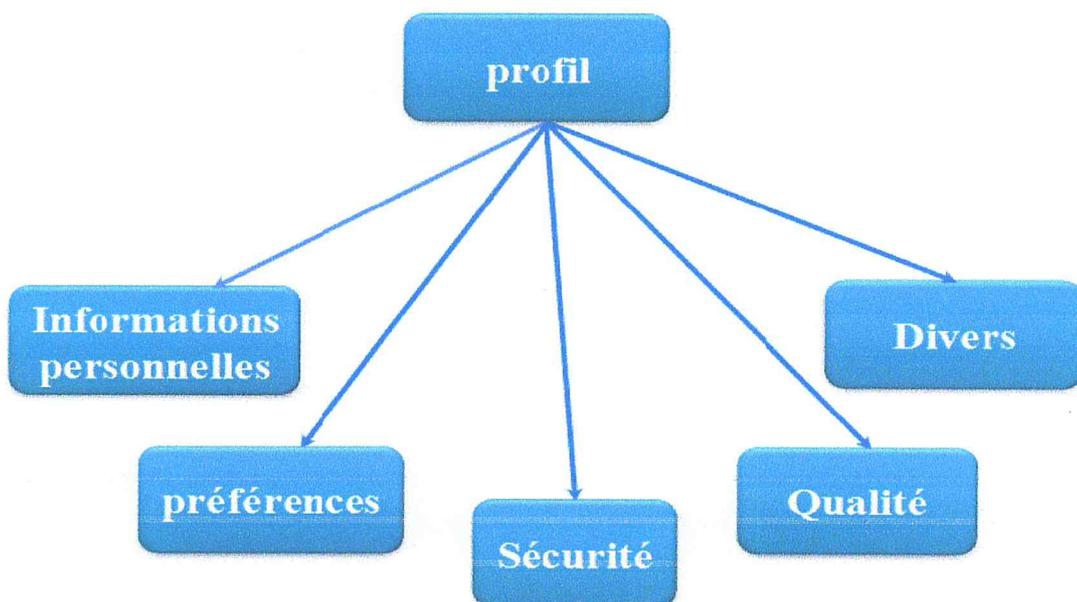


Figure n° III.1: représentation multidimensionnelle et hiérarchique d'un profil

### III.2 Les différentes dimensions, sous dimensions et attributs du profil

#### ➤ Les informations personnelles

Les informations personnelles sont la partie statique du profil. Elles contiennent des informations qui décrivent l'utilisateur, ces informations décrivent principalement des données personnelles tel que : nom, prénom, âge, profession, adresse, n° téléphone, e-mail, C.C.P, nationalité, sexe...etc.

#### ➤ Les préférences

Cette dimension traduit ce que veut ou ce que ne veut pas l'utilisateur comme information, Autrement dit ce que veut inclure ou exclure l'utilisateur comme informations durant le processus de recherche, Tel que les préférences de service : Ces préférences concernent l'information recherchée.

#### ➤ Sécurité

Cette dimension permet de sauvegarder les droits d'accès de chaque utilisateur aux données de la base de données, elle enregistre aussi les données sensibles concernant les clients comme par exemple le numéro du compte bancaire et le CCP ainsi que les mots de passe pour chaque utilisateur.

*Remarque : dans notre cas tous les clients ont les mêmes droits d'accès.*

#### ➤ Qualité (annotation)

Cette dimension permet de savoir le degré de satisfaction des clients concernant les informations proposées en utilisant les préférences stockées dans le profil de chaque utilisateur.

Elle va nous aider pour l'amélioration de notre prototype a travers :

- Une note donnée par le client.
- Un questionnaire destiné aux clients.

## ➤ Divers

L'intérêt de cette dimension est de ne pas restreindre le profil aux seules informations contenus dans les dimensions précédentes, car certaines applications demandent des informations spécifiques ne peuvent être inclus dans aucune des autres dimensions.

### III.3 La méta - description du modèle de profil

La méta-description du modèle fournit une représentation générale du profil, Cette représentation sera utilisée dans la définition des profils.

Nous avons choisi l'utilisation d'un formalisme par objets de type classe/association pour modéliser le profil utilisateur, car ce formalisme est largement connu et maîtrisé par les concepteurs. De plus, à travers cette représentation par objets, la relation d'agrégation est bien représentée.

La Figure III.2 présente une représentation par objets du modèle du profil. Cette représentation peut être structurée sous forme d'une hiérarchie de classes.

#### III.3.1 Identification des classes

##### ➤ Liste des classes d'objets

1. La classe *Profil* : cette classe décrit les critères qui caractérisent un utilisateur.
2. La classe *dimension* : cette classe décrit des informations caractérisant un profil telles que les préférences, les compétences, etc. En effet ces informations peuvent être regroupées et structurées sous forme d'un ensemble de dimensions. Donc nous pouvons dire qu'un profil est composé d'un ensemble de dimensions.
3. La classe *SousDimension* : les informations qui sont regroupées et structurées en dimensions seront à leur tour structurées en sous dimensions. Par exemple, la dimension *préférences* pourra être caractérisée par la sous dimension *Activité*, et la sous dimension *Présentation*, etc. Cette sous dimension peut être composée d'un ou plusieurs attributs.

4. **Classe *Attribut*** : cette classe permet d'identifier les attributs de chaque dimension et sous dimensions du profil.

5. **Classe *ValeurAttribut*** : cette classe décrit les valeurs possibles qui sont associées à chaque attribut.

### III.3.2 Liste des associations

En ce qui concerne la description des différentes associations, le modèle du profil comporte deux types d'association :

#### 1) Relation d'agrégation

Plusieurs relations de décomposition/d'agrégation existent entre les classes :

une relation d'agrégation entre la classe *profil* et la classe *dimension* pour signifier qu'un profil peut être composé d'une ou plusieurs dimensions, une relation d'agrégation entre la classe *dimension* et la classe *sous dimensions* pour signifier qu'une dimension est composée d'une ou plusieurs *sous dimensions* et une relation d'agrégation existe également entre la classe *sous dimensions* et la classe *Attribut* pour signifier qu'une sous dimensions peut être composé d'un à plusieurs attributs.

#### 2) Association *est Associé*

Cette association est définie entre la classe *Attribut* et la classe *valeur* pour signifier qu'un *Attribut* est associé à une ou plusieurs valeurs. En d'autres termes, ces valeurs peuvent être considérées comme des instances des *Attributs*. En principe, ces valeurs constituent les données qu'un profil donné peut contenir.

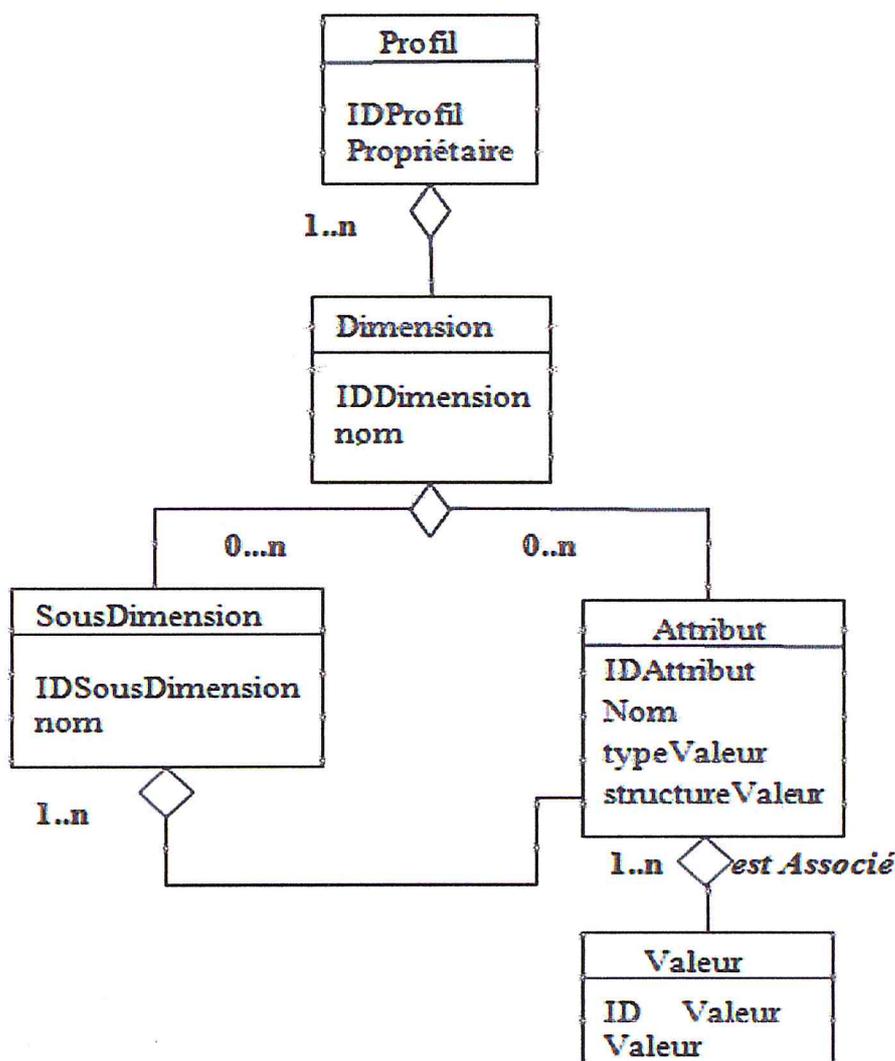


Figure n° III.2: Diagramme de classe pour la base de profil

Le profil utilisateur 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 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.

#### IV. Architecture globale de l'approche de la personnalisation

L'approche de la personnalisation que nous proposons est de pouvoir offrir à l'utilisateur des informations qui correspondent à ses besoins ainsi ses caractéristiques sont pris en considération.

Pour atteindre cet objectif nous avons conçu un système composé de plusieurs sous systèmes.

Ces sous-systèmes sont :

- Gestion du profil
- Appariement des profils.
- Mise en œuvre de la personnalisation.

Le schéma suivant donne un aperçu global de l'architecture fonctionnelle du système projeté.

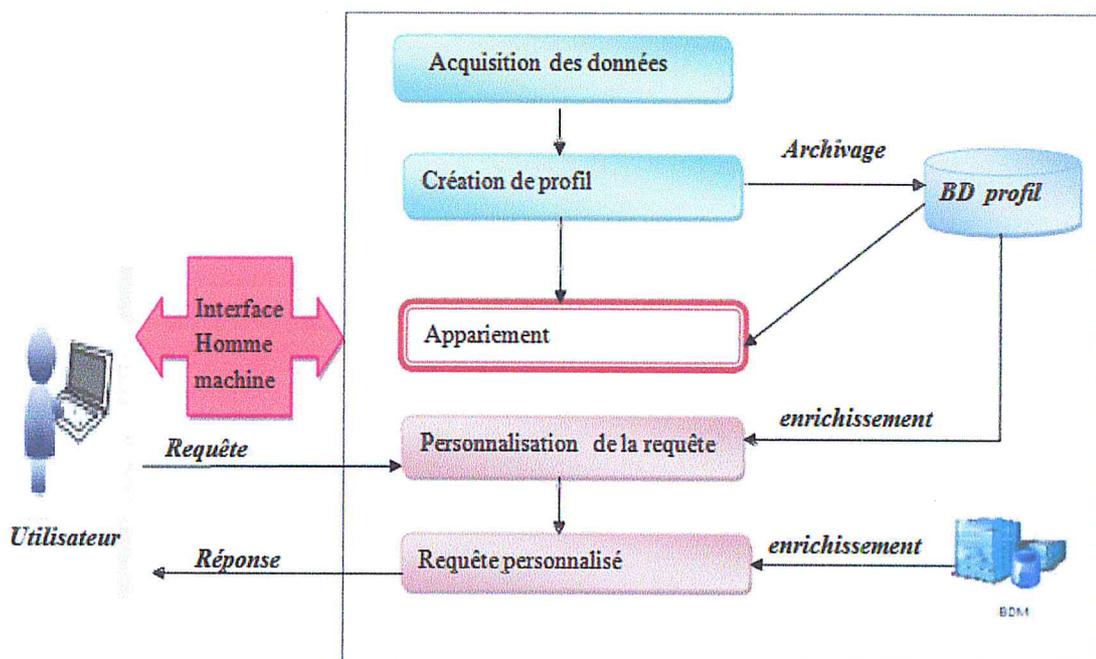


Figure n° IV.1 : schéma général du système de la personnalisation.

## IV.1 Description des différents sous-systèmes

### IV.1.1 Sous-système *de la gestion des profils*

Ce sous-système est composé du module d'acquisition basé sur explicite et du module de mise à jour du profil utilisateur et enfin l'archivage de ces profils dans une base. Le module d'acquisition explicite du profil se présente comme suit :

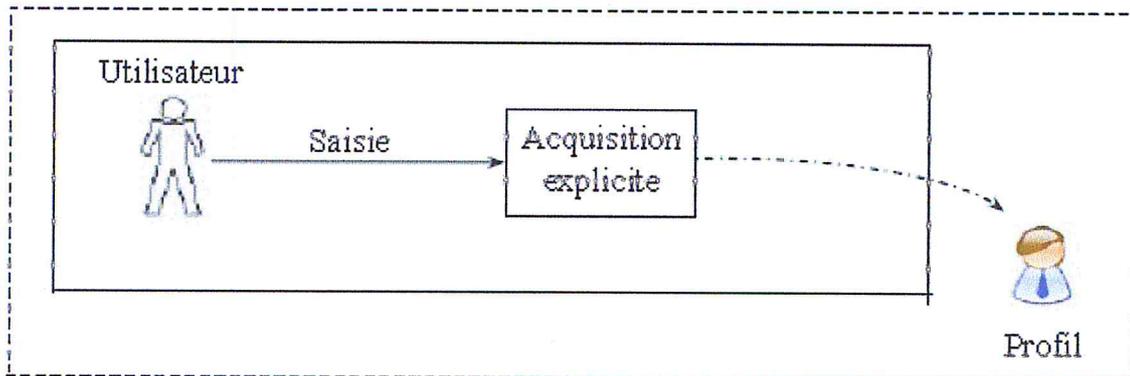


Figure n° IV.2 : Acquisition explicite du profil utilisateur

#### ➤ Acquisition explicite du profil

Au démarrage du système et dès son premier contact avec le système, l'utilisateur est invité à fournir explicitement des informations le concernant à travers le remplissage d'un formulaire.

Ces informations dont certaines sont obligatoires et d'autres facultatives, concernent:

- ✓ Son identification socio-professionnelle ;
- ✓ Ses préférences ;

On a choisi cette technique d'acquisition (explicite) parce que c'est difficile de définir un processus d'interprétation du comportement observé dans un contexte d'application spécifique. Le processus doit être capable d'inférer les prétentions (intention, centres d'intérêts, préférences) de l'utilisateur. Le plus souvent ces prétentions sont incertaines et doivent être validées par l'utilisateur parce que il ya pas mieux qu'un utilisateur qui connaît ses préférences.

➤ **Mise à jour du profil**

Cette fonctionnalité vise à rendre le profil une entité dynamique à l'intérieur du système, on parle ici de la maintenance du profil elle est faite d'une manière explicite par l'utilisateur ou implicite en observant le comportement de l'utilisateur lors de ses interaction avec le système.

Les opérations qui peuvent être effectuées sont :

1. Suppression d'un profil ;
2. Création d'un nouveau profil ;
3. Consultation, ajout, modification ou suppression du contenu de profil courant.

➤ **l'archivage des profils**

Chaque profil crée, doit être archivé dans une base de profils pour les utiliser par la suite dans l'appariement et dans le processus de la reformulation des requêtes  
Le schéma fonctionnel du sous-système de la gestion du profil se présente comme suit:

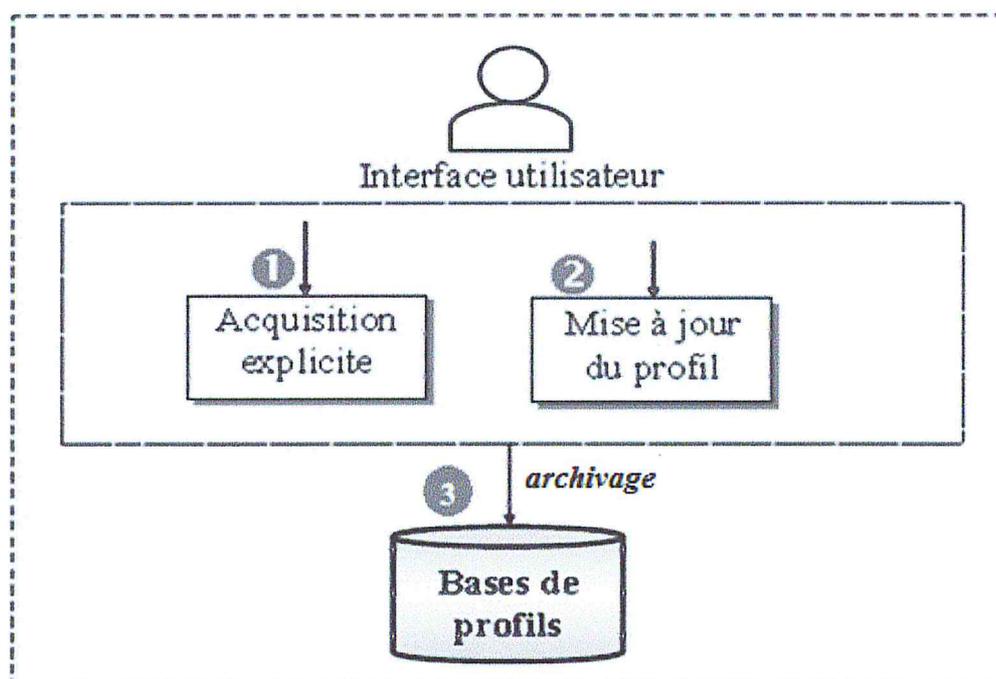


Figure n° IV.3 : Sous-système de la gestion du profil utilisateur

#### IV.1.2 Sous système d'Appariement des profils

Cette opération est constituée des primitives qui permettent de comparer deux profils pour s'assurer de leurs équivalences strictes, isoler leurs parties communes ou énumérer leurs différences. Ces primitives s'appliquent aussi bien au niveau modèle des profils qu'à leurs valeurs. Dans le premier cas, elles permettent de comparer la structure des profils qui est la représentation des profils au niveau modèle, tandis que dans le second cas, les primitives permettent de confronter les valeurs des attributs de deux profils. La primitive que nous avons identifiée est : *le matching (Match (P1, P2))*.

##### Qu'est ce qu'un matching ?

Le matching a pour but de calculer la similarité sémantique entre deux concepts:

**profil/profil:** L'appariement entre deux profils peut être appliqué dans le filtrage collaboratif afin de trouver des utilisateurs similaires.

**profil/contenu:** Cet appariement est utilisé dans les systèmes de recommandation basés sur le contenu, afin de savoir si ce contenu convient à l'utilisateur.

**contexte/contexte:** Cet appariement est utilisé dans systèmes de recommandation sensibles au contexte pour former les ensembles de contextes par exemple.

**contenu/contenu:** Cet appariement sert à trouver les produits similaires. Cette similarité peut par exemple, prendre en compte la structure des concepts, les votes ou bien les ontologies qui décrivent les concepts.

Les algorithmes de matching répondent à des questions essentielles telles que: quels ont les utilisateurs similaires à un utilisateur donné? Quels contenus pourraient l'intéresser?

Nous avons développé un des quatre algorithmes de matching cités au dessus: le matching **profil/profil**. La similarité entre deux profils prend en considération deux aspects: la structure des deux profils et leurs préférences.

Dans cette partie, nous présenterons notre approche qui calcule la similarité entre deux utilisateurs en se basant sur leurs profils mais aussi celle qui compare un profil utilisateur à un contenu, qui est considérée comme étant un cas particulier de la première approche.

#### IV.1.2 .1 Similarité entre deux profils (match profil/profil)

Aussi appelé le filtrage collaboratif, l'algorithme de similarité est appliqué sur un ensemble de profils et retourne une liste des utilisateurs jugés les plus similaires à un utilisateur donné. Soient P1, P2 deux profils utilisateurs (Figure IV.4). Les apparier revient à calculer la similarité entre ces deux profils en prenant en considération la structure du profil et les préférences des utilisateurs.

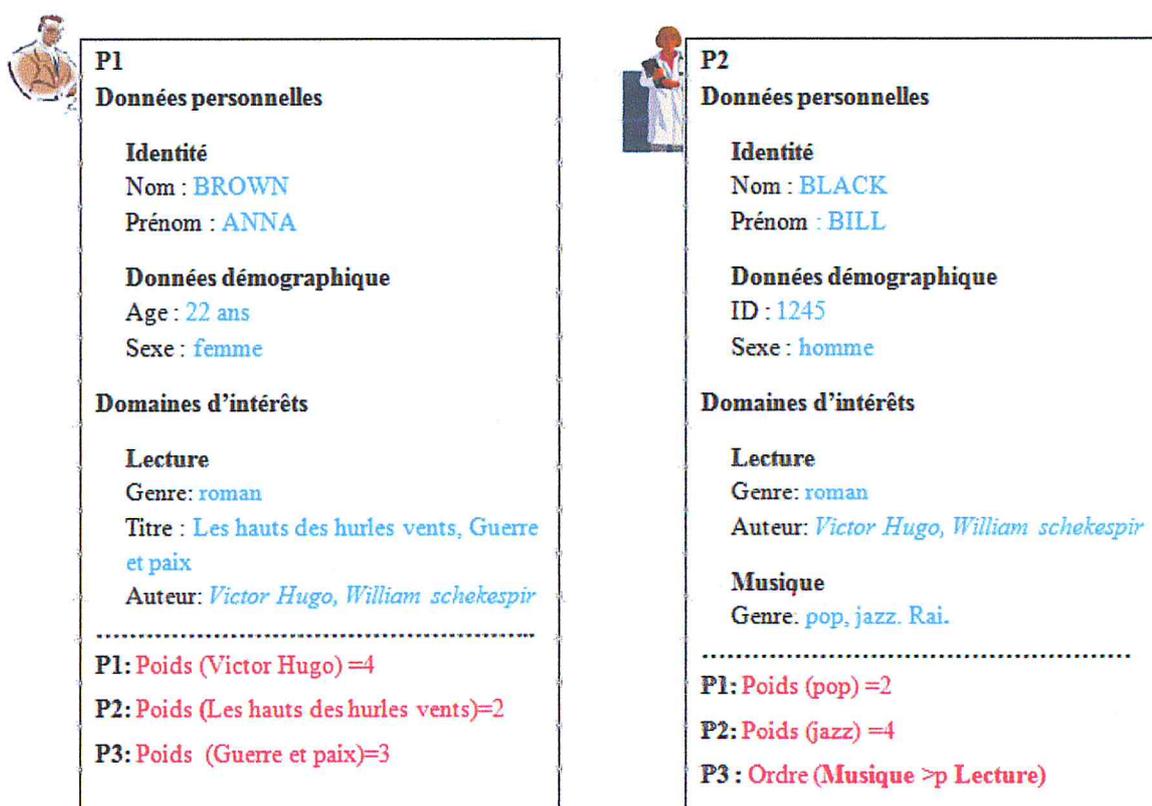


Figure N° IV.4 : Exemple de profils utilisateurs

On peut voir le matching comme une fonction qui attribut à deux similarités (structurelle et des préférences), et nous informe sur la similarité entre les deux utilisateurs P1 et P2.

Matching:  $P_u \times P_u \rightarrow \mathbb{R}$

$$(P1, P2) \rightarrow \alpha = \{\text{Sim}_{\text{Struct}}(P1, P2), \text{Sim}_{\text{Préf}}(P1, P2)\}$$

La fonction d'agrégation G peut s'agir de la moyenne simple:

$$G = 1/2 (\text{Sim}_{\text{Struct}}(P1, P2) + \text{Sim}_{\text{Préf}}(P1, P2))$$

Ou bien de la moyenne pondérée, en attribuant un poids  $\in [0,5]$ , à chaque similarité, selon son importance.

$$G = 1/2 (a \times \text{Sim}_{\text{Struct}}(P1, P2) + b \times \text{Sim}_{\text{Préf}}(P1, P2))$$

Dans ce qui suit, nous allons présenter l'algorithme de matching et sa méthodologie pour l'appariement des préférences du profil, car dans notre cas, les utilisateurs ont la même structure de profil.

#### ➤ Appariement par rapport aux préférences

Dans notre cas, nous nous référons aux données quantitatives pour effectuer l'appariement entre deux profils utilisateurs. Les préférences sont représentées par un vecteur de mots clés qui représentent les concepts et leur poids respectif qui représente l'importance d'un concept pour l'utilisateur.

Soient un ensemble  $P_u$  des préférences distinctes du profil utilisateur et ND l'ordre des préférences qui apparaissent dans ce profil. Une représentation vectorielle du profil  $P_u$  est un vecteur réel.

Calculer l'appariement entre deux préférences revient à calculer leur similarité avec des métriques telles que similarité Cosinus et la corrélation de Pearson. Avant de donner les algorithmes, rappelons la définition de ces métriques.

- **Similarité Cosinus**

La similarité cosinus (ou mesure cosinus) permet de calculer la similarité entre deux vecteurs à n dimensions en déterminant l'angle entre eux.

Soit deux vecteurs A et B, l'angle  $\theta$  s'obtient par le produit scalaire et la norme des vecteurs :

$$\theta = \arccos \frac{A \cdot B}{\|A\| \cdot \|B\|} \quad (1)$$

Comme l'angle  $\theta$  est compris dans l'intervalle  $[0, \pi]$ , la valeur  $\pi$  indiquera des vecteurs résolument opposés,  $\pi / 2$  des vecteurs indépendants (orthogonaux) et 0 des vecteurs colinéaires. Les valeurs intermédiaires permettent d'évaluer le degré de similarité.

Cette métrique se généralise dans un espace de dimension 2 à L :

$$V = (a_1, \dots, a_j, \dots, a_L) \text{ et } V' = (b_1, \dots, b_j, \dots, b_L).$$

$$Sim_{\cosinus}(V, V') = \frac{\sum_{j=1}^i a_j b_j}{\sqrt{\sum_{j=1}^i a_j^2} \sqrt{\sum_{j=1}^i b_j^2}} \quad (2)$$

- **Corrélation de Pearson**

Calculer le coefficient de corrélation entre 2 variables numériques revient à chercher à résumer la liaison qui existe entre les variables à l'aide d'une droite. On parle alors d'un ajustement linéaire.

Pour calculer le coefficient de corrélation entre deux séries, on suppose qu'on a les tableaux de valeurs suivants :  $V = (a_1, \dots, a_j, \dots, a_L)$  et  $V' = (b_1, \dots, b_j, \dots, b_L)$  pour chacune des deux séries. Alors, pour connaître le coefficient de corrélation liant ces deux séries, on applique la formule suivante :

$$Sim_{pearson}(V, V') = \frac{\sum_{j=1}^i (a_j - \bar{a})(b_j - \bar{b})}{\sqrt{\sum_{j=1}^i (a_j - \bar{a})^2 (b_j - \bar{b})^2}} \quad (3)$$

Deux cas se présentent: cas où les vecteurs sont homogènes et le cas où les vecteurs ne sont pas homogènes. On dit que deux vecteurs  $V, V'$  sont homogènes si et seulement si:

- $|V| = |V'| = N$ ,  $V$  et  $V'$  ont la même cardinalité.
  - $\forall i, 1 \leq i \leq N$ ,  $\text{prédicat}(pv[i]) = \text{prédicat}(pv'[i])$ .
- (4)

Dans le cas où les vecteurs ne sont pas homogènes, on peut prendre en considération que les concepts en commun. Comme on le remarque dans l'algorithme IV.5, on cherche à trouver les concepts qui sont en commun avec les deux ensembles de préférences (ligne2), puis on réorganise la liste de façon à ce qu'un concept dans le vecteur  $V$  correspond au même concept présent dans le vecteur  $V'$ , comme par exemple, faire une projection dans une base de données (ligne3). Enfin, on applique la métrique de similarité choisie (ligne4).

*En entrée :  $V, V'$  deux vecteurs de concepts avec leurs poids correspondant*

*En sortie : Similarité entre les deux vecteurs*

*1 :  $S1 \leftarrow \text{Concept}(V), S2 \leftarrow \text{Concept}(V')$*

*2 :  $S \leftarrow S1 \cap S2$*

*3 :  $V1 \leftarrow \Pi(V), V2 \leftarrow \Pi(V')$  avec  $\Pi$  représente la reconstitution des vecteurs*

*4 : Calculer  $\text{similarité} = \text{Vect}_{sim}(V, V')$ .*

*5 : Retourner similarité*

**Figure N° IV.5 : Algorithme de Matching**

Pour le calcul de la similarité, On a choisi la similarité cosinus parce que c'est la plus simple et la plus facile à implémenter que la corrélation de Pearson.

#### **IV.1.3 Sous système de la mise en œuvre de la personnalisation**

##### **IV.1.3.1 L'enrichissement des requêtes**

Ce service consiste à enrichir la requête de l'utilisateur par un ensemble de prédicats contenus dans son profil. L'enrichissement des requêtes exploite donc le profil utilisateur pour reformuler sa requête en y intégrant des éléments définis dans son profil (préférences, centre d'intérêt, etc.). La méthode la plus aboutie est celle de KOUTRIKA. Dans cette méthode, le profil utilisateur est composé d'un ensemble des prédicats pondérés. Le poids du prédicat exprime son intérêt relatif pour l'utilisateur. Il est spécifié par un réel compris entre 0 et 5.

Le processus d'enrichissement comporte deux phases principales :

- sélection des prédicats pertinents qui vont enrichir la requête.
- intégration de ces prédicats à la requête.

Dans ce qui suit, on résume les méthodes proposées dans les travaux de Koutrika et Ioannidis pour traiter ces deux étapes.

#### IV.1.3.1.1 Sélection des prédicats pertinents

La première étape du processus de personnalisation consiste à extraire les tops  $K$  préférences qui vérifient les deux propriétés suivantes :

propriété1 : la préférence est liée à la requête initiale.

propriété2 : la préférence n'est pas en conflit avec la requête.

- ✓ une requête est liée à des préférences, si elle peut être représentée comme étant un sous-ensemble de l'ensemble de profil. Ce sous-ensemble enferme tous les prédicats correspondant aux relations incluses dans la requête ainsi que tous les prédicats de sélection et de jointure correspondant à la requête. ~~Autrement dit, une préférence est liée à une requête, si elle trace un chemin qui s'attache à la requête.~~
- ✓ Une préférence est dite en **conflit** avec une requête, si elle est en conflit avec une condition qui existe déjà dans la requête.

#### IV.1.3.1.2 Intégration des préférences

Le but de cette étape est d'intégrer les top  $K$  préférences associées à la requête originale et de produire une nouvelle requête personnalisée qui donnera des résultats satisfaisants pour l'utilisateur, c'est-à-dire, ceux qui satisfait les  $M$  préférences parmi ces top  $K$  et au moins de  $L$  préférences restant du  $K - M$ . Afin d'atteindre ce but, on considère l'ensemble des  $K$  préférences et les  $M \in [0, K]$  premières préférences comme obligatoires indiquées explicitement à l'aide d'un critère d'intérêt. En outre, le nombre  $L \in [0, K - M]$ , indiqué explicitement à l'aide du degré d'intérêt minimum  $d$  pour chaque ligne dans les résultats (chaque ligne peut satisfaire une ou plusieurs préférences). Quand, on a récupéré les paramètres  $M$  et  $L$ , la requête personnalisée peut être construite selon deux approches différentes mais équivalentes.

**Note** :  $K$  : qui est le nombre de prédicats du profil devant être pris en compte;

$M$  : qui est le nombre de prédicats parmi les  $K$  qui doivent *obligatoirement* être satisfaits

$L$  : qui est le nombre minimal de prédicats parmi les  $K - M$  restants que chaque tuple du résultat doit satisfaire.

- *La première approche « une seule requête »* : elle est dénotée par SQ, elle établit une préférence qui est une conjonction de la conjonction des préférences obligatoires et de la disjonction de toutes les conjonctions possibles de L préférences parmi les K- M restant. La préférence est incluse dans la requête originale et toutes conditions répétées sont enlevées. En outre, toutes les variables additionnelles au tuple pour les conditions de préférence sont également incrustées à la requête.
- *La deuxième approche « requêtes multiples »* : elle est dénoté MQ, elle consiste à formuler un ensemble de K- M requêtes où chacune contient une préférence qui est une conjonction parmi les K-M. Chacune de ces requêtes est établie autour de la requête initiale prolongée avec la préférence correspondante ainsi que l'ensemble de nouvelles variables de tuples utilisées dans le dernier. Les résultats prévus sont obtenus en prenant l'union des résultats partiels, en la groupant par les attributs projetés de la requête initiale, et en excluant tous les groupes contenant moins que L lignes.

L'approche qu'on a utilisé pour la reformulation des requêtes est la première approche (SQ), ce choix est justifié par le fait qu'elle n'est pas compliqué et qu'elle est facile à mettre en œuvre. L'algorithme suivant illustre le principe de cette approche

*En entrée : profil P, requête initiale Q.*

*En sortie : requête personnalisée Q'.*

*1: Sélection des tops K prédicats pertinents.*

*2 :  $Q1 \leftarrow : Q \cup M$ .*

*3 :  $Q' \leftarrow : Q1 \cup L$ .*

*4 : Retourner la nouvelle requête Q'.*

Figure n° IV.6: Algorithme d'enrichissement des requêtes.

**Exemple 1 :**

```

Profil utilisateur _ p1 = {
Vente. Id_titre= Titres .Id_titre .....5 (a)
Vente. Id_auteur= Auteurs. Id_auteur.....5 (b)
Vente. Id_magasin = Magasins. Id_magasin..... 5 (c)
Titres. Titre= les misérables.....4 (d)
Titres. Type=roman.....3 (e)
Titres. Prix ≤ 1500 DA.....5 (f)
Titres. Nom_au= 'HUGO'..... 5 (g)
Titres. Prénom_au='VICTOR'.....5 (h)
Magasins .Nom_mag=' liberté '.....2 (i)
Magasins .cité = 'alger'.....1 (j)
}

```

Pour illustrer ce principe, considérons la requête Q1 de l'exemple 1 et fixons les paramètres K, L et M respectivement à 5, 2 et 2 sur le profil P1 de l'exemple 1. Le premier pas de l'algorithme va sélectionner les 5 prédicats de plus grand poids de P1 qui ne sont pas contradictoires avec ceux de la requête ; ce qui exclut le prédicat (f). Sur les prédicats sélectionnés {d, e, g, h, i}, les deux premiers (d et e) sont obligatoires et sont ajoutés à la requête Q1 qui devient Q'1.

**Requête initiale Q1**

```

Q1 = SELECT Id_titre, Id_auteur, Titre, Type, cité
FROM Vente V, Titres T, Auteurs A, Magasins M
WHERE V.Id_titre= T .Id_titre AND V. Id_auteur= A. Id_auteur AND
V. Id_magasin = M. Id_magasin AND prix='1700';

```

**Requête initiale enrichie avec les prédicats obligatoires Q'1**

```

Q'1 = SELECT Id_titre, Id_auteur, Titre, Type, cité
FROM Vente V, Titres T, Auteurs A, Magasins M
WHERE V.Id_titre= T .Id_titre AND V. Id_auteur= A. Id_auteur AND
V. Id_magasin = M. Id_magasin AND prix='1700' AND Titre= ' les misérables'
AND Type= 'roman';

```

Finalement, la disjonction des conjonctions de deux prédicats parmi les prédicats non obligatoires (g, h et i) est ajoutée à la requête; ce qui donne la requête enrichie Q1+.

**Requête initiale enrichie Q1+**

```

Q1+ = SELECT Id_titre, Id_auteur, Titre, Type, cité
FROM Vente V, Titres T, Auteurs A, Magasins M
WHERE V.Id_titre= T.Id_titre AND V. Id_auteur= A. Id_auteur AND
V. Id_magasin = M. Id_magasin AND prix='1700' AND Titre= ' les misérables'
AND Type= 'roman';
AND ((Nom_au= 'HUGO' AND Prénom_au='VICTOR')
OR (Nom_au= 'HUGO' AND Nom_mag=' liberté ')
OR (Prénom_au='VICTOR' AND Nom_mag=' liberté ' ));
    
```

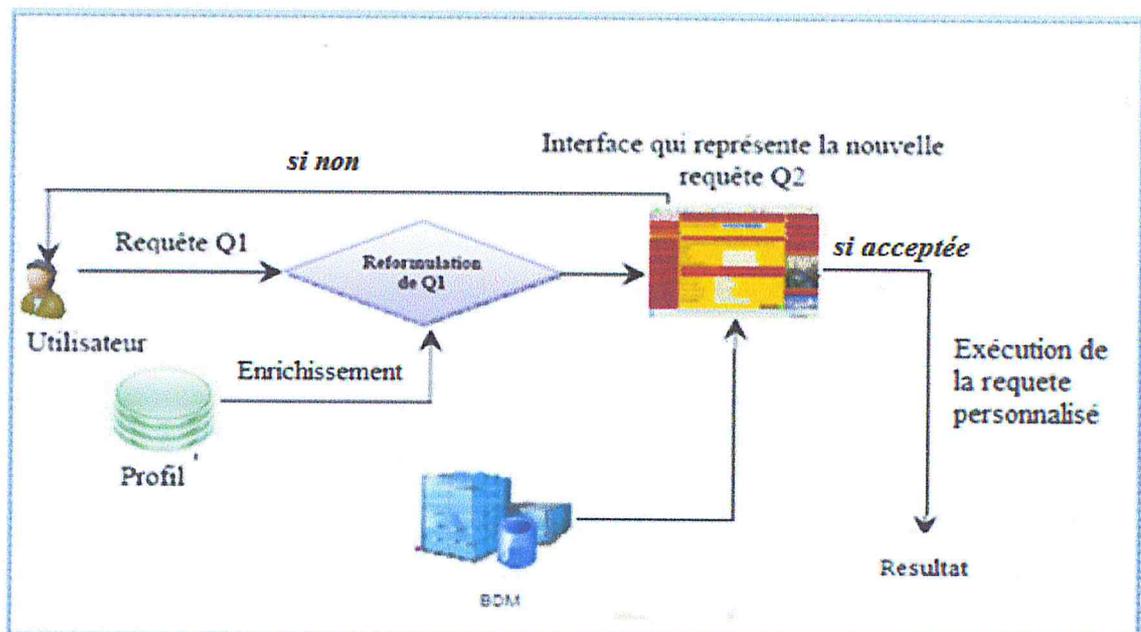


Figure n° IV.7 : Sous système de traitement de la requête.

L'utilisateur envoie sa requête Q1, elle va être enrichie (selon l'approche SQ) par des prédicats contenant dans son profil, ces prédicats sont incluse dans la requête originale sous forme de conditions (la clause WHERE), la nouvelle requête Q2 personnalisée s'affiche à l'utilisateur, si l'utilisateur accepte cette requête, elle s'exécute sur l'entrepôt de données et l'utilisateur va avoir un résultat pertinent selon ses préférences, si non le processus sera répété jusqu'à la satisfaction de l'utilisateur ou bien l'annulation du processus.

Pour l'exécution des requêtes et l'implémentation de notre prototype on va utiliser l'entrepôt de données qui va être décrit dans la section suivante :

## V. Présentation de l'entrepôt a utilisé

Nous pouvons modéliser les données de ventes en utilisant un cube qui est décrit sous forme de schéma en étoile par la figure IV.7.

Ce cube, nommé 'ventes', est composé d'une table de faits 'Ventes' et de trois dimensions qui sont 'Livre', 'Auteurs', et 'magasins'. Par exemple, pour une chaîne de magasins Chaque cellule de ce cube stocke la valeur d'une mesure donnée, dont cette mesure est la quantité d'un livre vendu dans un magasin donné (représenté par sa ville) et pour un auteur particulier.

**La table Ventes** contient les informations sur les ventes : livre vendu (id\_livre), point de vente (id\_magasin), auteur de ce titre (id\_auteur) et le chiffre d'affaire qui se calcule comme suit :  $\sum(Qty * Price)$ .

- Price : prix unitaire du titre.

- Qty : quantité vendue.

**La table Auteurs** : c'est une dimension qui contient toutes les informations concernant les auteurs qui existent.

**La table Magasins** : c'est la dimension qui représente le point de vente des titres.

**La table Livre** : c'est une table où on peut trouver les informations des titres.

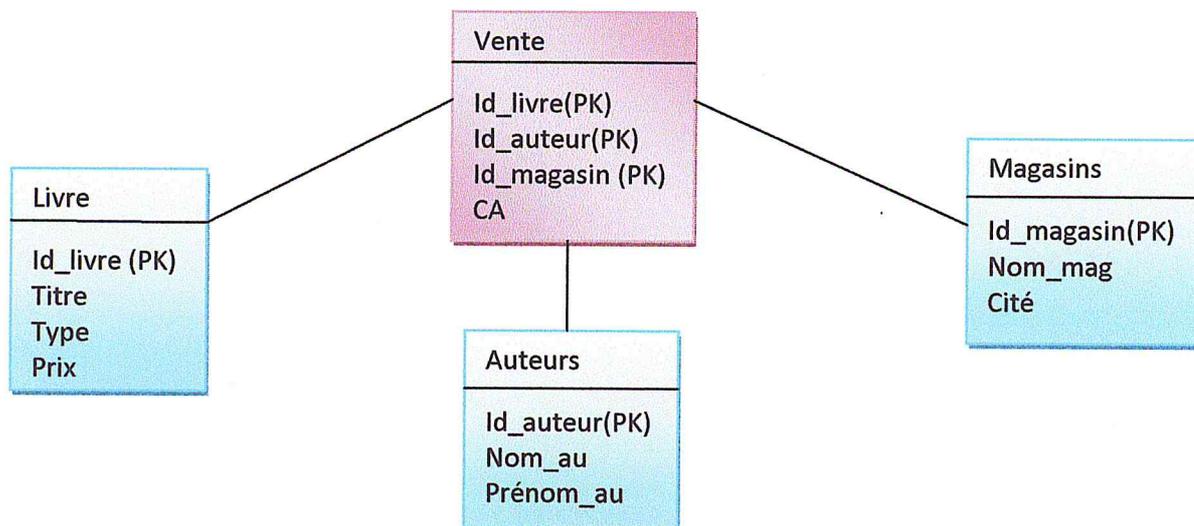


Figure n° IV.8 : Entrepôt de données à utiliser

## VI. Conclusion

Sur le plan théorique des arguments peuvent être avancés pour justifier le choix des différentes dimensions pour la modélisation du profil ainsi que leurs impacts sur les performances du système, mais seule une évaluation basée sur des tests réels, nous permettra de mieux cerner les mérites d'un tel choix pour l'amélioration du processus de la personnalisation. C'est un travail de longue haleine et exige un ajustement minutieux des différents paramètres du système. Toutefois, tout porte à croire que plus sont diversifiées les informations qui caractérisent le profil utilisateur et mieux sera la qualité de la personnalisation, puisque le système opère une multitude de traitements qui lui permettent d'avoir une meilleure visibilité de ce que veut réellement l'utilisateur comme informations.

Par ailleurs cette modélisation est flexible et évolutive pour admettre des améliorations futures, le profil ainsi modélisé est une entité dynamique à l'intérieur du système, transitant de phase en phase traduisant l'aspect changeant des centres d'intérêts de l'utilisateur. Dès lors la mise en place d'un mécanisme pour veiller à la maintenance continue du profil dès son intégration jusqu'au moment où il est volontairement supprimé du système, devient plus que nécessaire.

# Réalisation

## I. Introduction

Au cours de la partie précédente de ce mémoire, nous avons conçu et détaillé notre système ainsi que sa logique de fonctionnement. Nous avons donc décrit le module de gestion de profil utilisateur ainsi que son intégration dans le module de la personnalisation.

Afin de concrétiser cette conception, nous passons à la dernière partie du développement où nous allons présenter la réalisation de notre système.

## II. Architecture matérielle

### II.1 Schéma « un tiers »

Dans une application un tier (de l'anglais *tier* signifiant étage ou niveau), les trois couches applicatives sont intimement liées et s'exécutent sur le même ordinateur. On ne parle pas ici d'architecture client-serveur, mais d'informatique centralisée.

### II.2 Schéma « 2 tiers »

Dans une architecture deux tiers, encore appelée client/serveur de première génération ou client-serveur de données le poste client se contente de déléguer la gestion des données à un service spécialisé. C'est un modèle d'architecture applicative où les programmes sont répartis entre clients et serveurs communiquant par des requêtes avec réponses. La gestion des données est prise en charge par un SGBD centralisé, s'exécutant le plus souvent sur un serveur.

- Le client : c'est le programme qui provoque le dialogue.
- Le serveur : c'est le programme qui se contente de répondre au client.

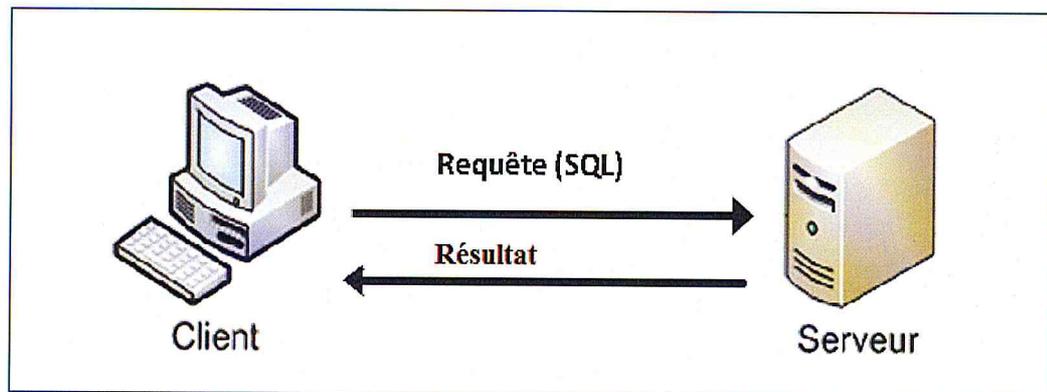


Figure n° II.1 : Schéma « 2 tiers » (Client/Serveur)

### III. Environnement de développement

#### III.1 Description du système de gestion de données

Concernant le SGBD que nous avons adopté pour notre projet (sachant que les requêtes sont formulées sur un entrepôt de données et une base de données), nous avons choisi le SQL Server version 2000 édition entreprise.

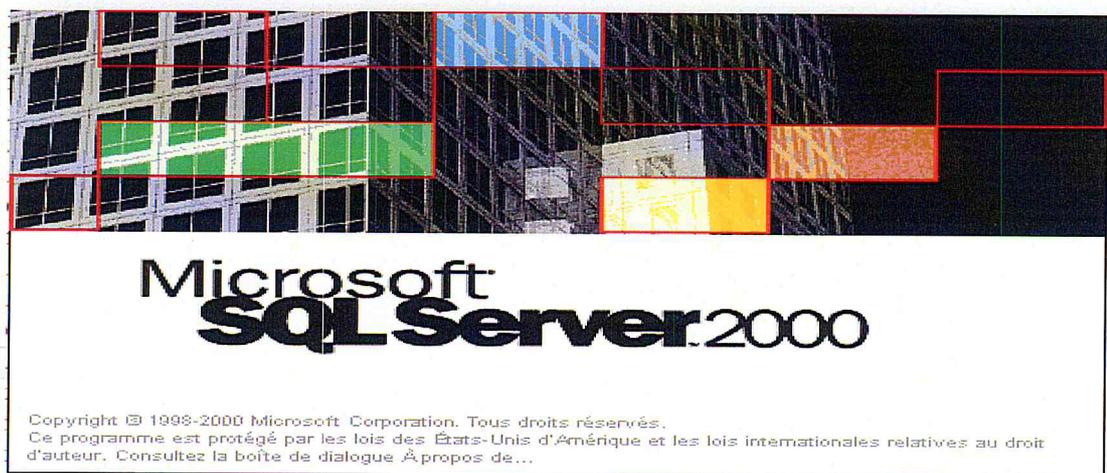


Figure n° III.1: le SGBD SQL Server 2005

Les raisons ayant motivé ce choix sont :

- une meilleure intégration à la plate-forme Microsoft.
- Il constitue une excellente plate-forme de gestion de base de données pour applications de commerce électronique, Data Warehousing et traitement transactionnel en ligne (OLTP, Online Transactional Processing) à grande échelle.
- Toutes les fonctionnalités indispensables pour les environnements les plus exigeants.
- production et gestion de cubes OLAP, datamart et datawarehouse (Analysis Services)
- Un jeu complet d'outils graphiques et d'utilitaires d'analyse et d'exploration de données.
- Sécurité et fiabilité.
- Échanges de données avec Oracle, MySQL...

Nous avons utilisé l'environnement Business Intelligence Development Studio 2000 qui est un environnement de développement intégré pour les solutions Analysis Services, Reporting Services et Integration Services.

### **III.2 Choix du langage de programmation**

Notre choix du langage de programmation s'est porté sur le langage JAVA et cela pour diverses raisons :

- JAVA est un langage orienté objet simple ce qui réduit les risques d'incohérence.
- JAVA est portable, Il peut être utilisé sous Windows, sur Macintosh et sur d'autres plates formes sans aucune modification. JAVA est donc un langage multiplateforme, ce qui permet aux développeurs d'écrire un code qu'ils peuvent exécuter dans tous les environnements.

- JAVA possède une riche bibliothèque de classes comprenant des fonctions diverses telles que les fonctions standards, le système de gestion de fichiers, les fonctions multimédia et beaucoup d'autres fonctionnalités.

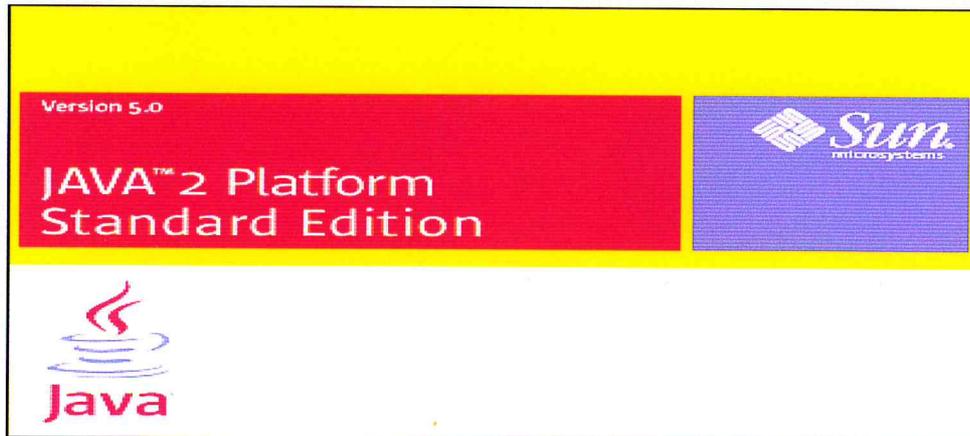


Figure n° III.2: La plateforme JAVA 2

Nous avons conçu notre application java en utilisant l'environnement de développement Eclipse 3.2 qui est un environnement de développement intégré. Eclipse offre de nombreuses fonctionnalités. Il comporte notamment une série d'outils permettant la conception, le développement, l'exécution, le débogage et le déploiement des applications JAVA.

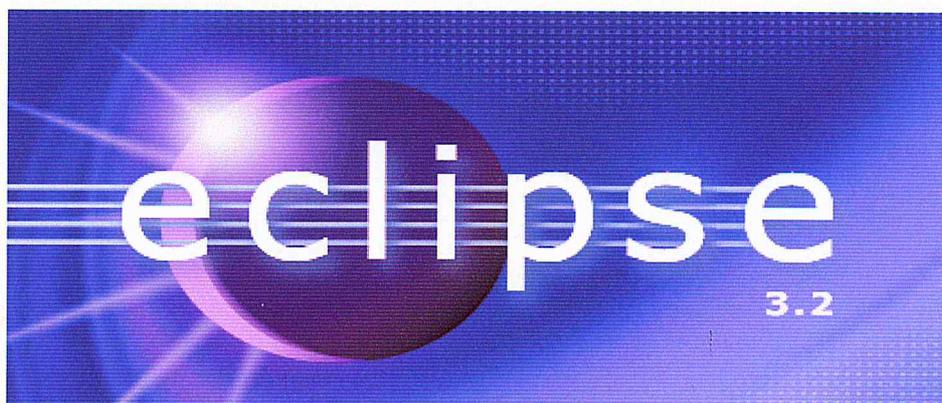


Figure n° III.3: Environnement de développement éclipse 3.2

## IV. Interface de test

Dans le cadre de nos travaux de recherche sur les systèmes de personnalisation, nous présenterons quelques interfaces de notre application afin de donner une vue générale sur les fonctionnalités de cette dernière.

### IV. 1 Interface d'authentification

L'utilisateur qui accède à l'application à partir d'un PC click simplement sur la fenêtre d'exécution. Une première interface lui sera affichée. Cette interface (Figure IV.1), contient une brève présentation du logiciel et permet à l'utilisateur d'introduire un nom d'utilisateur (utilisateur) et un mot de passe (Mot de passe). Ces derniers doivent être valides pour qu'il (l'utilisateur) puisse utiliser le système. Et aussi, cette fenêtre contient un bouton (Créer un compte) qui permet à un utilisateur non inscrit de créer un compte.



Figure n° IV.1 : Interface d'authentification

## IV. 2. Interface de création d'un nouveau compte

Afin qu'un nouvel utilisateur puisse accéder au système, il doit créer un compte en cliquant sur le bouton « Créer un compte » dans la page d'accueil. Le système affiche une nouvelle fenêtre (Figure n° IV.2) qui est sous la forme d'un formulaire contenant toutes les informations personnelles de l'utilisateur.

Créer un compte

Créer un nouveau compte

Identifiant :  \*

Mot de passe :  \*

Nom :  \*

Prénom :  \*

Age :  \*

Profession :  \*

Adresse :  \*

N° Telephone :  \*

Annuler Enregistrer

Figure n° IV.2 : Introduction de données personnelles

### IV. 3. Interface d'accueil

Une fois inscrit au système une fenêtre sera affichée (Figure n° IV.3), l'utilisateur à la possibilité de visualiser le résultat de ses préférences en choisissant le bouton « Afficher ». Cette fenêtre permet, également, à l'utilisateur de modifier ses préférences en cliquant sur le bouton « Modifier ».

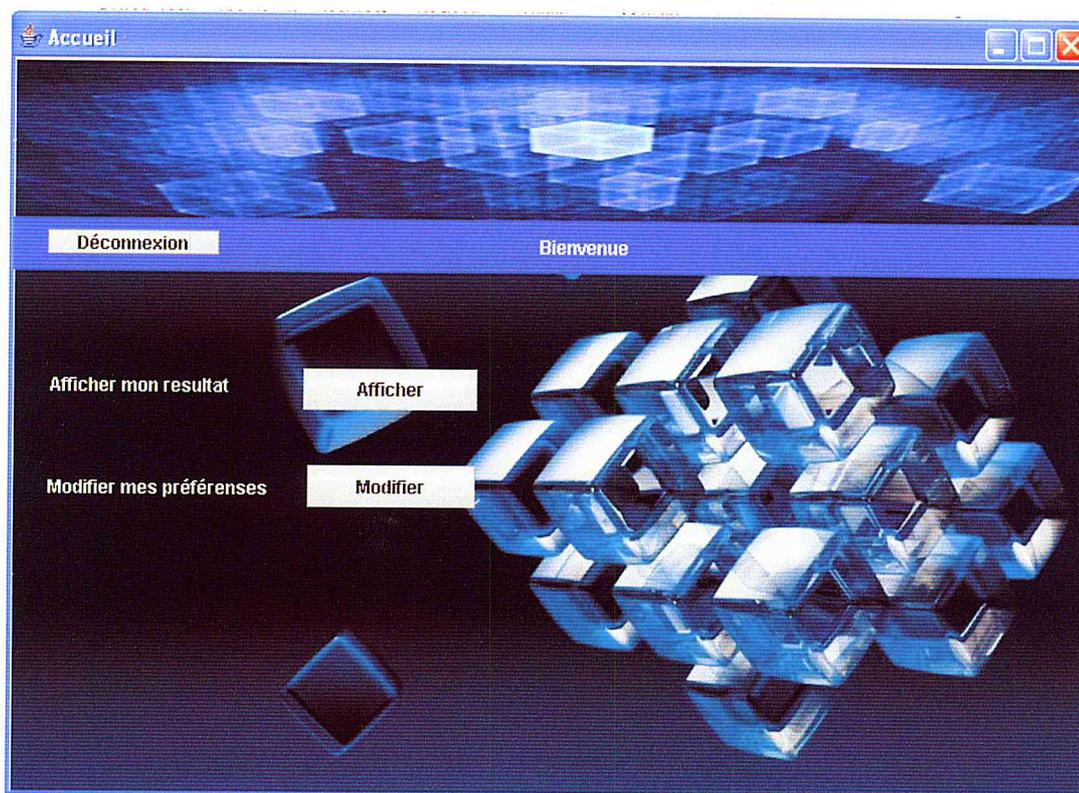


Figure n° IV.3: Interface d'accueil

#### IV. 4. Interface de gestion des préférences

En bref, cette fenêtre (Figure n° IV.4) présente à l'utilisateur les différentes dimensions, les sous dimensions et les membres qu'il devra noter selon ses préférences. La note est comprise entre 0 et 5 où la note minimale étant 0 et la note maximale est 5. Si l'utilisateur ne précise pas son degré d'intérêt ca veut dire qu'il n'a pas de préférence sur ces derniers, il suffit juste de cliquer sur le bouton « Valider ». Ce bouton se charge aussi de la sauvegarde de ses préférences.

The screenshot shows a window titled 'Préférence' with a blue background and a 3D cube graphic. The window is divided into three main sections: 'Préférences sur les dimensions', 'Préférences sur les niveaux', and 'Préférences sur les membres'. Each section contains various input fields and checkboxes for user preferences.

Section	Item	Value / Option
Préférences sur les dimensions	Table Livre	3
	Table auteur:	5
	Table magasin	4
Préférences sur les niveaux	<input checked="" type="checkbox"/> Titre	5
	<input checked="" type="checkbox"/> Type	4
	<input checked="" type="checkbox"/> Prix	3
	<input type="checkbox"/> Nom auteur	3
	<input type="checkbox"/> Prenom auteur	2
	<input checked="" type="checkbox"/> Nom magasin	5
Préférences sur les membres	popular_comp	
	Bookbeat	
	Portland	

Buttons: Valider, Annuler

Figure n° IV.4: Interface de gestion des préférences

#### IV.5. Interface de d'appariement

Cette fenêtre permet à l'utilisateur d'exécuter sa requête personnalisée ou bien d'exécuter la requête d'un profil similaire en cliquant sur le bouton « exécuter ».

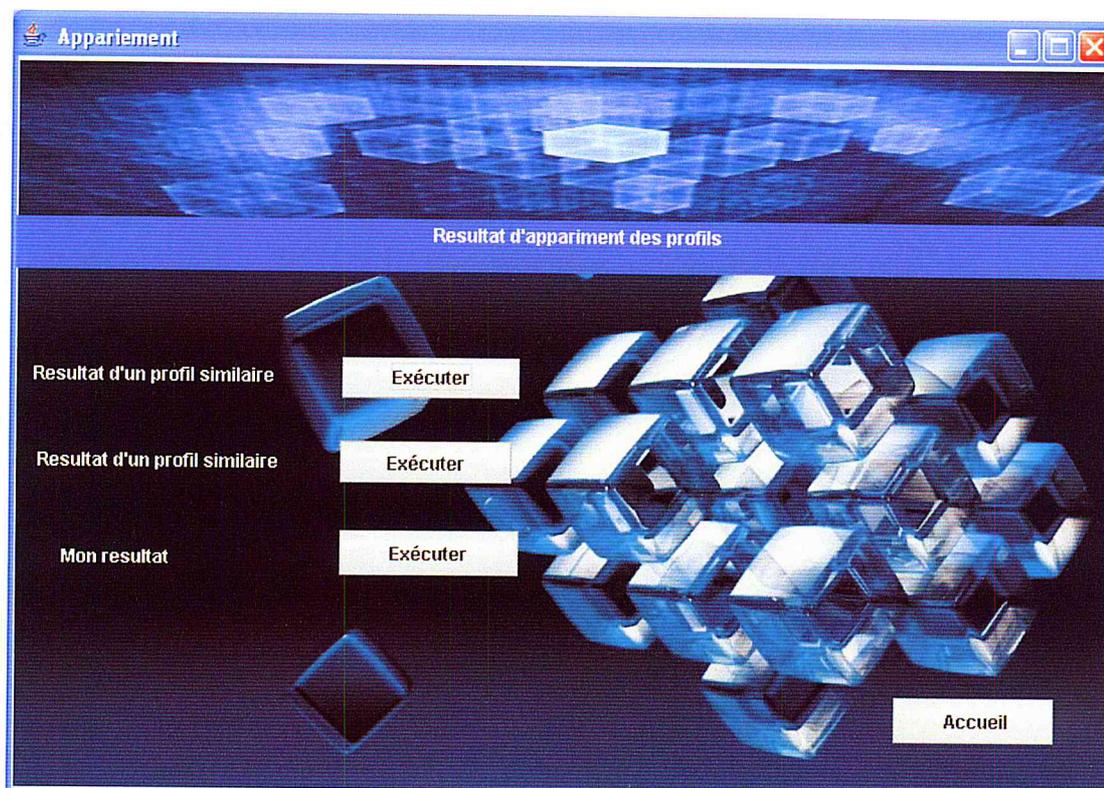


Figure n° IV.5: Interface d'appariement

#### IV.6. Interface d'affichage de résultats

Après avoir exécuté la requête de l'utilisateur, le système va lui afficher les résultats de l'exécution de cette requête, une interface est développée pour assurer cet affichage est représentée dans la Figure n° IV.6. Si l'utilisateur veut exécuter la requête d'un profil similaire la Figure n° IV.7 sera affichée.

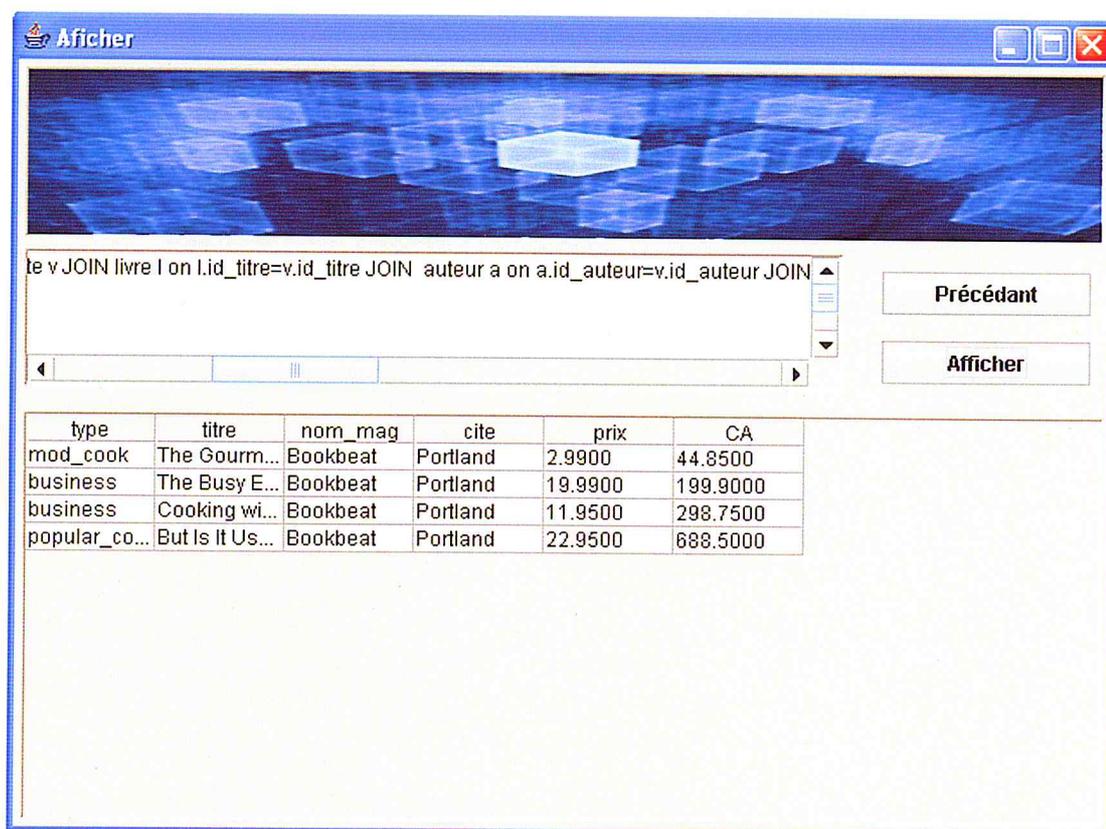
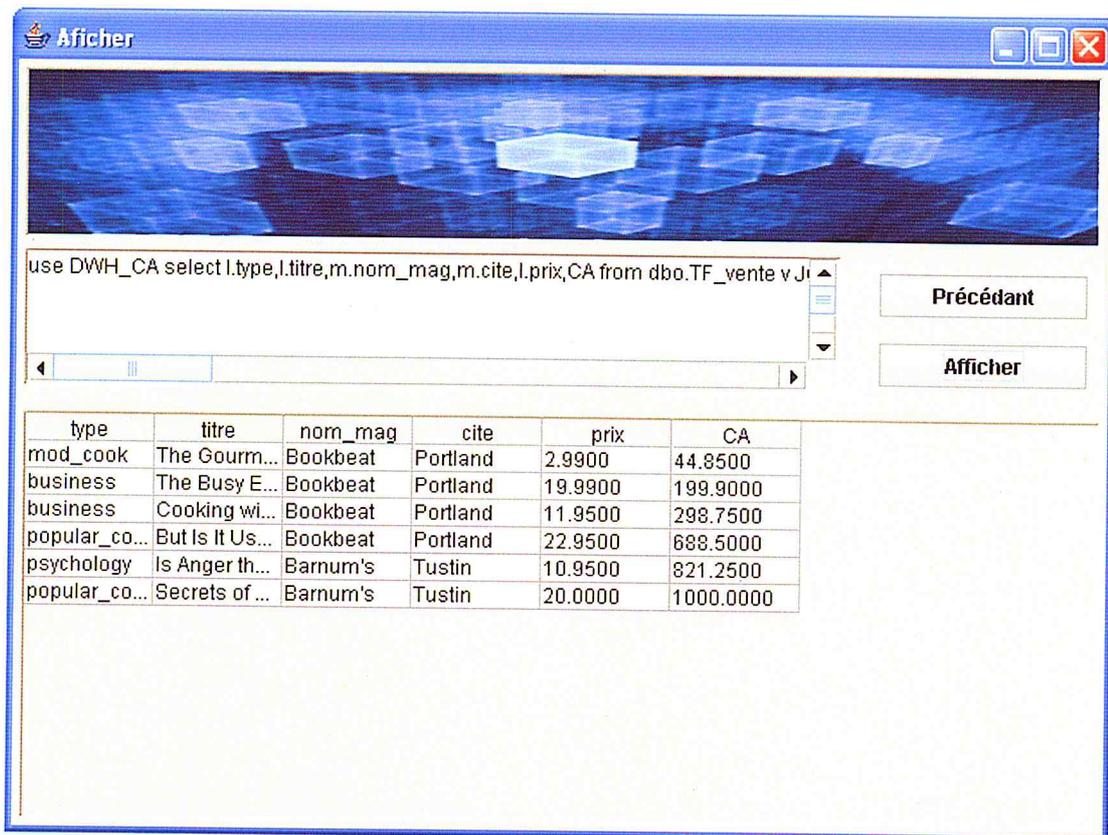


Figure n°IV.6: Interface d'affichage de résultats



The screenshot shows a window titled "Afficher" with a blue header. Below the header is a decorative blue background with a central glowing hexagon. A text area contains the SQL query: "use DWH\_CA select l.type,l.titre,m.nom\_mag,m.cite,l.prix,CA from dbo.TF\_vente v Ji". To the right of the text area are two buttons: "Précédant" and "Afficher". Below the text area is a table with the following data:

type	titre	nom_mag	cite	prix	CA
mod_cook	The Gourm...	Bookbeat	Portland	2.9900	44.8500
business	The Busy E...	Bookbeat	Portland	19.9900	199.9000
business	Cooking wi...	Bookbeat	Portland	11.9500	298.7500
popular_co...	But Is It Us...	Bookbeat	Portland	22.9500	688.5000
psychology	Is Anger th...	Barnum's	Tustin	10.9500	821.2500
popular_co...	Secrets of ...	Barnum's	Tustin	20.0000	1000.0000

**Figure n°IV.7:** Interface d'affichage de résultats d'un profil similaire

## V. Conclusion

Dans ce chapitre nous avons présenté en premier lieu les choix effectués pour les différents outils de développement ainsi nous avons montré les interfaces d'application, celles qu'on avait utilisées pour effectuer les tests sur l'approche de la personnalisation. Nous avons montré à travers ces évaluations que le profil est important pour effectuer la personnalisation et qu'il améliore les systèmes d'accès personnalisé à l'information en fournissant à l'utilisateur des résultats pertinents selon leurs besoins précis.

## **Conclusion et perspectives**

Ce travail a permis de contribuer à l'implantation d'un prototype de systèmes d'accès personnalisé à l'information. Il avait pour objectif la définition d'un modèle flexible de profil utilisateur pour un OLAP personnalisé.

Le but de la personnalisation est de faciliter l'expression du besoin d'utilisateur et de lui permettre d'obtenir des informations pertinentes lors de ses accès à un système d'information.

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.

La réalisation pratique de notre objectif s'est fait en deux étapes. La première étape consistait à la définition d'un modèle de profil utilisateur et la deuxième étape consistait à l'intégration de ce modèle dans un système d'accès personnalisé à l'information, de façon à obtenir des informations pertinentes selon les besoins de chaque utilisateur.

L'intégration du profil dans les systèmes de personnalisation, permet d'améliorer les performances de ces derniers.

Rappelons que nous avons utilisé l'approche d'enrichissement pour la reformulation des requêtes, l'approche explicite pour l'acquisition des données des utilisateurs et l'approche multidimensionnelles pour la représentation de profil utilisateur et on a utilisé l'appariement pour la comparaison des profils à l'aide d'un algorithme du matching avec similarité du Cosinus.

Finalement, le résultat du test effectué sur notre prototype, en vue de confirmer son efficacité, et en utilisant l'approche de personnalisation proposée, qui est basée sur l'enrichissement des requêtes, a montré l'importance du profil dans la personnalisation de l'information, dont il apporte une amélioration des performances de recherche par rapport au modèle d'accès à l'information classique.

Les résultats obtenus sont encourageants et ouvrent des perspectives intéressantes, que nous présentons dans ce qui suit :

## **Perspectives**

Nos propositions ouvrent plusieurs perspectives de recherche qui permettent d'améliorer en générale la personnalisation de l'accès à l'information dans les systèmes d'information et en particulier le fonctionnement de notre approche de personnalisation :

### **1) *Mise à jour automatique des profils***

La mise à jour est réalisée en partie manuellement dans notre système de personnalisation. Or, il existe des techniques utilisant des méthodes d'apprentissage automatiques qui permettent de mettre à jour automatiquement les profils, Ces techniques sont basé sur le principe d'observation du comportement de l'utilisateur lorsqu'il utilise le système.

### **2) *Amélioration de la personnalisation de la présentation***

La personnalisation de la présentation dans notre système dépend de la personnalisation du contenu. Cette personnalisation concerne essentiellement la présentation des informations personnalisées dans l'interface de navigation. Aucune adaptation, proprement dite, de l'interface n'est prévue dan *ce* système. Donc il est intéressant d'enrichir et d'approfondir cette personnalisation des outils qui permettent de personnaliser également l'interface (mise en forme du texte, organiser le positionnement des objets de l'interface, la structure de l'interface, etc.).

### **3) *Evaluation des performances***

Faire des tests pour savoir l'impact de la structure du profil sur les performances de système, c'est-à-dire valider notre système par des données réelles.

### **4) *Acquisition des données des profils***

Définition d'un prototype qui permet l'acquisition implicite des données des utilisateurs à travers l'analyse des historiques de leurs interactions avec le système en utilisant par exemple un fichier log.

### Bibliographie

- [AI MAKSSOUD, 08] Karine Al Makssoud épouse ABBAS « Système d'Accès Personnalisé à l'Information : Application au Domaine Médical », Institut National des Sciences Appliquées de Lyon, thèse doctorat, 9 décembre 2008.
- [BELLATRECHE et ALL, 05] L. Bellatreche, A. Giacometti, P. Marcel, H. Mouloudi, and D.Laurent. "A Personalization Framework for OLAP Queries". In 8th ACM International Workshop on Data Warehousing and OLAP (DOLAP 05), Bremen, Germany. ACM Press, 2005.
- [BELLATRECHE et AL, 06] L. Bellatreche, A. Giacometti, P. Marcel, and H. Mouloudi. "Personalization of MDX Queries", Dans la 22èmes journées Bases de Données Avancées (BDA 06), Lil, France, 2006.
- [BOUKHALFA, 02] Kamel BOUKHALFA « De la conception physique aux outils d'administration et de tuning des entrepôts de données », ESI, 2002.
- [BOULKRINAT, 07] Samia BOULKRINAT « Modélisation hybride du profil utilisateur pour un système de filtrage d'informations sur le WEB », 2006 / 2007.
- [BOUZEGHOUB ET KOSTADINOV, 05] Mokrane Bouzeghoub, Dimitre Kostadinov « une approche multidimensionnelle pour la personnalisation de l'information », université de versailles 45, avenue des Etats-Unis, 78035 Versailles, 2005.
- [BOUZIDI et SADKI, 10] BOUZIDI EL Hocine et SADKI Omar, « Conception et implémentation d'un entrepôt de données et application des techniques du Data Mining pour l'analyse des échecs/succès des étudiants de l'ESI », ESI, 2010.
- [BRADLEY et AL, 00] K. Bradley, R. Rafter, and B. Smyth. «Case-Based User Profiling for Content Personalisation ». Trento, Italy, Springer, 2000.
- [CHERNIACK et AL, 03] M. Cherniack, E.F. Galvez, M.J. Franklin, and S.B. Zdonik. «Profile-Driven Cache Management», 2003.

- [DJEMAI et GHOUALI, 08]** DJEMAI Khadidja GHOUALI Khadra « Conception et réalisation d'un système personnalisé pour l'interrogation d'une base de données », ESI, 2007/2008.
- [EISA, 09]** Elsa NEGERE, « Exploration collaborative de cubes de données », 01 décembre 2009.
- [FAVRE, 07]** Cécile Favre, "Evolution de schémas dans les entrepôts de données mise à jour des hiérarchies de dimension pour la personnalisation des analyses". Thèse de doctorat, Université Lumière-Lyon2, Lyon, France, 2007.
- [FERREIRA ET SILVA, 01]** MySDI:A Generic Architecture to Develop SDI Personalised Services, J. Ferreira, A. Silva, ICEIS (1) 2001: 262-270.
- [HACIANE, 07]** Ahmed HACIANE, «Conception d'un datawarehouse Orienté CRM» ESI, 09/01/2007
- [Kim, 97]** Ralph Kimball: « Entrepôt de données », International Thomson Publishing France, 1997.
- [KOUTRIKA et IOANNIDIS, 04]** G. Koutrika and Y. Ioannidis. "Personalization of Queries in Database Systems», Boston, Massachusetts,USA , 2004.
- [KOSTADINOV,08]** Personnalisation de l'information : une approche de gestion de profils et de reformulation de requêtes, Dimitre Kostadinov Thèse de doctorat, Université de Versailles ,2008.
- [MEZIANI et TEBANI, 09]** MEZIANI Nadir, TEBANI Ali « Conception et réalisation d'un outil de personnalisation pour l'interrogation d'un entrepôt de données », ESI, 2009.
- [MOULOUDI, 07]** Hassina MOULOUDI, « Personnalisation de requêtes et visualisations OLAP sous contraintes », thèse pour obtenir le grade de docteur, École Doctorale : Santé, Sciences et Technologies, Université François Rabelais Tours, 2007.
- [PRETSCHNER ET GAUCH, 99]** A. Pretschner and S. Gauch. «Ontology Based PersonalizedSearch ». Chicago, Illinois, USA, 1999.

## Bibliographie

---

[RAVAT et AL, 08] F. Ravat, O. Teste, G. Zurfluh « Personnalisation de bases de données multidimensionnelles», Toulouse, 2008.

[TAMINE et Ali ,06] Lynda Tamine-Lechani, Nesrine Zemirli et Wahiba Bahsoun «Définition d'un profil d'un multidimensionnel de l'utilisateur »,Acte de Corea , Lion, France ,2006.

[ZEMIRLI, 08] WAHIBA NESRINE ZEMIRLI « Modèle d'accès personnalisé à l'information basé sur les Diagrammes d'Influence intégrant un profil utilisateur évolutif » thèse de doctorat, Université Paul Sabatier de Toulouse III, le 12 Juin 2008.

