

Modèle conceptuel pour bases de données multidimensionnelles annotées

Guillaume Cabanac*, Max Chevalier* **, Franck Ravat*, Olivier Teste*

*IRIT (Institut de Recherche en Informatique de Toulouse) – UMR 5505
Université Paul Sabatier – 118 route de Narbonne – 31062 Toulouse cedex 9
{Guillaume.Cabanac, Max.Chevalier, Franck.Ravat, Olivier.Teste}@irit.fr

**LGC (Laboratoire de Gestion et Cognition) – EA 2043

IUT « A » Paul Sabatier – 129 avenue de Rangueil – BP 67701 – 31077 Toulouse cedex 4

Résumé. Nos travaux visent à proposer une mémoire d'expertises décisionnelles permettant de conserver et de manipuler non seulement les données décisionnelles mais aussi l'expertise analytique des décideurs. Les données décisionnelles sont représentées au travers de concepts multidimensionnels et l'expertise associée est matérialisée grâce au concept d'annotation.

1 Contexte et problématique

Les systèmes d'aide à la décision visent à transformer les données opérationnelles en informations facilement interprétables par les décideurs afin que ces derniers puissent effectuer des analyses complexes et prendre les meilleures décisions en temps utiles pour assurer la compétitivité et la pérennité de l'organisation considérée. Dans un tel contexte, plus que le patrimoine matériel, le patrimoine immatériel est important pour capitaliser un maximum d'informations, de connaissances et d'expertises afin de prendre les décisions adaptées. Nos travaux visent à proposer aux organisations plus qu'un système d'aide à la décision, un véritable outil de Mémoire d'Expertises Décisionnelles (MED).

1.1 Les données décisionnelles

Il est reconnu que les Bases de Données Multidimensionnelles (BDM) sont adaptées pour le stockage et la manipulation des données décisionnelles (Inmon, 1996). Les modèles conceptuels (Ravat et al., 2005) des BDM organisent les données en sujets et axes d'analyses au sein d'un **schéma en étoile** (Kimball, 1996). Tout sujet d'analyse est représenté par un **fait** composé de plusieurs **mesures** (indicateurs d'analyse). Les **dimensions** représentant les axes d'analyse sont formées de **paramètres** en fonction desquels les mesures sont étudiées. Les paramètres sont organisés en **hiérarchies**, de la granularité la plus fine (attribut racine servant d'identifiant à la dimension) à la plus générale (cet attribut est symbolisé par *All*).

1.2 Restitution de données décisionnelles : table multidimensionnelle

Les données d'une BDM sont couramment visualisées au travers d'une Table Multidimensionnelle (TM) : tableau affichant les valeurs d'un fait en fonction de deux dimensions

développées et d'un prédicat de sélection (Agrawal et al., 1997 ; Gyssens et al., 1997). La figure 1 représente un schéma de BDM composé d'un fait ventes et de trois dimensions (produit, magasin, temps) ainsi qu'une TM permettant d'analyser le montant des ventes (mesure **Montant** du fait **Ventes**) annuelles (paramètre **Année** de la dimension **Temps**) en fonction de la ville et du code du magasin (paramètres **Ville** et **CodeMag** de la dimension **Magasin**).

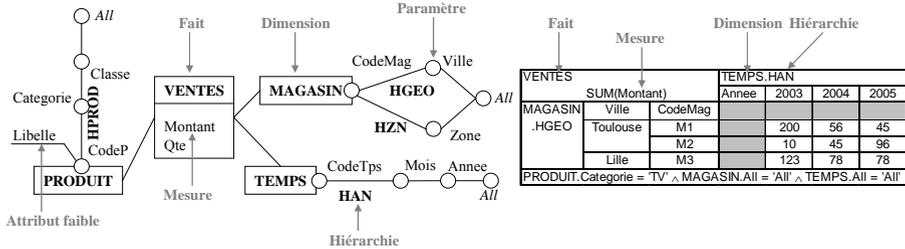


FIG. 1 – Schéma en étoile d'une BDM accompagné d'un exemple de TM.

1.3 Vers une mémoire d'expertises décisionnelles

Les TM ont pour objectif de faciliter les prises de décisions, mais elles s'avèrent parfois complexes et difficiles à interpréter. En effet, les prises de décisions reposent non seulement sur les données brutes mais également sur les réflexions, les commentaires des analystes voire la confrontation de différentes interprétations. À notre connaissance, il n'existe pas d'outil logiciel permettant aux décideurs d'analyser les données décisionnelles en intégrant les tâches qu'ils conduisent de manière manuelle sur des tableaux de bord : annoter, comparer... La mémorisation et la réutilisation de l'expertise des analystes permettent à l'organisation de préserver un patrimoine tout aussi important que les données elles-mêmes.

Notre objectif est de proposer un cadre informatique permettant d'exploiter, de partager les données multidimensionnelles tout en supportant des fonctionnalités d'annotation pouvant comprendre des fils de discussion (support de communication utilisé dans les forums). Ainsi, en permettant d'enrichir interactivement les composants des BDM et des TM, les décideurs deviendront des lecteurs actifs en insérant notamment des commentaires. Pour cela, nous proposons un outil informatique permettant :

- de visualiser les données décisionnelles sous la forme d'une TM ;
- de manipuler une TM au travers d'opérations multidimensionnelles ;
- d'annoter les schémas de BDM afin d'explicitier les composants décisionnels ;
- d'annoter les TM tout en permettant le dialogue sous forme de fil de discussion ;
- d'exploiter ces annotations au travers de fonctions d'exploration, de sélection, etc.

De nombreux travaux ont apporté une réponse aux deux premiers objectifs : des algèbres pour la définition et la manipulation de BDM ont été proposées (Gray et al., 1996 ; Agrawal et al., 1997 ; Cabibbo et Torlone, 1998 ; Marcel 1999 ; Abelló et al., 2003).

2 Existant : l'activité d'annotation

À notre connaissance, les systèmes d'annotation couplés aux BDM n'ont pas fait l'objet d'étude. Une première proposition consisterait à transposer aux BDM le principe des commentaires associés aux schémas de BD transactionnelles. Cette proposition est insuffisante car elle reste difficilement exploitable et très limitée (table, vue ou attribut). La deuxième

solution consiste à se baser sur les systèmes d'annotation existants pour la gestion électronique de documents. Dans ce cadre, les annotations sont qualifiées de commentaires, notes, explications ou autres types de remarques qui peuvent être associés à tout ou partie d'un document sans avoir à le modifier¹. Ces annotations sont dites informelles, contrairement aux annotations formelles du Web Sémantique reposant sur l'utilisation d'un langage formel (ex. : ontologie) pour cataloguer et indexer les documents. Dans cet article, seules les annotations informelles seront exploitées car nous ne souhaitons pas contraindre les décideurs à employer un vocabulaire normé et restreint.

La majorité des systèmes d'annotation informatisés (SAI) opèrent sur des documents textuels. Dans ce cadre, les annotations sont matérialisées sous différentes formes : texte ou marques libres (astérisques, etc.). Ces marques mettent en valeur des passages en y associant éventuellement un commentaire. À ce jour, on dénombre plus de vingt SAI tels que Amaya du W3C (Koivunen et Swick, 2001) ou Office Web Discussions de Microsoft (Brush, 2002). L'intérêt d'un SAI se situe à deux niveaux : à un niveau personnel, il aide l'utilisateur dans sa tâche de lecture active tandis qu'au niveau collectif il permet de partager ces annotations entre les utilisateurs. Au regard de nos problématiques, nous proposons d'adapter les fondements des SAI à l'analyse décisionnelle. L'intérêt des annotations dans un tel contexte est qu'elles seront utilisées pour véhiculer l'expertise des analystes.

3 Annotations décisionnelles

Les TM servent de support à des experts pour la prise de décisions. Cependant, afin de spécifier une MED, nous souhaitons permettre aux analystes d'enrichir ces TM avec des annotations qui visent à conserver les décisions et les commentaires formulés lors des analyses de la TM. L'expertise que véhiculent ces annotations est utilisée à des fins personnelles ou collectives et peut par conséquent contribuer à améliorer les analyses futures. Pour cela, les annotations apportent les fonctionnalités et avantages suivants :

- au niveau du schéma, elles permettent d'améliorer la compréhension des composants d'un schéma de BDM et ainsi de tirer des conclusions analytiques cohérentes ;
- au niveau des valeurs, elles peuvent contenir des informations expliquant un phénomène général, le contexte d'étude ou signaler la singularité d'une valeur spécifique. Les fils de discussion suscités par ces annotations permettent notamment aux experts de valider ou de compléter les commentaires de leurs collègues ;
- les liens spécifiés vers des annotations ou des documents permettent de construire un « dossier d'analyse ». Ils permettent également d'illustrer la réflexion des analystes pour expliquer et justifier les éventuelles conclusions qu'ils ont tirées ;
- les traitements automatisés tels que la classification permettent de faire émerger des liens implicites entre analyses permettant d'aiguiller les analystes dans leur tâche.

3.1 Modèle conceptuel d'annotation décisionnelle

Dans cette section, nous spécifions le modèle d'annotation décisionnelle associé aux TM. On pose $ETMA = \{tma_1, tma_2, \dots, tma_i\}$ un ensemble de TM annotées (TMA), $EAN = \{an_1, an_2, \dots, an_i\}$ un ensemble d'annotations, $FAnnoté : ETMA \rightarrow 2^{EAN}$ est l'application qui associe chaque tma_i à un ensemble d'annotations. Une TM annotée est définie par :

¹ cf. le site Web du projet Annotea initié par le W3C : <http://www.w3.org/2001/Annotea/>

Définition 1 : $\forall i \in [1;I] \ tma_i = (tm_i, FAnnote(tm_i))$

avec tm_i étant une TM telle que celle décrite dans la section 1.2.

et $FAnnote(tm_i)$ est l'ensemble des annotations associées à la table tm_i .

Une annotation est définie comme suit :

Définition 2 : $\forall g \in [1;I]$, une annotation décisionnelle ad_g est définie par le couple (IS_g, IO_g)

- $IS_g = \{is_1, is_2, \dots, is_k\}$ un ensemble d'informations subjectives,

- $IO_g = \{io_1, io_2, \dots, io_j\}$ un ensemble d'informations objectives.

Dans les sections suivantes, nous présentons les caractéristiques d'une annotation : informations subjectives (contenu textuel, etc.) et objectives (créateur, point d'ancrage, etc.).

3.1.1 Informations subjectives

Une annotation décisionnelle comporte les informations subjectives suivantes :

- **le contenu textuel** saisi par l'annotateur ;
- **le type de l'annotation** caractérisant son contenu textuel. Le type, défini par le créateur de l'annotation, peut être : *commentaire*, *question* (permet d'interroger les autres analystes directement en contexte au travers de la TM), *conclusion* (l'annotation présente les conclusions de l'analyse et les éventuelles décisions prises) et *référence* (références vers des documents ou d'autres TM). Ces types ne sont pas exclusifs : une annotation peut être un commentaire qui comprend des références ;
- **la portée de l'annotation** qui est soit locale soit globale. Une annotation est locale si elle n'est présentée qu'au travers d'une TM précise. Au contraire, une annotation globale est indépendante des TM. Cependant, la cible d'une annotation détermine parfois sa portée. Ainsi, une annotation sur les valeurs des mesures de la TM sera locale car fortement dépendante de l'analyse courante.

3.1.2 Informations objectives

Pour chaque annotation décisionnelle ad_g , le système définit également l'ensemble IO_g contenant les informations objectives suivantes :

- **son identification** : identifiant caractérisant sa position dans le fil de discussion ;
- **son créateur** : son identité (nom, prénom), sa fonction et son adresse email ;
- **son point d'ancrage** spécifiant la localisation précise de l'annotation dans la *tma*.

La définition du point d'ancrage ne peut se faire de manière naïve e.g. repérage en ligne et en colonne car la TM peut regrouper, en ligne et en colonne, une hiérarchisation de paramètres d'une même dimension. De plus, l'analyste peut réorganiser les axes des abscisses et des ordonnées grâce à l'opérateur de permutation de paramètres par exemple (Ravat et al., 2005).

Pour ces deux raisons, le point d'ancrage est défini par un CheminTM :

Définition 3 : Un CheminTM² CTM est défini par le quintuplet (F, M, C_1, C_2, R) où :

- F précise éventuellement le fait : $\lambda \mid \text{NomFait}$,
- M précise éventuellement le nom des mesures et les opérateurs d'agrégation associés :
 $\lambda \mid \langle (\text{op}(m)) \rangle + \rangle$,
- C_1 (resp. C_2) exprime le chemin de la première (resp. deuxième) dimension :
 $\lambda \mid \text{NomDimension} (. \text{NomHierarchie})^* (/ \text{NomParamètre} [= \text{Valeur}])^*$
- R précise éventuellement la condition de restriction : $\lambda \mid \text{ExpressionBooléenne}$,

² Notations exprimées en EBNF (Extended Backus-Naur Form)

avec: λ représente l'absence de valeur, $_{\text{valeur}}$ précise la valeur exploitée pour le paramètre considéré et / représente un opérateur de descendance entre paramètres dans la hiérarchie.

Seuls les éléments annotés figurent dans le chemin, ce qui explique que tous les champs sont facultatifs. Nous présentons ci-dessous quelques exemples de chemins de localisation représentés par la figure 2 (où R représente l'expression de restriction de cette TM).

VENTES		SUM(Montant)		TEMPS.HAN		
		Ville	CodeMag	Année	2003	2004
MAGASIN. HGEO	Toulouse	M1		200	56	45
		M2		10	45	96
	Lille	M3		123	78	78
PRODUIT.Categorie = 'TV' ^ MAGASIN.All = 'All' ^ TEMPS.All = 'All'						

① Concepteur – 29/11/2004
« Années civiles »

② Responsable M2 – 29/12/2004
« Montant expliqué par la fermeture 11 mois sur 12 du magasin »

③ Service des ventes – 28/11/2004
« Télévisions 16/9° »

FIG. 2 – Exemple d'une table multidimensionnelle annotée.

1. Le concepteur du schéma indique que le paramètre « Année » correspond aux années civiles. Le chemin de l'annotation est : $(\lambda, \lambda, \lambda, \text{TEMPS.HAN:Année}, R)$.
2. Le responsable du magasin M2 justifie le faible montant des ventes en 2003 : le magasin n'a été ouvert qu'un mois dans l'année. Le chemin de la cellule est : $(\text{VENTES}, <\text{SUM(Montant)}>, \text{MAGASIN.HGEO/Ville}='Toulouse' / \text{CodeMag}='M2', \text{TEMPS.HAN/Année}=2003, R)$.
3. Le service des ventes avertit les différents analystes : les ventes prises en compte sont celles des télévisions 16/9°. Le chemin de la cellule où l'annotation sera rattachée est : $(\text{VENTES}, \lambda, \lambda, \lambda, R)$.

3.2 Création, stockage et restitution d'une annotation décisionnelle

Pour créer une annotation, le décideur sélectionne directement dans la table la ou les cellules qu'il souhaite annoter. Pour compléter cette annotation il doit fournir toutes les informations qu'il souhaite inclure dans l'annotation. Nous stockons les annotations sur un serveur dédié i.e. indépendamment des TM pour faciliter le partage et la recherche des annotations.

La restitution des annotations est non intrusive et se fait en parallèle de la construction de la TM. Elles sont directement intégrées sous forme de pictogrammes au niveau de la TM.

4 Conclusion et perspectives

Nos travaux visent à proposer des solutions permettant la constitution d'une mémoire d'expertises décisionnelles pérenne. Cette mémoire permet de stocker et de restituer aussi bien les données nécessaires aux prises de décisions que les annotations qui véhiculent les commentaires des analystes. Les données décisionnelles sont représentées au travers de concepts multidimensionnels et restituées sous forme de tables (TM). Ces différents composants servent de support à la formulation, à l'utilisation et à la restitution d'annotations décisionnelles qui peuvent aussi bien servir à un usage personnel qu'à un usage collectif.

Nous proposons d'étendre ces travaux selon deux directions. La première consiste à étendre le modèle conceptuel des TMA en intégrant le concept de groupe d'utilisateurs avec les droits d'accès aux annotations associés. La seconde perspective consiste à coupler le principe

de discussion asynchrone avec un outil de workflow. Cet outil pourrait valider et organiser le circuit des annotations ainsi que de définir les délais pour une prise de décision fiable.

Références

- Abelló, A., J. Samos, et F. Saltor (2003). Implementing Operations to Navigate Semantic Star Schemas. In *DOLAP '03 : Proc. 6th ACM Int. Workshop on Datawarehousing and OLAP*, New York, NY, USA, pp. 56–62. ACM Press.
- Agrawal, R., A. Gupta, et S. Sarawagi (1997). Modeling Multidimensional Databases. In A. Gray et P.-Å. Larson (Eds.), *Proc. 13th Int. Conf. Data Engineering, ICDE*, pp. 232–243. IEEE Computer Society.
- Brush, A. J. B. (2002). Annotating Digital Documents for Asynchronous Collaboration. Technical report 02-09-02, Department of Computer Science and Engineering, University of Washington, USA.
- Cabibbo, L. et R. Torlone (1998). From a Procedural to a Visual Query Language for OLAP. In *SSDBM '98 : Proc. 10th Int. Conf. on Scientific and Statistical Database Management*, Washington, DC, USA, pp. 74–83. IEEE Computer Society.
- 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 S. Y. W. Su (Ed.), *Proc. 12th Int. Conf. on Data Engineering, February 26 - March 1, 1996, New Orleans, Louisiana*, pp. 152–159. IEEE Computer Society.
- Gyssens, M. et L. V. S. Lakshmanan (1997). A Foundation for Multi-Dimensional Databases. In M. Jarke, M. J. Carey, K. R. Dittrich, F. H. Lochovsky, P. Loucopoulos, et M. A. Jeusfeld (Eds.), *VLDB'97, Proc. 23rd Int. Conf. on Very Large Data Bases, August 25-29, 1997, Athens, Greece*, pp. 106–115.
- Inmon, W. H. (1996). *Building the Data Warehouse (2nd ed.)*. New York, NY, USA : John Wiley & Sons, Inc.
- Kimball, R. (1996). *The Data Warehouse Toolkit: Practical Techniques for Building Dimensional Data Warehouses*. New York, NY, USA : John Wiley & Sons, Inc.
- Koivunen, M. et R. Swick (2001). Metadata based annotation infrastructure offers flexibility and extensibility for collaborative applications and beyond. In *Proc. 1st Int. Conf. on Knowledge Capture (KCAP 2001)*, Victoria, B.C., Canada, Oct. 2001.
- Marcel, P. (1999). Modeling and Querying Multidimensional Databases: an Overview. *Networking and Information Systems* 2(5-6), 515–548.
- Ravat F., O. Teste, et G. Zurfluh (2005). Manipulation et fusion de données multidimensionnelles. In *Revue des Nouvelles Technologies de l'Information (RNTI-E-3) - Extraction et Gestion des Connaissances (EGC'2005)*, Volume I, pp. 349–354. Toulouse, France : Cepadues.

Summary

We intend to specify a decisional expert memory. The solution that we describe allows storing and manipulating decisional data as well as acquiring analysis and decisions of managers. Decisional data are defined through multidimensional concepts. The annotation concept models analysts' experience formulated on multidimensional tables.