
Recommandation de requêtes dans les bases de données multidimensionnelles annotées

Houssem Jerbi — Geneviève Pujolle — Franck Ravat — Olivier Teste

*Université de Toulouse, IRIT (UMR5505)
118, Route de Narbonne - F-31062 Toulouse cedex 9
{jerbi, pujolle, ravat, teste}@irit.fr*

RÉSUMÉ. La problématique traitée dans cet article consiste à personnaliser les systèmes OLAP annotés. Nous proposons de modéliser les données au sein d'une constellation supportant à la fois des annotations et des préférences. Les annotations sont utilisées pour représenter « l'expertise » immatérielle du décideur tandis que les préférences permettent d'individualiser les données durant les manipulations OLAP. Les préférences sont exploitées pour des recommandations contextuelles annotées qui assistent l'utilisateur dans son exploration de l'espace multidimensionnel.

ABSTRACT. This paper deals with personalization of annotated OLAP systems. Constellation data are extended with annotations and user preferences. Annotations reflect the decision-maker experience, whereas preferences enable users to focus on the most interesting data. User preferences allow annotated contextual recommendations that assist the decision-maker during his/her multidimensional navigations.

MOTS-CLÉS : Système d'aide à la décision, Bases de données multidimensionnelles, Personnalisation.

KEYWORDS : Decision-support system, Multidimensional databases, Personalization.

1. Introduction

Les systèmes OLAP (« On-Line Analytical Processing ») facilitent l'analyse en offrant un espace de représentation multidimensionnelle des données que les décideurs explorent interactivement (Kimbal, 1996). Cette approche a connu un développement important grâce à sa capacité à permettre un accès direct et interactif aux données décisionnelles en intégrant ces dernières dans un entrepôt de données (Teste, 2000). Cependant ces systèmes sont élaborés pour un groupe de décideurs ou un sujet d'analyse (« subject-oriented » (Inmon, 1994)) pour lesquels sont présumés des besoins parfaitement identiques. Cette simplification rend les systèmes OLAP parfois mal adaptés à un usage particulier. Le décideur se trouve alors confronté à un espace multidimensionnel, souvent très vaste, sur lequel il doit opérer un nombre important de manipulations afin d'obtenir le résultat le plus proche possible de son besoin.

La problématique traitée dans cet article consiste à personnaliser les systèmes OLAP en fonction de besoins analytiques individuels. Les mécanismes de personnalisation dans les systèmes OLAP, visant à mieux prendre en compte l'utilisateur, ne sont que très peu étudiés. La personnalisation peut prendre la forme de recommandations de données décisionnelles afin de simplifier voire d'optimiser les analyses OLAP. Mais les données brutes peuvent s'avérer insuffisantes pour la prise de décision. En effet, les systèmes OLAP sont généralement réduits à la mise à disposition des données décisionnelles tandis que les décideurs doivent analyser ces données en s'appuyant sur la base immatérielle de leur expertise. Il serait donc judicieux d'y ajouter des réflexions, des interprétations et des commentaires des décideurs. L'originalité de l'approche proposée réside dans le couplage des recommandations avec un processus d'annotations qui permet de prendre en compte non seulement des données brutes manipulées mais également les avis des décideurs.

2. Etat de l'art

Plusieurs travaux ont été menés sur la personnalisation en OLAP. Trois principales fonctionnalités ont été étudiées pour faciliter l'accès aux données OLAP : la **personnalisation du schéma des sources des données** (Bentayeb et al., 2009), (Garrigos et al., 2009) permettant d'adapter les structures des données aux besoins spécifiques des usagers, la **personnalisation de l'interrogation des données** qui consiste à personnaliser la visualisation de la requête (Bellatreche et al., 2005), ou bien du résultat renvoyé (Golfarelli & Rizzi, 2009), (Xin et al., 2006), (Jerbi et al., 2008), et la **recommandation de requêtes** visant à proposer à l'utilisateur des requêtes (Giacometti et al., 2008, 2009) additionnelles afin de l'aider dans l'exploration d'un contenu volumineux d'informations.

La personnalisation se base généralement sur des profils utilisateurs (Korfhage, 1997) pouvant s'apparenter à une modélisation de l'utilisateur/décideur. Aucun

consensus n'existe malgré la proposition d'un profil générique multidimensionnel visant à couvrir une majorité de contextes (Bouzeghoub *et al.*, 2005) ou encore, celle de profils flexibles interopérables (Chevalier, *et al.*, 2007). Dans le domaine OLAP, un profil se définit par un ensemble de préférences servant à configurer ou à adapter le système à l'utilisateur (Bellatreche *et al.*, 2005), (Golfarelli, Rizzi, 2009), (Jerbi *et al.*, 2008), éventuellement rattachées à des contextes définissant leurs cadres d'application (Jerbi *et al.*, 2008), (Garrigos *et al.*, 2009). Ces préférences peuvent être quantitatives dans le cas où elles sont définies par l'association de scores numériques aux éléments considérés, ou qualitatives pour des préférences spécifiées directement à l'aide de relations binaires. Certains travaux proposent de personnaliser en fonction de règles utilisateurs (Garrigós *et al.*, 2009) ou de scores (Xin *et al.*, 2006).

La personnalisation du schéma est assurée soit par l'ajout de nouveaux niveaux de granularité (Bentayeb *et al.*, 2009), ou bien par le filtrage des dimensions et/ou d'attributs (Garrigós *et al.*, 2009) pour adapter le schéma à l'évolution des besoins des usagers dans le temps. Ce processus est mis en place durant l'étape de conception de la BDM, ou même après son implantation (Bentayeb *et al.*, 2009). La personnalisation des requêtes OLAP vise à modifier la requête ou son résultat afin de mieux satisfaire l'utilisateur. Le travail de (Bellatreche *et al.*, 2005) consiste à personnaliser la visualisation du résultat d'une requête OLAP en fonction d'une contrainte de visualisation et d'un ensemble de préférences sur les valeurs de la BDM. Il s'agit d'extraire la meilleure sous-partie du résultat de la requête qui respecte la contrainte. D'autres travaux enrichissent la requête utilisateur par des préférences afin de générer une requête augmentée qui est exécutée. L'objectif de ces travaux est de trier le résultat généré (Golfarelli, Rizzi, 2009), (Xin *et al.*, 2006) pour réduire la charge de navigation de l'utilisateur (Jerbi *et al.*, 2008). Par ailleurs, (Giacometti *et al.*, 2008, 2009) proposent un cadre pour la recommandation des requêtes sur la base de calculs opérés sur l'historique des navigations réalisées par un groupe d'utilisateurs. Ces travaux recommandent des requêtes complètes déjà jouées par le groupe d'usagers, introduisant des approximations pouvant être importantes.

Ces travaux se focalisent sur la personnalisation des données brutes entreposées dans les systèmes OLAP sans prendre en compte l'expertise des décideurs. Dans (Eltabakh *et al.*, 2009), (Pérez *et al.*, 2009), les annotations sont intégrées sur les valeurs ainsi que sur les structures d'une BDM afin de mémoriser les réflexions des décideurs (Cabanac *et al.*, 2007). Lors de la restitution du résultat d'une requête, l'affichage des annotations relatives offre un cadre d'analyse personnalisé. La visualisation des annotations d'un groupe d'utilisateurs permet aux décideurs de profiter des expertises des autres ou bien de prendre des décisions d'une façon collaborative. Ces travaux proposent l'intégration d'annotations sans offrir à l'utilisateur une assistance dans la navigation de la BDM comme nous le proposons.

Tableau 1: Etude comparative sur la personnalisation OLAP.

		Beilatreche et al., 2005	Bentayeb et al., 2009	Etkabakh et al., 2009	Garrigós et al., 2009	Giacometti et al., 2009	Golfarelli & Rizzi, 2009	Jerbi et al., 2008	Pérez et al., 2009	Xin et al., 2006
Approche	Perso. Schéma BDM		x		x					
	Objectif	Résultat requête Visualis.		x	x		x	x		x
Modélisation utilisateur		Recomm. requêtes			x					
	Forme	Règles/fonctions			x					x
		Préférences	Qualita.				Qualita.	Qualita		
Modélisation utilisateur		Annotations		x					x	
	Niveau	Valeurs	x	x	x		x			x
		Structures		x	x		x	x		
	Contextualisation		non	non	oui	non	non	oui	oui	non
Algorithme	Timing (% exécution requête)	après	avant	après	avant	avant	avant	après	après	avant
		Profil / règles/ annota.	x	x	x			x	x	
	Entrée	Instance de la BDM	x		x		x			x
Algorithme		Schéma de la BDM	x		x		x			
		Autre								
	Sortie	Résultat requête	Attribut dimension	Annota.	Requêtes {Q _i }	Log requêtes	Requête G'	Contexte courant	Annota.	Résultat requête
		Résultat requête			Requêtes {Q _i }	Requêtes {Q _i }	Requête G'	Requête G'		

Le tableau 1 décrit ces travaux selon différentes catégories.

- L'approche générale (*objectif* des travaux) peut concerner la personnalisation du schéma ou de la requête (de son résultat ou de sa visualisation), ou peut être de la recommandation de requêtes.
- Le modèle utilisateur varie selon la *forme* (règles, scores, préférences, annotations) et selon le *niveau* auquel les caractéristiques des utilisateurs sont spécifiées (valeurs et/ou structures). Certains modèles représentent l'utilisateur par des informations qui dépendent également du *contexte*.
- Les algorithmes de personnalisation diffèrent selon l'instant (*timing*) d'exécution (avant ou après la requête). Ils se basent sur des paramètres différents (préférences utilisateurs, instances de la BDM, schéma de BDM...). Leur exécution permet de renvoyer des résultats de types différents, à savoir une requête enrichie Q' , un résultat de requête, un ensemble de requêtes supplémentaires $\{Q_i\}$, des annotations liées au résultat ou des attributs de dimension.

3. Modélisation

Notre objectif est de modéliser les données sous un format multidimensionnel en intégrant des mécanismes de personnalisation. Ces derniers permettent de matérialiser l'expertise des décideurs ainsi que leurs besoins individuels en données analysées. Cette section explicite les concepts OLAP de modélisation des données, les principes retenus pour modéliser une analyse multidimensionnelle, les mécanismes de matérialisation de l'expertise ainsi que les principes de personnalisation des données.

3.1. Schéma en constellation

Un schéma en constellation modélise les données décisionnelles au travers de faits (sujet d'analyse) et de dimensions (axes d'analyse). On pose : F l'ensemble des faits, D l'ensemble des dimensions, H l'ensemble des hiérarchies, M l'ensemble des mesures, A l'ensemble des attributs de dimension, P l'ensemble des paramètres, W l'ensemble des attributs faibles.

Définition 1. Une constellation personnalisée CP est définie par $(F; D; Star^{CP}; Preference^{CP}; Annotate^{CP})$ où

- $F = \{F_1, \dots, F_n\}$ est l'ensemble des faits,
- $D = \{D_1, \dots, D_m\}$ est l'ensemble des dimensions,
- $Star^{CP} : F \rightarrow 2^D$ associe chaque fait à un sous-ensemble des dimensions en fonction desquelles il est analysable,
- $Preference^{CP} = \{P^{CP}_1, P^{CP}_2, \dots\}$ est l'ensemble de préférences de personnalisation,
- $Annotate^{CP} = \{AD^{CP}_1, AD^{CP}_2, \dots\}$ est l'ensemble des annotations décisionnelles.

La constellation constitue une généralisation de la modélisation en étoile (Kimball, 1996) ; si $|F| = 1$ alors C est une étoile. Par rapport à la définition classique d'une constellation, l'extension réside dans la capacité d'une constellation de matérialiser l'expertise des décideurs par le mécanisme d'annotation (*Annotate^{CP}*) et par l'intégration de préférences (*Preference^{CP}*) permettant sa personnalisation.

Un fait est composé de mesures représentant les indicateurs d'analyse.

Définition 2. $\forall i \in [1..n]$ un fait F_i est défini par $(NF_i ; M_i)$ où

- NF_i est le nom identifiant le fait dans la constellation,
- $M_i = \{m_1, \dots, m_{x_i}\}$ est l'ensemble des mesures.

Exemple. La figure 1 donne une représentation graphique d'une constellation. Dans cette figure, le fait *PANIER* est composé de trois mesures *QTEC*, *MNT* et *NDPR*. Il est décrit de la manière suivante: (*FVentes*; $\langle QTEC, MNT, NDPR \rangle$).

Une dimension, ou axe d'analyse, est composée de hiérarchies représentant les différents niveaux d'agrégation des mesures. Chaque niveau est caractérisé par, au minimum, un paramètre et éventuellement des attributs faibles.

Définition 3. $\forall i \in [1..m]$ une dimension D_i est définie par $(ND_i ; A_i ; H_i)$ où

- ND_i est le nom identifiant la dimension dans la constellation,
- $A_i = \{Id_i, All_i\} \cup P_i \cup W_i$ est l'ensemble des attributs de la dimension. On distingue les *paramètres* $P_i \subseteq P$ représentant les graduations possibles, des *attributs faibles* $W_i \subseteq W$ représentant des informations additionnelles associées aux paramètres.
- $H_i = \{H_1, \dots, H_{p_i}\} \subseteq H$ est l'ensemble des hiérarchies.

Les attributs de D_i respectent les propriétés de disjonction $\forall (j_1, j_2) \in [1..m]^2$, si $j_1 \neq j_2$ alors $A_{j_1} \neq A_{j_2}$ et de recouvrement $A = \bigcup_{i=1}^m A_i$, $P = \bigcup_{i=1}^m P_i$ et $W = \bigcup_{i=1}^m W_i$.

Définition 4. $\forall H_j \in H_i$ une hiérarchie H_j est définie par $(NH_j ; P_{H_j} ; \prec_{H_j} ; Weak_{H_j})$ où

- NH_j est le nom identifiant la hiérarchie dans la constellation,
- $P_{H_j} = \{p_1, \dots, p_y\} \subseteq P$ est l'ensemble des paramètres de la hiérarchie,
- \prec_{H_j} est une relation d'ordre sur P_{H_j} telle que
 - l'ordonnancement des paramètres suit un ordre total $\forall p_{k1} \in P_{H_j}, p_{k2} \in P_{H_j}, k_1 \neq k_2, p_{k1} \prec_{H_j} p_{k2} \vee p_{k2} \prec_{H_j} p_{k1}$
 - il existe un paramètre *racine* $\forall p_{k1} \in P_{H_j}, Id_i \prec_{H_j} p_{k1}$
 - il existe un paramètre *extrémité* $\forall p_{k1} \in P_{H_j}, p_{k1} \prec_{H_j} All_i$
- $Weak_{H_j} : P_{H_j} \rightarrow 2^{W_{H_j}}$ associe les paramètres à un ensemble d'attributs faibles.

Les hiérarchies respectent les propriétés de disjonction $\forall i_1 \in [1..m], \forall i_2 \in [1..m]$, si $i_1 \neq i_2$ alors $H_{i_1} \neq H_{i_2}$ et de recouvrement $H = \bigcup_{i=1}^m H_i$.

Associés à l'ensemble des concepts, nous définissons, par extension des notations introduites dans (Golfarelli, *et al.*, 1998), les formalismes graphiques représentant le schéma conceptuel décrivant les structures d'une constellation. La figure 1 présente un exemple de constellation avec les formalismes graphiques associés aux concepts.

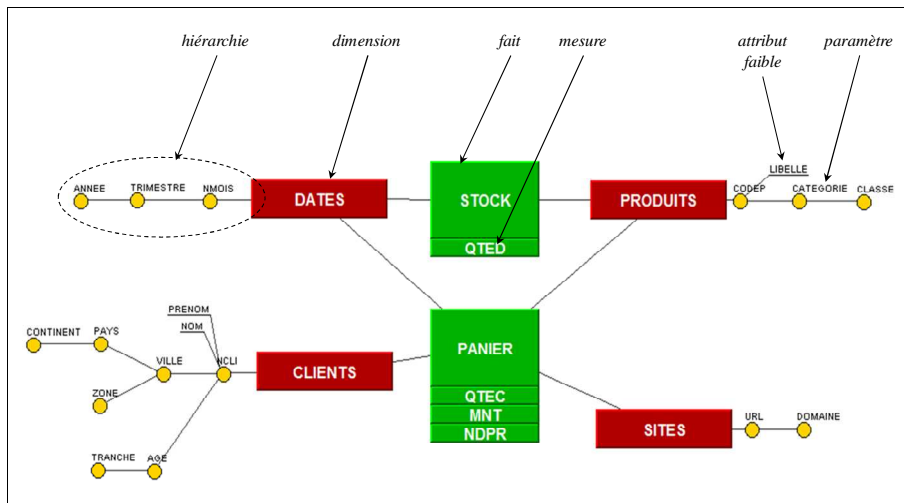


Figure 1. Exemple de constellation

Exemple. La constellation de la figure 1 est composée des éléments suivants : $F = \{PANIER, STOCK\}$ et $D = \{CLIENTS, PRODUITS, DATES, SITES\}$. Cette constellation comporte deux étoiles :

- $Star(PANIER) = \{CLIENTS, PRODUITS, DATES, SITES\}$ et
- $Star(STOCK) = \{PRODUITS, DATES\}$.

Dans l'exemple présenté à la figure 1, la dimension *CLIENTS* est décrite de la manière suivante: ('*Clients*' ; $\{NCLI, NOM, PRENOM, VILLE, PAYS, CONTINENT, AGE, TRANCHE, ZONE\}$; $\{HGEO, HAGE, HZON\}$). La dimension comprend

- sept paramètres *NCLI*, *VILLE*, *PAYS*, *CONTINENT*, *AGE*, *TRANCHE*, *ZONE*,
- deux attributs faibles *NOM*, *PRENOM*, et
- trois hiérarchies notées *HGEO*, *HAGE* et *HZON*.

Ces dernières sont définies de la manière suivante :

- ('*HGEO*' ; $\{NCLI, NOM, PRENOM, VILLE, PAYS, CONTINENT\}$; $N^{\circ}CLI \prec_{HGEO} VILLE \prec_{HGEO} PAYS \prec_{HGEO} CONTINENT$; $\{NCLI \rightarrow \{NOM, PRENOM\}\}$),

- ('HAGE' ; {NCLI, NOM, PRENOM, AGE, TRANCHE}; N°CLI <_{HAGE} AGE <_{HAGE} TRANCHE; {NCLI → {NOM, PRENOM}}), et,
- ('HZON' ; {NCLI, NOM, PRENOM, VILLE, ZONE}; NCLI <_{HZON} VILLE <_{HZON} PAYS <_{HZON} ZONE; {NCLI → {NOM, PRENOM}}).

3.2. Analyse OLAP

Les composants d'un schéma OLAP définis précédemment servent de supports aux analyses effectuées par les décideurs.

3.2.1. Contexte d'analyse

Les décideurs explorent interactivement l'espace multidimensionnel d'une constellation par une succession d'opérations de manipulation OLAP (Jerbi *et al.*, 2008) afin d'obtenir le résultat souhaité. Ce résultat peut prendre la forme d'une table multidimensionnelle (Ravat *et al.*, 2007), d'un graphique, etc.

Exemple. A partir de la constellation définie précédemment, considérons une analyse trimestrielle des montants des paniers ainsi que le panier maximum en fonction de la tranche d'âge des clients pour l'année 2009. Les données sont le plus souvent visualisées au sein de structures de visualisation facilitant la compréhension et l'analyse de ces données. La figure 2 visualise les données selon deux dimensions en ligne et en colonne au travers d'une table multidimensionnelle (Ravat *et al.*, 2007).

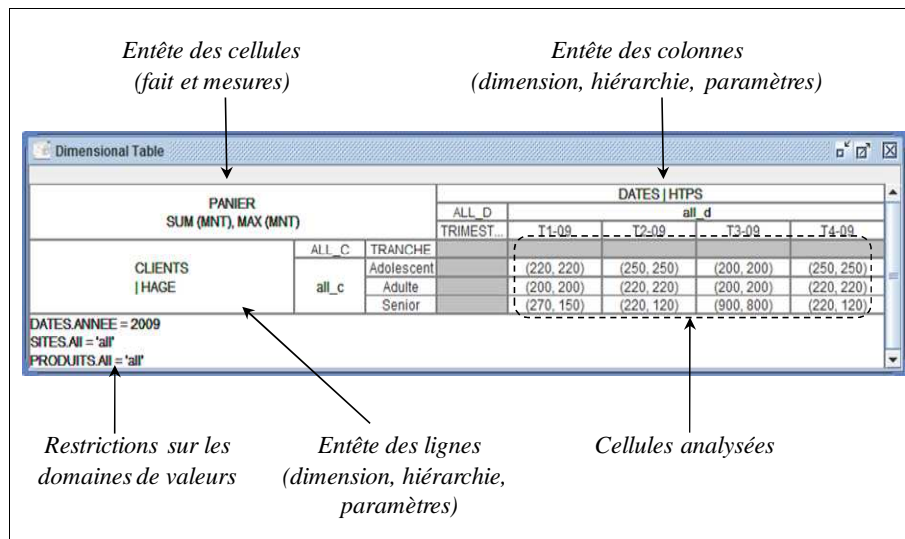


Figure 2. Exemple d'analyse visualisée par une table multidimensionnelle

Afin de rendre l'approche indépendante des structures de visualisation, nous définissons le concept de contexte d'analyse par une description arborescente des tables multidimensionnelles ou des représentations graphiques associées (Jerbi *et al.*, 2008). Dans les sections suivantes, nous définissons les contextes d'analyses et les opérations OLAP permettant de « naviguer » de contextes en contextes.

Définition 5. Un contexte d'analyse CA est défini par $(\mathcal{C}^F; \mathcal{C}^D; \mathcal{C}^R)$ où

- $\mathcal{C}^F = NF_i[f(m) \in \{\{val\}+\}]_+$ représente le sujet (fait F_i) en cours d'analyse,
- $\mathcal{C}^D = \{C^{D^1}, \dots, C^{D^u}\}$ représente les axes de l'analyse en cours avec $\forall i \in [1..u], C^{D^i} = ND_i[NH_i]?(p_{k1}, p_{k2}) \in \{\{(val_1, val_2)\}+\}]_+$ où $p_{k1} \in A^{D^i}, p_{k2} \in A^{D^i}$ et $(val_1, val_2) \in dom(p_{k1}) \times dom(p_{k2})$,
- $\mathcal{C}^R = \{pred^F, pred^{D^1}, \dots, pred^{D^u}\}$ est l'ensemble des prédicats définissant les restrictions sur les valeurs analysées.

Exemple. Pour illustrer le concept de contexte d'analyse, considérons l'analyse portant sur la visualisation trimestriellement les montants des paniers ainsi que le panier maximum en fonction de la tranche d'âge des clients, pour l'année 2009.

Le contexte d'analyse CA_4 se définit par l'expression $(C^F_4; \mathcal{C}^D_4; \mathcal{C}^R_4)$ où :

- $C^F_4 = PANIER/SUM(MNT) \in \{ 220;250;200;250; 200;220;200;220; 270;220;900;220\} /MAX(MNT) \in \{ 220;250;200;250; 200;220;200;220; 150;120;800;120\}$
- $\mathcal{C}^D_4 = \{ CLIENTS.HAGE/(ALL_C; TRANCHE) \in \{Adolescent; Adulte; Senior\}; DATES.HTPS/(ALL_D; TRIMESTRE) \in \{T1-09; T2-09; T3-09; T4-09\} \}$
- $\mathcal{C}^R_4 = \{ CLIENTS.ALL_C = 'all_c'; DATES.ANNEE = 2009; SITES.ALL_S = 'all_s'; PRODUITS.ALL_P = 'all_p' \}$

Afin de faciliter la gestion des contextes d'analyse, nous proposons de représenter ces derniers à l'aide d'une forme arborescente comme l'illustre la figure 3. L'arbre est constitué de trois sous-arbres :

- C^F_4 décrit l'entête des cellules (fait et mesures) et les cellules analysées (valeurs des mesures) ;
- C^D_4 décrit l'entête des lignes et des colonnes (dimensions, paramètres, attributs faibles et valeurs des attributs). On peut noter que la structure du contexte d'analyse peut être étendue pour décrire des structures d'analyse au-delà de deux dimensions ;
- C^R_4 décrit les éventuelles restrictions sur les domaines de valeurs (valeurs des mesures et/ou des paramètres).

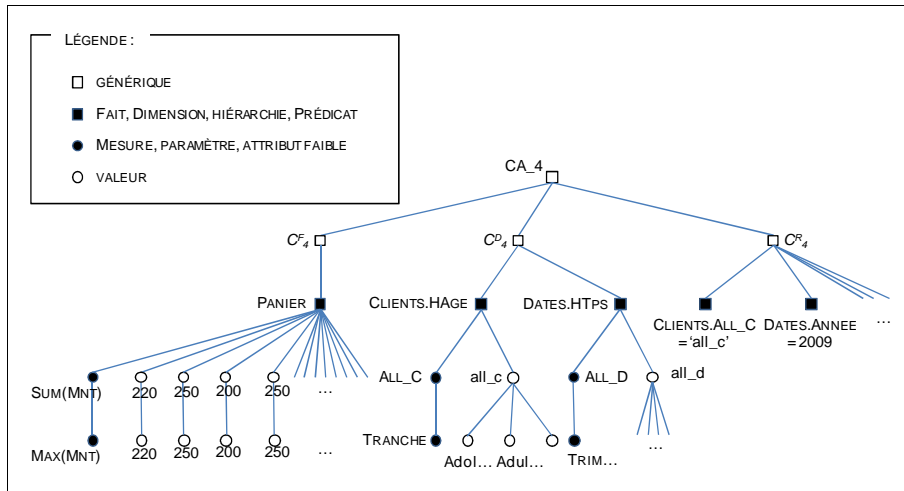


Figure 3. Représentation arborescente d'un contexte d'analyse

3.2.2. Analyse OLAP

Un contexte d'analyse représente l'état courant de l'analyse menée par un décideur. Une analyse OLAP consiste à explorer l'espace multidimensionnel de la BD multidimensionnelle par une succession de contextes d'analyse obtenus par l'application d'opérations OLAP (Ravat *et al.*, 2007). La figure 4 illustre ce principe de navigation par contextes d'analyse.

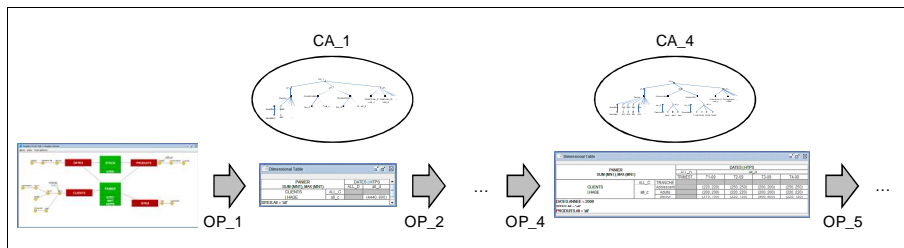


Figure 4. Principe de navigation par contextes d'analyse

Exemple. Le contexte d'analyse présenté à la figure 3 est obtenu en appliquant la succession d'opérations de manipulation OLAP suivante.

- La première opération (OP_1) permet d'afficher le montant des paniers et le panier maximum pour toutes les dates et les clients entreposés dans la constellation.

$DISPLAY(PANIER; \{SUM(MNT); MAX(MNT)\}, CLIENTS; HAGE, DATES, HTPS) = CA_1$

- La seconde opération (OP_2) permet d'affiner les données en obtenant les montants des paniers et le panier maximum par tranche d'âge des clients.
 $DRILLDOWN(CA_1; CLIENTS; TRANCHE) = CA_2$
- La troisième opération (OP_3) restreint l'analyse sur l'année 2009.
 $SELECT(CA_2; CLIENTS.ANNEE = 2009) = CA_3$
- La quatrième opération (OP_4) permet de détailler l'analyse en visualisant les données par trimestre de l'année sélectionnée préalablement.
 $DRILLDOWN(CA_3; DATES; TRIMESTRE) = CA_4$

Nous ne présentons pas de façon détaillée l'algèbre OLAP que nous utilisons. Toutefois, le lecteur pourra trouver une présentation exhaustive dans (Ravat *et al.*, 2007, 2008).

3.3. Annotations

3.3.1. Principes

Les prises de décisions reposent non seulement sur les données brutes mais également sur les réflexions des décideurs voire la confrontation de différentes interprétations. Notre proposition consiste à modéliser au travers d'annotations le capital immatériel mentalement associé aux données par les décideurs.

Les annotations visent à conserver les commentaires et les discussions formulés lors des analyses et du processus de prise de décisions. Ce cadre informatique permet d'exploiter et de partager les données multidimensionnelles tout en supportant des fonctionnalités d'annotation permettant d'enrichir interactivement les composants d'une constellation. Les décideurs sont alors des usagers actifs créant leur propre système de repérage au travers de signes graphiques (sur-lignage, cerclage...), de commentaires et de réponses (affirmation, infirmation) pouvant impliquer des fils de discussions (communication asynchrone) (Cabanac, *et al.*, 2010). L'expertise que véhiculent ces annotations est utilisée à des fins personnelles ou collectives et elles peuvent contribuer à améliorer les analyses futures. Les annotations contiennent des informations subjectives et des informations objectives (Cabanac, *et al.*, 2007).

Définition 6. Une annotation AD_x^{CP} est définie par $(IS_x^{AD}; IO_x^{AD})$ où

- IS_x^{AD} est un ensemble d'informations subjectives regroupant :
 - le contenu textuel saisi par le décideur qui annote,
 - le type de l'annotation (commentaire, question, réponse, conclusion).
- IO_x^{AD} est un ensemble d'informations objectives comportant :
 - son identifiant,
 - sa date de création permettant d'ordonner chronologiquement les annotations,
 - l'identifiant de son créateur (décideur),
 - référence à une annotation père,
 - son point d'ancrage spécifiant la localisation précise de l'annotation.

Le point d'ancrage peut être de deux natures :

- un point d'ancrage *global* localisé sur un concept dans une constellation (l'annotation sera présente dans tous les contextes d'analyse intégrant le concept annoté globalement), ou bien,
- un point d'ancrage *local* localisé sur un élément dans un contexte d'analyse (l'annotation n'est présente que dans ce contexte).

Définition 7. Un point d'ancrage α est défini par $(S; L; C)$ où

- $S = [CA_k] [NF_i [f(m) [=val] ?] ?] ?$ désigne un ancrage relatif au fait F_i ,
- $L = \lambda | ND_{i1} [.NH_{j1} [/pk_1 [=pos_1] ?] *] ?$ désigne un ancrage sur la dimension ligne,
- $C = \lambda | ND_{i2} [.NH_{j2} [/pk_2 [=pos_2] ?] *] ?$ désigne un ancrage sur la colonne.

Notons que CA_k désigne un contexte d'analyse, $f(m)$ est une mesure associée à une fonction d'agrégation, val représente une valeur prise par la mesure, pk_1 (respectivement pk_2) désigne un paramètre, pos_1 (respectivement pos_2) représente une valeur prise par le paramètre pk_1 (respectivement pk_2).

La conservation de l'expertise du décideur permet d'intervenir à deux niveaux :

- Au niveau schéma. Les annotations facilitent le processus de prise de décision par une plus grande compréhension de la sémantique des composants et des instances d'une base de données multidimensionnelle.
- Au niveau analyses décisionnelles. La spécification d'analyses décisionnelles au travers d'un contexte d'analyse enrichi d'une réflexion critique s'apparente au concept de lecture active (Adler *et al.*, 1972). La tâche des décideurs est facilitée par la conservation de la réflexion du décideur, mais également par le partage de ces réflexions entre les différents décideurs.

3.3.2. Exploitation des annotations

L'usage des annotations se décline selon des modalités soit personnelle, soit collective. Dans un *usage personnel*, les annotations matérialisent la réflexion et l'analyse de l'utilisateur décideur rendant leur réutilisation possible. Dans un *usage collectif* (Cabanac, *et al.*, 2010) plusieurs décideurs interviennent. Lorsque par exemple l'analyse est complexe, l'avis d'un autre expert est souvent sollicité, ce qui peut donner lieu à des débats argumentés visant à atteindre un consensus pour une prise de décision collégiale. Le support par des annotations de cet échange permet de sauvegarder et réutiliser les expertises.

Exemple. La figure 5 présente la constellation précédente comportant des annotations. La constellation comporte deux annotations globales (A_1, A_2) et deux autres locales (A_3, A_4) :

- A_1 : "Les analyses mensuelles sont suffisantes pour suivre les paniers"
- A_2 : "Ce paramètre permet des analyses par zone de chalandise"
- A_3 : "Pourquoi les montants ont exceptionnellement augmentés ?"
- A_4 : "Un panier exceptionnel (800) explique l'augmentation"

Au cours des analyses menées par les décideurs, les contextes d'analyses sont annotés. Dans notre exemple, un décideur U_1 annote le contexte d'analyse CA_4 (A_3)

tandis qu'un second décideur U_2 annote ce même contexte d'analyse (A_4) en réponse à l'annotation A_3 .

Les ancres de ces annotations sont définies par les expressions suivantes :

- A_1 : (PANIER; λ ; λ).
- A_2 : (PANIER; CLIENTS.HZON/ZONE; λ).
- A_3 : (CA_4. PANIER.SUM(MNT); CLIENTS.HAGE.TRANCHE = 'Senior'; DATES.HTPS.TRIMESTRE = 'T3-09').
- A_4 : même ancre que A_3 .

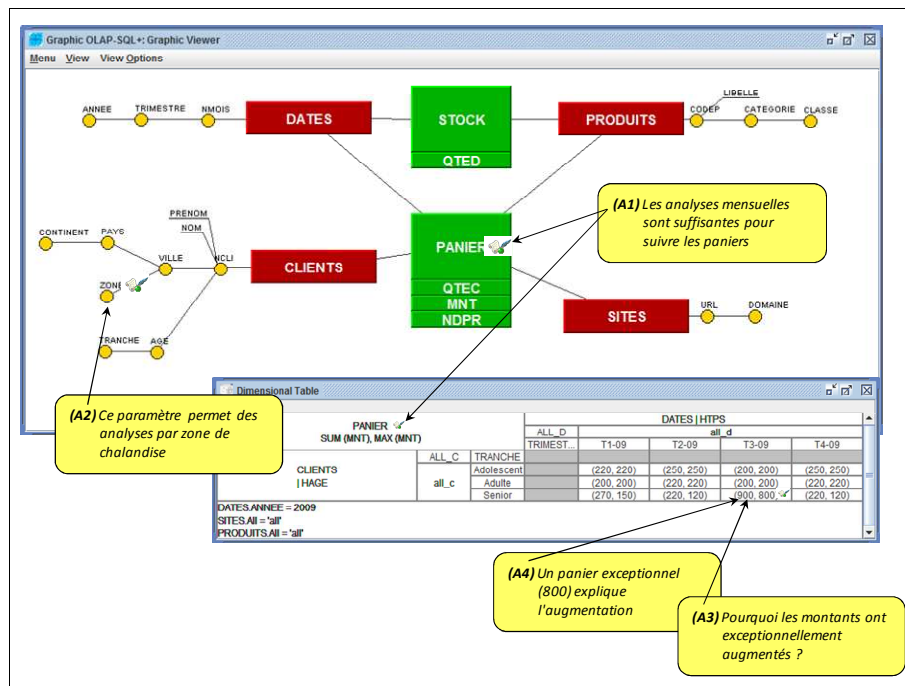


Figure 5. Exemple d'annotations globales et locales

Il est important de noter que l'annotation globale A_1 est rendue visible dans le contexte d'analyse CA_4 puisque son point d'ancrage α_1 appartient au contexte d'analyse. Par contre, A_2 n'est pas accessible au travers de CA_4 .

3.4. Personnalisation de la constellation

En complément des annotations donnant des informations sur les données manipulées, nous souhaitons limiter l'ensemble des données analysées aux données « préférées ». L'objet de cette section est de spécifier ce concept.

L'utilisateur peut exprimer des préférences, notées P_i , sur les éléments de structure d'une constellation et/ou sur les valeurs. Ces préférences peuvent être absolues ou contextuelles : une préférence absolue est toujours prise en compte par le système, tandis qu'une préférence contextuelle est prise en compte par le système lorsque le contexte d'analyse courant couvre le contexte de la préférence (Jerbi *et al.*, 2008).

Définition 8. Une préférence P_i est définie par $(\succ_{P_i}; C^{P_i})$ où

- \succ_{P_i} est une relation d'ordre sur un ensemble E d'éléments.
 - Si les éléments de E sont des éléments de structure de la constellation, $E \in \{F; D; H; M; A\}$, alors P_i est une préférence de structures,
 - Si les éléments de E sont des prédicats sur les propriétés de la constellation, alors P_i est une préférence de valeurs.
- $C^{P_i} = (C^F; C^D; C^R)$ est le contexte de la préférence. Le contexte de préférence est défini comme un contexte d'analyse, pouvant comporter des parties vides.

Exemple. Considérons l'ensemble de préférence $Preference^{CP} = \{P_1; P_2\}$.

- P_1 : "Je préfère visualiser les zones et les villes des clients lorsque j'analyse les paniers en fonction des clients"
- P_2 : "Je préfère l'année courante lorsque j'analyse les stocks de produits"
- P_3 : "Je préfère les deux dernières années lorsque j'analyse les montants (sommes) de paniers en fonction des dates"
- P_4 : "Je souhaite disposer des minima lorsque j'analyse les paniers d'un montant maximum"

Ces préférences sont définies de la manière suivante :

- $(P_1) : (ZONE \succ_{P_1} VILLE; (PANIER; \{ CLIENTS \}; \emptyset))$
- $(P_2) : (DATES.ANNEE = 2010; (STOCK; \emptyset; \emptyset))$
- $(P_3) : (DATES.ANNEE = 2010 \vee DATES.ANNEE = 2009; (PANIER.SUM(MNT); \{ DATES \}; \emptyset))$
- $(P_4) : (PANIER.MIN(MNT); (PANIER.SUM(MNT); \emptyset; \emptyset))$

4. Recommandations enrichies d'annotations

Notre objectif est d'assister le décideur dans sa navigation au sein de l'espace multidimensionnel. Ainsi à partir d'un contexte d'analyse courant CA_i , le décideur transforme CA_i en CA_{i+1} par l'application d'une opération OLAP Op_i . Comme l'illustre la figure 6, le système de recommandation OLAP cherche à déterminer un ensemble de recommandations \mathcal{R}_i :

- par anticipation en déterminant un futur contexte d'analyse CA_{i+j} que l'utilisateur est susceptible de construire, $\mathcal{R}_i = \{CA_{i+j}\}$,
- par alternatives en suggérant les contextes d'analyse CA_k alternatifs à la navigation de l'utilisateur.

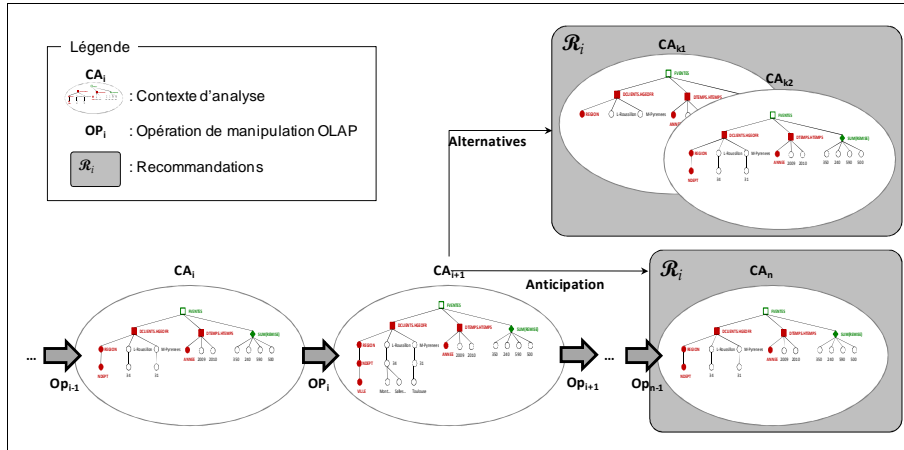


Figure 6. Principe des recommandations de contextes d'analyse

4.1. Préférences candidates et taux de couverture

Durant le processus de recommandation, le système construit l'ensemble des *préférences candidates* \mathcal{P}_i en sélectionnant parmi les préférences $Preference^{CP}$, celles dont le contexte est couvert par le contexte d'analyse CA_{i+1} . La couverture d'une préférence P_j correspond à la couverture entre l'arbre du contexte d'analyse et celui du contexte de la préférence. Nous ne considérons dans cet article que les cas de couverture totale : un contexte de préférence est couvert par le contexte d'analyse CA_{i+1} si et seulement si tous les arcs (v_{k1}, v_{k2}) appartiennent à CA_{i+1} .

Le *taux de couverture* χ^{P_j} d'une préférence candidate P_j correspond au nombre d'arcs en commun entre le contexte d'analyse et celui de la préférence :

$$\chi^{P_j} = N^{CP_j} / N^{CA_{i+1}}$$

avec N^{CP_j} : nombre d'arcs de P_j présents dans le contexte d'analyse CA_{i+1}

$N^{CA_{i+1}}$: nombre d'arcs du contexte d'analyse CA_{i+1}

Exemple. On suppose qu'une opération de rotation (OP_5) est déclenchée sur le contexte d'analyse CA_4 par le décideur afin de visualiser l'analyse des paniers par zone de chalandise.

$$ROTATE(CA_4; CLIENTS; CLIENTS; HZON) = CA_5$$

A partir des préférences $Preference^{CP} = \{P_1; P_2; P_3; P_4\}$ disponibles, le système détermine des préférences candidates $\mathcal{P}_i = \{P_1; P_3; P_4\}$. Comme l'illustre la figure 7, la préférence P_1 est candidate puisque son contexte de préférence ($PANIER; \{CLIENTS\}; \emptyset$) couvre totalement le contexte d'analyse CA_5 . Par contre, la préférence P_2 n'est pas candidate puisque son contexte de préférence ne couvre pas CA_5 .

4.2. Recommandation anticipée

L'anticipation vise à déterminer un contexte d'analyse futur que l'utilisateur devrait construire durant son processus d'analyse. Pour ce faire, le contexte d'analyse CA_{i+1} qui est calculé sert à déterminer un contexte d'analyse à recommander sur la base de préférences candidates \mathcal{P}_i . Parmi l'ensemble des préférences candidates \mathcal{P}_i , un sous-ensemble des préférences les plus couvrantes $\mathcal{P}_i^{\text{Max}}$ est déterminé. Chaque préférence $P_j \in \mathcal{P}_i^{\text{Max}}$ est intégrée dans CA_{i+1} formant ainsi le contexte d'analyse anticipé ; le symbole \oplus représente l'intégration d'une préférence dans un contexte d'analyse. Nous présentons ci-dessous l'algorithme permettant de calculer les recommandations basées sur l'anticipation.

Algorithme de recommandation par anticipation.

Entrées : $CA_i - Op_i$

Sortie : \mathcal{R}_i

Début

$CA_{i+1} \leftarrow \text{ConstruireContexte}(CA_i; Op_i);$

$\mathcal{P}_i \leftarrow \text{PreferencessCandidates}(CA_{i+1}; \text{Preference}^{\text{CP}});$

$\mathcal{P}_i^{\text{Max}} \leftarrow \text{PreferencessCouvrrantes}(\mathcal{P}_i);$

$CA_{i+j} \leftarrow CA_{i+1};$

Pour Chaque $P_j \in \mathcal{P}_i^{\text{Max}}$ **Faire**

$CA_{i+j} \leftarrow CA_{i+j} \oplus P_j;$

FinPour ;

$\mathcal{R}_i \leftarrow \{CA_{i+j}\};$

Fin.

Exemple. Soit $\mathcal{P}_i = \{P_1; P_3; P_4\}$. Le calcul des préférences les plus couvrantes $\mathcal{P}_i^{\text{Max}}$ repose sur les taux de couverture $\chi^{P_1} = 5 / 30$, $\chi^{P_3} = 6 / 30$ et $\chi^{P_4} = 5 / 30$. L'ensemble des préférences les plus couvrantes $\mathcal{P}_i^{\text{Max}}$ est obtenu en utilisant la couverture maximale χ^{Max} . Dans notre exemple, $\chi^{\text{Max}} = 6 / 30$, soit $\mathcal{P}_i^{\text{Max}} = \{P_3\}$.

On peut remarquer que le choix du calcul de la couverture basée sur le nombre d'arcs donne la priorité aux préférences les plus détaillées, c'est à dire dont le contexte d'analyse comporte le plus d'arcs sans distinction. Afin d'affiner le processus de recommandation, il est tout à fait possible d'introduire des heuristiques dans le calcul des préférences visant par exemple à définir une politique de priorité sur les structures de la constellation afin de pondérer les arcs (le fait analysé serait prioritaire sur les dimensions, elles-mêmes prioritaires sur les hiérarchies...).

4.3. Recommandations alternatives

La recommandation par alternatives consiste à proposer des contextes d'analyse considérés comme nouveaux dans l'analyse de l'utilisateur. Pour ce faire, le contexte d'analyse CA_{i+1} qui est calculé sert à déterminer un ensemble \mathcal{R}_i de contextes d'analyse à recommander sur la base de préférences candidates \mathcal{P}_i . A partir des préférences candidates $P_j \in \mathcal{P}_i$, le système utilise les préférences les plus couvrantes (\mathcal{P}^{Max}_i) pour générer les recommandations anticipatives. Les autres préférences ($\mathcal{P}_i \setminus \mathcal{P}^{Max}_i$, \mathbb{F}) sont intégrées dans CA_{i+1} formant un ensemble de recommandations alternatives \mathcal{R}_i .

Algorithme de recommandation d'alternatives.

Entrées : $CA_i - OP_i$

Sortie : \mathcal{R}_i

Début

$CA_{i+1} \leftarrow \text{ConstruireContexte}(CA_i; OP_i);$

$\mathcal{P}_i \leftarrow \text{PreferncesCandidates}(CA_{i+1}; \text{Preference}^{CP});$

$\mathcal{P}^{Max}_i \leftarrow \text{PreferncesCouvranes}(\mathcal{P}_i);$

$\mathcal{R}_i \leftarrow \emptyset;$

Pour Chaque $P_j \in \mathcal{P}_i \setminus \mathcal{P}^{Max}_i$ **Faire**

$CA_k \leftarrow CA_{i+1} \oplus P_j;$

$\mathcal{R}_i \leftarrow \mathcal{R}_i \cup \{CA_k\};$

FinPour;

Fin.

Exemple. Soient $\mathcal{P}_i = \{P_1; P_3; P_4\}$ et $\mathcal{P}^{Max}_i = \{P_3\}$. Les recommandations alternatives sont calculées à partir des préférences candidates $\mathcal{P}_i \setminus \mathcal{P}^{Max}_i = \{P_1; P_4\}$.

La figure 8 illustre les recommandations calculées durant l'opération de rotation (OP_5).

- Le contexte d'analyse anticipé CA_{5+j} est obtenu à partir des préférences de couverture maximale $\mathcal{P}^{Max}_i = \{P_3\}$.
- Les contextes d'analyse alternatifs CA_{k1} et CA_{k2} sont obtenus à partir des autres préférences candidates $\mathcal{P}_i \setminus \mathcal{P}^{Max}_i = \{P_1; P_4\}$.

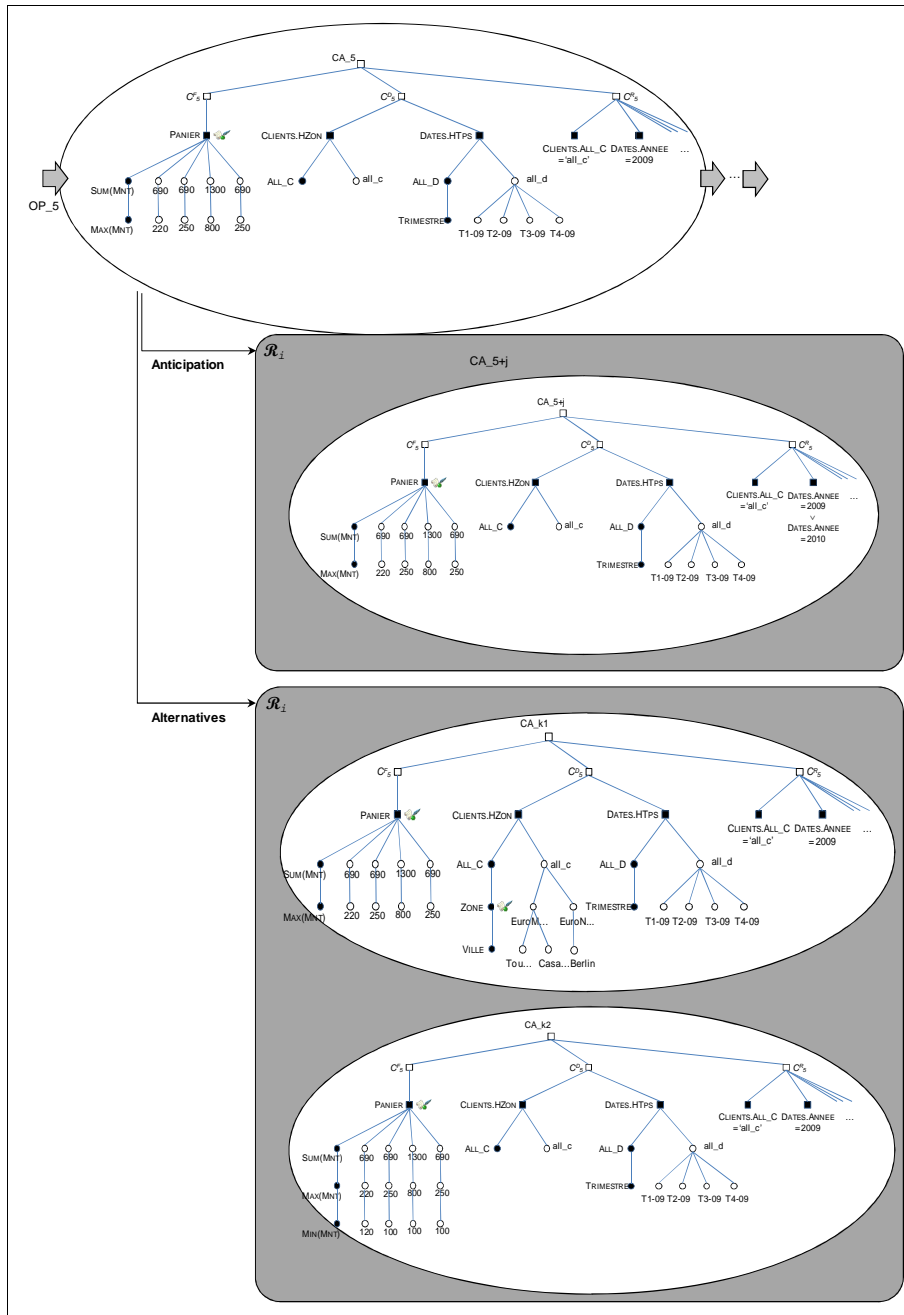


Figure 8. *Recommandations anticipées et alternatives*

4.4. Recommandations annotées

Le système OLAP personnalisé génère en plus du résultat des requêtes utilisateur, un ensemble de recommandations anticipées et alternatives. Les contextes d'analyses sont affichés sous forme de tables multidimensionnelles étendues par une zone d'affichage des recommandations. Comme l'illustre la figure 9, la table multidimensionnelle équivalente au nouveau contexte d'analyse est enrichie par une recommandation anticipée et deux recommandations alternatives. Le système affiche une description textuelle des contextes d'analyse recommandés que l'utilisateur peut sélectionner.

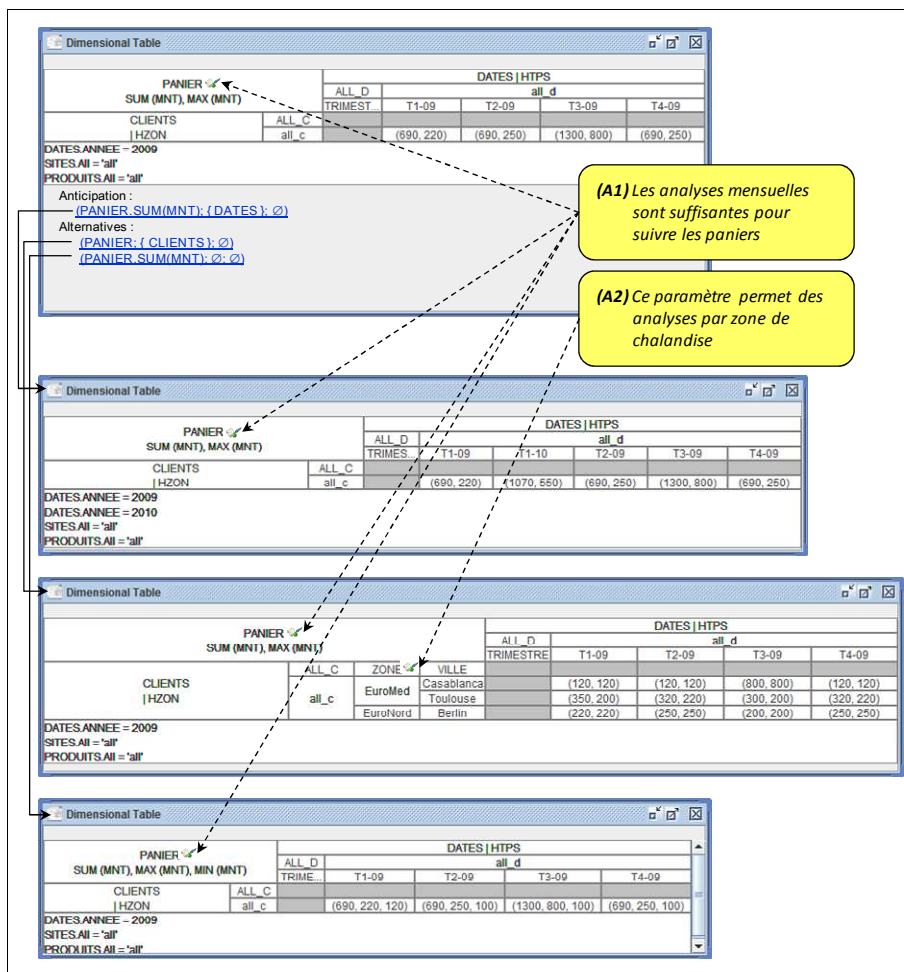


Figure 9. Exemple de tables multidimensionnelles restituées

Le résultat de la requête initiale est également enrichi par des annotations pour aider l'utilisateur à mieux interpréter et évaluer les données restituées. Les annotations ancrées dans chaque contexte d'analyse sont disponibles. Les annotations A_1 et A_2 sont restituées en fonction des contextes d'analyse recommandés. Les données recommandées sont donc simplement complétées par les annotations ancrées à ces dernières.

5. Prototype *Personal-GOLAP*

Afin de valider nos propositions, nous développons au sein du laboratoire le prototype *Personal-GOLAP* (les copies écrans des figures de l'article en sont issues). Les développements sont réalisés en Java au dessus du SGBD Oracle. Ce prototype repose sur une architecture modulaire présentée dans la figure 10.

- Le niveau « interfaces » offre l'éditeur de commandes pour saisir sous la forme de règles les préférences d'un utilisateur. Ces règles peuvent être définies soit de manière quantitative par des poids affectés aux propriétés de la constellation, soit de manière qualitative en ordonnant les propriétés en fonction de la priorité accordée par l'utilisateur à ces dernières. L'éditeur de commande permet également de saisir des requêtes d'interrogation via un ordre OLAP-SQL (Ravat, *et al.*, 2002). Parallèlement, l'éditeur d'annotations offre la possibilité de créer, éditer, supprimer et modifier des annotations ancrées dans la constellation. Enfin, les tables multidimensionnelles produites en réponses aux requêtes d'interrogation comportent les annotations ancrées aux propriétés de la table. Le système complète la table multidimensionnelle en recommandant des tables personnalisées.
- Le niveau « traitements » regroupe cinq composants. Chaque commande de définition de règles (LDR) ou d'interrogation (LID) est analysée par l'analyseur lexical et syntaxique. La commande valide est prise en charge par le moteur OLAP/SQL chargé des transcriptions vers le SGBD relationnel de stockage : les règles validées de personnalisation sont stockées dans la base de données relationnelles. Le module de mémoire d'expertise décisionnelle, appelé « Moteur MED » gère les différentes annotations exprimées par l'utilisateur. Les ordres d'interrogations sont calculés et le résultat est mis en forme par le générateur graphique dans une table multidimensionnelle intégrant les annotations associées au contexte. Le générateur graphique reçoit des recommandations sous forme de tables multidimensionnelles personnalisées venant s'ajouter à la table répondant à l'interrogation initiale.
- Le niveau « STOCKAGE » regroupe deux bases de données. La métabase regroupe la description des structures multidimensionnelles de la constellation ainsi que les préférences contextuelles exprimées par

l'utilisateur. La base R-OLAP contient les données décisionnelles de la constellation enrichies d'annotations de l'utilisateur.

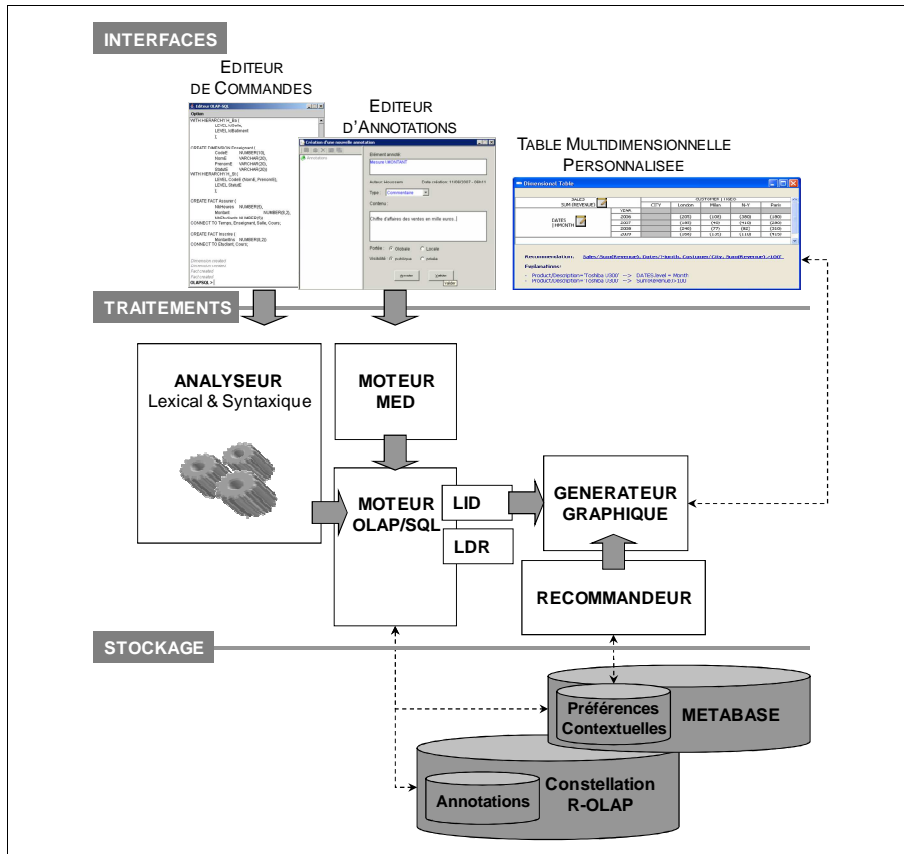


Figure 10. Architecture de Personal-GOLAP

6. Conclusion

L'objectif de nos travaux sur la personnalisation dans les systèmes OLAP est double. Il s'agit dans un premier temps de rendre disponible et accessible toute information ayant permis d'aboutir à une décision (données brutes et expertise des décideurs). Dans un deuxième temps, nous souhaitons mieux prendre en compte les préférences de l'utilisateur en termes de données. Ainsi, nous limitons l'ensemble des données multidimensionnelles aux données « préférées » par les décideurs.

Une première proposition repose sur le concept de mémoire d'expertise afin de conserver le patrimoine immatériel des décideurs au sein du système OLAP. En

effet, l'information utile lors du processus d'analyse décisionnelle ne se trouve pas uniquement dans les bases de données multidimensionnelles, mais une partie importante est habituellement immatérielle : il s'agit de « l'expertise » du décideur. Nous avons proposé de modéliser sous la forme d'annotations ancrées dans la base de données multidimensionnelles toutes ces informations immatérielles relevant de l'expertise de l'utilisateur décideur (commentaires, discussions, prises de décision...).

La seconde proposition consiste en la définition d'un modèle de préférence pour mieux représenter les besoins de l'utilisateur en matière de données analysées. Nos travaux reposent sur une approche qualitative représentant les préférences de l'utilisateur par une relation d'ordre exprimée sur les données. Ces préférences sont alors simplement définies les unes par rapport aux autres. Nous avons représenté son contexte d'analyse pour déterminer durant l'analyse les préférences relevant de l'analyse en cours. Cette « contextualisation » des préférences permet lors des manipulations OLAP des recommandations contextuelles annotées qui assistent l'utilisateur dans son exploration de l'espace multidimensionnel. L'assistance que nous proposons consiste à recommander à l'utilisateur une requête anticipée pour assister l'utilisateur durant son analyse, et des requêtes alternatives pour obtenir des perspectives d'analyses complémentaires à son analyse.

Les travaux présentés laissent de nombreuses perspectives à explorer. Le mécanisme d'annotations proposé pourrait être étendu pour mieux prendre en charge leurs évolutions (annotations devenant obsolètes, archivage d'annotations...). Afin de rendre l'approche interopérable, l'utilisation d'un standard tel que CWMI pour décrire les BDM pourrait être étendu aux annotations. Les mécanismes de recommandation présentés sont basés à la fois sur une structure de visualisation des données unique et sur le seul contexte d'analyse courant, sans tenir compte du chemin de navigation dans le graphe d'analyse. Il est donc envisageable de proposer différentes structures permettant de visualiser les données sous différentes facettes en fonction de préférences contextuelles. Enfin, les annotations sont dans notre approche simplement restituées en fonction des recommandations qui reposent seulement sur les données analysées. Pour aller vers un couplage plus important entre les annotations et la recommandation il est envisageable d'exploiter les annotations dans le processus de recommandation en analysant par exemple les contenus textuels par des techniques de fouille de texte et de recherche d'information. Une autre perspective concerne les extensions à apporter aux mécanismes de recommandation. Actuellement, notre mécanisme de recommandation est identique pour chaque type de recommandation : anticipation et alternatives. Une extension consiste à développer des mécanismes dédiés en fonction du type de recommandation. Par exemple, il serait intéressant d'étendre les mécanismes de recommandation par anticipation en exploitant l'historique des contextes d'analyse ainsi que les opérations permettant d'obtenir ces contextes d'analyse, tandis qu'il serait pertinent de développer des recommandations alternatives en utilisant des contextes d'analyses d'autres utilisateurs, formant ainsi un système OLAP

collaboratif (Cabanac, *et al.*, 2008) permettant à chaque usager de profiter de l'expertise d'un groupe.

7. Bibliographie

- Adler M.J., Van Doren C., *How to Read a Book*, Simon & Shuster, 1972.
- Bellatreche L., Giacometti A., Marcel P., Mouloudi H., Laurent D., A Personalization Framework for OLAP Queries, *8th International Workshop on Data Warehousing and OLAP, DOLAP'05*, pp. 9–18, 2005.
- Benammar A., Hubert G., Mothe G., Automatic Profile Reformulation Using a Local Document Analysis. *European Conference on IR Research (ECIR'02)*, p.124-134, 2002.
- Bentayeb F., Favre C., RoK: Roll-Up with the K-Means Clustering Method for Recommending OLAP Queries, *Database and Expert Systems Applications, DEXA'09*, pp.501-515, Linz, Austria, 2009.
- Bouzeghoub M., Kostadinov D., Personnalisation de l'information : aperçu de l'état de l'art et définition d'un modèle flexible de profils, *2ème Conférence en Recherche d'Informations et Applications, CORIA'05*, pp. 201–218, 2005.
- Cabanac G., Chevalier M., Chrisment C., Julien C. Social validation of collective annotations: Definition and experiment., *Journal of American Society for Information Science and Technology (JASIST)*, Wiley, Vol. 61 N. 2, pp. 271-287, 2010.
- Cabanac G., Chevalier M., Chrisment C., Julien C., Soule-Dupuy C., Tchienehom P. Web Information Retrieval: Towards Social Information Search Assistants. *Social Information Technology: Connecting Society and Cultural Issues. Terry Kidd, Irene Chen (Eds.)*, IGI Global, 16, pp. 218-252, March 2008.
- Cabanac G., Chevalier M., Ravat F., Teste O. An Annotation Management System for Multidimensional Databases. *International Conference on Data Warehousing and Knowledge Discovery (DaWaK 2007)*, Regensburg (Germany), Springer-Verlag, LNCS 4654, doi: 10.1007/978-3-540-74553-2_9, p. 89-98, septembre 2007.
- Chevalier M., Julien C., Soule-Dupuy C., Vallés-Parlangeau N. Personalized Information Access Through Flexible and Interoperable Profiles. *International Web Information Systems Engineering - International Workshop on Personalized Access to Web Information*, Vol. LNCS 4832 - ISBN 978-3-540-770, p. 374-385, décembre 2007.
- Eltabakh, M. Y. , Aref, W. G. , Ahmed K. Ouzzani, E.M., Silva, Y. N. : Supporting Annotations on Relations, *12th International Conference on Extending Database Technology, Saint Petersburg, Russia*, March 24-26, 2009.
- Favre C., Bentayeb F., Boussaid O., Evolution et personnalisation des analyses dans les entrepôts de données : une approche orientée utilisateur, *INFORSID'07*, pp. 308–323, 2007.
- Garrigós I., Pardillo J., Mazón J-N., Trujillo J., A Conceptual Modeling Approach for OLAP Personalization, *28th International Conference on Conceptual Modeling, ER'09*, pp. 401-414, Gramado, Brazil, November 9-12, 2009.

- Giacometti A., Marcel P., Negre E., A Framework for Recommending OLAP Queries, *11th International Workshop on Data Warehousing and OLAP, DOLAP'08*, pp. 73–80, 2008.
- Giacometti A., Marcel P., Negre E., Recommending Multidimensional Queries, *11th International Conference on Data Warehousing and Knowledge Discovery DAWAK'09*, pp.453-466, 2009.
- Golfarelli M., Maio D., Rizzi S., The Dimensional Fact Model: A Conceptual Model for Data Warehouses, *International Journal of Cooperative Information Systems*, 7(2-3), pp.215-247, 1998.
- Golfarelli M., Rizzi S., Expressing OLAP Preferences, *21st International Conference Scientific and Statistical Database Management, SSDBM'09*, Springer Verlag LNCS 5566, pp.83-91, New Orleans, LA, USA, 2009.
- Inmon W.H., *Building the Data Warehouse*, John Wiley & Sons, ISBN 0471-14161-5, 1994.
- Ioannidis, Y., Koutrika G., Personalized Systems: Models and Methods from an IR and DB Perspective, *31st International Conference on Very Large Data Bases, VLDB'05*, pp. 1365–1365, Trondheim, Norway, 2005.
- Jerbi H., Ravat F., Teste O., Zurfluh G., Management of context-aware preferences in multidimensional databases. *3rd International Conference on Digital Information Management (ICDIM'08)*, *IEEE*, doi: 10.1109/ICDIM.2008.4746714, p.669-675, Londres (UK), novembre 2008.
- Kimball R., *The data warehouse toolkit: practical techniques for building dimensional data warehouses*, John Wiley & Sons, ISBN 0-471-15337-0, 1996.
- Korfhage, R. R., *Information Storage and Retrieval*, JohnWiley & Sons, 1997.
- Pérez, J.M., Berlanga, R., Aramburu, M.J.: A relevance model for a data warehouse contextualized with documents *Information Processing and Management* 45, 2009.
- Ravat F., Teste O., Zurfluh G., Langages pour Bases Multidimensionnelles : OLAP-SQL. *Revue des Sciences et Technologies de l'Information, ISI (Ingénierie des Systèmes d'Information)*, *Hermès*, Vol.7, N°3, doi: 10.3166/isi.7.3.11-38, p.11-38, novembre 2002.
- Ravat F., Teste O., Tournier R., Zurfluh G., Querying Multidimensional Databases, *11th East-European Conference on Advances in Databases and Information Systems (ADBIS'07)*, Springer-Verlag, LNCS 4690, Y.E. Ioannidis, B. Novikov, B. Rachev, doi: 10.1007/978-3-540-75185-4_22, p.298-313, Varna (Bulgarie), septembre 2007.
- Ravat F., Teste O., Tournier R., Zurfluh G., Algebraic and graphic languages for OLAP manipulations, *International Journal of Data Warehousing and Mining*, IGI Group, Vol.4, N°1, doi: 10.4018/jdwm.2008010102, p.17-46, janvier 2008.
- Rizzi S., OLAP Preferences: a Research Agenda, *10th International Workshop on Data Warehousing and OLAP, DOLAP'07*, ACM, pp.99-100, Lisbon (Portugal), Nov. 2007.
- Sapia, C., PROMISE: Predicting Query Behavior to Enable Predictive Caching Strategies for OLAP Systems, *2nd International Conference on Data Warehousing and Knowledge Discovery, DaWaK'00*, LNCS 1874, pp. 224–233, London, UK, 2000.

- Thalhammer T., Schrefl M., Mohania M., Active DataWarehouses: Complement-ing OLAP with Analysis Rules, *Data and Knowledge Engineering* 39(3), 241–269, 2001.
- Teste O., Elaboration d'entrepôts de données complexes. INFORSID'00, Lyon (France), 16/05/2000-19/05/2000, INFORSID, p. 229-245, mai 2000.
- Xin, D., Han, J., Cheng, H., Li, X., Answering top-k queries with multi-dimensional selections: The ranking cube approach. *VLDB*, pp. 463–475, 2006.