

F.S.D N° D'ordre

Université Saad DAHLAB de Blida



Faculté des Sciences

Département d'Informatique

Mémoire présenter par :

**BRAHIMI Radia
MOUSSAOUI Hakima.**

En vue d'obtenir le diplôme de Master en Informatique

Option: Ingénierie des logiciels

Sujet :

**Génération automatique des contextes usuels et leurs
intégrations dans le processus de
personnalisation des requêtes de données multidimensionnell**

Encadré par :

M^{me} REZOUG Nachida

Soutenue le: 22/10/2013

devant le jury composé de :

Mr Sidoumou

M^{me} Zahra

M^{me} Laaroussi

Président

Examineur

Examineur

2012/2013

MA-004-181-1

Remerciement



Nous remercions avant tout Dieu tout puissant qui nous a donné la force, le courage et la volonté pour réaliser ce travail.

Nous tenons à remercier chaleureusement notre promotrice M^{me} REZOUG Nachida, d'avoir bien voulu diriger ce projet. Nous lui somme reconnaissants pour son appui et ses conseils, ainsi que pour sa disponibilité et la confiance qu'elle nous a accordée.

Nous remercions les membres de jury pour nous avoir fait l'honneur de juger notre travail.

Nous sommes reconnaissantes à tous nos enseignants qui nous ont facilité la compréhension et la maîtrise.

Nous remercions aussi tous nos amis et nos collègues de travail qui nous encouragée.

Nous tenons un grand merci à tous ceux qui de près ou de loin ont participé à l'aboutissement de ce travail.

MOUSSAOUI Hakima et BRAHIMI Radia.

Dédicaces

C'est avec un immense plaisir que je dédie ce travail

A mes chers parents,

A mes chers frères, ma sœur Mounira,

*Ma belle sœur Farida et mes nièces Arwa, Maria et
Marame et toute la famille Brahimi*

Ainsi que mes chères cousines Fouzia et Salima,

Tantes Tawesse et Zineb

*A tous mes amies et surtout Ichrak, Akila, Khadidja et
Imene et leurs familles.*

A mon binôme Hakima et à toute sa famille,

A tous mes amis

BRAHIMI RADIA.





Dédicace

*Louange à Dieu le tout puissant, je dédie ce modeste
Travail aux deux êtres les plus chers dans ma vie, pour
Tout ce qu'ils ont fait pour que je pusse réussir.*

A mes chers frères

Abdelkader, Mohamed et Ali

A mes chères sœurs

Fatiha, Faiza et Razika

A mes cousins et cousines

Akila, Kamel et Amina

Et à toute ma famille.

A tous mes camarades de ma promotion et tous mes amis et surtout

Mari, Imene, Hanane, Hayet, Souad, Anissa, Lamia

Et bien sur à mon binôme Radia

HAKIMA

Sommaire

Introduction général

Introduction générale.....	1
Problématique.....	2
Objectif.....	2

Chapitre 01 : Entrepôt de données et système OLAP

Introduction.....	3
I. System décisionnel.....	3
1. Définition.....	3
2. Processus de prise de décision.....	3
II. Entrepôt de données.....	4
1. Définition.....	4
2. Domaines d'utilisation des DW.....	5
3. Architecture d'un système décisionnel.....	5
4. Classes de données.....	6
4.1. Métadonnées.....	6
4.2. Datamart (magasin de données).....	6
4.3. Cubes de données.....	7
4.4. Données agrégées.....	7
4.5. Données détaillées.....	8
5. Différents modèles de données.....	8
5.1. Modèle en étoile (Star schema).....	8
5.2. Modèle en flocons (Snowflake schema).....	8
5.3. Modèle en constellation (Factflake schema).....	9
III. Système OLAP.....	10
1. Définition d'un système OLAP.....	10
2. OLAP pour quoi faire ?.....	11
3. OLAP pour qui ?.....	11
4. Approche multidimensionnelle dans le datawarehouse.....	12
4.1. Fait.....	12
4.2. Mesure.....	12
4.3. Dimension.....	12
5. Outils de navigation OLAP.....	13
6. Différentes implémentations OLAP.....	13
6.1. Relational OLAP(ROALP).....	13
6.2. Multidimensional OLAP (MOLAP).....	14
6.3. Hybride OLAP(HOLAP).....	14
6.4. Desktop OLAP(DOLAP).....	14
7. Requête OLAP.....	14
Conclusion.....	15

Chapitre 02 : Personnalisation dans les systèmes OLAP

Introduction.....	16
I. Notion de personnalisation	16
1. Définition.....	16
2. Domaines intéressés par la personnalisation	17
3. Profil utilisateur	18
3.1. Notion de profil.....	18
3.2. Exemples de profils	19
3.3. Modélisation de profil.....	22
3.4. Contenu de profil.....	23
3.5. Construction de profil	23
3.6. Exploitation de profil	24
3.7. Evolution de profil utilisateur	24
4. Préférences utilisateur.....	25
4.1. Notion de préférence	25
4.2. Niveau des préférences	25
4.3. Expression des préférences	26
II. Personnalisation des systèmes OLAP	26
1. Personnalisation de l'interrogation des données	26
2. Personnalisation des requêtes	27
2.1. Définition	27
2.2. Approches de personnalisation des requêtes	27
III. Recommandation de requêtes.....	28
1. Système de recommandation.....	28
1.1. Définition	28
1.2. Classification des approches de recommandation.....	29
IV. Personnalisation VS recommandation	29
V. Travaux existants dans les systèmes OLAP.....	29
Synthèse.....	30
Conclusion.....	31

Chapitre 03 : Contextualisation dans les systèmes OLAP

Introduction.....	32
I. Définition du contexte	32
II. Différents types d'éléments contextuels	32
III. Systèmes sensible au contexte.....	33
IV. Contextualisation.....	35
V. Acquisition de l'information contextuelle	35
1. Acquisition explicite.....	35
2. Acquisition implicite.....	36
3. Déduction de l'information contextuelle.....	36
VI. Méthodes d'intégration de l'information contextuelle.....	37
1. Pré-filtrage contextuel	37
1.1.Pré-filtrage Exact.....	38
1.2.Pré-filtrage généralisé.....	38
2. Post-filtrage contextuel.....	39
3. Modélisation contextuelle.....	41
VII. Systèmes OLAP et Datamining.....	42
1. Définition.....	42
2. Technique du Datamining.....	42
3. Datamining et Systèmes OLAP.....	43
VIII. Contextualisation et Systèmes OLAP	44
Synthèse.....	44
Conclusion.....	44

Chapitre 04: Architecture d'un système personnalisé et contextualisée

Introduction.....	46
I. Génération automatique de contexte.....	47
1. Définition d'un fichier Log.....	47
2. World Wide Web Consortium (W3C).....	48
3. Format d'un fichier Log W3C.....	48
4. Pourquoi ce type ?.....	49
5. Exemple d'un fichier Log.....	49
6. Classification automatique.....	50
II. Contextualisation.....	51
III. Recommandation.....	53
Conclusion.....	53

Chapitre 05: Application et test

Introduction.....	54
1. Présentation de l'entrepôt utilisé.....	54
2. Environnement de développement Visual Studio 2008.....	56
2.1.Langage de programmation.....	56
2.2.SQL Server Analysis Services 2008.....	57
2.3.Langage MDX.....	58
3. Module de l'application.....	58
Conclusion.....	62

Conclusion Générale

Liste des Figures

Figure01 : Processus de prise de décision.....	3
Figure02 : Architecture d'un système décisionnel.....	5
Figure03 : Un datawarehouse vs des datamart.....	6
Figure04 : Exemple de cube de données.....	7
Figure05 : Exemple d'une modélisation en étoile.....	8
Figure06 : Exemple d'une modélisation en flocon.....	9
Figure07 : Exemple d'une modélisation en constellation.....	9
Figure08 : Exemple de fait et de mesure.....	12
Figure09 : Exemple de dimension.....	13
Figure10 : Principes de la personnalisation.....	17
Figure11 : Schéma des éléments contextuels.....	33
Figure12 : Pré-filtrage Contextuelle.....	37
Figure13 : Post-filtrage Contextuelle.....	40
Figure14 : Modélisation contextuelle.....	41
Figure15 : Architecture global du système.....	46
Figure16 : Fichier Log.....	49
Figure17 : Modèle de fichier Log.....	58
Figure18 : Matrice des transformations.....	59
Figure19 : Classification des données.....	60
Figure20 : Interface authentification.....	61
Figure21 : Interface lancement des requêtes.....	61
Figure22 : Interface des testes.....	62

Liste ses Tableaux

Tableau 01 : Exemples de profil dans le domaine RI.....	19
Tableau 02 : Exemples de profil dans le domaine d'IHM.....	20
Tableau 03 : Exemples de profil dans le domaine BD.....	21
Tableau 04 : Comparaison des travaux de personnalisation.....	31
Tableau 05 : Tableau descriptive des transformation des données.....	60

Résumé

Le travail mené dans le cadre de cette thèse repose sur une problématique centrale : la contextualisation dans les systèmes OLAP. En effet, face à la croissance considérable des données, à l'hétérogénéité des rôles et des besoins, il devient important de proposer un système capable de fournir à l'utilisateur une information pertinente et utilisable. Ce système doit prendre en considération les différentes caractéristiques de l'utilisateur ainsi que l'ensemble des situations contextuelles qui influence son comportement lors de son interaction avec le système d'information.

Notre projet se divise en deux parties : la première c'est de trouver une approche qui permet la génération automatique des contextes a partir de fichier log de l'utilisateur en utilisant des techniques du Datamining, et la deuxième c'est de trouver la manière d'intégrer le contexte dans le processus de personnalisation c-à-d découvrir un lien naturel entre le profil et le contexte ce qui nous a permet d'obtenir des préférences contextualisée qui enrichiraient les requêtes multidimensionnelles.

Alors notre système réalisée va donnerait une pertinence réel aux résultats parc qu'il va fournir des résultats qui ne sont pas seulement personnalisées.

Mots clés

Contextualisation, Système OLAP, Contexte, Personnalisation, Profil, Préférences contextualisée.

Abstract

The work done in this subject is based on a central issue: contextualization in OLAP systems. Indeed, faced with the enormous growth of data, the heterogeneity of roles and needs, it is important to provide a system able to providing the user with relevant and useful information. This system must take into account the different characteristics of the user as well as the contextual situations that influence his behavior during his interaction with the information system together.

Our project is divided into two parts: the first is to find an approach that allows the automatic generation of contexts from the log file of the user by using data mining techniques and the second is to find the way integrate the context in the personalization process or to discover a natural link between the profile and the context that allows us to obtain a contextual preferences that enrich contextualized multidimensional queries.

So our system is made to give a real relevance results park that will provide results that are not only customized.

Key words

Contextualization, OLAP systems, context, personalization, profile, contextual preferences.

Introduction Générale

Introduction Générale

Les systèmes d'information actuels donnent accès à un grand nombre de sources de données. Au fur et à mesure que les sources se multiplient et que le volume de données disponibles s'accroît, l'utilisateur se voit confronté à une surcharge informationnelle dans laquelle il est difficile de distinguer l'information pertinente de l'information secondaire et même du bruit. En outre, l'évaluation d'une requête se fait généralement sans tenir compte du contexte et/ou des besoins spécifiques de l'utilisateur qui l'a émise. La même requête, faite par deux utilisateurs différents, produit les mêmes résultats même si ces utilisateurs n'ont pas les mêmes attentes.

Les systèmes d'accès d'information basé sur le processus de personnalisation de système OLAP permet la mise en œuvre de système centré utilisateur non dans le sens d'un utilisateur générique mais d'un utilisateur spécifique pour cela en vue adapter son fonctionnement à son contexte.

En clair, le problème n'est pas dans la disponibilité d'information mais sa pertinence relativement à un contexte particulière. Pour cela, les travaux sont orientés vers la conception d'une nouvelle génération des systèmes contextualisée.

De plus il a été prouvé expérimentalement que l'intégration du contexte dans le processus de personnalisation enrichirait plus les requêtes personnalisées et donnerait une pertinence réelle aux résultats. Peu de travaux existent dans la littérature intégrant le contexte dans le processus de personnalisation des systèmes OLAP et s'il existe il l'intègre par un paramètre seulement qui est donné en plus explicitement.

Le but de ce travaille est d'obtenir des préférences contextualisée qui permet l'enrichissement des requêtes multidimensionnelles.

Problématique

Comment améliorer pour un utilisateur donné ses possibilités d'interaction avec un système donné et retrouver l'information pertinente ?

L'extraction d'une telle information pose un certain nombre de problèmes :

- Le temps de réponse élevé : Ce problème est d'autant plus crucial que l'utilisateur ne sachant pas a priori ce qu'il cherche, pose plusieurs requêtes pour trouver ce qui l'intéresse.

- La taille importante des réponses : Le nombre d'informations trouvées est généralement très important. L'exploitation de ces réponses n'est pas toujours possible. On peut se retrouver dans des situations où des réponses ne peuvent pas être toutes explorées pour être utilisées.

C'est pourquoi on fait appel aux outils de personnalisation et de contextualisation pour trouver rapidement un ensemble réduit de réponses intéressantes et pouvant satisfaire l'utilisateur.

Objectif

On propose dans ce travail de :

- Trouver une approche permettant de générer automatiquement le contexte en utilisant une ou plusieurs technique(s) du datamining.

- Trouver une manière d'intégrer le contexte obtenu dans le processus de personnalisation.

Le résultat obtenu serait des préférences contextualisée qui enrichiraient les requêtes multidimensionnelles.

Chapitre01 :

*Entrepôt de données et système
OLAP*

Introduction

Aujourd'hui, la majorité des entreprises transforme leur système d'information qui avait une vocation de production à un système d'information décisionnel dont la vocation de pilotage devient majeure.

Le service attendu d'un système décisionnel est d'avoir une vue synthétique de l'entreprise, ceci dans le but de pouvoir prendre des décisions stratégiques concernant les directions et engagement à prendre.

En général, le marché des solutions décisionnelles basées sur les entrepôts de données et les outils OLAP.

I. System décisionnel

1. Définition

Un système d'information décisionnel (SID) est un ensemble de données organisées de façon spécifique, facilement accessible et appropriées à la prise de décision ou encore une représentation intelligente de ces données au travers d'outils spécialisés. La finalité d'un système décisionnel est le pilotage de l'entreprise [FAN, 08].

2. Processus de prise de décision [LYD, 08]

1. Définir le problème.
2. Champs d'application des systèmes décisionnels :
 - 1.2. Rassembler les données.
 - 1.3. Analyser les données
3. Établir des solutions
4. Décider

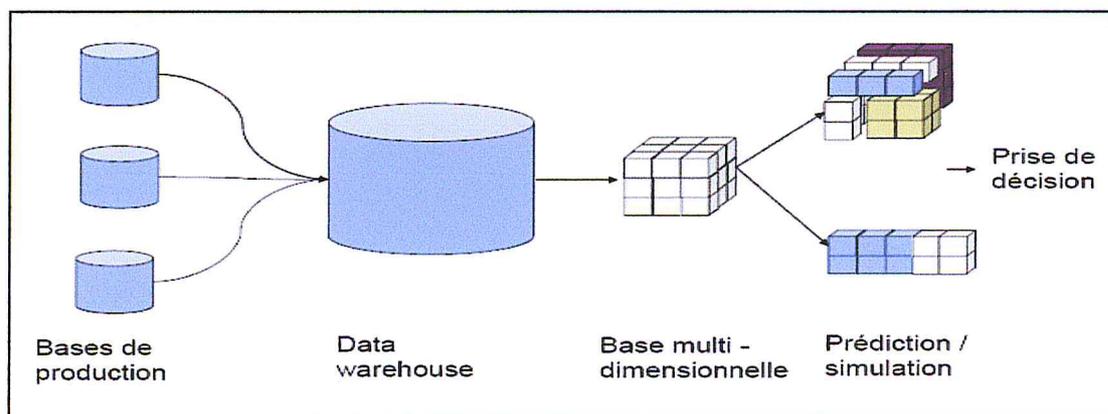


Figure 01 : Processus de prise de décision [LYD, 08].

II. Entrepôt de données

1. Définition

En 1996, Bill Inmon définit un entrepôt de données comme étant une « Collection de données orientées sujet, intégrées, non volatiles et historisées, organisées pour le support du processus d'aide à la décision ».

Les données sont « *orientées sujet* » dans la mesure où elles sont organisées par thèmes, l'entrepôt de données est organisé autour des sujets majeurs et des métiers de l'entreprise. Il permet une vision transversale des différentes activités de l'entreprise.

Le fait que les données soient « *intégrées* » exprime leur provenance de sources différentes. Cette intégration nécessite une bonne connaissance des sources de données, des règles de gestion, de la sémantique des données, etc.

En outre, les données sont « *historisées* » afin de rendre possible la réalisation d'analyses au cours du temps, nécessitant un recours à un référentiel temporel associé aux données.

De plus, les données sont dites « *non volatiles* ». Cela signifie que les données stockées au sein de l'entrepôt de données ne peuvent pas être supprimées. Une requête émise sur les mêmes données à différents intervalles de temps doit donner le même résultat. Cela doit permettre de conserver la traçabilité des informations et des décisions prises.

Enfin, les données sont « *organisées pour le support du processus d'aide à la décision* »; il s'agit en l'occurrence d'une organisation multidimensionnelle. Cette organisation est en effet propice à l'analyse et, en particulier, à l'agrégation [INM, 96].

Selon Jean-François Goglin : un datawarehouse est un entrepôt de données. Il s'agit d'un stockage intermédiaire des données issues des applications de production, dans lesquelles les utilisateurs finaux puisent avec des outils de restitution et d'analyse [GOG, 01].

2. Domaines d'utilisation des DW [LYD, 08]

- ✚ **Banque** : Risques d'un prêt, prime plus précise.
- ✚ **Santé** : Épidémiologie, risque alimentaire.
- ✚ **Commerce** : Ciblage de clientèle, déterminer des promotions.
- ✚ **Logistique** : Adéquation demande/production.
- ✚ **Assurance** : Risque lié à un contrat d'assurance (voiture).

3. Architecture d'un système décisionnel

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). L'interrogation est l'opération la plus utilisée dans le contexte d'entrepôt de données où la mise à jour consiste seulement à alimenter l'entrepôt [STA, 12].

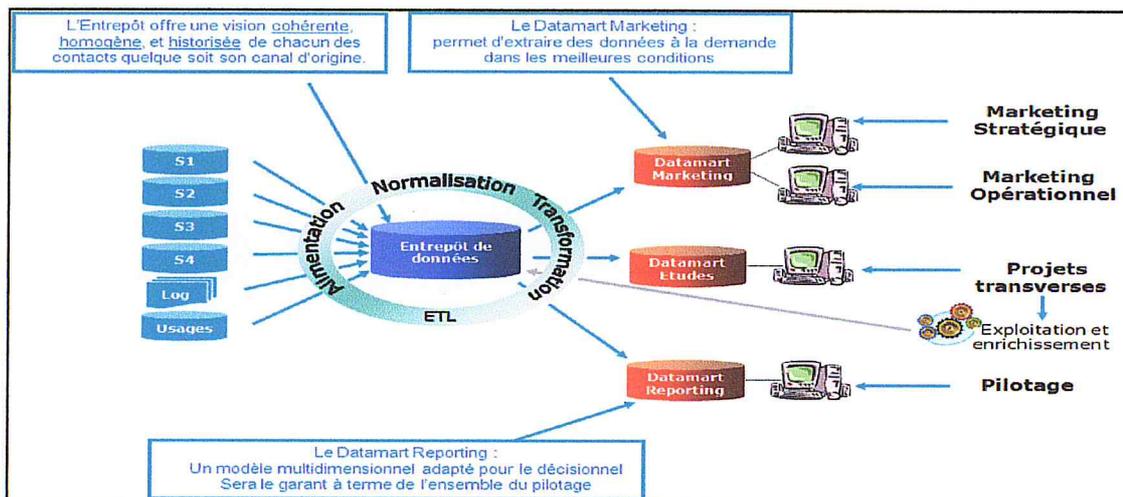


Figure 02: Architecture d'un système décisionnel [W, 01].

Le processus de construction d'un entrepôt de données est composé de trois principales phases (Figure02) :

✚ Extraction

L'extraction peut se faire à travers un outil qui doit travailler de façon native avec les SGBD qui gèrent les données sources. Les données source sont extraites de système, de base de données et de fichiers.

✚ Transformation de données

C'est une suite d'opérations qui a pour but de rendre les données cibles homogènes et puissent être traitées de façon cohérente. Les données source sont nettoyer, transformer et intégrées avant d'être stockées dans l'entrepôt.

✚ Chargement des données

C'est l'opération qui consiste à charger les données nettoyées et préparées dans le DW. Les utilisateurs accèdent à l'entrepôt à partir d'interfaces et d'applications (clients).

4. Classes de données

4.1. Métadonnée

La métadonnée représente de l'information structurée décrivant, expliquant et localisant la ressource et en facilitant la recherche, l'usage et la gestion [BET, XX].

4.2. Datamart (magasin de données) [LYD, 08]

Est un sous-ensemble d'un entrepôt de données, destiné à répondre aux besoins d'un secteur ou d'une fonction particulière de l'entreprise. Point de vue spécifique selon des critères métiers.

Datamart contient moins de données que DW, elle est plus facile à comprendre et à manipuler aussi bien améliorer le temps de réponse.

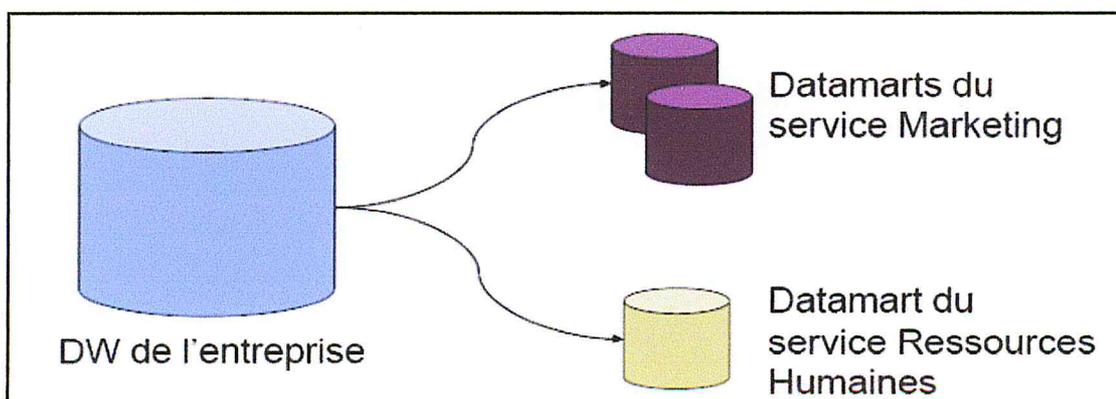


Figure 03: Un datawarehouse vs des datamart [LYD, 08].

4.3. Cube de donnée

Le cube est une modélisation d'une base de données multidimensionnelle. Cela permet aux utilisateurs de l'entrepôt de données de pouvoir trouver, extraire et évaluer les données dont ils ont besoin sans l'aide des informaticiens. Chaque côté du cube dispose d'une dimension et les caractéristiques correspondantes.

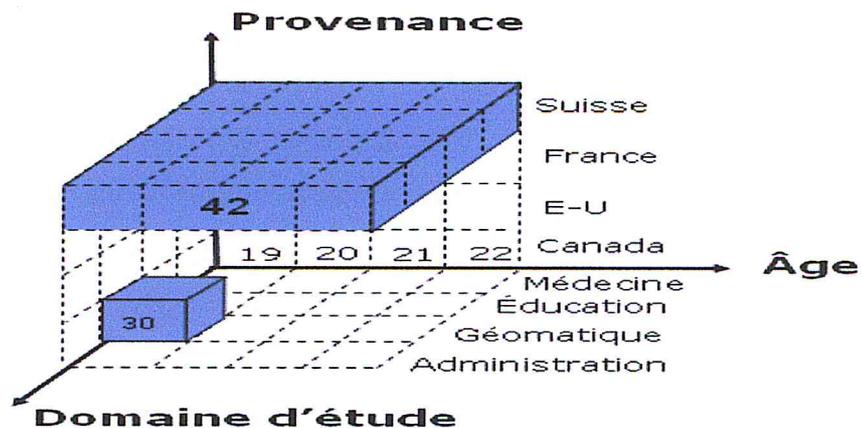


Figure 04 : Exemple de cube de données [TJL, 02].

- Un cube est une représentation multidimensionnelle des données dans un entrepôt. Il est décrit par les tables de dimensions (les axes d'analyse) et une table de faits (les mesures) [MOU, 07].
- Un cube de données peut aussi être vu comme un ensemble de cellules où chacune d'elles représente une association entre un membre de chaque dimension et une mesure (ou contenu de cellule) [INM, 96].

4.4. Données agrégées

Les données agrégées correspondent à des éléments d'analyse représentatifs des besoins utilisateurs, Ce sont donc des données déjà traitées par le système et représentant un premier résultat d'analyse et de synthèse des données contenues dans les systèmes de production. Elles présentent des avantages tels que facilité d'analyse, rapidité d'accès, moindre volume.

Les données agrégées sont préconçues pour répondre rapidement et efficacement à un besoin ciblé et connu avant de charger les données dans l'entrepôt. On agrège les données dans la BD par souci de performances, car il est plus rentable et efficace de les interroger déjà agrégées à travers un outil de rapport que de les calculer dans l'outil à chaque fois qu'une requête est demandée [DEM, 10].

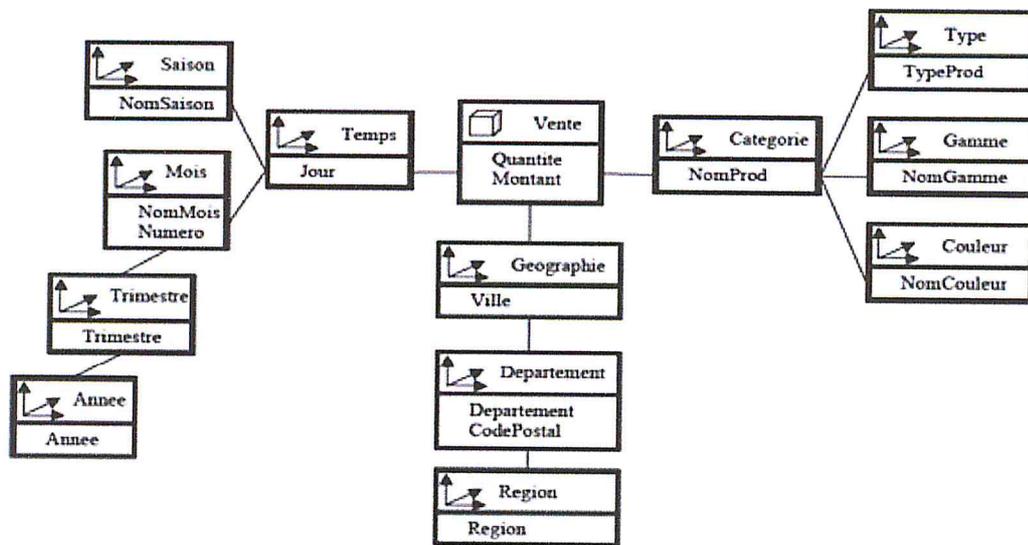


Figure 06: Exemple d'une modélisation en flocon [W, 02].

5.3. Modèle en constellation (Factflake schéma)

Une autre technique de modélisation, issue du modèle en étoile, est la modélisation 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 des dimensions communes ou non [STA, 12].

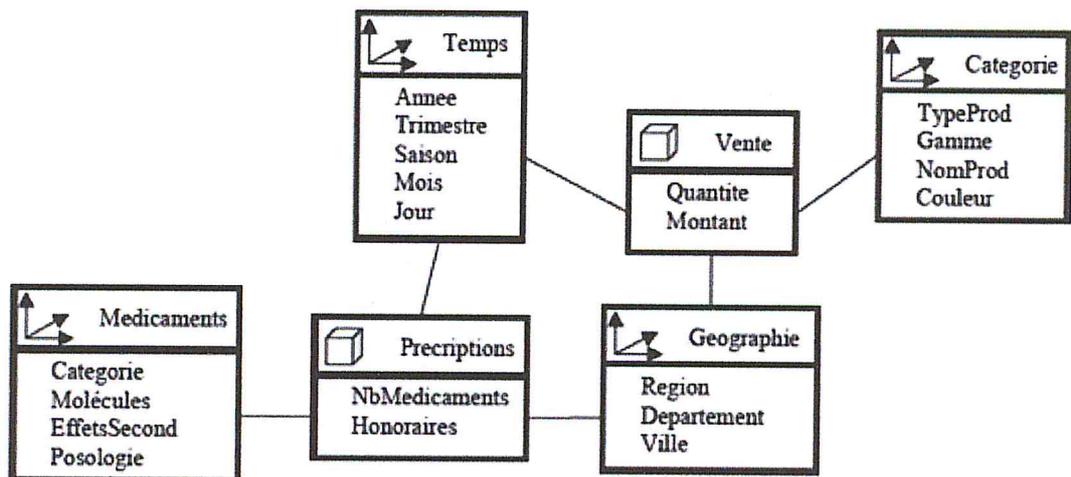


Figure 07: Exemple d'une modélisation en constellation [W, 02].

- Dimensions génériques : essayer, autant que possible, d'avoir une unicité dans la définition des dimensions. Ne pas avoir deux dimensions client.
- Gestion des matrices creuses : en mathématiques, les matrices creuses sont des matrices qui contiennent beaucoup de zéros. En informatique, il existe des algorithmes qui utilisent cette spécificité pour optimiser le stockage de ce type de matrices. Les performances sont, en général, au rendez vous. Les outils OLAP doivent avoir cette capacité d'optimisation d'espace de stockage par la gestion des matrices creuses.
- Multi-utilisateurs : les outils OLAP sont, par définition, destinés à un accès concurrent.
- Croisement inter dimensions illimité : l'utilisateur ne doit avoir aucune restriction quand au nombre de croisements qu'il fait entre les dimensions.
- Intuitifs : les utilisateurs d'outils OLAP ne sont pas forcément informaticiens. Il est donc nécessaire d'offrir des solutions adaptées à leur style cognitif.
- Affichage flexible : l'utilisateur doit pouvoir aisément " arranger " son résultat au format désiré.
- Nombre illimité de dimensions et de niveaux d'agrégation.

2. OLAP pour quoi faire ?

La technologie OLAP peut être utilisée dans un grand nombre de domaines :

- ✚ Analyse de ventes et de marketing : C'est le domaine de prédilection. L'exemple proposé ici est souvent le point de départ des applications de ce type. On étudie les volumes de vente par produit, région et suivant le temps.
- ✚ Consolidation de données : On utilise directement l'un des atouts des outils OLAP.
- ✚ Aide à la décision : On tente de prévoir l'évolution des recettes et dépenses.
- ✚ Analyse de la qualité de service.

3. OLAP pour qui ?

Tous les gestionnaires d'une entreprise sont susceptibles d'être intéressés par une application OLAP.

4.5. Données détaillées(ou données élémentaires)

Elles sont des données qui reflètent les événements les plus récents, Les données élémentaires présentent des avantages évidents (profondeur et niveau de détail, possibilité d'appliquer de nouveaux axes d'analyse et même de revenir à posteriori sur le passé) mais représentent un plus grand volume et nécessitent donc des outils plus performants [DEM, 10].

5. Différents modèles de données

5.1. Modèle en étoile (Star schéma)

A partir du fait et des dimensions, il est possible d'établir une structure de données simple qui correspond au besoin de la modélisation multidimensionnelle. Cette structure est constituée du fait central et des dimensions. Ce modèle représente visuellement une étoile, on parle de modèle en étoile [STA, 12].

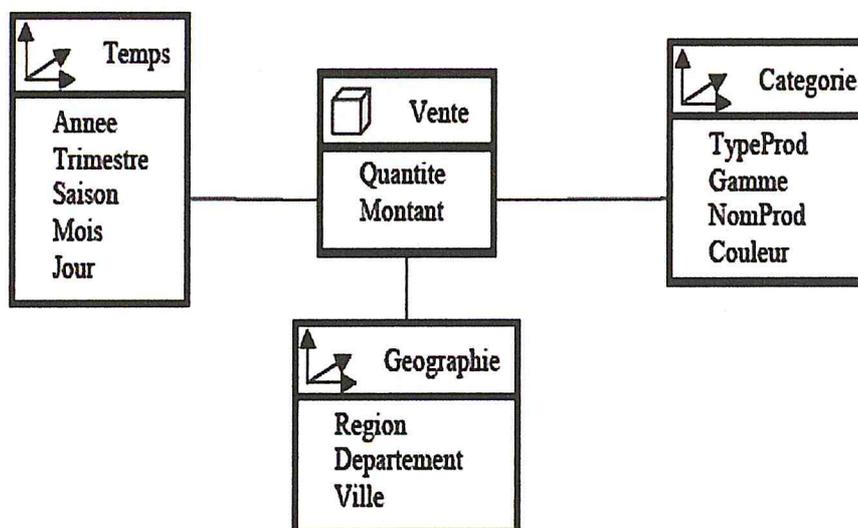


Figure 05: Exemple d'une modélisation en étoile [W, 02].

5.2. Modèle en flocons (Snowflake schéma)

Il existe d'autres techniques de modélisation multidimensionnelle, notamment la modélisation en flocon (snowflake). Une modélisation en flocon consiste à décomposer les dimensions du modèle en étoile en sous hiérarchies (normalisation) [STA, 12].

- ✚ Grâce à la consultation des données consolidées, les responsables commerciaux peuvent comparer les performances et les objectifs présents et passés, en fonction des produits, des clients, des services.
- ✚ Grâce au volume important de données mémorisées, les chefs de produit ou les techniciens peuvent étudier quotidiennement les volumes et ratios, ou découvrir les problèmes en déclarant des exceptions.

4. Approche multidimensionnelle dans le datawarehouse

4.1. Fait : Un fait représente le sujet ou le thème analysé. Il présente un centre d'intérêt de l'entreprise et est considéré comme un concept clé sur lequel repose le processus de prise de décision. Un fait est formé de « mesures » ou attributs du fait (atomiques ou dérivés) qui correspondent aux informations liées au thème analysé. Les mesures sont stockées dans des tables de faits qui contiennent les valeurs des mesures et les clés vers les tables de dimensions [INM, 96].

4.2. Mesure : Une mesure est la valeur qui associe un fait à un axe d'analyse (dimension) [NAO, 06].

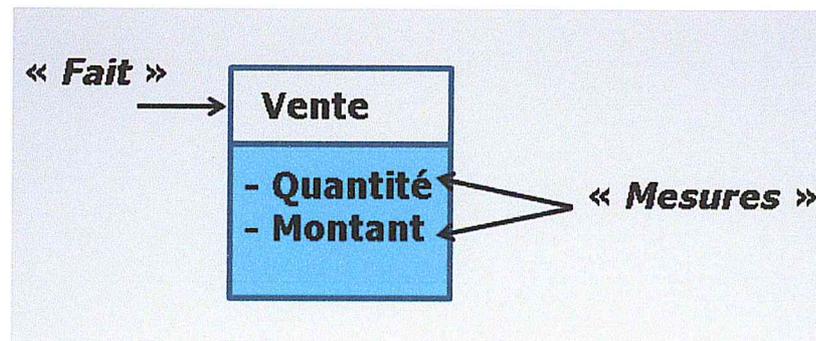


Figure 08 : Exemple de fait et de mesure [W, 02].

4.3. Dimension : La structure de base de toute application multidimensionnelle est la dimension.

Une dimension est une liste complète d'éléments d'entrée qui contiennent des données et d'éléments calculés ou dérivé à l'aide d'une formule quelconque [GOG, 01].

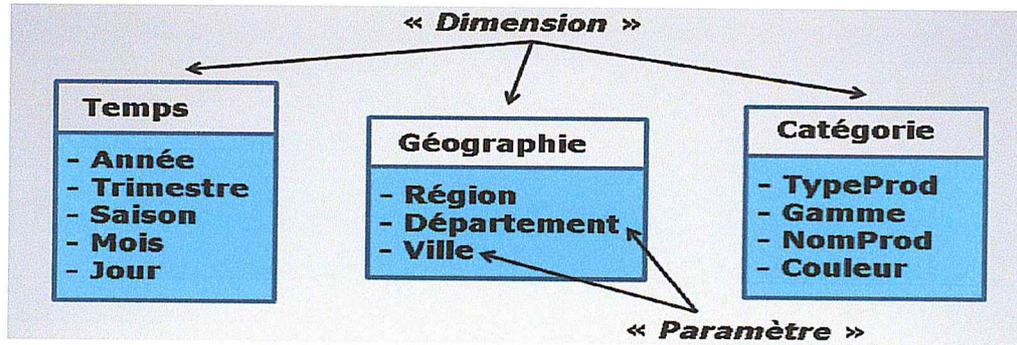


Figure 09 : Exemple de dimension [W, 02].

5. Outils de navigation OLAP [RIZ, 07]

- ✚ **Drill up /down** : Cela permet d'aller vers les informations détaillées dans une hiérarchie ou au contraire de remonter d'un niveau de granularité. Il s'agit donc de « zoomer ou de dé-zoomer » sur une dimension.
- ✚ **Rotate** : Consiste à effectuer une rotation de l'hyper cube afin de présenter une face différente. Il s'agit donc de modifier une dimension de lecture.
- ✚ **Slice** : Consiste à ne travailler que sur une tranche de l'hyper cube. Une des dimensions est alors réduite à une seule valeur.
- ✚ **Scoping** : Consiste à ne travailler que sur un sous-cube. On s'intéressera alors seulement à une partie des données.
- ✚ **Dice** : Extraction d'un bloc de données : S'élection de deux ou plusieurs dimensions.
- ✚ **Drill-across** : Exécution de requêtes impliquant plus d'un cube ayant une dimension commune.

6. Différentes implémentations OLAP

6.1. Relational OLAP(ROALP) : sont des bases de données OLAP supportant une architecture relationnelle.

Un outil ROLAP est ainsi capable de simuler le comportement d'un SGBD multidimensionnelle en exploitant un SGBDR classique. L'intérêt de cette technique réside d'une part dans le fait que l'outil ROLAP s'appuie sur un SGBDR traditionnel et ne nécessite donc pas l'acquisition d'un SGBD multidimensionnelle, et d'autre part, qu'il est plus facile de mettre en œuvre des requêtes complexe de type OLAP sur des informations provenant directement des applications de production ou du Datawarehouse[GOG, 01].

6.2. Multidimensionnel OLAP (MOLAP)

MOLAP, pour Multidimensional On Line Analytical Processing, stocke physiquement les données dans une base multidimensionnelle. Le cube est alors fourni directement. La structure multidimensionnelle utilisée dans les systèmes MOLAP, est un tableau à n dimensions. Nous trouvons également l'appellation d'hypercube ou de cube pour désigner cette structure [NAO, 06].

6.3. Hybride OLAP(HOLAP)

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

6.4. Desktop OLAP(DOLAP)

Desktop OLAP. Le cas spécial, Il ne s'agit en fait pas d'une technologie particulière mais plutôt d'un mode de fonctionnement.

D-OLAP permet à l'utilisateur d'enregistrer une partie de la base de données multi dimensionnelle en local.

On voit très vite l'utilité d'une telle solution pour les commerciaux et les " nomades " de l'entreprise. Cela permettrait à un commercial, par exemple, de faire des analyses sur les ventes, conserver ses résultats, et vérifier l'évolution de ses analyses, une fois revenu de son voyage d'affaire [TJL, 02].

7. Requête OLAP

La recherche des données disponibles dans un contexte d'entrepôt nécessite une identification non ambiguë des données de l'entrepôt. Nous proposons une approche qui s'inspire de la façon dont les requêtes OLAP référencent les données de l'entrepôt. En effet, une requête OLAP est constituée d'une sélection de membres pour une ou plusieurs dimensions ainsi qu'un choix de mesures ou de leurs agrégats.

Une requête OLAP est une requête complexe impliquant plusieurs opérations de sélection, de jointure ainsi que des calculs d'agrégation [WEH, 09].

IV. Conclusion

Dans ce chapitre nous avons commencé par définir les principaux concepts liés à la modélisation dimensionnelle et à la manipulation des entrepôts de données où nous avons vu l'architecture d'un système décisionnel, les domaines d'utilisation de ce dernier et les différents types de données.

En suite nous avons cité les différentes implémentations et les outils de navigation OLAP.

Chapitre02:

*Personnalisation dans les systèmes
OLAP*

Introduction

Les recherches qui réalisent des services d'accès au volume important d'informations ne reflètent pas les intérêts de l'utilisateur. Ces recherches utilisent souvent le concept d'analyse des données multidimensionnels à travers des requêtes (requêtes d'analyse) qu'ils ont une structure complexe, et peuvent parfois fournir des résultats indésirable (champs vide, des informations brutes ou secondaires, etc.) et sauvant ne réponds pas aux besoins de l'utilisateur.

Pour améliorer la qualité des services d'accès à l'information l'intervention de l'utilisateur sera donc nécessaire au cours de l'exécution de la requête dans le but de délivrer une information pertinente adaptée à son besoin et préférence précis. Cette vue rentre dans le cadre de la personnalisation de l'information qui est un domaine récent.

La personnalisation consiste à diminuer les informations inutiles pour rendre la navigation et l'exploration de masse importante d'information facile et rapide grâce à la notion de profil, préférence et contexte afin de répondre au mieux aux attentes de chaque utilisateur.

I. Notion de personnalisation

1. Définition

- La personnalisation de l'information est définit comme l'action d'adapter l'accès à l'information en fonction d'informations sur un usager ou sur un groupe d'utilisateurs [JER, 12].
- 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 [TZB, 07].
- 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 [BOK, 05].

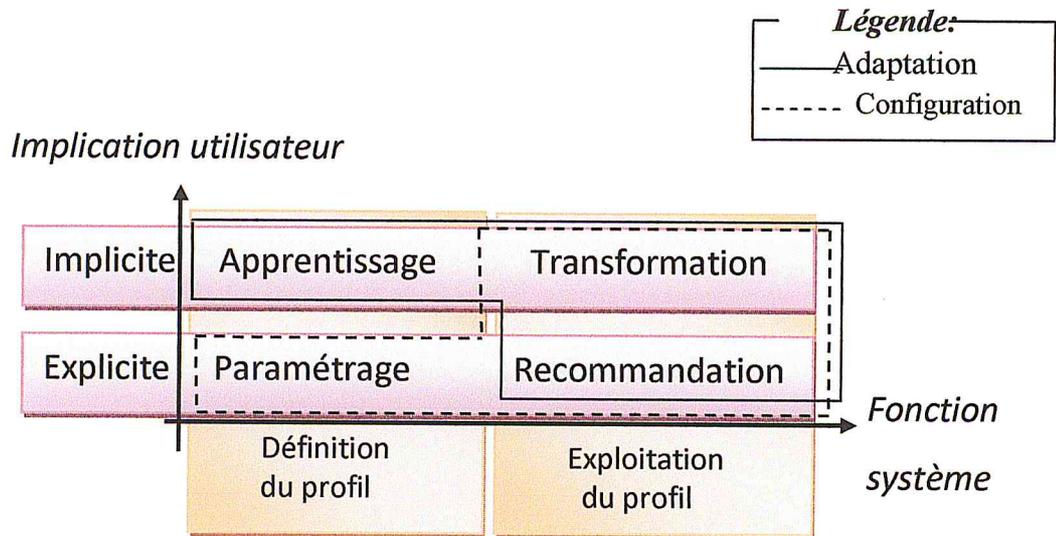


Figure 10: Principes de la personnalisation [BBF, 08].

La **figure 10** présente les principes mis en jeu lors de la personnalisation. La définition d'un profil réalisée de façon explicite ou implicite.

- La configuration (« customisation » ou « user modeling ») consiste donc pour l'utilisateur à paramétrer explicitement son profil. Le profil ainsi déterminé est exploité à travers des mécanismes de transformation ou de recommandation. Par exemple, dans le logiciel Word, l'opération consistant à placer manuellement un bouton dans la barre d'outils est une tâche qui s'apparente à une configuration.
- L'adaptation (« user profiling ») consiste pour le système à définir implicitement le profil de l'utilisateur, puis à l'exploiter selon les principes de transformation ou de recommandation. Par exemple, dans le logiciel Word, depuis sa version 97, les items des menus sont rendus automatiquement visibles en fonction de l'usage qui est fait du traitement de texte.

2. Domaines intéressés par la personnalisation [BOK, 05]

- ✚ Dans le domaine des Interface Homme Machine (IHM), la notion de profil, souvent appelé modèle d'utilisateur, se focalise plus sur le niveau d'expertise et le métier de l'utilisateur afin de déterminer le type de dialogue que le système va avoir avec lui, les métaphores graphiques les plus appropriées ainsi que les modalités de livraison des résultats qu'il attend du système d'information.
- ✚ Dans le domaine de la Recherche d'Information (RI), l'utilisateur fait partie du processus de personnalisation. L'évaluation d'une requête se fait généralement de façon interactive et incrémentale; à chaque itération, le système tient compte des informations collectées à partir des interactions précédentes avec l'utilisateur ou

profite de l'expérience des autres utilisateurs (filtrage collaboratif). La personnalisation est ainsi définie comme un apprentissage réalisé à partir des préférences rendues par les utilisateurs à l'issue de la présentation des résultats successifs du système.

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

3. Profil utilisateur

3.1. Notion de profil

- Un profil utilisateur regroupe l'ensemble des connaissances nécessaires à une évaluation efficace des requêtes et à une production d'une information pertinente adaptée à chaque utilisateur [BOK, O4].
- un profil est un modèle utilisateur qui décrit les préférences d'un utilisateur qui le différencie des autres [JER, 12].
- Un profil utilisateur est une collection d'informations sur l'utilisateur. Cette collection peut être vue comme un ensemble de caractéristiques avec des valeurs associées contenant par exemple ce que l'utilisateur préfère, ce qu'il est capable de faire...etc.

On doit également prendre en compte l'historique des actions de l'utilisateur, voir leur évolution dans le temps [W, 03].

Chapitre 02 : Personnalisation dans les systèmes OLAP

3.2.Exemples de profile

- Exemple sur le domaine de la RI

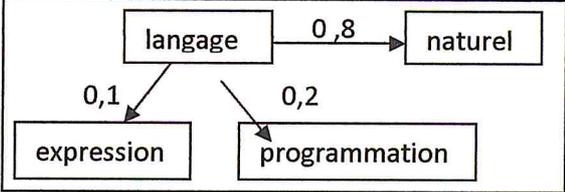
Références	Approches	Exemples																
[FES, 01], [PRG, 99], [SOC, 98], [CCL, 01]	Le profil de l'utilisateur décrit le plus souvent son centre d'intérêt et, de ce fait, est souvent confondu avec la requête de l'utilisateur. Ce profil est généralement défini à l'aide d'un vecteur de mots clés avec éventuellement un poids associé à chaque mot	Par exemple le profil d'un utilisateur intéressé par la personnalisation des données peut être présenté par le vecteur à trois termes suivant; le poids de chaque terme correspond généralement à sa fréquence d'apparition dans les documents: profil vectoriel {(personnalisation, 0.7), (profil, 0.9), (modèle, 0.5)}.																
[SOE, 95]	le profil est représenté sous forme d'un graphe orienté (Figure) où les sommets sont les mots et les arcs expriment les poids entre ces mots. Chaque poids correspond à une probabilité qu'un mot apparaisse après un autre dans le texte.	Sur l'exemple de la figure, l'arc entre les nœuds « langage » et « naturel » signifie que dans 80% des cas le mot naturel est précédé par le mot langage : 																
CASPER [BRS, 00]	Le profil d'un utilisateur est défini sous la forme de statistiques des actions effectuées par l'utilisateur sur les offres d'emplois. Le projet de CASPER présente un moteur de recherche d'emploi, Dans ce cas, 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.	par exemple un utilisateur qui a lu une annonce de travail, a postulé pour une autre et a envoyé une troisième à un ami : Profil de Paul <table border="1" data-bbox="829 1317 1426 1626"> <thead> <tr> <th>Annonce</th> <th>Action</th> <th>Nombre de Clicks</th> <th>Temps de lecture</th> </tr> </thead> <tbody> <tr> <td>Job 5</td> <td>lire</td> <td>1</td> <td>234</td> </tr> <tr> <td>Job 5</td> <td>candidater</td> <td>2</td> <td>186</td> </tr> <tr> <td>Job 5</td> <td>Envoyer à un ami</td> <td>1</td> <td>54</td> </tr> </tbody> </table> <p>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.</p>	Annonce	Action	Nombre de Clicks	Temps de lecture	Job 5	lire	1	234	Job 5	candidater	2	186	Job 5	Envoyer à un ami	1	54
Annonce	Action	Nombre de Clicks	Temps de lecture															
Job 5	lire	1	234															
Job 5	candidater	2	186															
Job 5	Envoyer à un ami	1	54															

Tableau 01 : Exemples de profile dans le domaine de RI.

Chapitre02: Personnalisation dans les systèmes OLAP

- Exemple sur le domaine des IHM

Références	Approches	Exemples
[BOK, 05]	le profil contient par exemple, des informations permettant au système d'adapter l'affichage des résultats selon les préférences de l'utilisateur.	Un exemple simple d'un tel profil est celui utilisé par les fournisseurs de services Web. Dans ces systèmes le profil d'un utilisateur représente un ensemble de données personnelles (nom, prénom, genre, date de naissance, code postal, e-mail, profession, poste etc. dans le cas de Yahoo par exemple) et des catégories d'intérêts qui constituent sa page d'accueil (ex. météo, football, jeux, etc.). Le contenu de certaines catégories du centre d'intérêt peut être déduit à partir des données personnelles (par exemple si l'utilisateur est né au début du mois de Juin et s'il a choisi la catégorie 'Horoscope', le zodiaque affiché sera Gemmaux).
[MDL, 02]	consiste à regrouper les transactions similaires d'un utilisateur dans des clusters.	Le profil (noté PRc) d'un utilisateur est obtenu sur la base d'un cluster c et contient les identifiants d'un ensemble de pages dont le poids associé à chaque page dépasse un seuil μ : $PRc = \{(p, \text{poids}(p, PRc)), \text{poids}(p, PRc) > \mu\}$. Le poids d'une page p est calculé en fonction de la fréquence de son apparition dans le cluster.

Tableau 02 : Exemples de profile dans le domaine de d'IHM.

- Exemple sur le domaine des BD

Dans le domaine des BD, le profil de l'utilisateur contient des données qui expriment ses habitudes, des prédicats fréquemment utilisés dans ses requêtes ou des définitions de l'ordre de préférences de ces prédicats [KOI, 04].

Prenons par exemple une base de données dont le schéma est le suivant :

```
TRANSPORT (idT, moyen)
HOTEL (idH, nombre_étoiles, région)
VOYAGE (idV, prix, lieu_départ, lieu_arrivée, nombre_jours,
idH, idT)
DEPART (idD, idV, date, heure)
```

Chapitre 02 : Personnalisation dans les systèmes OLAP

Références	Approches	Exemples																		
[KOI, 04]	<p>Le profil de l'utilisateur contient des données qui expriment ses habitudes, des prédicats fréquemment utilisés dans ses requêtes ou des définitions de l'ordre de préférences de ces prédicats.</p> <p>L'intérêt de l'utilisateur pour chacun de ces éléments est exprimé par un degré qui est un nombre réel compris entre 0 et 1.</p>	<p>Si un utilisateur qui habite à Paris et qui aime voyager pendant le week-end, descend d'habitude dans des hôtels au centre ville et préfère voyager en train plutôt qu'en car, son profil peut être écrit sous la forme :</p> <div style="border: 1px solid black; padding: 5px; margin: 10px 0;"> <p>Exemple 5 de profil :</p> <pre>{ TRANSPORT.idT = VOYAGE.idT 1 (a) HOTEL.idH = VOYAGE.idH 1 (b) VOYAGE.lieu_départ = 'Paris' 1 (c) HOTEL.région = 'centre ville' 0.9 (d) VOYAGE.nombre_jours = 2 0.7 (e) TRANSPORT.moyen = 'train' 0.7 (f) TRANSPORT.moyen = 'car' 0.5 (g) }</pre> </div> <p>Sur chaque expression du profil, considérée comme une sous requête, est ajouté un nombre compris entre 0 et 1 pour exprimer l'importance relative de cette expression par rapport aux autres. Ainsi la valeur 1 sur les trois premières expressions signifie que ces conditions doivent être toujours satisfaites.</p> <p>Les autres expressions expriment le fait que l'utilisateur a une plus forte préférence pour les hôtels situés au centre ville (d) que pour les voyages de deux jours (e) et qu'il préfère voyager en train (f) plutôt qu'en car (g).</p>																		
[SCA, 00]	<p>Dans cet article, il y a deux types de profils : (i) le profil statistique qui est composé des attributs souvent utilisés dans les requêtes de l'utilisateur auxquels on associe la fréquence de leur apparition et (ii) le profil déduit qui représente les attributs qui apparaissent dans les requêtes que l'utilisateur exécute avec une certaine répétition.</p>	<p>par exemple les profils d'un utilisateur qui interroge un système d'information sur les voyages</p> <table border="1" style="margin: 10px 0;"> <thead> <tr> <th colspan="2">Profil statistique :</th> </tr> </thead> <tbody> <tr> <td>Attribut <i>lieu_arrivée</i></td> <td>98%</td> </tr> <tr> <td>Attribut prix</td> <td>70%</td> </tr> <tr> <td>Heure étant donnée date</td> <td>60%</td> </tr> </tbody> </table> <table border="1" style="margin: 10px 0;"> <thead> <tr> <th colspan="2">Profil déduit :</th> </tr> <tr> <th>Requête</th> <th>Répétition</th> </tr> </thead> <tbody> <tr> <td>Lieu arrivée, prix</td> <td>Tous les jours 11h</td> </tr> <tr> <td>Date, heure, idV</td> <td>Lundi 16h</td> </tr> <tr> <td>Moyen, prix</td> <td>Fin du mois</td> </tr> </tbody> </table> <p>L'attribut 'lieu_arrivée' figure dans 98% des requêtes de l'utilisateur et une requête portant sur les attributs 'lieu_arrivée' et 'prix' est émise tous les jours vers 11h.</p>	Profil statistique :		Attribut <i>lieu_arrivée</i>	98%	Attribut prix	70%	Heure étant donnée date	60%	Profil déduit :		Requête	Répétition	Lieu arrivée, prix	Tous les jours 11h	Date, heure, idV	Lundi 16h	Moyen, prix	Fin du mois
Profil statistique :																				
Attribut <i>lieu_arrivée</i>	98%																			
Attribut prix	70%																			
Heure étant donnée date	60%																			
Profil déduit :																				
Requête	Répétition																			
Lieu arrivée, prix	Tous les jours 11h																			
Date, heure, idV	Lundi 16h																			
Moyen, prix	Fin du mois																			

Tableau 03: Exemples de profile dans le domaine des BD.

3.3. Modélisation de profil

L'introduction de la dimension utilisateur dans un processus d'accès à l'information, nécessite une réflexion sur la modélisation de l'entité utilisateur [KOB, 01]. La fiabilité ou qualité des profils est en effet d'une importance bien connue dans le domaine de la modélisation utilisateur (User modeling).

En effet, on constate que l'une des principales raisons du manque de performances des techniques de personnalisation est typiquement l'application d'un profil utilisateur hors contexte [GSC, 07].

La modélisation de l'utilisateur est un processus à différentes étapes à savoir, la représentation du profil utilisateur qui peut être :

- **Vectorielle** : le profil est constitué d'un ou plusieurs vecteurs définis dans un espace de termes d'indexation [GOW, 03].
- **Hierarchique** : les caractéristiques d'un utilisateur sont organisées dans une structure hiérarchique de concepts représentant les domaines d'intérêt [GLR, 02].
- **Multidimensionnelle** : le profil est représenté par un modèle structuré de dimensions prédéfinies (données personnelle, domaine d'intérêt, préférences de livraison, etc.) [KOS, 03].

De ce fait, les questions fondamentales posées pour modéliser le profil utilisateur sont le « Quoi », le « Comment » et le « Quand » [TLD, 05] :

- Quoi ?
 - Quelles propriétés décrivent un l'utilisateur ?
 - Quelle représentation ou quel modèle de l'utilisateur ?
 - Quel contexte d'utilisation ?
- Comment ?
 - Comment construire le modèle de l'utilisateur ?
 - Comment découvrir son intention courante ?
 - Comment exploiter le modèle utilisateur lors de processus de recherche ?
 - Comment évaluer l'impact de la personnalisation sur le processus de recherche d'information ?
- Quand ?
 - Quand faut-il faire évoluer le profil de l'utilisateur ?

3.4. Contenu de profil

Dans un processus d'accès à des informations structurées comme les bases de données ou les entrepôts de données, un profil utilisateur peut être composé de :

- (1) Des informations personnelles, telles que le nom, le prénom, l'âge, etc.
- (2) Préférences sur le contenu de la réponse à la requête.
- (3) Préférences sur la présentation de la réponse, et/ou des conditions d'exploitation (les conditions matérielles et/ou les contraintes utilisateur) [BOU, 10].

3.5. Construction de profil

La construction du profil utilisateur repose sur des outils et des algorithmes permettant de collecter les sources d'information pertinentes servant à l'acquisition et la collecte des données utilisateur en première phase, et des techniques de déploiement de ces sources d'information dans le but de construire le profil final en deuxième phase [DAO, 09].

Le processus d'acquisition peut se faire explicite et/ou implicite :

3.5.1. Acquisition explicite (Paramétrage)

Cette technique constitue une approche simple pour obtenir des informations sur l'utilisateur. Dans le cadre d'une implication explicite, l'utilisateur doit effectuer des interactions directes avec le système [TES, 09]. On interroge directement l'utilisateur ou on lui demande par exemple de remplir des formulaires pour collecter ses préférences sur les dimensions, les membres ainsi que des informations décrivant son environnement à savoir la taille de son écran, la vitesse de son processeur et la taille de sa mémoire, etc.

En effet, l'utilisateur émet directement son jugement d'intérêt en donnant une valeur de pertinence sur une échelle graduée allant du moins intéressant au plus intéressant [BOU, 10].

3.5.2. Acquisition implicite (Apprentissage)

L'acquisition implicite consiste à évaluer les objets sans les interventions des utilisateurs et collecter les informations décrivant l'utilisateur, en observant les dimensions et les membres fréquemment sollicités et en scrutant les caractéristiques de l'environnement à partir duquel il intervient (les capacités et les limites du dispositif utilisé lors de ces interactions). Et ce, en se basant sur l'historique de ses interactions avec le système [BOU, 10].

3.6. Exploitation de profil

Le domaine d'application conditionne fortement les techniques de personnalisation employées par le système. Cela a en effet un impact direct sur l'exploitation du profil dans la chaîne d'accès à l'information et par conséquent sur les mécanismes mis en œuvre. Les principales interrogations posées concernent le « Quoi », le « Comment » de la mise en œuvre [ZEM, 08] :

- Quoi ?
 - Quels services de personnalisation proposés : de la recommandation et/ou du filtrage, de l'aide à la navigation, un assistant personnel de recherche ?
 - Dans quelles étapes du cycle de vie de la requête faut-il intégrer le profil ?
 - Quelles informations du profil exploiter lors de l'accès à l'information ?
- Comment ?
 - Comment intégrer le profil de l'utilisateur dans le processus de personnalisation ?
 - Comment évaluer l'impact de la personnalisation sur le processus de recherche ?

L'intégration du profil utilisateur dans le processus de RI revient à l'exploiter dans l'une des principales phases de l'évaluation de la requête : reformulation, calcul du score de pertinence de l'information ou présentation des résultats de recherche. Une approche de personnalisation basée sur le raffinement des requêtes des utilisateurs [DTB, 08].

3.7. Evolution de profil utilisateur

L'évolution des profils désigne leur adaptation à la variation des centres d'intérêt des utilisateurs qu'ils décrivent, et par conséquent, de leurs besoins en information au cours du temps. La phase d'évolution ne prend un sens que lorsque le profil a une structure pérenne, ce qui permet de distinguer les besoins à court terme, construits à partir de la session d'interaction courante, des besoins à long terme qui sont une réelle représentation des centres d'intérêt persistants de l'utilisateur.

A notre connaissance, peu de travaux ont exploré le problème de l'évolution du profil de l'utilisateur sous l'angle de la dimension temporelle (court terme, long terme).

L'évolution est davantage abordée comme un problème de représentation de la diversité des domaines d'intérêt de l'utilisateur en utilisant des techniques de

classification [PMB, 96], [MIT, 02], [GOW, 03] ou heuristiques liées la notion de cycle de vie artificielle d'un centre d'intérêt [CHS, 98].

4. Préférences utilisateur

4.1. Notion de préférence

- ✚ Une préférence est une expression permettant de hiérarchiser l'importance des informations dans un profil ou un contexte [ZEM, 08].
- ✚ Les préférences utilisateur correspondent à un ensemble de critères permettant pour un utilisateur spécifique :
 - De mesurer la pertinence d'une information, et
 - D'évaluer si une information est plus pertinente qu'une autre information [MOU, 07].

Dans le contexte des base de données, les préférences utilisateur permettent d'ordonner les tuples de la réponse selon leur importance et ainsi de déterminer quels sont les tuples les plus intéressants.

Les préférences dans ce domaine sont exprimées sur le contenu (attributs, tuples ou requêtes) [BGM, 05].

4.2. Niveau des préférences

les préférences doivent spécifier le chemin des données que l'utilisateur désire analyser [GRB, 11]. Ainsi, les modèles des préférences OLAP ont porté sur le schéma ainsi que les valeurs de la base de données multidimensionnel.

Les préférences sur le schéma de la BDM sont définies sur deux niveaux :

- Les préférences sur les dimensions décrivent l'ensemble des dimensions pertinentes pour l'analyse d'un fait [BGM, 06].
- Les préférences sur les paramètres spécifient les niveaux de granularité préférés au long d'une dimension [GRE, 09], [RAT, 08].

4.3. Expression des préférences

Ils existent plusieurs approches de classification des préférences. Selon le type de données utilisées pour la comparaison de deux éléments on distingue des préférences intrinsèques et extrinsèques.

- **Préférences intrinsèques** : portent uniquement sur les valeurs des attributs d'un élément.
- **Préférences extrinsèques** : prennent en compte des facteurs extérieurs comme l'origine d'un élément [BOU, 10].

Une autre manière de classer les approches de personnalisation est basée sur la manière de comparer deux éléments. On retrouve ici :

- **Approche qualitative** où la préférence est spécifiée directement par des relations binaires de préférence [KOS, 07].
- **Approche quantitative** où les préférences sont exprimées indirectement par le biais de fonctions qui attribuent un nombre à chaque élément et la pertinence d'un élément dépend du nombre attribué [KOS, 07].

II. Personnalisation des systèmes OLAP

Les SID sont censés être centrés utilisateur, l'OLAP classique ne dispose pas d'outils permettant de guider l'utilisateur vers les faits les plus intéressants du cube.

La prise en compte de l'utilisateur dans les SID est une problématique nouvelle, connue sous le nom de personnalisation, qui pose plusieurs enjeux peu ou pas étudiés.

1. Personnalisation de l'interrogation des données

Consiste à adapter l'évaluation d'une requête par rapport aux caractéristiques et aux préférences de l'utilisateur qui l'a émise. Dans ce contexte, le système réagit à une demande spécifique de l'utilisateur en transformant sa requête afin de la rendre plus précise, en personnalisant l'affichage des résultats [BGM, 06].

Les travaux de personnalisation de l'interrogation des données se situent au niveau « restitution et analyse » du système OLAP. Ainsi, ces travaux permettent de personnaliser l'interrogation du schéma et/ou des instances de la base de données multidimensionnelles.

Deux catégories de travaux peuvent être distinguées [JER, 12]:

- Des travaux permettant la personnalisation des requêtes de l'utilisateur

- Des travaux visant à assister l'utilisateur dans la définition des requêtes, appelés communément des travaux de recommandation.

2. Personnalisation des requêtes

La personnalisation de requête est basée sur le constat que des usagers peuvent juger des résultats différents pertinents lors du requêtage des données [PSC, 02]. L'objectif de cette approche est de restituer les données les plus pertinentes pour chaque usager.

2.1. Définition : La personnalisation de requête est un mécanisme effectué avant ou après l'évaluation de la requête afin de changer la requête ou l'ordre du résultat [JER, 12].

2.2. Approches de personnalisation des requêtes

Deux approches de personnalisation des requêtes OLAP ont été proposées:

2.2.1. Expansion de requêtes

Les méthodes d'expansion des requêtes supposent l'existence d'un profil de l'utilisateur. Des éléments du profil sont utilisés pour étendre la requête de l'utilisateur. Ceci se traduit par l'ajout de conditions de sélection [BGM, 05], [BGM, 06], d'attributs d'agrégation [RTZ, 07]. La version étendue de la requête est ensuite exécutée en substitution de la requête initiale afin de générer un résultat adapté à l'utilisateur.

Le processus de personnalisation de requêtes par expansion se déroule en deux étapes:

- Sélection d'un sous-ensemble des éléments du profil ou des préférences définies en ligne qui sera utilisé pour personnaliser la requête.
- Intégration des éléments sélectionnés dans la requête initiale.

2.2.2. Requêtes de tri

La catégorie des requêtes de tri la plus étudiée est celles des requêtes Top-K qui permettent de renvoyer seulement les k meilleurs objets du résultat [LWL, 07], [LYA, 07], [XIH, 08]. Les critères de tri sont définis en tant que conditions faibles afin de rendre la sélection des n-uplets flexible.

Les requêtes top-k permettent de trier le résultat selon des fonctions de score. Ces fonctions portent sur un ou plusieurs attributs (par exemple, les 10 appartements avec le loyer le moins cher et la surface la plus grande), d'une ou plusieurs tables de la base

(par exemple, les 10 appartements avec le loyer et le coût de vie de la ville les moins chers).

III. Recommandation de requêtes

La recommandation de requête est l'action de proposer à l'utilisateur une requête ou des parties de requête d'une manière adaptée à ses intérêts et/ou à son analyse en cours afin de l'assister dans l'exploration des données [JER, 12].

La recommandation de requête fournit deux fonctionnalités :

- 1) L'assistance à la définition de requête par la proposition de parties de requête.
- 2) La proposition de requêtes complètes afin de faciliter l'exploration de l'espace multidimensionnel. Cependant, seule la deuxième fonctionnalité a été étudiée par la communauté des bases de données multidimensionnelles [GMN, 08], [GMN, 09].

1. Système de recommandation

La recommandation de requêtes dans les bases de données est un axe de recherche prometteur [KBG, 09], [CEP, 09] en particulier dans les systèmes OLAP où l'utilisateur navigue interactivement dans un cube en lançant une séquence de requêtes sur un entrepôt de données [SAR, 00].

1.1. Définition

Robin Burke définit d'une manière générale un système de recommandation comme « tout système capable de générer des recommandations ou permettant de guider l'utilisateur vers des objets utiles au sein d'un espace de données important » [BUR, 02].

La recommandation est un service de personnalisation qui consiste à proposer à l'utilisateur des éléments vis à vis de ses préférences ou en se servant de l'expérience des autres utilisateurs [BOU, 10].

1.2. Classification des approches de recommandation

Il existe plusieurs familles de systèmes de recommandation, en fonction de la manière dont la recommandation est effectuée et de la nature des données :

1.2.1. Recommandation basée sur le contenu

Consiste à proposer à l'utilisateur des objets qui sont similaires à ceux qu'il a appréciés dans le passé [MAE, 94], [PAB, 07]. Une mesure de similarité entre les

objets est souvent utilisée afin d'identifier ceux qui sont susceptibles d'être utiles pour l'utilisateur.

1.2.2. Recommandation basée sur le filtrage collaboratif

Exploite les appréciations d'une communauté d'utilisateurs sur les objets afin de découvrir des corrélations entre les utilisateurs. L'utilisateur courant se verra recommander des objets que des utilisateurs similaires ont appréciés [KMM, 97].

Ainsi une mesure de similarité entre utilisateurs est généralement établie [SER, 06].

1.2.3. Recommandation hybride

Des approches qualifiées d'hybrides combinent entre le calcul basé sur le contenu et le filtrage collaboratif [BAS, 97].

IV. Personnalisation VS recommandation

Les deux processus de recommandation et de personnalisation de l'information sont souvent fortement liés. Cependant, ils diffèrent sur le contenu de la réponse [MOU, 07]:

- La personnalisation permet de réduire le contenu du résultat qui consiste en un sous ensemble de la réponse à la requête initiale posée par l'utilisateur,
- La recommandation permet de compléter le résultat personnalisé. Elle consiste à enrichir le contenu du résultat avec de nouvelles informations qui ne sont pas demandées par la requête initiale.

V. Travaux existants dans les systèmes OLAP

Plusieurs travaux sont en cours pour la personnalisation de l'information. Ces travaux visent à personnaliser l'information.

Travaux de [BGM, 05a]

La proposition de Travaux de [GMN, 08] qui permet la recommandation de requêtes pour anticiper sur une séquence de requêtes d'un utilisateur grâce à l'analyse des historiques de navigations réalisées par les autres utilisateurs.

Chapitre 02 : Personnalisation dans les systèmes OLAP

Travaux de [GPM, 09]

La personnalisation selon [GPM, 09] présentent une approche pour personnaliser les bases de données multidimensionnelles à un niveau d'abstraction conceptuel, la personnalisation est basée sur l'utilisation d'évènements (fonctions représentant les opérations OLAP telles que roll up, sélection, rotate, drill down...) et de conditions (fonctions mettant à jour les données en fonction des évènements et du profil utilisateur) cette approche permet de créer des cubes de données dans les deux types de personnalisation (1) statique (créés divers cubes d'OLAP pour différents utilisateurs dans le design-time), (2) dynamique (créé un cube de données pour chaque utilisateur pendant le temps d'exécution en tenant compte les besoins et les actions pris par l'utilisateur).

Travaux de [RAT, 08]

Ravat [RAT, 08] a proposé un modèle conceptuel, un modèle de question et une personnalisation des bases de données multidimensionnels. Ce modèle est basé sur les concepts de base de multidimensionnel (fait, dimension, hiérarchie, mesure, attribut de poids) et règles de personnalisation. Les règles sont basées sur le formalisme de l'événement-condition-action (ECA) et affectent des poids de la priorité aux attributs de schéma multidimensionnels (approche quantitative) afin d'exprimer des préférences d'utilisateur. Ces poids sont chargés et stockés dans la base de données, les valeurs de ces derniers sont explicitement exprimées par des utilisateurs.

Synthèse

Afin de comparer les différents travaux, nous positionnons ces travaux, dans le tableau 04, par rapport aux principes de la personnalisation présentés dans la (figure 10).

Chapitre 02 : Personnalisation dans les systèmes OLAP

Références	Définition du profil		Exploitation du profil	
	Apprentissage	Paramétrage	Transformation	Recommandation
[Bellatreche et al., 2005]		Contrainte de visualisation.	Modification table résultat.	
[Espil et Vaisman, 2001]		Règles d'exception.	Modification restitution.	
[Favre et al., 2007]		Règle si-alors.	Modification schéma entrepôt.	
[Giacometti et al., 2008]	Suite de requêtes.			Requêtes suivantes.
[Jerbi et al., 2008-2009]		Ordre des préférences selon contexte.	Modification table résultat.	Requêtes suivantes, enrichies, alternatives
[Ravat et al., 2008]		Poids sur les éléments du schéma.	Modification table résultat.	
[Thalhammer et al., 2001]		Scénario d'analyse, règles ECA.	Génération de rapports.	

Tableau 04 : comparaison des travaux de personnalisation [BBF, 09].

Conclusion

La personnalisation vise à trouver rapidement un ensemble réduit de réponses intéressantes et pouvant satisfaire l'utilisateur par rapport à ses attentes, mais il s'est avéré que la personnalisation à elle seule ne peut pas être toujours suffisante car on pourrait donner la même réponse personnalisée à plusieurs usagers ayant le même profil mais qui ne se trouvent pas dans le même contexte. C'est pour quoi les travaux sont orientés vers la génération des nouveaux systèmes qui ont appelés les systèmes contextualisés.

Chapitre03:

*Contextualisation dans les systèmes
OLAP*

Introduction

Les systèmes d'accès d'information basé sur le processus de personnalisation des systèmes OLAP permet la mise en œuvre de système centré utilisateur non dans le sens d'un utilisateur générique mais d'un utilisateur spécifique pour cela en vue adapter son fonctionnement à son contexte.

En clair, le problème n'est pas dans la disponibilité d'information mais sa pertinence relativement à un contexte particulière. Pour cela, les travaux sont orientés vers la conception d'une nouvelle génération des systèmes contextualisés.

Contextualiser le traitement de l'information. Cette initiative a été largement encouragée ces dernières années, par l'explosion phénoménale des nouveaux types de services. En effet, la situation est actuellement paradoxale : la masse d'informations est telle que l'accès à une information pertinente, adaptée aux besoins spécifiques d'un utilisateur donné, devient à la fois difficile et nécessaire.

Une préférence peut être associée à un contexte. Dans ce cas, elle est dite contextuelle (ou conditionnelle). Le contexte d'une préférence définit sa portée, c'est à-dire l'environnement dans lequel elle doit être prise en compte.

I. Définition du contexte

Le contexte est une question clé dans l'interaction entre l'humain et l'ordinateur, décrivant les faits environnants qui ajoutent la signification. Les divers domaines de l'informatique avaient étudié ce concept au cours des 40 dernières années, pour relier le traitement de l'information et la communication aux aspects des situations dans lesquelles un tel traitement se produit.

Le contexte selon **Dey et al. 2001** : toute information pouvant être utilisée pour caractériser la situation d'une entité (personne, objet physique ou informatique). Et plus généralement tout élément pouvant influencer le comportement d'une application [DEY, 01].

II. Différents types d'éléments contextuels

Pour structurer le concept du contexte [LLT, 05] proposons le modèle suivant :

En termes de ce modèle, un espace hiérarchiquement organisé de dispositif pour le contexte peut être développé. Au niveau supérieur nous proposons de distinguer le contexte lié à l'environnement physique, et le contexte lié aux facteurs humains dans le sens le plus large puis le contexte informatique et par la suite le contexte temporel.

Dans chaque catégorie (les quatre catégories générales), des dispositifs appropriés peuvent être identifiés, encore hiérarchiquement, dont les valeurs déterminent le contexte comme représenté dans le schéma.

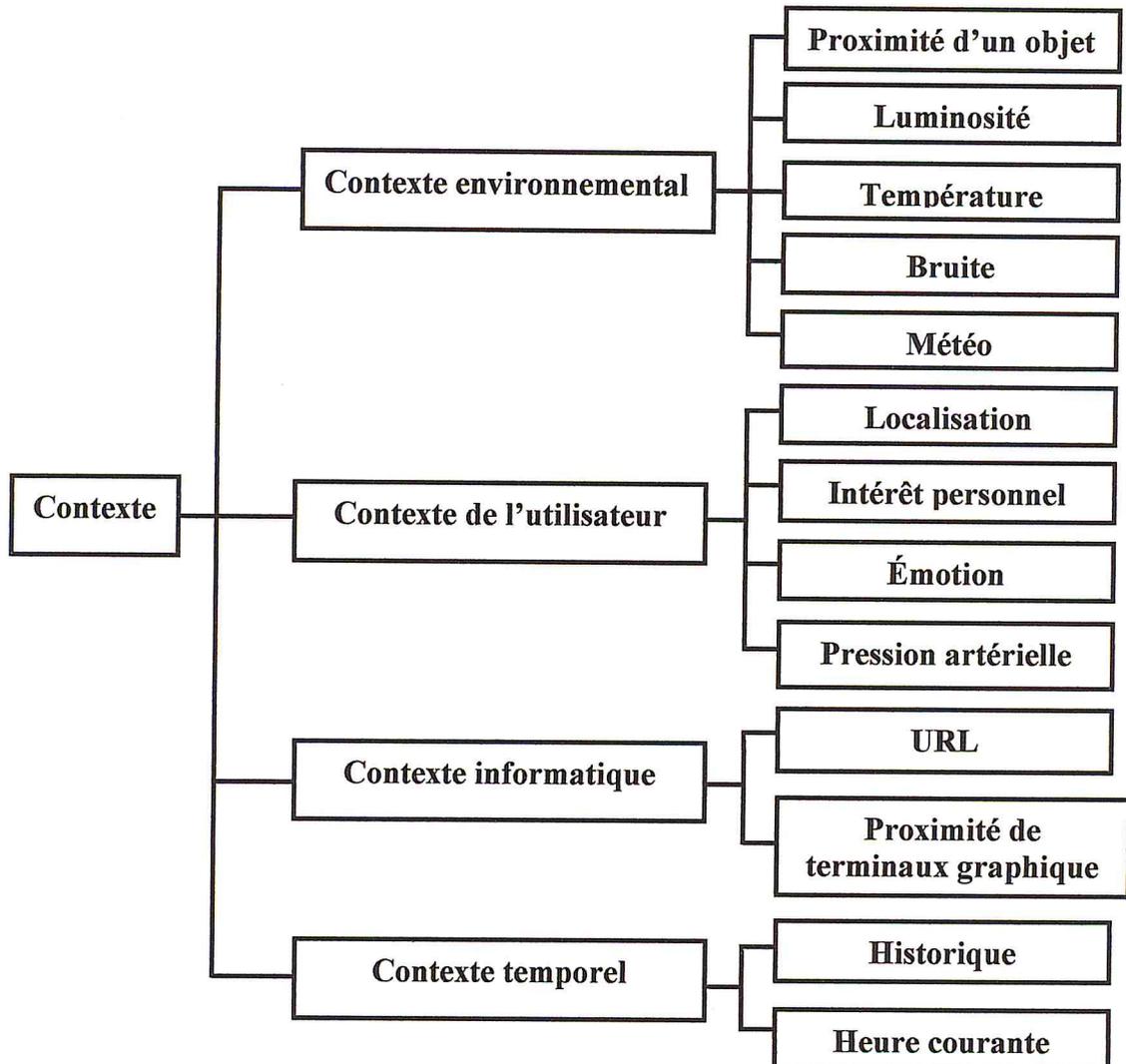


Figure 11: schéma des éléments contextuels [LLT, 05].

III. Système sensible au contexte

Puisque le concept général du contexte est très large, nous essayons de concentrer sur ces domaines qui sont directement liés aux systèmes de recommandation contextualisés, tels que la personnalisation d'e-commerce, les bases de données, recherche documentaire, Systèmes sensible au contexte Ubiquitaire et mobile, le marketing, et management :

1. Personnalisation d'e-commerce

Des systèmes de Recommandation sont également liés à la personnalisation d'e-commerce, puisque des recommandations personnalisées de divers produits et de services sont fournies aux clients. L'importance d'inclure et d'employer l'information contextuelle dans des systèmes de recommandation a été démontrée dedans [ADT, 05].

Palmisano et al [PTG, 08] emploient l'intention d'un achat fait par un client dans une application d'e-commerce en tant qu'information contextuelle. Les différentes intentions d'achat peuvent mener à différents types de comportement. Par exemple, le même client peut acheter par le même compte différents produits pour différentes raisons: un livre pour améliorer ses qualifications personnelles de travail, un livre comme cadeau. Pour traiter différentes intentions d'achat, Palmisano et al [PTG, 08] établissent un profil séparé d'un client pour chaque contexte d'achat, et ces profils séparés sont utilisées pour établir des modèles séparés de prévision de comportement du client dans des contextes spécifiques et pour des segments spécifiques des clients.

2. Recherche documentaire

L'information contextuelle est utilisée dans la recherche documentaire et l'accès aux documents [JGG, 05], bien que la plupart des systèmes existants basent leurs décisions de récupérer seulement des requêtes et des collections de document, tandis que des informations sur le contexte de recherche sont souvent ignorées [ALG, 95].

La plupart des techniques de contextualisation d'accès et de récupération de l'information courantes se concentrent sur les problèmes à court terme et les intérêts et les demandes immédiats d'utilisateur, et ne sont pas conçues pour modeler les intérêts et les préférences à long terme d'utilisateur.

3. Systèmes sensible au contexte Ubiquitaire et mobile

Brown et al [BCB, 05] présentent une autre application intéressante qui laisse des touristes partager interactivement leurs expériences guidées avec les utilisateurs à distance, démontrant la valeur que les techniques de contextualisation peuvent fournir en soutenant des activités sociales.

4. Base de données

Des possibilités contextuelles ont été ajoutées aux certains systèmes de gestion de base de données en saisissant des préférences d'utilisateur et en renvoyant différentes réponses aux requêtes selon le contexte en lequel les requêtes ont été exprimées et les préférences particulières d'utilisateur correspondant aux contextes spécifiques.

Stephanidis et al [SPV, 07] présentent une extension contextualisation de SQL pour adapter de telles préférences et des informations contextuelles. Plus spécifiquement, un ensemble de paramètres contextuels est présenté et des préférences sont définies pour chaque combinaison des attributs relationnels réguliers et de ces paramètres contextuels.

5. Marketing et Management

Les chercheurs de vente ont maintenu que le processus d'achat est dépendant du contexte dans lequel la transaction est met en place, puisque le même client peut adopter différentes stratégies de décision et préférer différents produits ou marques selon le contexte [BLP, 91], [LOT, 79].

Selon Lilien et al [LKM, 92], les clients changent dans leurs règles de prise de décision en raison de la situation d'utilisation, de l'utilisation du bon ou du service (pour la famille, pour le cadeau, pour l'individu) et de la situation d'achat (la vente par correspondance, et l'achat par facilité).

IV. Contextualisation

Les préférences des utilisateurs peuvent changés selon le contexte où elles sont utilisées. Pour cela il est impératif de faire dépendre les préférences avec le contexte.

V. Acquisition de l'information contextuelle

L'acquisition de l'information contextuelle ce fait par deux manières :

1. Acquisition explicite

Le contexte est acquis explicitement en exigeant de l'utilisateur de l'indiquer.

Il existe plusieurs manières pur fournir ce contexte soit par :

1.1. Interaction manuelle

Ce type d'interaction consiste à laisser le soin à l'utilisateur de juger de ce qui l'intéresse ou non. Le système propose par exemple la liste de tous les données qu'il a pu recenser en réponse à une demande de recherche, charge à l'utilisateur de valider ensuite ceux qui correspondent à ses besoins [CHS, 98] [SHM, 95].

1.2. Interaction semi-automatique

Le système effectue un filtrage des informations : il propose uniquement celles qui lui semblent répondre aux besoins de l'utilisateur, lequel confirme ou infirme ensuite explicitement, en fonction de ce qui lui convient effectivement.

2. Acquisition implicite

Le contexte est acquis implicitement en surveillant l'utilisateur et l'activité sur ordinateur, pour déterminer ce contexte en doit basé sur :

Interaction automatique : Le système utilise l'historique des actions de l'utilisateur (sites fréquemment visités, liens suivis...) pour enrichir le modèle par inférence [ALG, 95] [GOS, 00]. L'utilisateur n'intervient pas explicitement dans la construction du modèle.

3. Déduction de l'information contextuelle (Statistiques et de Data Mining)

Afin de déduire cette information contextuelle, il est nécessaire de construire un modèle prédictif (ie, un classificateur) et former sur les données appropriées. Le succès de déduire cette information contextuelle dépend beaucoup de la qualité de ce classificateur, et elle varie aussi considérablement selon les différentes applications. Il a été démontré dans [PTG, 08] que les différents types d'informations contextuelles peuvent être déduits avec un degré raisonnable de précision élevé dans certaines applications et l'utilisation de certaines méthodes de fouille de données, tels que les classificateurs de Naïve Bayes et réseaux bayésiens.

VI. Méthodes d'intégration de l'information contextuelle

Le processus de contextualisation de recommandation peut prendre l'un des trois formes :

1. Pré-filtrage contextuel

Dans ce paradigme de recommandation, l'information contextuelle conduit le choix de données ou la construction de données pour ce contexte spécifique. En d'autres termes, des informations sur le contexte courant c sont employées pour choisir ou construire l'ensemble approprié d'enregistrements (c.-à-d., estimations). Puis, des estimations peuvent être prévues en utilisant n'importe quel système traditionnel de recommandation sur les données choisies.

Comme représenté dans la figure 12, l'approche pré-filtrage contextuelle emploie l'information contextuelle pour choisir ou construire les 2D données (User \times Item) les plus appropriées pour produire des recommandations. Un avantage principal de cette approche est qu'il permet le déploiement d'un quelconque des nombreuses techniques traditionnelles de recommandation. En particulier, dans une utilisation possible de cette approche, le contexte c sert essentiellement comme une requête de sélection (filtrage) des données d'estimation pertinentes [ADT, 11].

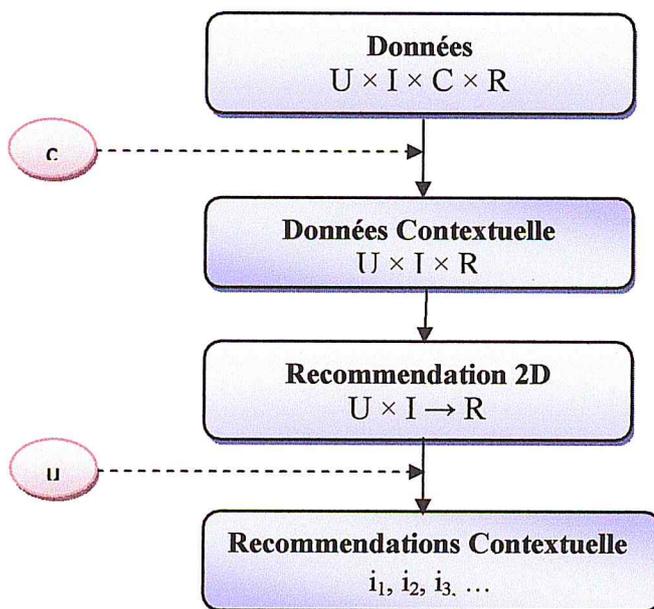


Figure 12: pré-filtrage Contextuel [ADT, 11].

1.1. Pré-filtrage exact

Un exemple d'un filtrage contextuel de données pour un système de recommandation de film serait : si une personne veut voir un film le samedi, seulement les données d'estimation de samedi sont employées pour recommander des films. Noter que cet exemple représente un **pré-filtrage exact**. En d'autres termes, la question de filtrage de données a été construite en utilisant exactement le contexte indiqué.

L'utilisation de ce **contexte exact** comme une requête de filtrage des données peut être une problématique pour plusieurs raisons.

Premier, certains aspects du contexte trop spécifique ne peuvent pas être significatifs. Par exemple, les préférences de l'utilisateur d'observation du film avec un ami dans un théâtre le samedi peuvent être exactement les mêmes que le dimanche, mais différentes de mercredi. Par conséquent, il peut être plus approprié d'employer des spécifications plus générales de contexte, c.-à-d., week-end au lieu de samedi.

Et en second, le **contexte exact** peut ne pas avoir assez de données pour la prévision précise d'estimation, qui est connue comme problème de " sparsity " en littérature de systèmes de recommandation. En d'autres termes, le système de recommandation peut ne pas avoir assez de points de repères au sujet des préférences d'observation de film passé d'un utilisateur donné avec une petite amie dans un théâtre le samedi [ADT, 11].

1.2. Pré-filtrage généralisé

Adomavicius et al [ADT, 05] présentent la notion du pré-filtrage généralisé, qui laisse généraliser la requête de données filtré obtenue basée sur un contexte indiqué. Plus formellement, nous laisser définissent $c' = (c'_1, \dots, c'_k)$ pour être une généralisation de contexte $c = (c_1, \dots, c_k)$ si et seulement si $c_i \rightarrow c'_i$ pour le chaque $i = 1, \dots, k$ dans la hiérarchie correspondante de contexte. Puis, c' (au lieu de c) peut être employé comme requête de données pour obtenir des données mises dans un contexte d'estimations.

Après l'idée de la généralisation de contexte, Adomavicius et al ont proposé d'employer pas un pré-filtrage simple [temps = t], qui représente le contexte exact t de l'estimation (u, i, t), mais plutôt un pré-filtrage généralisé [temps $\in S_t$], où S_t représente un certain ensemble du contexte t (s'appelle un segment contextuel).

Par exemple, si nous voudrions prévoir combien de John Doe voudrait voir le film de "gladiateur" lundi, c.-à-d., pour calculer $R_{User \times Item \times Time}^D(\text{JohnDoe}, \text{Gladiateur}, \text{Lundi})$, nous pourrions employer non seulement d'autres estimations personnalisées par l'utilisateur de lundi pour la prévision, mais estimations de jour de la semaine (weekday) en général. En d'autres termes, pour chaque (u, i, t) où t \in Jour de la semaine (weekday), nous pouvons prévoir l'estimation comme

$$R_{User \times Item \times Time}^D(u, i, t) = R_{User \times Item}^{D[Time \in Weekday](User, Item, AGGR(Rating))}(u, i). \text{ Plus}$$

généralement, afin d'estimer une certaine estimation R (u, i, t), nous pouvons utiliser une certaine S_t contextuelle spécifique de segment comme:

$$R_{User \times Item \times Time}^D(u, i, t) = R_{User \times Item}^{D[Time \in S_t](User, Item, AGGR(Rating))}(u, i).$$

Note, celle nous avons employé la notation d'AGGR(Rating) dans les expressions ci-dessus, puisqu'il peut y avoir plusieurs estimations personnalisées par l'utilisateur avec les mêmes valeurs d'utilisateur et d'items pour différents instances de temps dans l'ensemble de données D appartenant à une certaine S_t contextuelle de segment (par exemple, différentes estimations pour lundi et mardi, appartenant tout au jour de la semaine (weekday) de segment).

2. Post-filtrage contextuel

Comme représenté sur la figure 13, l'approche de post-filtrage contextuelle ignore l'information de contexte dans les données d'entrée en produisant des recommandations, c.-à-d., quand produire la liste rangée de tous les Items de candidat à partir desquels tout nombre de recommandations de top-N peut être fait, selon des valeurs spécifiques de N. Puis, l'approche de post-filtrage contextuelle ajuste la liste obtenue de recommandation à chaque utilisateur en utilisant l'information contextuelle [ADT, 11].

Les ajustements de liste de recommandation peuvent être faits par :

- Filtrant hors des recommandations qui sont non pertinentes (dans un contexte donné), ou
- Ajustement du rang des recommandations concernant la liste (basée sur un contexte donné).

Par exemple, dans une application de recommandation de film, si une personne veut voir un film un week-end, et des week-ends où elle observe seulement des comédies, le système peut filtrer dehors tous les non-comédies de la liste recommandée de film. Plus généralement, l'idée fondamentale pour des approches de post-filtrage contextuelles est d'analyser les données de préférence contextuelles pour un utilisateur donné dans un contexte donné pour trouver les habitudes d'utilisation d'items spécifiques (par exemple, l'utilisateur Jane observe seulement des comédies des weekends) et puis utiliser ces modèles pour ajuster la liste d'items, résultant en plus des recommandations contextuels [ADT, 11].

Les approches de post-filtrage peuvent établir les modèles prédictifs qui calculent la probabilité avec laquelle l'utilisateur choisit un certain type d'item dans un contexte donné, c.-à-d., probabilité de l'importance (par exemple, probabilité de choisir des films d'un certain genre dans un contexte donné), et puis emploient cette probabilité pour ajuster les recommandations.

Panniello et al [PTG, 09] fournissent une comparaison expérimentale de la méthode de pré-filtrage exacte contre deux méthodes de post-filtrage différentes **Pondéré et filter**.

2.1. Méthode de post-filtrage Pondéré (Weight) réordonne les items recommandés par la pondération de l'estimation prédit avec la probabilité de pertinence dans ce contexte spécifique.

2.2. Méthode de post-filtrage Filtré (Filter) filtre les items recommandés qui ont de faibles probabilités de pertinence dans le contexte spécifique.

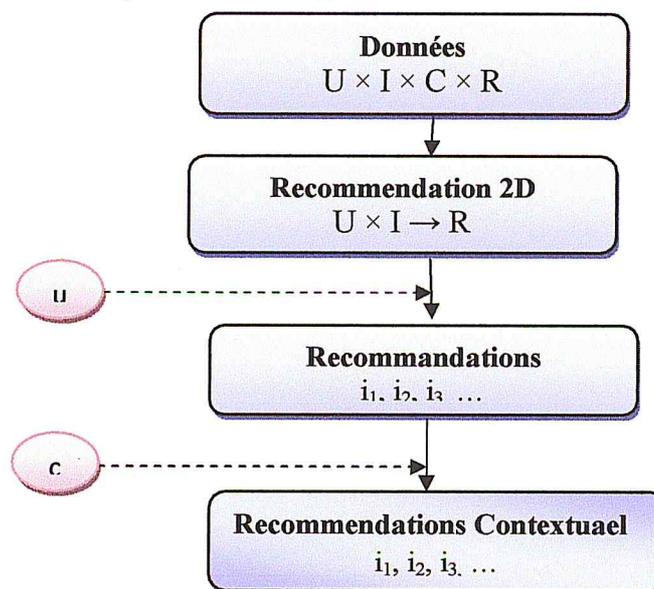


Figure 13: post-filtrage Contextuel [ADT, 11].

3. Modélisation contextuelle

Comme représenté dans la figure 14, l'approche de modélisation contextuelle utilise des informations contextuelles directement dans la fonction de recommandation comme un indicateur explicite de l'estimation d'un utilisateur pour un item. Alors que les approches de pré-filtrage et de post-filtrage contextuels peuvent utiliser des fonctions de recommandation 2D traditionnels, l'approche de modélisation contextuelle donne lieu à des fonctions de recommandation véritablement multidimensionnelle, qui représentent essentiellement des modèles de prévision (construit en utilisant l'arbre de décision, régression, modèle probabiliste, ou autre technique) ou calculs heuristiques qui intègrent des informations contextuelles en plus des données de l'utilisateur et d'item, c'est, Estimation = R (User, Item, contexte) [ADT, 11].

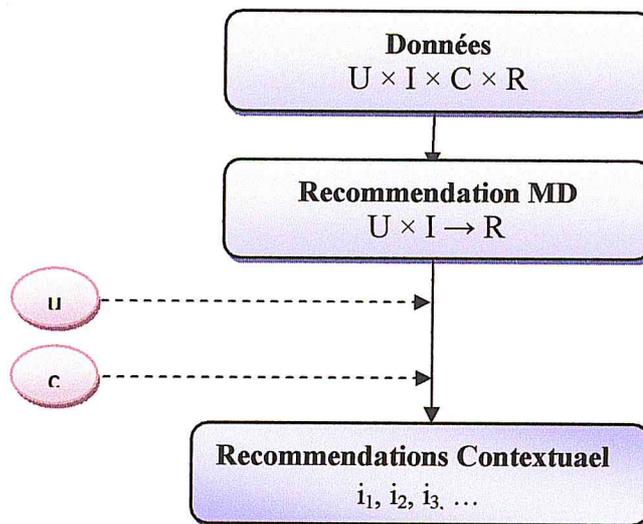


Figure 14: Modélisation Contextuel [ADT, 11].

VII. Systèmes OLAP et Datamining

L'exploitation de données, qui vise à extraire l'information intéressante à partir de grandes collectes des données, a été largement répandue comme outil efficace de prise de décision.

L'intérêt de couplage entre la fouille de données et OLAP réside principalement dans les caractéristiques des modèles multidimensionnels, spécialement conçue pour traiter de gros volumes de données agrégées et de nature multidimensionnel, des requêtes complexes, et de niveau de hiérarchie dans un but décisionnel.

1. Définition

Le datamining signifie littéralement « fouille de données » ou « forage de données ». Ce procédé, basé sur une série d'algorithmes, permet d'extraire des informations à partir de données, informations qui grâce à l'analyse, se convertissent en connaissances [CAN, 06].

2. Techniques du datamining

Le Datamining met en œuvre un ensemble de techniques issues des statistiques, de l'analyse de données et de l'informatique pour explorer les données.

2.1. Techniques descriptives (non supervisées)

On parle d'apprentissage non supervisé lorsque l'on cherche à extraire des informations nouvelles et originales d'un ensemble de données dont aucun attribut n'est plus important qu'un autre [GRS, 08].

Ils visent à structurer et à simplifier les données issues de plusieurs variables, sans privilégier l'une d'entre elles en particulier, il s'agit notamment de l'analyse en composantes principales (ACP), l'analyse factorielle des correspondances (AFC), l'analyse des correspondances multiples (ACM) et des méthodes de classification automatiques [BAC, 08].

Exemple d'algorithmes [GAM, 13]

- k -moyennes.
- Classification hiérarchique.
- Partitionnement de graphes et modularité.

2.2. Techniques prédictives (supervisées)

L'objectif général de la classification est d'être capable d'étiqueter des données en leur associant une classe. Si les classes possibles sont connues et si les exemples sont fournis avec l'étiquette de leur classe, on parle d'apprentissage supervisé ou d'analyse discriminante. Dans ce cas, il s'agit alors d'utiliser les exemples fournis et déjà classés pour apprendre un modèle qui permette ensuite d'associer à tout nouvel exemple rencontré sa classe la plus adaptée [LIA, XX].

Exemple d'algorithme [GAM, 13]

- Approche du centroïde.
- k -plus proches voisins.
- Classifieurs linéaires.

3. Data Mining et système OLAP (Travaux)

Le sujet de couplage entre la fouille de données et l'analyse en ligne OLAP est tout récent, peu sont les travaux réalisés dans ce contexte.

[GMN, XX]: Présentent un cadre générique qui permet de recommander des requêtes OLAP basé sur le journal de requête du serveur OLAP. Ce cadre est générique en ce sens que la modification de ses paramètres modifie la façon dont les recommandations sont calculées. Ils montrent comment utiliser ce cadre pour recommander des requêtes MDX simples. Ils utilisent un algorithme de classification simple K-medoids pour la division de l'ensemble de requête.

[AST, 11]: proposent une méthode automatique d'élucidation de préférence basée sur des techniques d'extraction. La méthode consiste à extraire un profil utilisateur à partir d'un ensemble d'échantillons de préférence d'utilisateur.

Dans leur arrangement, un profil est indiqué par un ensemble de règles contextuelles de préférence vérifiant quelques propriétés importantes (solidité et concision). Ils ont évalué l'efficacité de la méthode proposée dans une série d'expériences exécutées sur une base de données réelle des préférences d'utilisateur au sujet des films.

[JPL, 08]: Dans ce travail les auteurs proposent d'employer des échantillons de préférence fournis par l'utilisateur pour impliquer un ordre sur n'importe quelle paire de tuples dans la base de données. De tels sorte que les échantillons sont classifiés en deux catégories, supérieur et inférieur et contiennent des informations sur quelques tuples préférés et certains sont non préféré. De ces règles, un ordre est impliqué sur les tuples dans la base de données.

Après cette petite vue sur certains travaux réalisés dans le contexte de Datamining et les systèmes Olap nous remarquons que le Datamining touche seulement le domaine de personnalisation mais pas la contextualisation.

VIII. Contextualisation et système OLAP(Travaux)

Travaux de [JRT, 08] : Une solution de personnalisation qualitative est introduite par Jerbi et al. (2008). Il s'agit non plus d'exploiter des poids, mais plutôt des ordres (représentation qualitative des préférences), ce qui rend la tâche plus aisée pour l'utilisateur. En outre, ces ordres ne sont pas exprimés de façon absolue, mais par rapport à un contexte d'analyse donné. Ceci permet de prendre en compte le fait que les préférences peuvent varier d'un contexte d'analyse à l'autre.

Travaux de [TSM, 01] : Présentent un système à base d'entrepôt actif au sein duquel l'utilisateur doit spécifier ses scénarios d'analyse au travers d'un mécanisme de type ECA (Événement - Condition - Action). L'objectif est, par une meilleure connaissance des analyses effectuées sur l'entrepôt, d'améliorer le prétraitement des données. Au-delà de l'amélioration des performances, les auteurs proposent d'exploiter les résultats obtenus lors des analyses pour induire des changements dans les données opérationnelles.

Travaux de [BGM, 05] : L'approche de Bellatreche [BGM, 05] permet de personnaliser la requête ainsi que la visualisation du résultat. Elle consiste à transformer la requête de l'utilisateur par des prédicats issus du profil, l'utilisateur est interrogé afin de donner ses préférences et une contrainte de visualisation. Ces préférences utilisateurs qui classent les membres, les dimensions et la contrainte de visualisation qui contrôlent les résultats et qui sont stockées dans le profil utilisateur. Et une définition du profil utilisateur dans le contexte OLAP est proposée. Cette approche s'applique à une requête ponctuelle sans considérer les requêtes précédentes de l'analyse courante.

Synthèse

La contextualisation n'a été jamais utilisé dans les travaux précédents et le contexte dans les travaux de **Bellatreche [BGM, 05]** signifie un contexte de visualisation (taille de l'écran, etc.) tel que chaque dimension lui affecte un poids selon son importance puis les ordonne dans un ordre croissant.

Conclusion

Dans ce chapitre nous avons essayé de couvrir certains travaux, mais malheureusement ces travaux ne représentent pas la contextualisation qui présente des défis avec l'augmentation des nouvelles exigences et jusqu'à aujourd'hui il n'existe aucun travail dans le domaine des systèmes OLAP.

Dans notre travail nous avons trouvé une manière de contextualisation de systèmes OLAP, et dans le chapitre suivant nous allons présenter notre système expliquant en détail les différents modules de notre architecture.

Chapitre04:

Architecture d'un système personnalisé et contextualisée

Chapitre 04: Architecture d'un système personnalisé et contextualisée

Introduction

La contextualisation dans les systèmes OLAP est un domaine assez récent et présente plus de challenge. Ceci est du à plusieurs facteurs tel que : les problèmes d'extraction des données via une connexion, débit par exemple, ainsi que la masse importante des données par rapport les sources disponibles.

Alors nous devons prendre en charge ces facteurs pour réaliser un système performant capable au moins de diminuer quelque difficulté et de fournir à l'utilisateur des résultats pertinents et exploitables.

Pour cela nous proposant la figure ci-dessous qui représente l'architecture de notre système contextualisés.

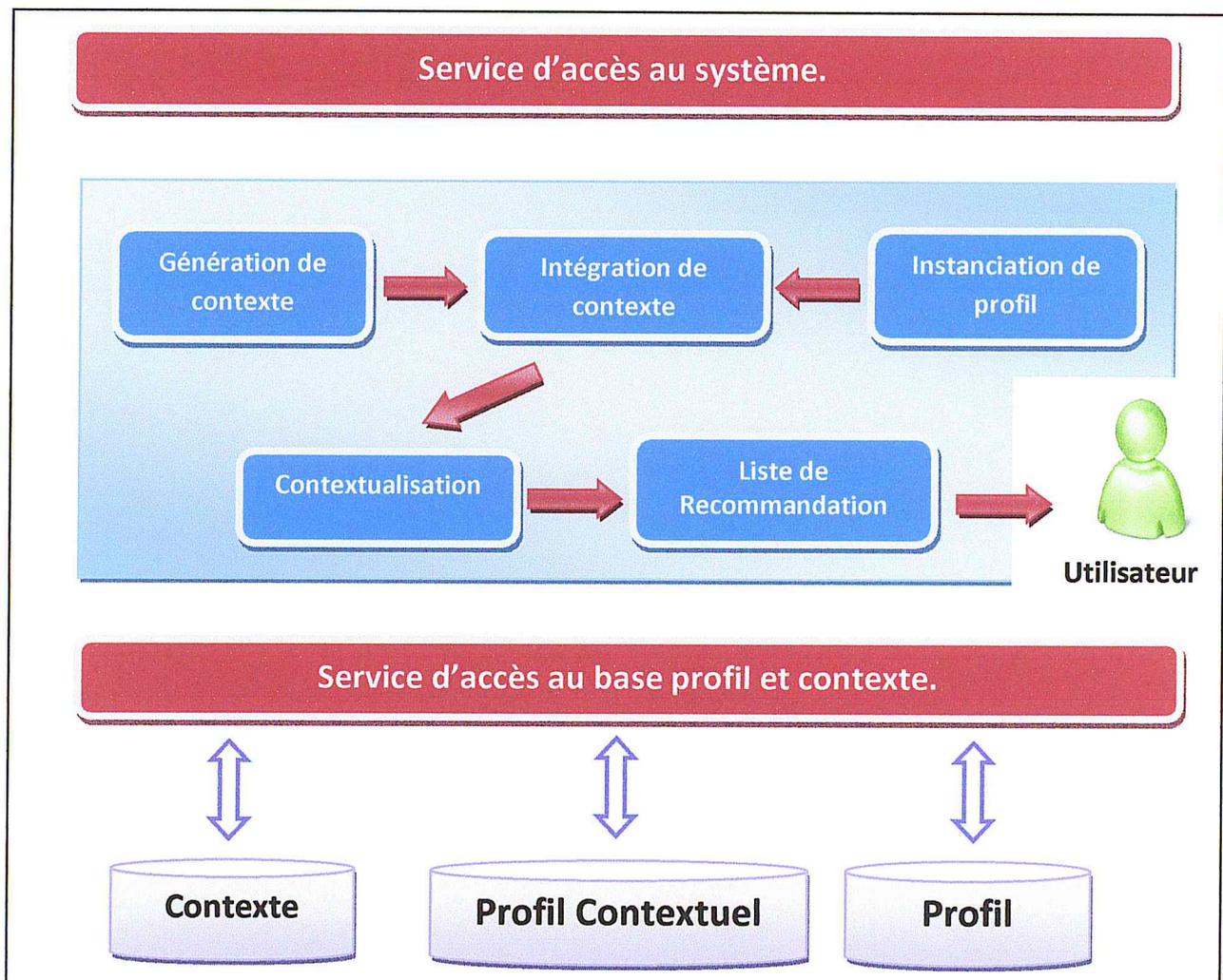


Figure 15: Architecture global du système.

Chapitre 04: Architecture d'un système personnalisé et contextualisée

Dans le schéma ci-dessus nous représentons les principaux modules de notre architecture dans lesquels chaque module a sa propre fonctionnalité comme suit :

- Module de stockage des profils et des contextes : ce module est constitué de trois bases de données dans le but de stocker les profils utilisateur, les contextes générés ainsi que les profils contextualisés (représentent le résultat de l'intégration de profil avec le contexte qui lui correspondre).
- Module de traitement : qui regroupe les phases suivantes :
 1. La génération automatique de contexte qui se fait par l'extraction des données à partir de fichier log de l'utilisateur courant.
 2. Instanciation de profil correspondant à l'utilisateur courant à partir de base de profils
 3. Intégration de contexte avec le profil qui nous a permet d'obtenir des préférences contextualisés permettant d'enrichir la requête de l'utilisateur.

I. Génération automatique de contexte

L'approche de la génération des contextes que nous proposons de découvrir est basée en premier (1) sur l'analyse d'un fichier Log W3C (fichier journal) pour capturer des contextes réguliers et en deuxième (2) choisir l'un des méthodes de classification on donne comme entré le fichier log et qui nous permettant de sortir un ensemble de cluster différents. Ci-après, la définition d'un fichier Log (W3C) et par la suite un exemple de motivation est donné pour comprendre l'approche.

1. Définition d'un fichier Log

Un fichier log est un fichier journal au format texte dans lesquels on peut stocker beaucoup d'information utile, et pouvant trier ces information ainsi que trouver les événements principaux qui représente la partie cruciale dans l'analyse d'un fichier log. Puisque les fichiers logs sont la source la plus critique des données pour la plupart des systèmes de technologie d'information, ils doivent être correctement contrôlés [CER, 08].

Chapitre 04: Architecture d'un système personnalisé et contextualisée

La découverte de contexte à partir des logs nécessite la manipulation du fichier log bien structuré et l'analyse de fichier log le plus connu compose comme Apache CLF, IIS LF et W3C LF [PHB, 08] indique que la plupart d'entre eux contient déjà l'information contextuelle (champs) qui correspondent au méta modèle de contexte proposé dedans [ABK, 08].

2. World Wide Web Consortium (W3C)

World Wide Web Consortium (W3C) a été créé en octobre 1994 pour mener le World Wide Web à sa pleine capacité en développant les protocoles communs qui favorisent son évolution et assurent son interopérabilité. W3C a plus de 400 organismes de membre dans le monde et a gagné l'identification internationale pour ses contributions à la croissance de web [ORE, 03].

3. Format d'un fichier Log W3C (World Wide Web Consortium)

Le format d'un fichier Log que nous avons employé est basée sur le format d'un fichier Log W3C qui sont assez extensibles (grâce à dix champs tout usage de W3C) pour l'adapter à de diverses applications. Parmi ces champs :

Date: le domaine de la date est organisé en hiérarchie, de façon que nous puissions savoir si c'est un jour ouvrable ou un week-end, vacances ou pas, etc.

Time: un jour est divisé dans des périodes qui sont organisées en hiérarchie. Le temps peut informer au sujet de la période de jour : matin, après-midi, nuit, etc.

Périphérique: caractérise le dispositif utilisé pour l'interaction d'utilisateur, dispositifs peut être segmenté dans une hiérarchie, on prend par exemples de périphérique: Laptop, Mobile, PC, etc.

La conjonction des ces attribut représente les contextes dans lesquels les utilisateurs agit l'un sur l'autre avec le système. Nous proposons de grouper ces contextes dans des clusters, chacun représente un contexte ou une situation régulière telle que la maison, bureau, etc.

Chapitre 04: Architecture d'un système personnalisé et contextualisée

4. Pourquoi ce type ?

- Accorder l'interopérabilité.
- Accès universel.
- Maintenir l'extensibilité.
- Une bonne fonctionnalité.

5. Exemple d'un fichier Log

Supposer qu'un utilisateur donnée, agit avec le système. Puis, grâce aux informations contenues dans le fichier Log, le système peut capturer les préférences de l'utilisateur qui changent selon la date l'heure et le dispositif, etc. ... Ceci peut être dirigé par la consultation de son fichier Log.

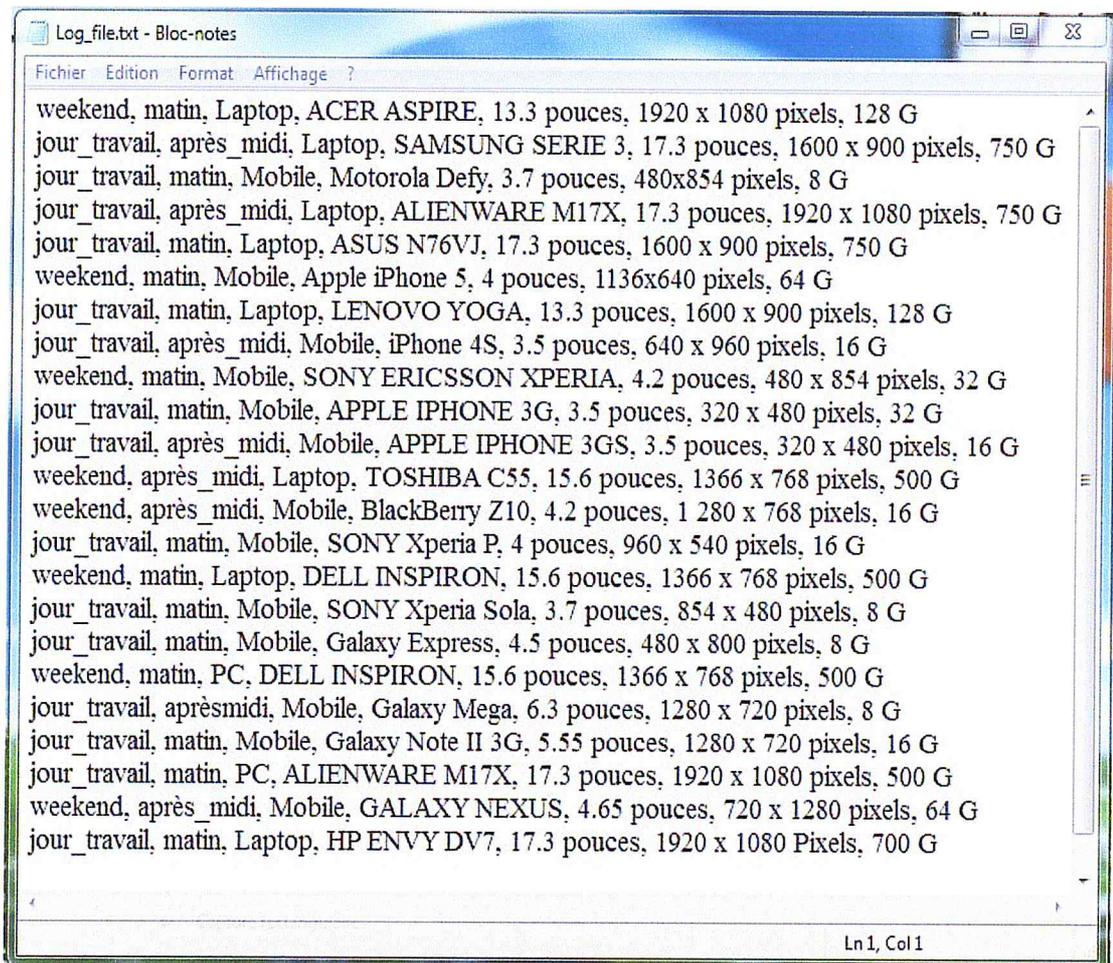


Figure 16: Fichier Log.



6. Classification automatique

Ce type de classification de données est l'un des méthodes statistiques largement utilisées dans le Data Mining. Elle est dans un cadre d'apprentissage non supervisé, qui a pour but de d'obtenir des informations sans aucune connaissance préalable sur les classes, au contraire de l'apprentissage supervisé [XAV, 03].

Dans notre approche on utilise un fichier log qui contient un ensemble d'information bien structurées, le type de ces données est un mélange de données numérique et catégorique. Pour analyser ces données on a choisi la classification K-means qui fonctionne bien le type de données numérique.

Méthode des K-moyennes (K-Means)

La méthode des k-moyennes est une méthode de classification qui permet de mettre au jour une éventuelle structure de groupes dans un ensemble de données [RUW, 12].

Algorithm K-means (log file L, seuil α) [ABL, 09].

Entré : fichier Log L, seuil α

Sortie : Ensemble de contexte $C = \{c_1, c_2, \dots, c_n\}$.

Initialisation : $CP_u \leftarrow \emptyset, \varepsilon \leftarrow \infty$.

CI \leftarrow **Contexte (L)** // Extraction des instances des contexte

C \leftarrow **CHA (CI, α)** // Appelez la classification hierarchique.

For all cluster $c_i \in C$ do

Calculer le centroïde du cluster

Repeat

$\varepsilon_{old} \leftarrow \varepsilon$

For all context instance $t_j \in CI$ do

Assiger chaque t_j au cluster qui lui le plus proche

For all cluster $c_i \in C$ do

Recalculer le centroïde du cluster

Return C

II. Contextualisation

L'idée principal du contextualisation est de vérifier s'il ya des corrélations entre les éléments de profil d'utilisateur et le fichier log (L) de l'utilisateur.

$M(P_u, C)$: Contextualisation (P_u, L).

Le processus de contextualisation prend comme entrer un profil utilisateur P_u et log d'utilisateur L correspond au contexte d'utilisateur dans les contextes C générés précédemment, il renvoie un ensemble de traces contextuels M représentent les dépendances entre les éléments dans P_u et les contextes C.

1. Profil Contextuel

Dans les systèmes de recommandation traditionnels un profil utilisateur est un ensemble d'estimation $P_u = \{E_1, E_2, \dots, E_n\}$, où chaque estimation E_i se compose d'un attribut att_i et un score sc_i (poids) :

$E_i = (att_i, sc_i)$ tel que :

Le score sc_i est un nombre réel exprimant l'importance de l'attribut att_i pour un utilisateur.

L'intégration de contexte dans le processus de personnalisation est basée sur ce concept, un profil utilisateur mis dans un contexte CP_u est défini comme l'ensemble M des traces (Mapping) contextuel qui relier des estimations d'utilisateur avec un ensemble de contexte C.

$$CP = \{M_{ij}(E_i, c_j, sc_{ij}) / E_i \in P_u, c_j \in C, sc \in [-1, 1]\}. \text{ [ABK, 09]}$$

Chaque trace contextuelle est défini par une estimation E_i , un contexte c_j et un score sc_{ij} . Le score est un nombre réel exprimant l'importance de tenir compte l'estimation E_i lorsque l'utilisateur interagit avec le contexte c_j .

Ci-après, nous décrivons la manière qu'un profil utilisateur contextuel est construit en utilisant un fichier log.

Algorithme Mapping [ABK, 09].

Entrer : Profil utilisateur $P_u = \{E_1, E_2, \dots, E_n\}$, Fichier log L.

Sortie : Profil Contextuel CP_u .

Initialisation : $CP_u \leftarrow \emptyset$, $sp \leftarrow 0$, $cf \leftarrow 0$.

$C \leftarrow$ Contexte (L).

- For all $E_i \in P_u$ do
 - For all $c_j \in C$ do
 - Calculer $sp^+ \leftarrow$ Support ($att_i \rightarrow +, c_j, L$).
 - Calculer $cf^+ \leftarrow$ Confiance ($att_i \rightarrow +, c_j, L$).
 - Calculer $sc_{ij}^+ \leftarrow$ Merge (cf^+, sp^+)
 - If $sc_{ij}^+ \geq \gamma$ Then
 $CP_u \leftarrow CP \cup (E_i, c_j, sc_{ij}^+)$
 - Calculer $sp^- \leftarrow$ Support ($att_i \rightarrow -, c_j, L$).
 - Calculer $cf^- \leftarrow$ Confiance ($att_i \rightarrow -, c_j, L$).
 - Calculer $sc_{ij}^- \leftarrow$ Merge (cf^-, sp^-)
 - If $sc_{ij}^- \geq \gamma$ Then
 $CP_u \leftarrow CP \cup (E_i, c_j, sc_{ij}^-)$
 - return CP_u .
-

Algorithme de mapping (Description)

Le profil contextuel pour chaque estimation d'utilisateur (plus précisément « attribut »), son score dans chaque contexte utilisateur. Noter que dans chaque contexte, les activités d'utilisateur peuvent être de deux sortes :

Positive (+) (Appartient au contenu aimé) et/ou négative (-) (relativement au contenu non désirer), selon le type d'action appliqué par l'utilisateur sur le contenu.

Par conséquent le score de chaque attribut att_i est calculé pour l'activité positive (resp. négative) dans un contexte donné en applique des règles d'association de la forme ($att_i \rightarrow +$) (resp. $att_i \rightarrow -$) puis l'importance att_i est obtenue par la combinaison de support (sp) et confiance (cf).

Chapitre 04: Architecture d'un système personnalisé et contextualisée

Recommandation

Les résultats fournis par l'algorithme de mapping (profil contextualisé) est combiné avec la requêtes de l'utilisateur pour construire une liste de recommandation qui va aider l'utilisateur lors de sa interaction avec le système.

Conclusion

Dans ce chapitre nous avons montré notre architecture de système contextualisée ainsi que les différents modules qui la regroupent.

Nous avons aussi cité en détail les algorithmes utilisés dans les différents modules de l'architecture et dans le chapitre qui suit, on entamant l'implémentation et les tests.

Chapitre05:

Application et Test

Introduction

Pour fournir aux utilisateurs des réponses à ces besoins on va utiliser un langage de requête MDX qui nous permettant de faciliter la navigation les informations et qui interroge avec une quantité importantes de données. Pour exécuter ce type de requêtes on doit utiliser l'environnement de développement Visual Studio 2008.

Le langage C# dans notre application est utilisé pour l'implémentation des algorithmes montrés dans le chapitre précédent.

1. Présentation de l'entrepôt utilisé

Les données à manipuler sont celles de « Adventure Works 2008 ». Entrepôt de données de Microsoft utilisé dans la documentation en ligne de SQL Server. L'entrepôt contient des données relatives à deux sous-domaines d'activité, les finances et les ventes.

1.1 Tables de fait de l'entrepôt

L'entrepôt contient des données relatives à deux sous-domaines d'activité, les finances et les ventes.

1.1.1 Finance

Dans l'entrepôt de données, les données financières sont réparties dans deux schémas qui présentent les caractéristiques suivantes :

Finance

- Contient les données financières de la société Adventure Works et de ses filiales.
- Contient les données en devise locale de l'organisation à laquelle elles sont associées.
- Prend en charge le groupe de mesures Finance Analysis Services.

Currency Rates

- Contient les données de conversion de devise, y compris les cours moyens journaliers et les cours en clôture par rapport au dollar US.
- Prend en charge le groupe de mesures Currency Rates Analysis Services.

1.1.2 Sales

Les données relatives aux ventes sont réparties en quatre schémas qui présentent les caractéristiques suivantes :

Reseller Sales

- Contient les ventes uniquement des revendeurs.
- Contient uniquement les commandes expédiées.
- Contient les données en dollar US et conserve ces données dans leur devise d'origine.
- Prend en charge le groupe de mesures Reseller Sales Analysis Services.

Sales Summary

- Contient une vue de synthèse des données des ventes effectuées par les revendeurs et des ventes Internet.
- Les dimensions sont limitées par rapport aux schémas des revendeurs et des ventes Internet.

Internet Sales

- Contient les commandes individuelles des clients Internet et des données détaillées.
- Contient uniquement les commandes expédiées.
- Contient les données en dollar US et conserve ces données dans leur devise d'origine.
- Prend en charge le groupe de mesures Internet Sales Analysis Services.

Sales Quota

- Contient les données des quotas de ventes des représentants.
- Prend en charge le groupe de mesures Sales Quotas Analysis Services.

2. Environnement de développement Visual Studio 2008

Visual Studio 2008 est un environnement de programmation riche en outils comportant toutes les fonctionnalités nécessaires pour créer des projets C# de toute taille. Vous pouvez même créer des projets qui combinent harmonieusement des modules compilés utilisant des langages de programmation différents.

2.1.Langage de programmation

C# est un langage récent. Il a été disponible en versions beta successives depuis l'année 2000 avant d'être officiellement disponible en février 2002 en même temps que la plate-forme .NET 1.0 de Microsoft à laquelle il est lié. C# ne peut fonctionner qu'avec cet environnement d'exécution. Celui-ci rend disponible aux programmes qui s'exécutent en son sein un ensemble très important de classes.

En première approximation, on peut dire que la plate-forme .NET est un environnement d'exécution analogue à une machine virtuelle Java. On peut noter cependant deux différences :

- Java s'exécute sur différents OS (windows, unix, macintosh) depuis ses débuts. En 2002, la plate-forme .NET ne s'exécutait que sur les machines Windows. Depuis quelques années le projet Mono [<http://www.mono-project.com>] permet d'utiliser la plate-forme .NET sur des OS tels qu'Unix et Linux. La version actuelle de Mono (février 2008) supporte .NET 1.1 et des éléments de .NET 2.0.
- La plate-forme .NET permet l'exécution de programmes écrits en différents langages. Il suffit que le compilateur de ceux-ci sache produire du code IL (Intermediate Language), code exécuté par la machine virtuelle .NET. Toutes les classes de .NET sont disponibles aux langages compatibles .NET ce qui tend à gommer les différences entre langages dans la mesure où les programmes utilisent largement ces classes. Le choix d'un langage .NET devient affaire de goût plus que de performances.

2.2. SQL Server Analysis Services 2008

Il fournit des fonctions OLAP et d'exploration de données pour les applications décisionnelles. Analysis Services prend en charge OLAP en permettant de concevoir, de créer et de gérer des structures multidimensionnelles qui contiennent des données agrégées provenant d'autres sources de données, telles que des bases de données relationnelles.

Serveurs OLAP

Le serveur utilisé dans notre application est SQL Server Analysis Services (SSAS).

SQL Server Analysis Services est la solution de traitement analytique en ligne (OLAP) de Microsoft, et c'est un outil de SQL Server. Des solutions OLAP sont au cœur de l'intelligence d'affaires à l'aide de données sur les activités à comprendre l'entreprise et à prendre des décisions qui feront la réussite en affaires.

SSAS permet aux utilisateurs de rassembler les données de plusieurs bases de données relationnelles dans un emplacement central et présente alors que les données d'une façon qui le rend facile pour les utilisateurs d'interroger les données, afficher les données à différents niveaux de détail et effectuer des calculs avancés contre les données, avec l'objectif de dérivation des informations utiles à partir des millions ou des milliards de transactions qui se produisent dans une entreprise sur une base quotidienne.

Outre ses fonctionnalités OLAP, fournit un ensemble riche de fonctionnalités d'exploration de données qui permettent aux utilisateurs de construire des modèles statistiques avancées de leurs données et d'utiliser ces modèles pour classer les données et faire des prédictions basées sur les données .

SSAS exécute des requêtes utilisant le langage MDX, Ce langage permet de créer des requêtes dont l'équivalent en langue SQL nécessiterait un grand nombre de requêtes et de temps d'exécution beaucoup plus longs.

2.3. Langage MDX

MDX est un langage de requêtes pour les bases de données multidimensionnelles, de la même manière que SQL est utilisé pour les requêtes sur les bases de données relationnelles. Dans son approche, MDX est proche du SQL sur son aspect select et where même si la similarité ne va pas plus loin. Le but des expressions multidimensionnelles MDX est de rendre aisé et intuitif l'accès aux données de différentes dimensions.

MDX est fait pour naviguer dans les bases multidimensionnelles et pour définir des requêtes sur tous les objets (dimensions, hiérarchies, niveaux, membres et cellules) afin d'obtenir une représentation sous forme de tableaux croisés.

Il en découle une approche très hiérarchisée. Tout d'abord un cube est composé de dimensions. Une dimension peut contenir une ou plusieurs hiérarchies.

3. Module de l'application

1. Génération automatique de contexte

1.1. Modèle de données (Log)

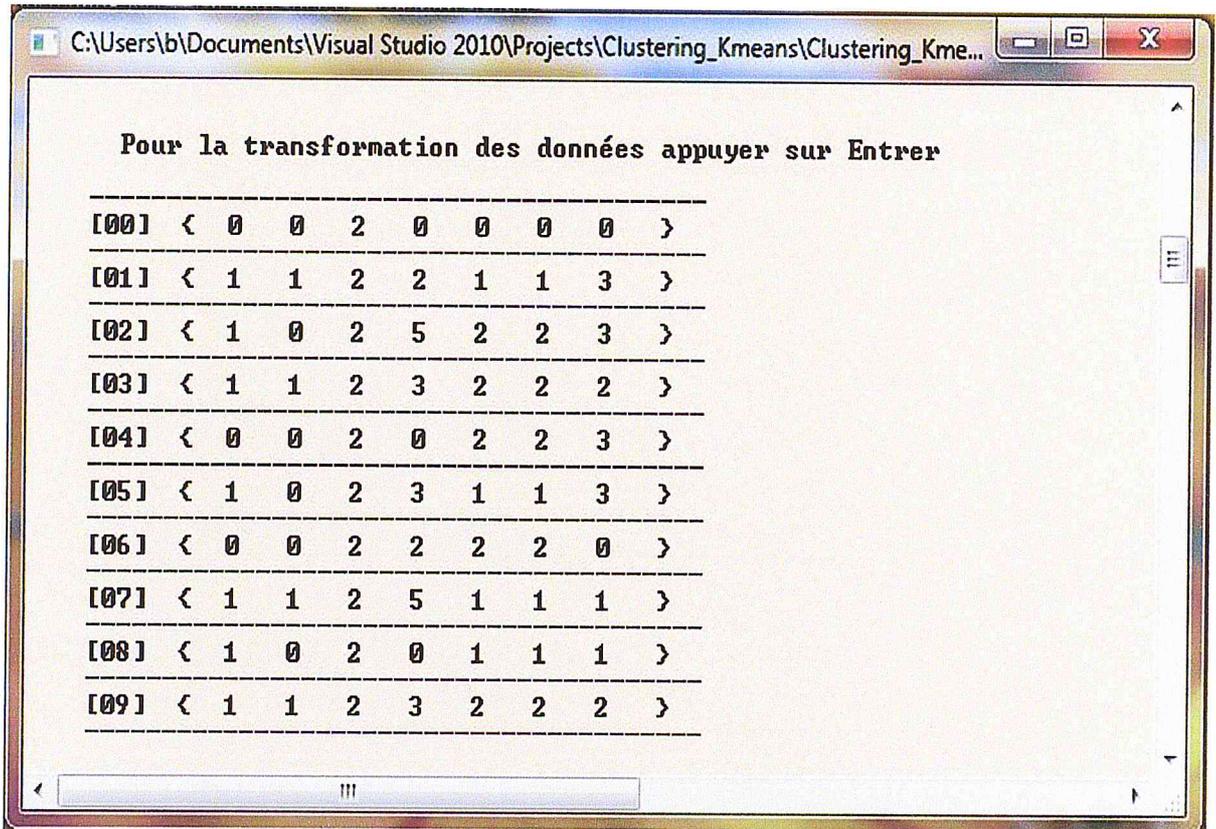
```
*****
* -----|| Le Clustering (K-means) ||----- *
*****
Pour Afficher les données taper Entrer:
-----
Jour      Heure      Type      Gamme      D_Screen      Resol      TM
-----
[00]==> < weekend      Matin      Laptop      ACER ASPIRE      13.3      1020*568      128      >
-----
[01]==> < jourtravail  AprèsMidi  Laptop      SAMSUNG SERIE 3  15.6      1366*768      750
-----
[02]==> < jourtravail  Matin      Laptop      ASUS N76UJ      17.3      1600*900      750      >
-----
[03]==> < jourtravail  AprèsMidi  Laptop      HP ENVY DU7      17.3      1600*900      700      >
-----
[04]==> < weekend      Matin      Laptop      ACER ASPIRE      17.3      1600*900      750      >
-----
[05]==> < jourtravail  Matin      Laptop      HP ENVY DU7      15.6      1366*768      750      >
-----
[06]==> < weekend      Matin      Laptop      SAMSUNG SERIE 3  17.3      1600*900      128      >
-----
[07]==> < jourtravail  AprèsMidi  Laptop      ASUS N76UJ      15.6      1366*768      500      >
-----
[08]==> < jourtravail  Matin      Laptop      ACER ASPIRE      15.6      1366*768      500      >
-----
[09]==> < jourtravail  AprèsMidi  Laptop      HP ENVY DU7      17.3      1600*900      700      >
-----
```

Figure 17: Modèle de fichier Log.

Description

Cet affichage illustre un exemple d'un fichier log qui compose un ensemble d'information sur l'interaction de l'utilisateur avec le système tel que le contexte date, contexte heure et le dispositif (type, gamme, display screen, résolution, taille mémoire).

1.2. Transformation des données



Pour la transformation des données appuyer sur Entrer

[00]	<	0	0	2	0	0	0	0	>
[01]	<	1	1	2	2	1	1	3	>
[02]	<	1	0	2	5	2	2	3	>
[03]	<	1	1	2	3	2	2	2	>
[04]	<	0	0	2	0	2	2	3	>
[05]	<	1	0	2	3	1	1	3	>
[06]	<	0	0	2	2	2	2	0	>
[07]	<	1	1	2	5	1	1	1	>
[08]	<	1	0	2	0	1	1	1	>
[09]	<	1	1	2	3	2	2	2	>

Figure 18: Matrice des transformations.

Chapitre 05 : Application et test

Description

Dans cet exemple on a ce tableau qui résume tous les transformations de données :

Nom de champ	Valeur de tuples	Type de transformation	Valeur après la transformation
Date	Weekend	Binaire.	0
	jour travail		1
Heure	Matin	Binaire.	0
	après midi		1
Type	Mobile	Discret (Ordinal).	0
	PC		1
	Laptop		2
Gamme	ACER ASPIRE, DELL NSPIRON 15-3521, SAMSUNG SERIE 3, etc	Discret (Ordinal).	0, 1, 2, 3, 4,etc.
Display screen	13.3, 15.6, 17.3, etc.	Discret (Ordinal).	0, 1, 2, 3, 4,etc.
Résolution	1020*568, 1366*768, 1600*900, etc.	Discret (Ordinal).	0, 1, 2, 3, 4,etc.
Taille mémoire	128, 500, 700, 750, etc.	Discret (Ordinal).	0, 1, 2, 3, 4,etc.

Tableau 05: Tableau descriptive des transformation des données.

2. Classification des données

jourtravail	AprèsMidi	Laptop	HP ENVY DU7	17.3	1600*900	700
jourtravail	AprèsMidi	Laptop	HP ENVY DU7	17.3	1600*900	700
jourtravail	Matin	Laptop	ASUS N76UJ	17.3	1600*900	750
weekend	Matin	Laptop	ACER ASPIRE	17.3	1600*900	750
weekend	Matin	Laptop	SAMSUNG SERIE 3	17.3	1600*900	128
jourtravail	AprèsMidi	Laptop	SAMSUNG SERIE 3	15.6	1366*768	750
jourtravail	Matin	Laptop	HP ENVY DU7	15.6	1366*768	750
jourtravail	AprèsMidi	Laptop	ASUS N76UJ	15.6	1366*768	500
jourtravail	Matin	Laptop	ACER ASPIRE	15.6	1366*768	500
weekend	Matin	Laptop	ACER ASPIRE	13.3	1020*568	128

Figure 19 : Classification des données.

Description

Pour la classification des données on donne en entrant la matrice de transformation et on applique l'algorithme de classification. On obtient un ensemble de clusters qui représente des contextes.

1. Authentification

L'utilisateur qui accède au système, il faut passer par l'interface d'authentification pour introduire son nom d'utilisateur et son mot de passe.

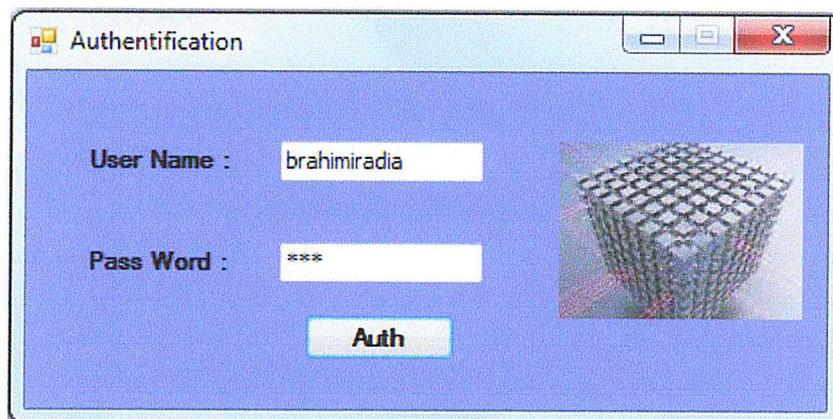


Figure 20 : Interface authentification.

2. Interface lancement des requêtes

Dans cette interface l'utilisateur va choisir le dispositif lors de son accès au système.

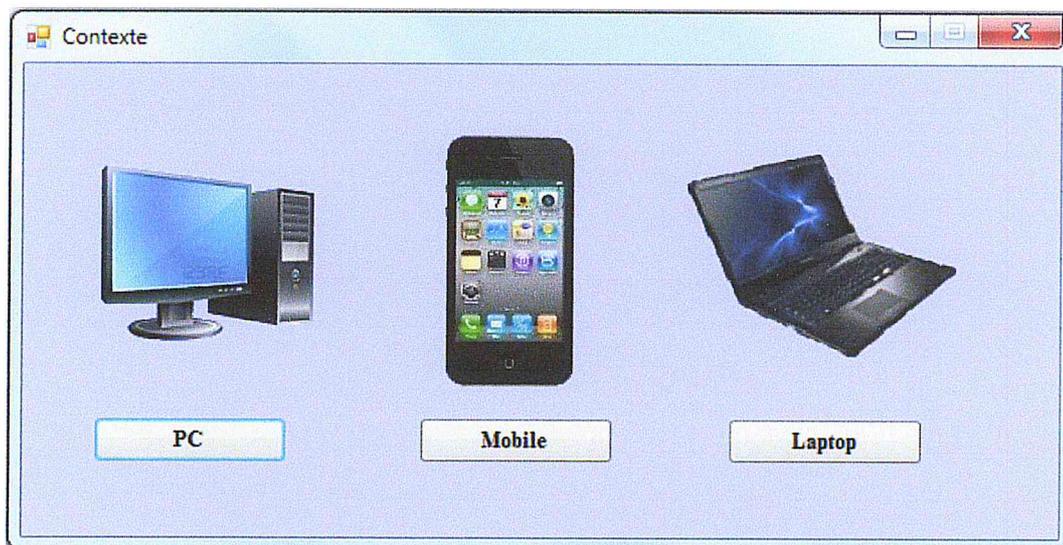


Figure 21 : Interface lancement des requêtes.

3. Interface des testes

Dans cette interface on vas lancé des requêtes MDX qui ont adapter au préférences d'utilisateur.

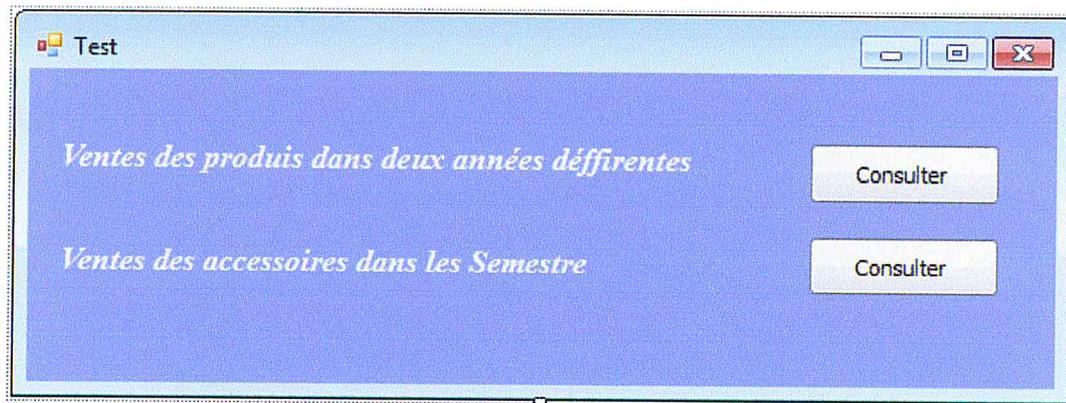


Figure 22 : Interface des testes.

Conclusion

Dans ce chapitre nous avons met en ouvre notre système contextualisée, ainsi qu'on a montré les différentes interfaces qui représentent les étapes que l'utilisateur doivent les suivit lors de sa interaction avec le système réalisé.

Conclusion Générale

Conclusion générale

Conclusion

L'origine de ce travail de thèse partait d'un constat simple à savoir : La croissance considérable de volume des données ainsi que la diversité des besoins des utilisateurs rendaient l'accès à une information pertinente de plus en plus difficile. Nous avons vu qu'un utilisateur, en recherchant une information, est submergé par des données pas forcément toutes intelligibles.

Donc il existait un réel besoin de personnaliser et de contextualisée l'information recherchée par l'utilisateur en fonction de ses attentes, ses préférences ainsi que son contexte. Ce besoin de contextualisation a été accentué par le large développement des systèmes qui offrent aux utilisateurs mobiles un accès à des informations quelque soit la localisation et le dispositif.

Devant cet état de fait, il était nécessaire de proposer un système qui tenait en compte à la fois des besoins de l'utilisateur en termes de personnalisation et du contexte d'utilisation.

A partir de la problématique exposée ci-dessus, un système de contextualisation de l'information a été proposé dans ce mémoire dont les contributions s'articulent autour de deux axes de travail interconnectés :

- La génération automatique des contextes usuels à partir de fichier journal de l'utilisateur en utilisant des techniques d'extraction de Datamining.
- L'intégration de contexte dans le processus de personnalisation.

Notre système réalisée va donner plus de performance et de pertinence aux résultats des requêtes multidimensionnels.

Perspectives

- La génération automatique du profil.
- Contextualisation spécifique au système OLAP base de profil et contexte.

Bibliographie.

- [**ABK, 08**]: S. Abbar, M. Bouzeghoub, D. Kostadinov, S. Lopes, A. Aghasaryan, and S. Betge-Brezetz. A personalized access model: concepts and services for content delivery platforms. In Proceedings of the 10th iiWas, Linz, Austria, pages 41-47, 2008.
- [**ADT, 05**]: G. Adomavicius and A. Tuzhilin, Incorporating context into recommender systems using multidimensional rating estimation methods. In Proceedings of the 1st International Workshop on Web Personalization, Recommender Systems and Intelligent User Interfaces, 2005.
- [**ADT, 11**]: Gediminas Adomavicius and Alexander Tuzhilin, Context-Aware Recommender Systems, 2011.
- [**ALG, 95**]: D. D'Aloisi et V. Giannini : The Info Agent: an Interface for Supporting Users in Intelligent Retrieval ; In C. Stephanidis (Ed.), Proceedings of the ERCIM Workshop « Towards Interfaces for all: Current Trend and Future Efforts » (UI4ALL Crete, October 30-31), pp. 143-155, 1995.
- [**AST, 11**]: Sandra Amol, Mouhamad Saliou Diallo, Cheikh Talibouya Diop, Arnaud Giacometti, Haoyuan D. Li, and Arnaud Soulet, Mining Contextual Preference Rules for Building User Profiles, 2011.
- [**BAC, 08**]: Alpha Oumar BAH, Aurel CRECIUN, Recherche des profils patients dépassant la durée normale de séjour, 2008.
- [**BAS, 97**]: M. Balabanovic, Y. Shoham, Fab: Content-based, collaborative recommendation. Communications of the ACM, Vol. 40, No. 3, pages 66–72, 1997.
- [**BBF, 08**]: F. Bentayeb, O.Boussaid, C.Favre, F.Ravat, O.Teste, Personnalisation dans les entrepôts de données, 2008.
- [**BBF, 09**] : F. Bentayeb, O.Boussaid, C.Favre, F.Ravat, O.Teste, Personnalisation dans les entrepôts de données, 2009.
- [**BCB, 05**]: B. Brown, M. Chalmers, M. Bell, M. Hall, I. MacColl, and P. Rudman, Sharing the square: collaborative leisure in the city streets. In H. Gellersen, K. Schmidt, M. Beaudouin-Lafon, and Mackay, W.E., editors, Proceedings of the ninth conference on European Conference on Computer Supported Cooperative Work, pages 427–447. Springer, 2005.

- [**BET, XX**] : Anthony Betremieux, Les métadonnées: définition et présentation générale.
- [**BGM, 05**]: L. Bellatreche, A. Giacometti, P. Marcel, H. Mouloudi, D. Laurent, A personalization framework for OLAP queries, Intl. Workshop on Data Warehousing and OLAP (DOLAP), pages 9–18, 2005.
- [**BGM, 06**]: L. Bellatreche, A. Giacometti, P. Marcel, H. Mouloudi, Personalization of MDX Queries, Journées Bases de Données Avancées (BDA), 2006.
- [**BLP, 91**]: J.R. Bettman, M.F. Luce, and J.W. Payne, Consumer decision making: A constructive perspective. In M. Tedeschi (editor), Consumer Behavior and Decision Theory pages 1–42, 1991.
- [**BOU, 10**]: L. Boucetha, Thèse: Personnalisation des requêtes OLAP, 2010.
- [**BOK, 04**]: M. Bouzeghoub, D. Kostadinov. Une approche multidimensionnelle pour la personnalisation de l'information, « INRIA Rocquencourt et laboratoire PRiSM, université de versailles 45, avenue des Etats-Unis, 78035 Versailles », 2004.
- [**BOK, 05**]: M. Bouzeghoub, D. Kostadinov, Personnalisation de l'information: aperçu de l'état de l'art et définition d'un modèle flexible de profils, 2005.
- [**BRS, 00**]: K. Bradley, R. Rafter, B. Smyth, Case-Based User Profiling for Content Personalisation, In: Proceedings of the International Conference on Adaptive Hypermedia and Adaptive Web-based Systems, Trento, Italy, August, 2000.
- [**BUR, 02**]: R. Burke, Hybrid Recommender Systems: Survey and Experiments. User Modeling and User-Adapted Interaction, Vol. 12, No. 4, pages 331–370, 2002.
- [**CAN, 06**]: L. Candillier, Contextualisation, visualisation et évaluation en apprentissage non supervisé, Thèse, Université Charles de Gaulle - Lille 3, 2006.
- [**CCL, 01**]: W. B. Croft, St. Cronen-Townsend, V. Lavrenko, Relevance Feedback and Personalization: A Language Modeling Perspective, In: Proceedings of the Second DELOS Network of Excellence Workshop on "Personalisation and Recommender Systems in Digital Libraries", Dublin City University, Ireland, 18-20, 2001.
- [**CEP, 09**]: G. Chatzopoulou, M. Eirinaki, N. Polyzotis, « Query Recommendations for Interactive Database Exploration », SSDBM, p. 3-18, 2009.
- [**CER, 08**]: CERT-In Security Guidelines (CISG), Guidelines for Auditing and Logging version 2, 2008.

- [CHS, 98]:** L. Chen and K. Sycara : WebMate: A Personal Agent for Browsing and Searching ; In Proceedings of the 2nd International Conference on Autonomous Agents and Multi Agent Systems (AGENTS'98 -- Minneapolis, USA, May 10-13), pp. 132-139, 1998.
- [CSW, 98]:** K. Chen, L. Sycara, Webmate: A personal agent for browsing and searching, In Proceedings of the 2nd international conference on autonomous agents and multi agent systems, Minneapolis, pages 10–13, 1998.
- [DAO, 09]:** M.DAoud, Accès personnalisé à l'information : approche basée sur l'utilisation d'un profil utilisateur sémantique dérivé d'une ontologie de domaines à travers l'historique des sessions de recherche, Université Paul Sabatier de Toulouse, 2009.
- [DEM, 10] :** Louisa Demmou, (Exploration de problèmes de performance d'un entrepôt de données), Sherbrooke, Québec, Canada, avril 2010.
- [DEY, 01]:** A. K. Dey. Pattern Recognition Letters, 5(1):4{7. (Diane Lingrand and Michel Riveill, Input interactions and context component based modelisations: differences and similarities), 2001.
- [DTB, 08]:** M. Daoud, L. Tamine-Lechani, M. Boughanem, B.Chebaro, Construction des profils utilisateurs à base d'ontologie pour une recherche d'information personnalisée, 2008.
- [FAN, 08]:** Sébastien FANTINI, Business Intelligence avec SQL Server 2008 R2, Copyright Edition ENI-2010, ISBN : 978-2-7460-5566-7, ISSN : 1650-2230, pp 6 8.
- [FES, 01]:** J. Ferreira, A. Silva, MySDI: A Generic Architecture to Develop SDI Personalised Services, In: Proceedings of the 3rd International Conference on Enterprise Information Systems, Setubal, Portugal, 2001.
- [GAM, 13]:** Philippe Gambette, Classification supervisée et non supervisée, 2013.
- [GLR, 02]:** G. Gonzalez, B. Lopez, J. Rosa, The Emotional Factor: An Innovative Approach to User Modelling for Recommender, SystemsProceedings of Second International Conference Adaptive Hypermedia and Adaptive Web-Based Systems AH, 2002.
- [GMN, 08]:** A. Giacometti, P. Marcel, E. Negre, A Framework for Recommending OLAP Queries, Intl. Workshop on Data Warehousing and OLAP (DOLAP), pages 73–80, 2008.

- [GMN, 09]:** A. Giacometti, P. Marcel, E. Negre, Recommending Multidimensional Queries, Intl. Conf. on Data Warehousing and Knowledge Discovery (DaWaK), Springer, pages 453–466, 2009.
- [GMN, XX]:** Arnaud Giacometti, Patrick Marcel, Elsa Negre, A Framework for Recommending OLAP Queries, XX.
- [GPM, 09]:** I. Garrigós, J. Pardillo, J. Mazón, J. Trujillo. A Conceptual Modeling Approach for OLAP Personalization. Intl. Conf. on Conceptual Modeling (ER), pages 401–414, 2009.
- [GOG, 01] :** Jean-François Goglin, (La construction du datawarehouse) ,2^e édition revue, 2001, ISBN 2-7462-0242-5.
- [GOW, 03]:** J. Gowan, A multiple model approach to personalised information access, Master thesis in computer science, Faculty of science, Université de College Dublin, February, 2003.
- [GRB, 11]:** M. Golfarelli, S. Rizzi, P. Biondi, my OLAP: an Approach to Express and Evaluate OLAP Preferences, IEEE Transactions on Knowledge and Data Engineering (TKDE), Vol. 23, No. 7, pages 1050–1064, 2011.
- [GRE, 09]:** M. Golfarelli, S. Rizzi , Expressing OLAP Preferences, Intl. Conf. on Scientific and Statistical Database Management (SSDBM), IEEE Computer Society, pages 83–91, 2009.
- [GRS, 08]:** Géraldine Graf, Julien Stern, Le CRM analytique Les outils d’analyse OLAP et le Datamining, Fribourg, 2008.
- [GSC, 07]:** S. Gauch, M. Speretta, A. Chandramouli, A. Micarelli, User proles for persona-lized information access, The Adaptive Web, 4321 :54–89, 2007.
- [INM, 96]:** W H Inmon. Building the Data Warehouse. Third ed. John Wiley and Sons, 1996.
- [JER, 12]:** H. Jerbi , Personnalisation d’analyses décisionnelles sur des données multidimensionnelles, 2012.
- [JGG, 05]:** G. Jones, J.F. D. Glasnevin, and I. Gareth, Challenges and opportunities of context-aware information access. In International Workshop on Ubiquitous Data Management, pages 53–62, 2005.
- [JPL, 08]:** B. Jiang, J. Pei, X. Lin, D. W. Cheung, and J. Han. Mining preferences from superior and inferior examples. In KDD, pages 390–398. ACM, 2008.

- [KBG, 09]: N. Khoussainova, M. Balazinska, W. Gatterbauer, Y. Kwon, D. Suciu, « A Case for A Collaborative Query Management System. », 2009.
- [KMM, 97]: J.A. Konstan, B.N. Miller, D. Maltz, J.L. Herlocker, L.R. Gordon, J. Riedl, GroupLens: Applying Collaborative Filtering to Usenet News. Communications of the ACM, Vol. 40, No. 3, pages 77–87, 1997.
- [KOB, 01]: A. Kobsa. Generic user modeling systems. User Modeling and User Adapted Inter-action Journal, (11):49–63, 2001.
- [KOI, 04]: G. Koutrika, Y. Ioannidis, Personalization of Queries in Database Systems, In: Proceedings of the 20th International Conference on Data Engineering, Boston, Massachusetts, USA, 2004.
- [KOS, 03]: D. Kostadinov, La personnalisation de l'information : définition de modèle de profil utilisateur, Master's thesis, Université de Versailles, France, 2003.
- [KOS, 07]: D. Kostadinov, Personnalisation de l'information et gestion des profils utilisateurs, 2007.
- [LIA, XX]: Bertrand LIAUDET, Cours de datamining, XX.
- [LKM, 92]: G.L. Lilien, P. Kotler, and K.S. Moorthy, Marketing models. Prentice Hall, 1992.
- [LLT, 05]: Stéphane Lavirotte, Diane Lingrand, Jean-Yves Tigli, « A propos du contexte unifié », Réunion Rainbow, 2005.
- [LOT, 79]: D.A. Lussier, and R.W. Olshavsky, Task complexity and contingent processing in brand choice. Journal of Consumer Research, pages 154–165, 1979.
- [LWL, 07]: C. Li, M. Wang, L. Lim, H. Wang, K. C. Chang, Supporting ranking and clustering as generalized order-by and group-by, ACM SIGMOD Intl. Conf.on Management of Data (SIGMOD), ACM Press, pages 127–138, 2007.
- [LYA, 07]: H-G. Li, H. Yu, D. Agrawal, A. El Abbadi , Progressive ranking of range aggregates. and Knowledge Engineering (DKE), Elsevier Science Publishers, Vol. 63, No. 1, pages 4–25, 2007.
- [LYD, 08]: Lydie Soler, Les entrepôts de données, 2008.
- [MAE, 94]: P. Maes, Agents that reduce work and information overload. Communications of the ACM, Vol. 37, No. 7, pages 31–40, 1994.

- [MDL, 02]: B. Mobasher, H. Dai, T. Luo, M. Nakagawa, Discovery and Evaluation of Aggregate Usage Profiles for Web Personalization, *Data Mining and Knowledge Discovery*, 6(1):61-82, 2002.
- [MIT, 02]: S. Mizarro et C. Tasso, Ephemeral and persistent personalisation in adaptive information access to scholarly publications on the web, In *Proceedings of the 2nd International Conference on adaptive hypermedia and adaptive Web-based systems*, pages 306–316, 2002.
- [MOU, 07]: Hassina MOULOUDI, (Personnalisation de requêtes et visualisations OLAP sous contraintes), 2007.
- [NAO, 06]: Lamiaa NAOUM, (Un modèle multidimensionnel pour un processus d'analyse en ligne de résumés flous), 2006.
- [ORE, 03]: Oreste Signore, *W3C Technologies: a Key for Interoperability*, 2003.
- [PAB, 07]: M. Pazzani, M. Billsus, Content-Based Recommendation Systems. *The Adaptive Web*, pages 325–341, 2007.
- [PHB, 08]: B. B. Phillip M. Hallam-Baker. Extended log file format. Technical report, W3C Draf WD- logfile -960323, 1996.
- [PMB, 96]: M. Pazanni, J. Muramatsu, D. Billsus, Syskill and webert : Identifying interesting web sites, In *Proceedings of the 13th National Conference on Artificial intelligence*, pages 54–61, 1996.
- [PRG, 99]: A. Pretschner, S. Gauch, Ontology Based Personalized Search, 1999, In: *Proceeding of the 11th IEEE Intl. On Tools with Artificial Intelligence*, pp. 391-398, Chicago.
- [PSC, 02]: J. E. Pitkow, H. Schutze, T. A. Cass, R. Cooley, D. Turnbull, A. Edmonds, E. Adar, T. M. Breuel, Personalized Search. *Communications of the ACM*, 2002, Vol. 45, No. 9, pages 50–55.
- [PTG, 08]: Palmisano, C., Tuzhilin, A., and Gorgoglione, M., Using context to improve predictive modeling of customers in personalization applications. *IEEE Transactions on Knowledge and Data Engineering*, 20(11):1535–1549, 2008.
- [PTG, 09]: U. Panniello, A. Tuzhilin, M. Gorgoglione, C. Palmisano, and A. Pedone, Experimental comparison of pre-vs. post-filtering approaches in context-aware recommender systems. In *Proceedings of the 3rd ACM conference on Recommender systems*, pages 265–268, ACM, 2009.

- [**RAT, 08**]: F. Ravat, O. Teste, Personalization and OLAP Databases, *Annals of Information Systems*, Vol. 3, Numéro spécial “New Trends in Data Warehousing and Data Analysis”, pages 1–22, 2008.
- [**RAW, 12**]: Christel Ruwet, Efficacité de classification de la méthode desk-moyennes tronquées, 2012.
- [**RIZ, 07**]: S. Rizzi, OLAP Preferences : a Research Agenda 10th, International Workshop on Data Warehousing and OLAP(DOLAP07), ACM, pp.99-100, Lisbon (Portugal), Novembre, 2007.
- [**RTZ, 07**]: F. Ravat, O. Teste, G. Zurfluh, Personnalisation de bases de données multidimensionnelles, Congrès INformatique des ORganisations et Systèmes d'Information et de Décision (INFORSID), pages 121–136, 2007.
- [**SAR, 00**]: S. Sarawagi, « User-Adaptive Exploration of Multidimensional Data », VLDB, p. 307-316, 2000.
- [**SCA, 00**]: S. Schiaffino, A. Amandi, User profiling with Case-Based Reasoning and Bayesian Networks, In Open Discussion Track Proceedings - International Joint Conference, 7th Ibero-American Conference on AI, 15th Brazilian Symposium on AI, IBERAMIA-SBIA 2000, Atibaia, SP, Brazil, pp19-22, 2000.
- [**SER, 06**]: B. Satzger, M. Endres, W. Kießling, A reference-Based Recommender System. Intl. Conf. on Electronic Commerce and Web Technologies (EC-Web), Springer, Heidelberg, pages 31–40, 2006.
- [**SHM, 95**]: U. Shardanand and P. Maes: Social Information Filtering: Algorithms for Automating 'Word of Mouth'; In Proceedings of the CHI-95 Conference. Denver, CO, USA, ACM Press, 1995.
- [**SOC, 98**]: S. Soltysiak, B. Crabtree, Automatic learning of user profiles-towards the personalisation of agent services, *BT Technol J.* Vol 16 No 3, pp 110-117, 1998.
- [**SOE, 95**]: H. Sorensen, M. Mc Elligott, PSUN : A Profiling System for Usenet News, Conference on Information and Knowledge Management (CIKM'95), Intelligent Information Agents Workshop, Baltimore, 1995.
- [**SPV, 07**]: K. Stefanidis, E. Pitoura and P. Vassiliadis, A context-aware preference database system. *International Journal of Pervasive Computing and Communications*, 3(4):439–600, 2007.



- [STA, 12]: STAMBOULI Mouhamed Fateh, (Conception et Réalisation d'un Système personnalisé dans l'analyse en ligne(OLAP), 2012.
- [TES, 09]: O. Teste, Modélisation et manipulation des systèmes OLAP de l'intégration des documents à l'utilisateur, Institut de Recherche en Informatique de Toulouse, 2009.
- [TJL, 02]: Thomsen and James A. Landay . A visual language for sketching large and complex interactive design. In: CHIACM, p307-314, 2002.
- [TLD, 05]: L. Tamine Lechani, M. Doughanem, accès personnalisés à l'information : Approches et Techniques, 2005.
- [TSM, 01] : T. Thalhammer, M. Schrefl et M. Mohania. Active DataWarehouses: Complementing OLAP with Analysis Rules. Data and Knowledge Engineering 39(3), 241–269, 2001.
- [TZB, 07]: L. Tamine-Lechani, N. Zemirli, W. Bahsoun. Approche statistique pour la définition du profil d'un utilisateur de système de recherche d'information. Information – Interaction - Intelligence, Laboratoire PRiSM, Université de Versailles, 2007.
- [WEH, 09]: Pascal Wehrle « Modèle multidimensionnel et OLAP sur architecture de grille », N° d'ordre 2009-ISAL-0002, 2009.
- [XAV, 03]: François-Xavier Jollois, Contribution de la classification automatique à la Fouille de Données, 2003.
- [XIH, 08]: D. Xin, J. Han, P-cube: Answering preference queries in multidimensional space, Intl. Conf. on Data Engineering (ICDE), IEEE Computer Society, pages 1092–1100, 2008.
- [ZEM, 08]: W.N. ZEMIRLI, Modèle d'accès personnalisé à l'information basé sur les Diagrammes d'Influence intégrant un profil utilisateur évolutif, Toulouse III, 2008.
- [W, 01]: <http://www.d-oneconsulting.com/vos-besoins-et-attentes>
- [W, 02]: <http://www.igm.univmlv.fr/XPOSE2010/chainedecisionnelle/modelisation.html>
- [W, 03]: www.maia.loria.fr/lexique/profil_utilisateur.html.