

Personnalisation dans les entrepôts de données : bilan et perspectives

Fadila Bentayeb*, Omar Boussaid*, Cécile Favre*
Franck Ravat**, Olivier Teste**

* Université de Lyon, ERIC Lyon 2
5 av. Pierre Mendès-France - 69676 Bron Cedex
bentayeb@eric.univ-lyon2.fr, {cecile.favrelomar.boussaid}@univ-lyon2.fr
<http://eric.univ-lyon2.fr>

** Université de Toulouse, IRIT
118, Route de Narbonne - F-31062 Toulouse cedex 9
{ravatlteste}@irit.fr
<http://www.irit.fr>

Résumé. Nous étudions dans cet article les enjeux et les opportunités relevant de la prise en compte des utilisateurs au sein des entrepôts de données. Pour ce faire, nous présentons tout d'abord un panorama des travaux sur la personnalisation dans les domaines connexes des bases de données et de la recherche d'information. Nous présentons ensuite les quelques travaux qui émergent dans le domaine des entrepôts de données et menons une étude comparative selon différents angles d'approche. Ceci nous permet de faire émerger des perspectives de recherche sur la personnalisation dans les entrepôts de données.

1 Introduction

L'approche d'entreposage de données (Kimball, 1996) (« data warehousing ») constitue un champ de recherche important dans lequel de nombreux problèmes restent à résoudre. Les entrepôts de données sont généralement intégrés dans un système d'aide à la prise de décision où l'on distingue deux espaces de stockage : l'entrepôt de données et les magasins de données (Figure 1). Un entrepôt de données est vu comme un espace de stockage centralisé homogène regroupant des données utiles aux analyses et à la prise de décision, extraites de sources et pouvant faire l'objet d'historisation (Inmon, 1996). Un magasin de données est un extrait de l'entrepôt, où les données sont préparées de manière spécifique pour faciliter leur analyse et leur exploitation par des utilisateurs appelés décideurs. Les données au sein d'un magasin sont souvent organisées de manière multidimensionnelle (Lehner, 1998) : les données analysées sont modélisées comme des points dans un espace à plusieurs dimensions formant un cube de données (Gray et al., 1996). Le cube de données peut alors être exploité par différents outils d'analyse. Il est en particulier possible, pour le décideur, de naviguer dans les données grâce à la technologie OLAP (On Line Analytical Processing).

Les travaux menés ces quinze dernières années ont essentiellement porté sur des aspects techniques tels que les processus ETL (« Extract-Transform-Load ») (Vassiliadis et al., 2002 ; Trujillo et Luján-Mora, 2003), les techniques d'optimisation et de sélection des vues matérialisées (Widom, 1995 ; Kotidis et Roussopoulos, 1999). Bien que les architectures décisionnelles soient considérées centrées utilisateurs, la prise en compte de ces derniers dans les systèmes décisionnels a finalement été peu étudiée. Des propositions ont été faites dans les méthodes de conception, notamment dans les méthodes descendantes, pour prendre

Personnalisation dans les entrepôts de données

en compte les besoins utilisateurs (Luján-Mora et Trujillo, 2003). Néanmoins ces travaux ne se focalisent pas réellement sur l'utilisation individualisée d'un magasin de données ; le magasin est conçu à un instant donné pour un groupe d'utilisateurs. Toute adaptation du magasin de données nécessite de réitérer le processus de conception. Or la constitution d'un magasin reste une tâche complexe qui met en jeu des processus d'extraction et de transformation qui rendent impossible la constitution de magasins individualisés en fonction des spécificités et des usages particuliers de chaque décideur. De plus, l'usage fait par un utilisateur n'est pas statique. Ces évolutions doivent être intégrées dans le système.

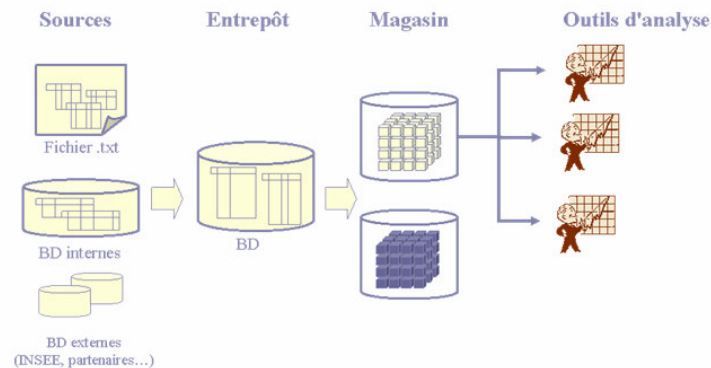


Fig. 1 – Architecture d'un système décisionnel à base d'entrepôt de données

La prise en compte des préférences, des usages et des interactions du décideur par le système, appelée personnalisation, constitue un champ de recherche qui reste à explorer dans le domaine des systèmes décisionnels intégrant des magasins multidimensionnels (Rizzi, 2007).

Après avoir présenté les différentes approches de la personnalisation dans le contexte des systèmes d'information, de la Recherche d'Information (RI) et des Bases de Données (BD) principalement, notre objectif est d'étudier les approches de personnalisation proposées dans le contexte des entrepôts de données. En effet, dès lors qu'il s'agit de répondre à des besoins utilisateurs peut se poser la question de la personnalisation vis-à-vis de ces derniers (Oulasvirta et Blom, 2008), en particulier dans le contexte des magasins de données au sein desquels les utilisateurs ont un rôle central à jouer lors de la navigation dans les données. Cet article vise à établir les différents enjeux que pose la personnalisation dans les systèmes d'aide à la décision fondés sur des entrepôts.

Cet article s'articule de la manière suivante. La section 2 présente les définitions préalables. La section 3 étudie les travaux sur la personnalisation dans les systèmes d'information. La section 4 présente quelques travaux sur la personnalisation dans le domaine des systèmes d'aide à la décision et propose une étude comparative de ces derniers. Enfin la section 5 énonce les différentes perspectives sur les travaux de personnalisation dans les systèmes d'aide à la décision.

2 Définitions préalables

Y. Ioannidis et G. Koutrika définissent la personnalisation en ces termes « ...providing an overall customized, individualized user experience by taking into account the needs, prefer-

ences and characteristics of a user or group of users » (Ioannidis et Koutrika, 2005). Généralement, la personnalisation d'un système consiste à définir, puis à exploiter un profil utilisateur (Korfhage, 1997) qui regroupe un ensemble de caractéristiques servant à configurer ou à adapter le système à l'utilisateur, afin de fournir des réponses plus pertinentes à l'utilisateur (Domshlak et Joachims, 2007). Ce profil peut s'apparenter à une modélisation de l'utilisateur (identification, antécédents, droits d'accès, préférences...). Cet axe de recherche a fait l'objet de travaux en RI (Tchienehom, 2006) ainsi qu'en BD.

Aucune proposition de profil n'a encore fait l'objet d'un consensus. Cette absence de standard rend cette notion parfois ambiguë ; le concept même n'est pas toujours clairement formalisé dans les solutions proposées (certaines approches ne faisant pas appel explicitement à un profil, mais recueillant néanmoins les préférences d'un utilisateur). Bouzeghoub et Kostadinov (2005) ont alors apporté un modèle générique multidimensionnel de profil, convenant ainsi à une majorité de contexte. De son côté, Kobsa (2007) a proposé un état de l'art sur la modélisation des utilisateurs en fonction des besoins des systèmes.

Nous proposons de caractériser un profil selon deux perspectives : l'implication de l'utilisateur et les fonctions systèmes liées au profil.

- L'implication de l'utilisateur peut être soit explicite, soit implicite. Dans le cadre d'une implication explicite, l'utilisateur doit effectuer des interactions directes avec le système tandis que lors d'une implication implicite, le système s'adapte automatiquement à l'utilisateur.
- Les fonctions systèmes liées au profil consistent dans un premier temps à définir le profil et dans un second temps à exploiter ce dernier pour une meilleure prise en compte de l'utilisateur.

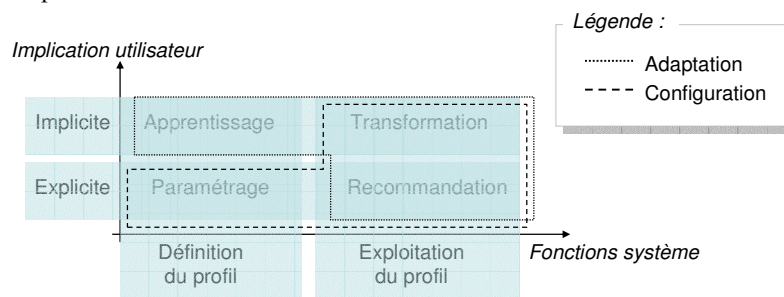


Fig. 2 – Principes de la personnalisation

La figure 2 présente les principes mis en jeu lors de la personnalisation. La définition d'un profil réalisée de façon explicite correspond au paramétrage d'un système tandis que la définition implicite s'apparente à l'apprentissage. L'exploitation du profil peut soit nécessiter l'intervention explicite de l'utilisateur qui réalise un choix par rapport à la recommandation du système, soit induire une transformation automatique du système. Les termes de configuration et d'adaptation sont caractérisés en fonction de ces quatre principes.

- La configuration (« customisation » ou « user modeling ») consiste donc pour l'utilisateur à paramétrer explicitement son profil. Le profil ainsi déterminé est exploité au travers de 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.

Personnalisation dans les entrepôts de données

- 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.

3 La personnalisation dans les SI

La personnalisation a fait l'objet de nombreux travaux dans les systèmes d'information (Ioannidis et Koutrika, 2005). Reprenant les principes de la personnalisation évoqués précédemment, nous présentons brièvement les travaux sur la configuration et l'adaptation.

3.1 Configuration

Différents travaux se sont focalisés sur la définition du profil dans une logique de configuration par un paramétrage du profil. Dans le contexte de la RI, le profil utilisateur consiste souvent à regrouper un ensemble plus ou moins structuré de mots-clés, définis par l'utilisateur, avec éventuellement des poids qui leur sont associés (Pretschner et Gauch, 1999). Le profil utilisateur peut également correspondre à des fonctions d'utilités sur un domaine d'intérêt, exprimant l'importance relative des sujets de ce domaine, les uns par rapport aux autres (Cherniack et al., 2003). Le profil est ensuite exploité de diverses manières :

- Les systèmes de filtrage d'information (« information filtering ») visent à distribuer des informations de façon personnalisée en comparant les caractéristiques de l'utilisateur fournies par ce dernier avec une collection d'informations (Aas, 1997).
- Les systèmes de recommandation collaboratifs (« collaborative filtering ») ont le même objectif de mise à disposition d'informations pertinentes en faisant reposer le processus de sélection d'information sur une approche collaborative (comparaison d'un utilisateur avec d'autres en fonction d'éléments fournis par ceux-ci) (Goldberg et al., 1992 ; Adomavicius et Tuzhilin, 2008).
- Les systèmes temps réel fondés sur les stratégies d'interactions sociales pour la recherche et l'accès à l'information (Castagnos, 2008).

La notion de préférence a également été introduite dans le domaine plus structuré des BD (Lacroix et Lavency, 1987). Ces travaux ont essentiellement porté sur la personnalisation de requêtes, notamment à travers l'intégration des préférences utilisateurs (Koutrika et Ioannidis, 2005 ; Agrawal et Wimmers, 2000 ; Chomicki, 2003 ; Kiessling, 2002 ; Lacroix et Lavency, 1987). Dans ce contexte, deux approches principales ont émergé : quantitative et qualitative. L'approche quantitative consiste à exprimer les préférences d'une façon indirecte par l'utilisation de fonctions de score qui associent un score numérique à chaque n-uplet du résultat d'une requête. Dans l'approche qualitative, les préférences sont spécifiées directement à l'aide de relations binaires. L'intégration consiste alors à un enrichissement des requêtes opérées sur la BD par de nouveaux prédicats.

3.2 Adaptation

La particularité des systèmes adaptatifs réside dans la détermination « automatique » des caractéristiques de l'utilisateur et dans une certaine mesure de leur évolutivité. La difficulté réside alors dans la modélisation de l'utilisateur (de ses caractéristiques, de ses préférences).

On peut distinguer les différents travaux concernant les systèmes adaptatifs selon le type d'apprentissage des caractéristiques, des préférences. En effet, l'apprentissage peut se faire soit à partir d'éléments fournis par l'utilisateur, soit à partir du comportement même de l'utilisateur, *i.e.* de l'interaction entre ce dernier avec le système.

Dans la première catégorie, on retrouve les travaux sur les systèmes éducatifs qui se basent sur un questionnaire préliminaire rempli par l'apprenant pour « apprendre » le niveau de celui-ci ou sur des tests au fur et à mesure de l'apprentissage ; c'est le cas des systèmes ACE (Specht et Oppermann, 1998) et ARTHUR (Gilbert et Han, 1999). On trouve également des travaux se basant sur la programmation par l'exemple où les utilisateurs sont invités à fournir des exemples pour la personnalisation d'applications Web (Macías et Paternò, 2008).

Dans la deuxième catégorie, on retrouve également des travaux sur les systèmes éducatifs, mais ceux-ci vont exploiter l'interaction système-utilisateur. C'est le cas dans le système iMANIC (Stern et Woolf, 2000), où les données concernant l'interaction avec l'étudiant sont analysées pour déterminer quelles ressources doivent lui être recommandées. Poursuivant cette idée d'exploiter le comportement de l'utilisateur, Lieberman (1995) a proposé, dans le contexte de la RI, un agent nommé Letizia, qui enregistre les URL parcourues par l'utilisateur, lit les pages et détermine au fur et à mesure le profil de l'utilisateur. Sur cette base, il effectue une recherche pour recommander d'autres pages susceptibles d'intéresser l'utilisateur. Bradley et al. (2000) proposent une personnalisation de contenu dans le projet CASPER en définissant les profils et en les mettant à jour par rapport au comportement des usagers à travers les statistiques d'interaction entre le système et l'utilisateur (nombre de clics, etc.).

3.3 Bilan

Les travaux relatifs à la configuration et à l'adaptation se distinguent essentiellement par la manière de déterminer le profil de l'utilisateur. Pour l'exploitation du profil, le choix entre la transformation ou la recommandation dépend essentiellement de l'objectif poursuivi.

Concernant la définition du profil lui-même, la configuration permet de répondre à une certaine forme de personnalisation, dans laquelle l'utilisateur doit intervenir pour fixer les paramètres de cette personnalisation. Néanmoins, nous avons pu noter que la détermination de paramètres peut constituer une tâche difficile, dans la mesure où elle relève parfois d'une certaine subjectivité. Cette étude a alors mis en avant l'intérêt des systèmes adaptatifs. Outre le fait d'éviter à l'utilisateur de configurer le système en spécifiant les paramètres permettant la personnalisation, il s'avère qu'il est parfois préférable de déduire ces paramètres de façon implicite. Le cas des systèmes éducatifs en est un parfait exemple, dans la mesure où la détermination d'un paramètre tel que le niveau de l'apprenant peut être difficile à déterminer par l'apprenant lui-même. Néanmoins, précisons que cette étude ne tranche pas forcément sur le fait que telle ou telle alternative est meilleure, étant donné que cela dépend beaucoup de l'objectif poursuivi par la personnalisation.

Par la suite, nous allons étudier les travaux s'intéressant à la prise en compte de l'utilisateur dans le domaine des entrepôts de données. Comme nous allons le montrer dans notre étude comparative, ces travaux relèvent davantage d'un processus de configuration pour la définition du profil. Ainsi, nous pensons que l'aspect d'adaptabilité en recourant à un apprentissage du profil présente un intérêt certain et devra être pris en compte dans les perspectives de recherche que nous envisageons, l'objectif poursuivi étant de faciliter la tâche du décideur.

4 L'utilisateur dans les entrepôts de données

Bien que les travaux dans les domaines de la RI et des BD pour la personnalisation aient été beaucoup développés, nous nous intéressons ici aux travaux moins nombreux et plus récents dans le domaine des entrepôts de données. La personnalisation dans le cadre des entrepôts de données présente un réel intérêt dans un contexte où les analyses pour permettre la prise de décision sont réalisées par l'utilisateur lui-même. Ce dernier a sa propre connaissance des données visualisées lors de la navigation et donc des besoins qui peuvent lui être propre, d'où cet intérêt pour la personnalisation. Nous présentons tout d'abord les différentes notions liées aux entrepôts de données à travers un exemple qui illustrera la présentation des différents travaux. Nous comparons ensuite les différentes approches dans le but de faire émerger les enjeux qu'il reste à explorer dans ce domaine.

4.1 Exemple illustratif

Pour illustrer les différents travaux, nous choisissons d'utiliser une étude de cas simplifiée définie avec l'établissement bancaire LCL-Le Crédit Lyonnais (LCL). Le contexte d'analyse multidimensionnelle est représenté graphiquement dans la figure 3. Il s'agit d'analyser un fait correspondant aux performances de l'établissement bancaire. Cette analyse est faite à travers une mesure : le Produit Net Bancaire (PNB). Ce PNB représente ce que rapporte la gestion des comptes des clients à l'établissement bancaire. Cette mesure est analysée selon différentes dimensions : CLIENT, TEMPS et AGENCE. Certaines dimensions peuvent être hiérarchisées, comme c'est le cas de la dimension AGENCE, dont la hiérarchie représente la structure commerciale de LCL (H-StructCommerciale); les agences sont regroupées en UC (Unité commerciale), elles-mêmes regroupées en DPP (Direction Particuliers Professionnels), les DPP étant regroupées en DE (Direction d'Exploitation). Notons que les hiérarchies sont déterminées grâce à des attributs que l'on nomme paramètres (UC), la sémantique de ces derniers étant complétée par des attributs dits faibles (Nom_UC).

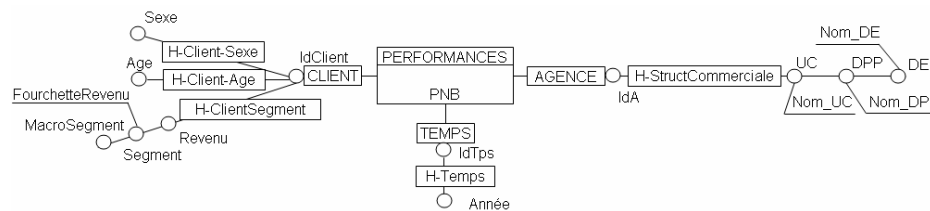


Fig. 3 – Modèle initial de l'entrepôt de données du LCL considéré

4.2 Présentation et illustration des travaux

Initialement, les entrepôts étaient destinés à la consultation de données organisées en fonction d'une vision d'analyse. L'utilisateur ne pouvait interagir avec les valeurs et les structures mises à sa disposition autrement que par des requêtes d'interrogation.

Face à la nécessité d'offrir davantage de flexibilité pour répondre au mieux aux besoins des utilisateurs, une première approche (Espil et Vaisman, 2001) a consisté à proposer le langage à base de règles nommé IRAH (« Intensional Redefinition for Aggregation Hierar-

chies ») visant à transformer les appariements de valeurs entre les niveaux d'agrégation au sein des hiérarchies. Cette approche permet à l'utilisateur de construire ses propres chemins de navigation en réorganisant les instances mises à sa disposition. Si l'on considère notre exemple illustratif en nous focalisant sur la hiérarchie H-ClientSegment. La hiérarchie signifie que chaque client a un certain revenu qui permet de déterminer le segment de ce client. Un conseiller commercial peut alors ré-affecter un client dans un autre segment, pour répondre à ses besoins, en exprimant une règle avec le langage IRAH, produisant ainsi une révision du chemin d'agrégation pour la dimension CLIENT.

Afin d'enrichir les possibilités d'analyse d'un entrepôt de données, une proposition permettant la création de niveaux d'analyse supplémentaires dans les hiérarchies de dimension ou définissant de nouvelles hiérarchies de dimension a été faite (Favre et al., 2007). Cette approche, dénommée WEDriK (« data Warehouse Evolution Driven by Knowledge »), permet aux utilisateurs d'intégrer leurs propres connaissances sur la façon d'agréger les données sous la forme de règles de type « si-alors ». Une évolution du modèle de l'entrepôt permet le partage des chemins d'agrégation nouvellement créés entre les utilisateurs du système. Si l'on considère le cas de LCL, supposons qu'un utilisateur veuille analyser le PNB, non pas en fonction de la structure commerciale de l'établissement, mais en fonction du type d'agence, information qui n'est pas présente dans l'entrepôt. L'utilisateur peut alors exprimer ses connaissances sur les types d'agence, afin de créer un niveau type d'agence, correspondant à une nouvelle hiérarchie de la dimension agence. L'utilisateur pourra ainsi réaliser des analyses du PNB en fonction du type d'agence.

Thalhammer et al. (2001) 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. Dans notre exemple, on peut imaginer une règle d'analyse qui permettrait de lancer une action marketing pour améliorer les ventes, pour l'UC d'Annonay, lorsque l'analyse du PNB faite à la fin de l'année est inférieure à 100 000 euros pour cette UC.

Concernant la personnalisation vis-à-vis des analyses multidimensionnelles, Bellatreche et al. (2005) se sont inspirés des techniques de filtrage d'information en fonction de profil utilisateur pour affiner des requêtes en y ajoutant des prédicats. L'objectif de ces travaux est de pouvoir fournir à l'utilisateur un résultat focalisé sur son centre d'intérêt, tout en prenant en compte des contraintes de visualisation pour adapter le résultat à l'utilisateur. Par exemple, supposons qu'un utilisateur veuille connaître le PNB total détaillé par UC, année et segment de clients. Si le nombre de segments et/ou celui d'UC et/ou celui des années sont importants, la réponse complète ne peut être visualisée sur l'écran. Selon l'utilisateur, l'intérêt pour telles ou telles UC, tels ou tels segments peut varier. Si l'utilisateur est un responsable de la clientèle « haut de gamme », certains segments sont moins « intéressants » que d'autres. Concernant la segmentation, il s'agirait de créer un profil « responsable clientèle haut de gamme » dans lequel seraient classés les segments client par ordre de préférence pour la visualisation.

Pour faciliter la navigation de l'utilisateur dans les données, Ravat et al. (2008) proposent une approche de personnalisation des données multidimensionnelles manipulées. Un poids fixé par l'utilisateur est associé aux données exprimant ainsi les préférences de ce dernier. Un système basé sur des règles ECA permet de générer des tables multidimensionnelles contenant uniquement les données identifiées comme pertinentes en fonction des poids. Cette solution quantitative permet de simplifier l'expression des requêtes d'analyse. Par exemple,

Personnalisation dans les entrepôts de données

un utilisateur peut exprimer ses préférences pour indiquer que, lors de l'analyse du PNB, les paramètres DE et DPP sont prioritaires, tandis que les paramètres UC et AGENCE ne le sont pas. Ainsi, si l'affichage de la mesure du fait « PERFORMANCES » est demandé en fonction de la dimension AGENCE, le système affiche automatiquement les attributs DE et DPP.

Pour aller au-delà de cette approche quantitative, 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. Par exemple, un utilisateur peut exprimer comme précédemment que les paramètres DE et DPP sont prioritaires par rapport à UC simplement par un ordre entre les paramètres des préférés au moins préférés. En outre, cette préférence peut s'exprimer dans un contexte d'analyse tel que l'analyse du PNB par l'affichage de la mesure du fait « PERFORMANCES ». Ainsi l'ordonnement est pris en compte uniquement dans ce contexte, tandis que d'autres préférences peuvent être exprimées dans un autre contexte sur cette même dimension AGENCE. Jerbi et al. (2009) poursuivent ces travaux en présentant un environnement OLAP intégrant des mécanismes de recommandation contextuelle des requêtes.

Citons aussi la proposition de Giacometti et al. (2008) 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. Par exemple, supposons qu'un utilisateur a réalisé une analyse du PNB par année et DE, puis avec un forage vers le bas par année et DPP, et enfin par année et UC (avec un second forage vers le bas). Si un nouvel utilisateur réalise une analyse du PNB par année et DE, puis une analyse du PNB par année et DPP, une analyse par année et UC lui sera recommandée, sa navigation étant similaire à une navigation réalisée précédemment.

4.3 Synthèse

Afin de comparer les différents travaux, nous positionnons ces travaux, dans la figure 4, par rapport aux principes de la personnalisation présentés dans la figure 2.

Références	Définition du profil		Exploitation du profil	
	Apprentissage	Paramétrage	Transformation	Recommandation
[Bellatreche et al., 2005]		Contraintes de visualisation	Modification table résultat	
[Espil et Vaisman, 2001]		Règles d'exception	Modification restitution	
[Favre et al., 2007]		Règles 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]		Scénarii d'analyse, règles ECA	Génération de rapports	

Fig. 4 – Positionnement des travaux sur les entrepôts de données par rapport aux principes de personnalisation définis précédemment

Nous constatons que l'apprentissage est peu exploité dans les entrepôts de données, alors même que cela allège la tâche de l'utilisateur. De la même façon, la recommandation n'est exploitée que très récemment par peu de travaux. Finalement, les travaux faits en matière d'entrepôts de données relèvent davantage de la configuration que de l'adaptation.

Pour affiner la lecture des différents travaux portant sur les entrepôts de données, nous proposons une grille d'analyse basée sur des critères de comparaison que nous avons jugés pertinents pour étudier les différentes facettes de la personnalisation. Ces critères, inspirés en partie de l'étude de Anli (2006), doivent nous aider à déterminer des voies intéressantes sur la personnalisation dans les entrepôts de données.

Profil. Ce critère permet d'identifier si le système de personnalisation se base ou non sur le concept de profil. Le profil d'un objet est vu comme un ensemble de caractéristiques permettant de l'identifier ou de le représenter (Tchienehom, 2005). Ainsi un profil utilisateur est vu comme une collection plus ou moins structurée d'informations sur l'utilisateur.

Collecte d'information. Il s'agit de déterminer comment sont collectées les informations liées aux utilisateurs et à leurs besoins. Cette collecte peut se faire de manière explicite dans le cas où l'utilisateur est amené à fournir ces informations directement. Par opposition, cette collecte peut se faire de façon implicite lorsque ces informations sont récoltées sans que l'utilisateur n'ait besoin de préciser ces informations (comportement de l'utilisateur vis-à-vis du système). La façon dont est collectée l'information induit le mode de personnalisation du système : mode configuration dans le contexte de la collecte explicite de données et mode adaptation pour la collecte implicite.

Type de personnalisation. Ce critère représente la portée prise en compte pour la personnalisation (cognitif ou social). Dans le cadre cognitif, il s'agit de prendre en compte l'utilisateur de façon individuelle, en fonction de ses besoins. Dans le cadre social ou collaboratif, il s'agit de baser la personnalisation en prenant en compte le contexte des autres utilisateurs qui auraient des préoccupations similaires (réseau social).

Objet de la personnalisation. Il s'agit de préciser à quel niveau se situe la personnalisation. Elle peut porter sur le contenu, la présentation (ou contenant), l'interaction. La personnalisation du contenu impacte les données restituées tandis que la personnalisation de la présentation impacte la manière dont sont restituées les données. Enfin, la personnalisation de l'interaction consiste à modifier (généralement simplifier) les actions effectuées par un utilisateur sur le système.

Personnalisation évolutive. Cette caractéristique indique si l'évolution du profil est possible. Nous évoquons deux aspects. Tout d'abord le mode de mise à jour. En effet, le profil est établi à un instant donné en fonction des spécificités de l'utilisateur. Ce profil peut ensuite être modifié selon deux modes : soit les valeurs des critères existants sont modifiables, soit la structure même des critères est transformable (il est possible d'en ajouter/enlever). Par exemple, dans un système de RI, le profil peut être vu comme un vecteur de mots-clés pondérés :

- si les mots-clés et les poids associés peuvent être modifiés à taille de vecteur constante, on parlera d'évolution en mode valeur,
- si la taille du vecteur évolue par l'ajout et/ou la suppression de mots-clés pondérés, on parlera d'évolution en mode structure. Il faut noter que le mode structure englobe généralement le mode valeur.

Personnalisation dans les entrepôts de données

Le deuxième aspect est l'extensibilité du profil qui désigne la capacité du système à étendre le profil et à l'exploiter avec de nouveaux critères non pré-établis.

Type d'approche. L'intégration des préférences de l'utilisateur peut s'effectuer de manière quantitative ou qualitative. L'approche quantitative consiste à exprimer les préférences d'une façon indirecte en fonction de score numérique sur les données tandis que l'approche qualitative définit les préférences directement à l'aide de relations binaires entre les données.

Exploitation du profil. Le profil peut être utilisé soit pour une transformation du système, soit pour une recommandation faite par le système.

Le tableau de la figure 5 permet de caractériser chacun des travaux présentés en les positionnant par rapport aux critères définis précédemment.

Références	Profil	Collecte information	Type perso	Objet perso	Perso évolutive		Type approche	Exploitation profil
					Mode mise à jour	Extensibilité		
[Bellatreche et al., 2005]	Oui	Explicite	Cognitif	Présentation	Structure	Non	Quantitative	Transformation
[Espil et Vaisman, 2001]	Non	Explicite	Cognitif	Contenu	Valeur	Non	-	Transformation
[Favre et al., 2007]	Non	Explicite	Cognitif	Contenu	Structure	Non	-	Transformation
[Giacometti et al., 2008]	Oui	Implicite	Social	Contenu	Structure	Non	-	Recommandation
[Jerbi et al., 2008-2009]	Oui	Explicite	Cognitif	Interaction	Structure	Non	Quantitative	Transformation
[Ravat et al., 2008]	Oui	Explicite	Cognitif	Interaction	Structure	Non	Quantitative	Transformation
[Thalhammer et al., 2001]	Non	Explicite	Cognitif	Interaction	Valeur	Non	-	Transformation

Fig. 5 – Comparaison des travaux dans le domaine des entrepôts

Nous remarquons que les techniques de collecte d'information implicite n'ont été exploitées que dans un des travaux, de façon très récente au niveau des entrepôts de données. Ceci implique que dans la plupart des cas, l'utilisateur est amené à paramétrer l'entrepôt de données en spécifiant ses préférences ; cette approche explicite peut s'avérer parfois ardue, en particulier lorsque les préférences doivent être quantifiées. À ce jour, l'aspect social de la personnalisation dans les entrepôts de données n'a été également exploité que par ce seul travail. L'approche proposée est la plupart du temps une approche cognitive où l'utilisateur est considéré de manière individuelle, sans tirer profit de l'expérience des autres utilisateurs. Par ailleurs, notons que les travaux se proposent souvent d'exploiter le profil pour transformer le système, l'aspect recommandation restant presque inexploré dans le contexte des entrepôts de données, à l'exception d'un des récents travaux. Enfin, nous notons que les travaux proposés jusque là n'ont pas abordé la possibilité d'extensibilité du profil avec des critères non prédéfinis. Ces remarques servent de support à la section suivante dans laquelle nous traçons différentes perspectives de recherche.

5 Discussion et perspectives

L'étude précédente montre que les travaux qui étudient la prise en compte des utilisateurs au sein des entrepôts de données sont en pleine émergence. Les travaux proposés jusque là ont consisté principalement à étendre les approches habituelles en permettant de modifier les structures multidimensionnelles et/ou les mécanismes de présentation des données par une meilleure connaissance des utilisateurs. Pour ce faire, ces travaux exigent des efforts cog-

tifs de la part de l'utilisateur qui doit souvent exprimer de manière explicite les préférences qui le caractérisent.

La prise en compte de l'utilisateur est une problématique nouvelle qui pose plusieurs enjeux peu ou pas étudiés. Une plus grande interaction de l'utilisateur avec le système permettrait d'envisager des bénéfices à deux niveaux :

- du point de vue système, la connaissance accrue de l'utilisateur ou du groupe d'utilisateurs doit pouvoir servir à mieux paramétrer celui-ci, et par conséquent, doit permettre un fonctionnement plus proche des utilisateurs ;
- du point de vue utilisateur, un système mieux adapté doit permettre une réduction des efforts nécessaires pour accéder, manipuler et structurer une information pertinente afin de faciliter davantage le processus décisionnel qui en découle. Cela doit permettre au décideur de se centrer sur le processus d'analyse décisionnelle en lui simplifiant la manipulation des données.

Plus précisément, nous traçons cinq axes majeurs de recherche.

Un premier axe concerne la construction de profils utilisateurs dans les entrepôts de données. Cet aspect a été largement étudié, notamment en RI, mais, à l'heure actuelle, aucune solution unanimement reconnue ne semble émerger (Bouzeghoub et Kostadinov, 2005). L'étude des travaux a montré qu'un premier point intéressant est la définition de mécanismes rendant la collecte des caractéristiques utilisateur automatique et transparente à ce dernier. Cette approche vise à définir un entrepôt de données adaptatif.

Pour remplir cet objectif, il est nécessaire de définir des solutions permettant l'évolution des profils, leur extensibilité. Les caractéristiques de ces derniers sont constituées principalement des préférences utilisateur sur les données et les structures de l'entrepôt. Un aspect original pour la constitution du contenu d'un profil serait d'ajouter l'aspect navigation opérée par l'utilisateur sur les données à l'image de la navigation Internet. Une meilleure connaissance des navigations OLAP sur les données peut par exemple servir à l'entrepôt de données pour préparer des données (agrégations, ...) qui anticipent la demande de l'utilisateur.

Plus généralement, la définition d'un profil englobant les caractéristiques complètes couvrant l'ensemble des besoins de présentation et d'interaction peut être envisagé. Un dernier point, qui n'a fait l'objet que d'un seul travail dans le cadre des entrepôts de données alors qu'il est largement exploité d'ores et déjà en RI, est la prise en compte de l'aspect social dans les entrepôts en tenant compte du groupe d'utilisateurs auquel appartient un décideur. Dans ce contexte, l'apport du web social peut présenter un réel intérêt dans une perspective de personnalisation sociale des entrepôts de données. Les travaux proposés par Morfonios et Koutrika (2008), ainsi que Aouiche et al. (2008) constituent des pistes intéressantes. Ainsi, la définition d'un entrepôt de données proposant des recommandations pourrait servir à aider un décideur en tirant bénéfice de l'expertise des autres membres du groupe d'utilisateurs.

Cet axe couvre des problématiques récurrentes dans les systèmes d'informations et se heurte à la définition d'une (des) mesure(s) de similarité dans les entrepôts de données. Ce verrou scientifique pose en effet le problème de la constitution d'une mesure de similarité permettant de comparer et d'aligner les profils des utilisateurs ainsi que les contextes d'analyse formés de sujets et d'axes d'analyses.

Une des manières d'aborder ce problème de recommandation est également de se placer dans le contexte des travaux concernant les requêtes de type « top k » ou « skyline » (Brando et al., 2007). Ces travaux traités dans le contexte des bases de données tentent également d'inclure un processus de personnalisation (Lee et al., 2009). Il serait intéressant d'étudier les requêtes OLAP et leurs éventuelles spécificités par rapport à cette problématique.

Personnalisation dans les entrepôts de données

La navigation, qui est propre aux entrepôts de données, constitue également un enjeu par rapport à la construction de profil. La détermination de poids proposée par Ravat et al. (2008) pour personnaliser la navigation constitue une première approche de construction de profil. Mais il est vrai que la détermination de ces profils de façon explicite par l'utilisateur peut s'avérer fastidieuse. Un travail sur l'apprentissage de ces préférences en termes de navigation au fur et à mesure de la navigation de l'utilisateur pourrait alléger le processus, et par là même devenir un processus davantage évolutif. La proposition de Giacometti et al. (2008) pourrait constituer un point de départ intéressant pour cette réflexion.

Le deuxième axe concerne la gestion des données. L'analyse multidimensionnelle consiste à restituer des données agrégées en fonction de différents axes d'analyse. L'objectif de l'OLAP est une restitution rapide de ces données agrégées. La solution envisagée repose sur le principe de vue matérialisée. Si le profil enregistre les différentes navigations qu'effectue un décideur sur un cube de données, voire précise les composants des axes d'analyse souhaités, ceci aura un impact sur la définition des vues matérialisées.

Un des enjeux qui en découle est alors de pouvoir définir des algorithmes de sélection de vues à matérialiser avec prise en compte du profil des utilisateurs (ou groupe d'utilisateurs). Classiquement, les algorithmes de sélection se basent sur une charge de requêtes qui représente l'utilisation de l'entrepôt (Kotidis et Roussopoulos, 1999), et donc en quelque sorte les préférences des utilisateurs en matière de requêtes. Il s'agit d'aller au-delà, en prenant également en compte l'aspect enchaînement des requêtes. Ceci pourrait avoir non seulement un impact sur le choix des vues à matérialiser, mais également sur l'usage de ces vues. Il s'agit par exemple de s'intéresser à la comparaison du treillis de vues matérialisables et de l'arbre de navigation au sein des vues.

Concernant l'aspect gestion des performances, on peut imaginer qu'en connaissant mieux la navigation des utilisateurs, il soit possible, dans un contexte d'une architecture client serveur, que le système soit pro-actif dans cette gestion. Par exemple, grâce à la connaissance sur la navigation des utilisateurs, il est envisageable de charger certaines vues en mémoire pour anticiper l'usage de l'utilisateur en s'inspirant des travaux de Cherniack et al. (2003).

Un autre aspect spécifique aux entrepôts de données concerne la personnalisation de la phase d'ETL (Extract, Transform and Load). Une perspective envisageable serait d'étudier l'apport de la personnalisation dans le cadre de ce processus. Par exemple, une meilleure connaissance des profils utilisateurs pourrait permettre une adaptation de la stratégie de rafraîchissement afin de déterminer quelles données rafraîchir à quels moments en fonction des utilisateurs, de leur usage des données.

Un troisième axe porte sur l'influence de la personnalisation dans la modélisation multidimensionnelle. Des travaux ont étudié cet aspect en proposant d'enrichir et de transformer les hiérarchies du modèle en tenant compte de spécificités exprimées par l'utilisateur (Favre et al., 2007). Ces travaux peuvent être complétés notamment par la génération de hiérarchies basée sur la collecte automatique des caractéristiques de l'utilisateur. Ces nouvelles hiérarchies peuvent être proposées à l'utilisateur dans un système de recommandation. Une autre voie concerne le couplage de techniques de fouille de données et de la modélisation multidimensionnelle pour inférer des hiérarchies à recommander (Bentayeb, 2008). Il est intéressant de pouvoir exploiter les connaissances cachées dans les données, ce qui constitue un des objectifs de la fouille de données. Les techniques d'apprentissage non supervisé peuvent trouver dans ce contexte un intérêt. Le couplage avec la fouille de données constitue un axe prometteur, car cet apprentissage peut concerner les données elles-mêmes, mais également

l'utilisateur si l'on applique des techniques d'apprentissage sur les usages individuels qu'il fait du système.

Un quatrième axe d'étude repose sur la définition de structures de visualisation alternatives à la visualisation traditionnelle sous forme de table à deux dimensions (Gyssens et Lakshmanan, 1997). La présentation des données est un point important car les utilisateurs réclament des moyens d'accès facilitant l'analyse des données. Différents travaux proposent de nouvelles formes de visualisation (Mansmann et Scholl, 2007 ; Choong et al., 2006 ; Sifer, 2003 ; Stolte et al., 2002), mais au-delà des structures de visualisation, des travaux relatifs au résumé de données tels que ceux proposés par Choong et al. (2008) semblent prometteurs. La problématique est alors d'inclure le processus de personnalisation par rapport à la visualisation et au résumé d'informations. Outre la définition de différentes structures de visualisation, une problématique nouvelle est l'adaptation de la structure de présentation des données en fonction des données analysées. Une première approche existante consistait à réorganiser les modalités dans les dimensions selon les ordonnancements fournis par l'Analyse en Correspondances Multiples (ACM). Grâce à cette réorganisation, les modalités d'une dimension sont agrégées selon l'ordre de leur proximité et non selon l'ordre de leur appartenance hiérarchique établi (Ben Messaoud et al., 2007). L'objectif sera alors de prendre en compte, dans la définition et l'utilisation de structures de visualisation, des caractéristiques propres à l'utilisateur, par exemple, son niveau d'expertise.

Le cinquième et dernier axe est transversal par rapport aux axes cités précédemment. En effet, il s'agit de poser les problèmes relatifs à la sécurité. Il est possible de considérer que la personnalisation et la sécurité, dans certains contextes, peuvent reposer sur des concepts similaires, celui de profil entre autres. La principale différence est alors la suivante : dans un cas l'utilisateur préfère ou non accéder à telles informations, et dans l'autre l'utilisateur a le droit ou non d'y accéder. Selon la solution de personnalisation proposée, celle-ci peut se baser sur des données plus ou moins « personnelles ». Se pose alors le problème de la protection des données servant à la personnalisation. Ce problème a été évoqué dans le contexte du Web (Wang et Kobsa, 2007). Il peut également se poser dans le contexte des entrepôts de données au niveau de l'accès sécurisé aux données (dont les stratégies présentent des particularités en entrepôt de données, vis-à-vis des niveaux de granularité de l'information).

Références

- Aas, K. (1997). A Survey on Personalised Information Filtering Systems for the World Wide Web. Technical report, Norwegian Computing Center.
- Adomavicius, G. et A. Tuzhilin (2008). Context-aware Recommender Systems. In *RecSys 08*, pp. 335–336.
- Agrawal, R. et E. L. Wimmers (2000). A Framework for Expressing and Combining Preferences. In *SIGMOD 00*, pp. 297–306.
- Aouiche, K., D. Lemire, et R. Godin (2008). Collaborative OLAP with Tag Clouds: Web 2.0 OLAP Formalism and Experimental Evaluation. In *WEBIST 08*.

Personnalisation dans les entrepôts de données

- Anli, A. (2006). *Méthodologie de développement des systèmes d'information personnalisés. Application à un système d'information au service des usagers des transports terrestres de personnes*. Thèse de doctorat, Université de Valenciennes et du Hainaut-Cambresis.
- Bellatreche, L., A. Giacometti, P. Marcel, H. Mouloudi, et D. Laurent (2005). A Personalization Framework for OLAP Queries. In DOLAP 05, pp. 9–18.
- Ben Messaoud R., O. Boussaid, et S. Loudcher Rabaséda (2007). A Multiple Correspondence Analysis to Organize Data Cubes. *Databases and Information Systems IV* 155(1), 133–146.
- Bentayeb, F. (2008). K-Means based Approach for OLAP Dimension Updates. In ICEIS 08, pp. 531–534.
- Bouzeghoub, M. et D. Kostadinov (2005). Personnalisation de l'information : aperçu de l'état de l'art et définition d'un modèle flexible de profils. In CORIA 05, pp. 201–218.
- Bradley K., R. Rafter et B. Smyth (2000). Case-Based User profiling for Content Personalisation. In AH 00, Volume 1892 of LNCS, pp. 62–72.
- Brando C., M. Goncalves, et V. González (2007). Evaluating Top-k Skyline Queries over Relational Databases, *Database and Expert Systems Applications*, Volume 4653 of LNCS, pp. 254–263.
- Castagnos S. (2008). *Modélisation de comportements et apprentissage stochastique non supervisé de stratégies d'interactions sociales au sein de systèmes temps réel de recherche et d'accès à l'information*. Thèse de doctorat, Université Nancy 2.
- Cherniack M., E. F. Galvez, M. J. Franklin et S. Zdonik (2003). Profile-driven Cache Management. In ICDE 03, pp. 645–656.
- Chomicki, J. (2003). Preference Formulas in Relational Queries. *ACM Transactions on Database Systems (TODS)* 28(4), 427–466.
- Choong, Y. W., A. Laurent et D. Laurent (2008). Mining Multiple-Level Fuzzy Blocks from Multidimensional Data. *Fuzzy Sets and Systems* 159(12), 1535–1553.
- Choong, Y. W., A. Laurent, et D. Laurent (2006). Pixelizing Data Cubes: A Block-Based Approach. In VIEW 06, pp. 63–76.
- Domshlak C. et T. Joachims (2007). Efficient and Non-Parametric Reasoning over User Preferences. *User Modeling and User-Adapted Interaction* 17(1-2), 41–69.
- Espil, M. M. et A. A. Vaisman (2001). Efficient Intensional Redefinition of Aggregation Hierarchies in Multidimensional Databases. In DOLAP 01, pp. 1–8.
- Favre, C., F. Bentayeb, et O. Boussaid (2007). Évolution et personnalisation des analyses dans les entrepôts de données : une approche orientée utilisateur. In INFORSID 07, pp. 308–323.
- Giacometti, A., P. Marcel, et E. Negre (2008). A Framework for Recommending OLAP Queries. In DOLAP 08, pp. 73–80.
- Gilbert, J. E. et C. Y. Han (1999). Arthur : Adapting Instruction to Accommodate Learning Style. In WebNet 99, pp. 433–438.

- Goldberg, D., D. Nichols, B. M. Oki, et D. Terry (1992). Using Collaborative Filtering to Weave an Information Tapestry. *Communications of the ACM* 35(12), 61–70.
- Gray, J., A. Bosworth, A. Layman, et H. Pirahesh (1996). Data Cube : A Relational Aggregation Operator Generalizing Group-By, Cross-Tab, and Sub-Total. In *ICDE 96*, pp. 152–159.
- Gyssens, M. et L. V. S. Lakshmanan (1997). A Foundation for Multi-Dimensional Databases. In *VLDB 97*, pp. 106–115.
- Inmon, W. H. (1996). *Building the Data Warehouse*. John Wiley & Sons.
- Ioannidis, Y. et G. Koutrika (2005). Personalized Systems : Models and Methods from an IR and DB Perspective. In *VLDB 05*, pp. 1365–1365.
- Jerbi, H., F. Ravat, O. Teste, et G. Zurfluh (2008). Management of Context-Aware Preferences in Multidimensional Databases. In *ICDIM 08*, pp. 669-675.
- Jerbi, H., F. Ravat, O. Teste, et G. Zurfluh (2009). Applying Recommendation Technology in OLAP Systems. In *ICEIS 09*, pp. 220-233.
- Kiessling, W. (2002). Foundations of Preferences in Database Systems. In *VLDB 02*, pp. 311–322.
- Kimball, R. (1996). *The Data Warehouse Toolkit*. JohnWiley & Sons.
- Kobsa, A. (2007). Generic User Modeling Systems. *The Adaptive Web*, 136–154.
- Korfhage, R. R. (1997). *Information Storage and Retrieval*. JohnWiley & Sons.
- Kotidis, Y. et N. Roussopoulos (1999). DynaMat: A Dynamic View Management System for Data Warehouses. *SIGMOD Record* 28(2), 371–382.
- Koutrika, G. et Y. Ioannidis (2005). Personalized Queries under a Generalized Preference Model. In *ICDE 05*, pp. 841–852.
- Lacroix, M. et P. Lavency (1987). Preferences; Putting More Knowledge into Queries. In *VLDB 87*, pp. 217–225.
- Lee, J., G. You, et S. Hwang (2009). Personalized Top-k Skyline Queries in High-Dimensional Space. *Information Systems* 34(1), 45–61.
- Lehner, W. (1998). Modelling Large Scale OLAP Scenarios. In *EDBT 98*, Volume 1377 of LNCS, pp. 153–167.
- Lieberman, H. (1995). Letizia: An Agent That Assists Web Browsing. In *IJCAI 95*, pp. 924–929.
- Luján-Mora, S. et J. Trujillo (2003). A Comprehensive Method for Data Warehouse Design. In *DMDW 03*, Volume 77 of CEUR Workshop Proceedings.
- Mansmann, S. et M. H. Scholl (2007). Exploring OLAP Aggregates with Hierarchical Visualization Techniques. In *SAC 07*, pp. 1067–1073.
- Macías, J. A. et F. Paternò (2008). Customization of Web Applications through an Intelligent Environment Exploiting Logical Interface Descriptions. *Interacting with Computers* 20(1), 29–47.

Personnalisation dans les entrepôts de données

- Morfonios, K. et G. Koutrika (2008). OLAP Cubes for Social Searches: Standing on the Shoulders of Giants? In WebDB 08.
- Oulasvirta, A. et J. Blom (2008). Motivations in Personalisation Behaviour. *Interacting with Computers* 20(1), 1–16.
- Tchienehom, P. (2006). Modélisation et Exploitation de Profils : Accès Sémantique à des Ressources. Thèse de doctorat, Université de Toulouse.
- Pretschner A. et S. Gauch (1999). Ontology Based Personalized Search. In ICTAI 99, pp. 391–398.
- Ravat, F. et O. Teste (2008). Personalization and OLAP Databases. *New Trends in Data Warehousing and Data Analysis*, 3, 71–92.
- Rizzi, S. (2007). OLAP Preferences: a Research Agenda. In DOLAP 07, pp. 99–100.
- Sifer, M. (2003). A Visual Interface Technique for Exploring OLAP Data with Coordinated Dimension Hierarchies. In CIKM 03, pp. 532–535.
- Specht, M. et R. Oppermann (1998). ACE - Adaptive Courseware Environment. *The New Review of Hypermedia and Multimedia* 4, 141–162.
- Stern, M. et B. P. Woolf (2000). Adaptive Content in an Online Lecture System. In AH 00, Volume 1892 of LNCS, pp. 227–238.
- Stolte, C., D. Tang, et P. Hanrahan (2002). Polaris: A System for Query, Analysis, and Visualization. *IEEE Transactions on Visualization and Computer Graphics* 8(1), 52–65.
- Tchienehom, P. L. (2005). Modèle générique de profils pour la personnalisation de l'accès à l'information. In INFORSID 05, pp. 269–284.
- Thalhammer, T., M. Schrefl, et M. Mohania (2001). Active DataWarehouses: Complementing OLAP with Analysis Rules. *Data and Knowledge Engineering* 39(3), 241–269.
- Trujillo, J. et S. Luján-Mora (2003). A UML Based Approach for Modeling ETL Processes in Data Warehouses. In ER 03, Volume 2813 of LNCS, pp. 307–320.
- Vassiliadis, P., A. Simitsis, et S. Skiadopoulos (2002). On the Logical Modeling of ETL Processes. In CAiSE 02, pp. 782–786.
- Wang, Y. et A. Kobsa (2007). Respecting Users' Individual Privacy Constraints in Web Personalization. In UM 07, pp. 157–166.
- Widom, J. (1995). Research Problems in Data Warehousing. In CIKM 95, pp. 25–30.

Summary

Data warehouses aim at collecting, transforming and preparing data issued from operating systems, in order to support decision process. Although data warehouses are considered user-centered, personalization within data warehouses is an emergent topic. Thus, in this paper, we study issues and opportunities about the personalization in data warehouses.