

La fouille intelligente des règles d'association à partir des données décisionnelles

Hanen BRAHMI¹, Nadia ARAOUR¹, Sadok BEN YAHIA¹



¹ FACULTÉ DES SCIENCES DE TUNIS, TUNISIE.
LIPAH, DÉPARTEMENT DES SCIENCES DE L'INFORMATIQUE.
CAMPUS UNIVERSITAIRE 1060.

EGC

EGC-31 Janvier 2012



Atelier AiDE

Plan

- Introduction et Motivations
- Les ontologies
- Les ontologies dans le processus d'Extraction des Connaissances à partir des Données (ECD)
- Les ontologies et les Entrepôts de Données (ED)
- Conclusion et perspectives

Qui sont mes
meilleurs
clients?

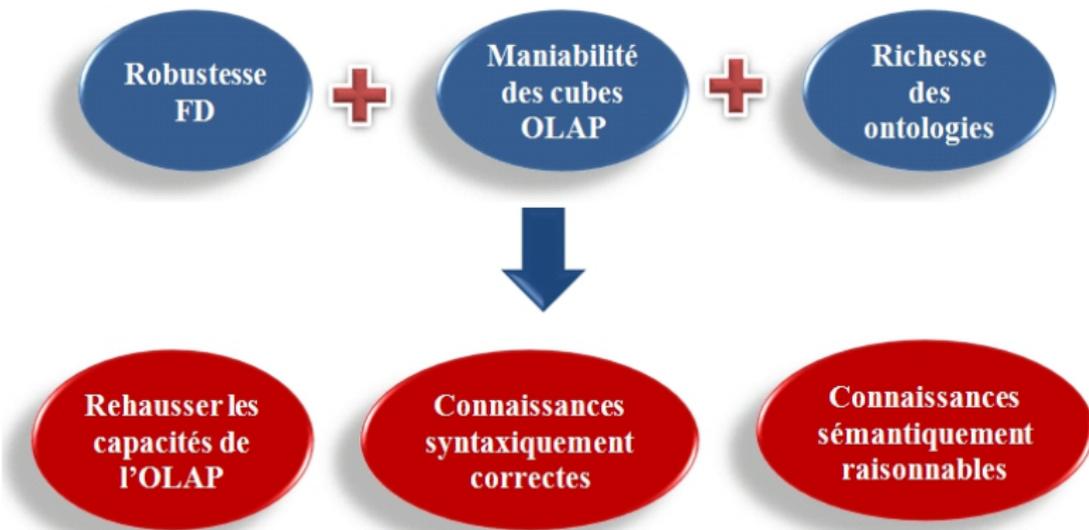
Pourquoi le
chiffre
d'affaire a
baissé?



Qui sont
mes fidèles
clients?

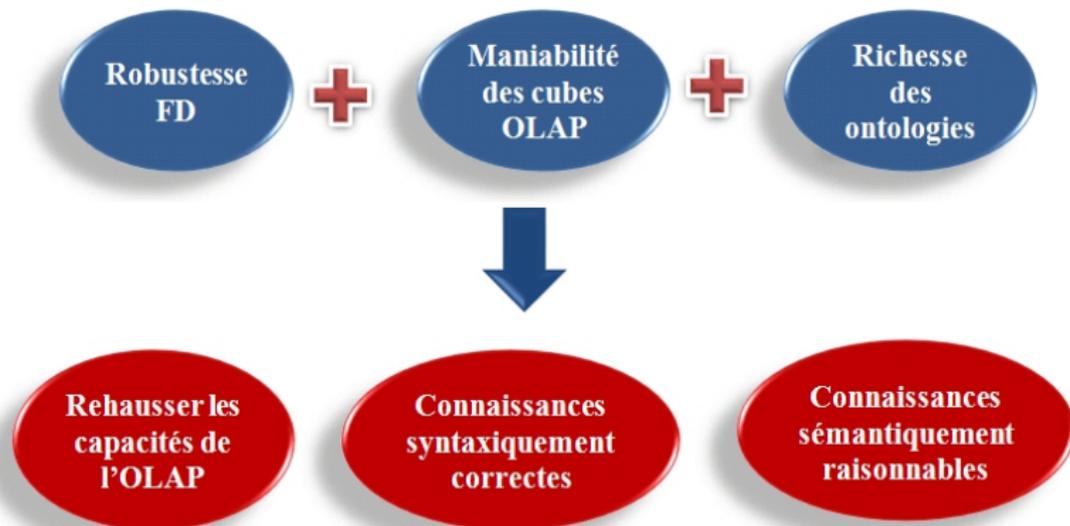
A combien
s'élèvent
mes ventes ?





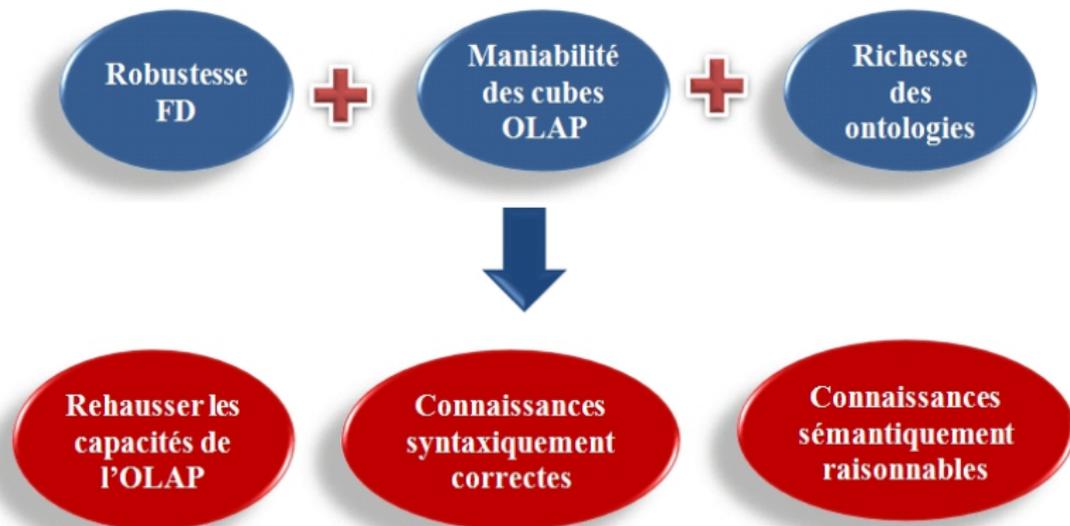


La fouille intelligente



La fouille intelligente

- L'expression, la structuration et la formalisation des connaissances expertes (à priori) et leur confrontation aux connaissances extraites.



La fouille intelligente

- L'expression, la **structuration** et la **formalisation** des **connaissances expertes** (à priori) et leur **confrontation** aux connaissances extraites.
- L'**intégration** des **ontologies** dans l'**ECD** ainsi que dans les **EDs**.

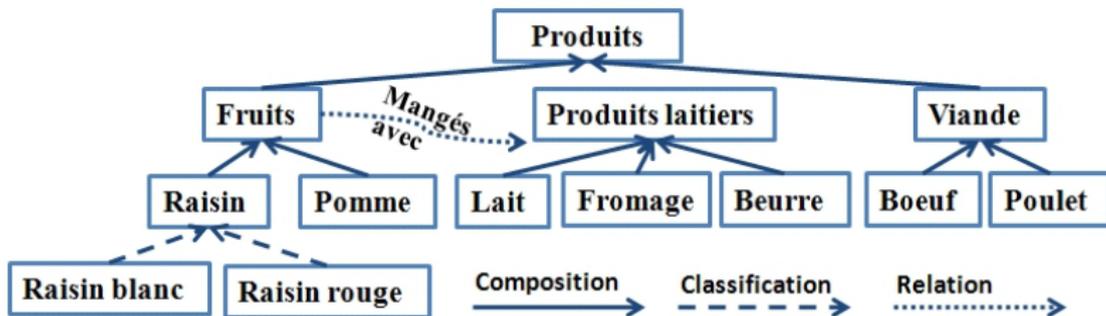
Qu'est-ce qu'une ontologie ?

- Une représentation **formelle**, **référéncable** et **consensuelle** de l'ensemble des concepts partagés d'un domaine (Fankam et al. 2009 [5]).

Qu'est-ce qu'une ontologie ?

- Une représentation **formelle**, **référéncable** et **consensuelle** de l'ensemble des concepts partagés d'un domaine (Fankam et al. 2009 [5]).

Exemple d'ontologie "Produits".



Composants d'une ontologie

- Les **concepts** ou les **classes** : **abstractions** pertinentes du domaine.

Composants d'une ontologie

- Les **concepts** ou les **classes** : **abstractions** pertinentes du domaine.
- Les **relations** existent entre les concepts.

Composants d'une ontologie

- Les **concepts** ou les **classes** : **abstractions** pertinentes du domaine.
- Les **relations** existent entre les concepts.
- Les **fonctions** sont les cas particuliers des relations.

Composants d'une ontologie

- Les **concepts** ou les **classes** : **abstractions** pertinentes du domaine.
- Les **relations** existent entre les concepts.
- Les **fonctions** sont les cas particuliers des relations.
- Les **axiomes** : **signification** des composants, **restrictions** sur les attributs, etc.

Composants d'une ontologie

- Les **concepts** ou les **classes** : **abstractions** pertinentes du domaine.
- Les **relations** existent entre les concepts.
- Les **fonctions** sont les cas particuliers des relations.
- Les **axiomes** : **signification** des composants, **restrictions** sur les attributs, etc.
- Les **instances** : connaissances **statiques**, **factuelles**.

Composants d'une ontologie

- Les **concepts** ou les **classes** : **abstractions** pertinentes du domaine.
- Les **relations** existent entre les concepts.
- Les **fonctions** sont les cas particuliers des relations.
- Les **axiomes** : **signification** des composants, **restrictions** sur les attributs, etc.
- Les **instances** : connaissances **statiques**, **factuelles**.

Typologie des ontologies

Composants d'une ontologie

- Les **concepts** ou les **classes** : **abstractions** pertinentes du domaine.
- Les **relations** existent entre les concepts.
- Les **fonctions** sont les cas particuliers des relations.
- Les **axiomes** : **signification** des composants, **restrictions** sur les attributs, etc.
- Les **instances** : connaissances **statiques**, **factuelles**.

Typologie des ontologies

- Les ontologies **de haut niveaux** décrivent des concepts **très généraux** ou des connaissances **de sens commun**.

Composants d'une ontologie

- Les **concepts** ou les **classes** : **abstractions** pertinentes du domaine.
- Les **relations** existent entre les concepts.
- Les **fonctions** sont les cas particuliers des relations.
- Les **axiomes** : **signification** des composants, **restrictions** sur les attributs, etc.
- Les **instances** : connaissances **statiques**, **factuelles**.

Typologie des ontologies

- Les ontologies **de haut niveaux** décrivent des concepts **très généraux** ou des connaissances **de sens commun**.
- Les ontologies **spécifiques spécialisent** un sous-ensemble d'ontologies génériques (**domaine**, d'**application**, de **méthodes** ou de **tâches partagées**).

Extraction des règles d'association

Phase d'ECD : Pré-traitement

Approches	Type d'ontologie			Type de règles			Mesures		
	Domaine	Application	Floue	Généralisées	sous contrainte	Floue	Support	Confiance	Autres
Zeman et al. (2009) [17]	x			x			x	x	
Hou et al. (2005) [7]	x			x			x	x	
Bellandi et al. (2008) [2]	x			x			x	x	
Ferraz et Garcia (2008) [6]	x			x			x	x	x
Brisson et Collard (2009) [3]	x			x			x	x	
Escovar et al. (2006) [4]			x	x		x	x	x	x
Miani et al. (2009) [10]			x	x		x	x	x	

Extraction des règles d'association

Phase d'ECD : Traitement

Approches	Type d'ontologie			Type de règles			Mesures		
	Domaine	Application	Floue	Généralisées	sous contrainte	Floue	Support	Confiance	Autres
Escovar et al. (2006) [4]			×	×		×	×	×	×
Miani et al. (2009) [10]			×	×		×	×	×	
Antunes et al. (2007) [1]	×			×	×		×	×	

Extraction des règles d'association

Phase d'ECD : Post-traitement

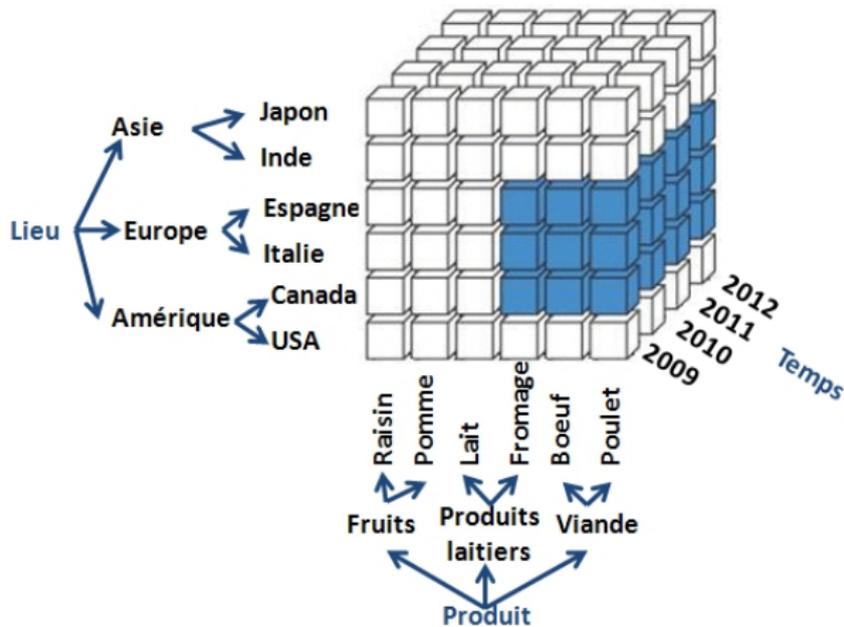
Approches	Type d'ontologie			Type de règles			Mesures		
	Domaine	Application	Floue	Généralisées	sous contrainte	Floue	Support	Confiance	Autres
Ferraz et Garcia (2008) [6]	×			×			×	×	×
Brisson et Collard (2009) [3]	×			×			×	×	
Marinica et Guillet (2010) [9]	×			×			×	×	×
Mansingh et al. (2011) [8]		×		×			×	×	×

Extraction des règles d'association

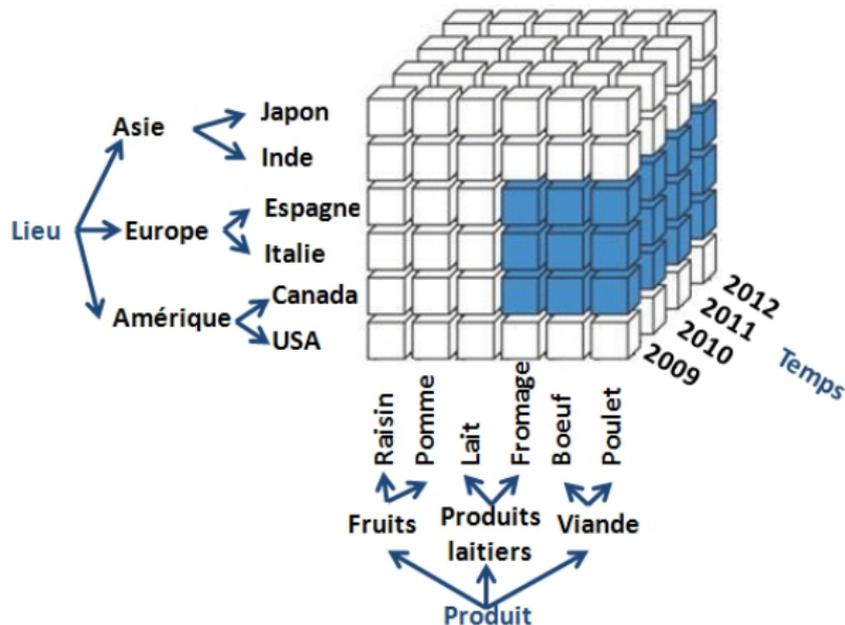
Phase d'ECD : Évaluation

Approches	Type d'ontologie			Type de règles			Mesures		
	Domaine	Application	Floue	Généralisées	sous contrainte	Floue	Support	Confiance	Autres
Hou et al. (2005) [7]	×			×			×	×	
Bellandi et al. (2008) [2]	×			×			×	×	
Brisson et Collard (2009) [3]	×			×			×	×	
Mansingh et al. (2011) [8]		×		×			×	×	×

Cube de données "Ventes"



Cube de données "Ventes"



L'intégration des ontologies dans la fouille des EDs.

Couplage entre les ontologies et les EDs

- **Stoffel et al. (1997)** [13] : Schéma sémantique d'indexation pour la construction d'un ED médicale.

Couplage entre les ontologies et les EDs

- **Stoffel et al. (1997)** [13] : Schéma sémantique d'indexation pour la construction d'un ED médicale.
- **Priebe et Pernul (2003)** [12] : Un portail intelligent de connaissances qui intègre le système OLAP et la recherche des informations.

Couplage entre les ontologies et les EDs

- **Stoffel et al. (1997)** [13] : Schéma sémantique d'indexation pour la construction d'un ED médicale.
- **Priebe et Pernul (2003)** [12] : Un portail intelligent de connaissances qui intègre le système OLAP et la recherche des informations.
- **Nabli et al. (2009)** [11] : Une architecture de construction d'ontologie décisionnelle à base de concepts multidimensionnels.

Couplage entre les ontologies et les EDs

- **Stoffel et al. (1997)** [13] : Schéma sémantique d'indexation pour la construction d'un ED médicale.
- **Priebe et Pernul (2003)** [12] : Un portail intelligent de connaissances qui intègre le système OLAP et la recherche des informations.
- **Nabli et al. (2009)** [11] : Une architecture de construction d'ontologie décisionnelle à base de concepts multidimensionnels.
-

Couplage entre les ontologies et les EDs

- **Stoffel et al. (1997)** [13] : Schéma sémantique d'indexation pour la construction d'un ED médicale.
- **Priebe et Pernul (2003)** [12] : Un portail intelligent de connaissances qui intègre le système OLAP et la recherche des informations.
- **Nabli et al. (2009)** [11] : Une architecture de construction d'ontologie décisionnelle à base de concepts multidimensionnels.
-

Une nouvelle tendance d'approches

La fouille des **règles d'association multidimensionnelles** à partir des **EDs** en présence d'**ontologies**.

Fouille des règles multidimensionnelles

Approches

- 1 **Fouille évolutive** des règles d'association : Tseng et al. (2007) [15].

Fouille des règles multidimensionnelles

Approches

- 1 **Fouille évolutive** des règles d'association : Tseng et al. (2007) [15].
- 2 **Maintenance** avec des **informations ontologiques** : Tseng et Lin (2009) [14].

Fouille des règles multidimensionnelles

Approches

- 1 **Fouille évolutive** des règles d'association : Tseng et al. (2007) [15].
- 2 **Maintenance** avec des **informations ontologiques** : Tseng et Lin (2009) [14].
- 3 Fouille basée sur des **requêtes intelligentes** : Wu et al. (2011) [16].

Fouille des règles multidimensionnelles

Approches

- 1 **Fouille évolutive** des règles d'association : Tseng et al. (2007) [15].
- 2 **Maintenance** avec des **informations ontologiques** : Tseng et Lin (2009) [14].
- 3 Fouille basée sur des **requêtes intelligentes** : Wu et al. (2011) [16].

Comparaison des approches

Approches	Relations ontologiques			Type de stockage		Niveau d'abstraction		Intéraction avec l'utilisateur		Domaine d'application	
	Classification	Composition	Autres	ROLAP	MOLAP	Unique	Multiple	Oui	Non	Général	Spécifique
Tseng et al. (2007)	x	x		x			x		x		x
Tseng et Lin (2009)	x	x		x			x		x		x
Wu et al. (2011)	x	x		x		x		x			x

Fouille des règles multidimensionnelles

Le manque de représentations sémantiques des données

Fouille des règles multidimensionnelles

Le manque de représentations sémantiques des données

- Les schémas (en étoile ou en flocon de neige) ne reflètent pas les relations sémantiques ;

Fouille des règles multidimensionnelles

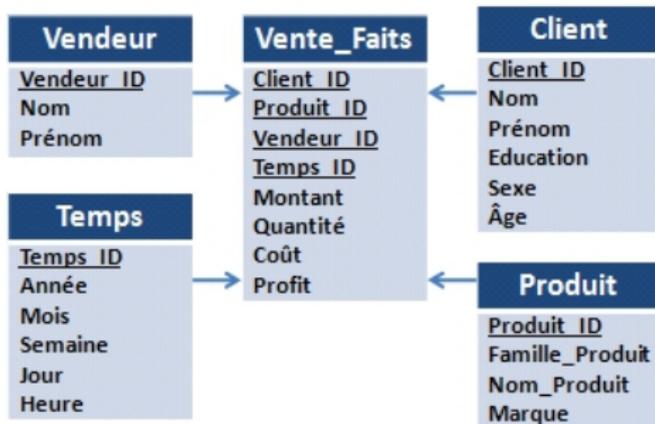
Le manque de représentations sémantiques des données

- Les schémas (en étoile ou en flocon de neige) **ne reflètent pas** les relations sémantiques ;
- **Les contraintes sur les mesures ne sont pas exprimées.**

Fouille des règles multidimensionnelles

Le manque de représentations sémantiques des données

- Les schémas (en étoile ou en flocon de neige) **ne reflètent pas** les relations sémantiques ;
- Les contraintes sur les mesures **ne sont pas exprimées**.



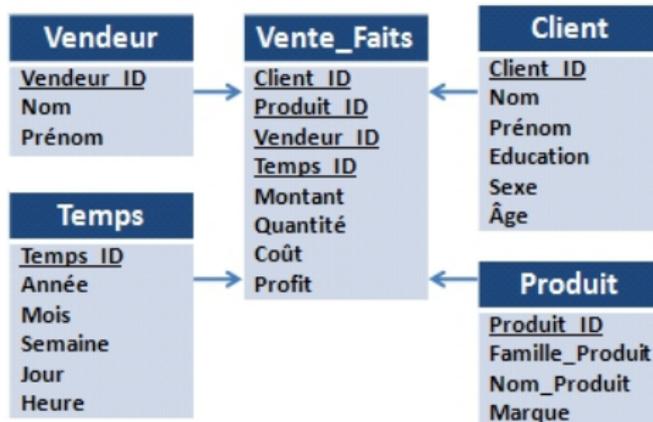
Exemple : Limites du schéma en étoile

- Des relations structurelles entre la table de faits et les tables des dimensions ;

Fouille des règles multidimensionnelles

Le manque de représentations sémantiques des données

- Les schémas (en étoile ou en flocon de neige) **ne reflètent pas** les relations sémantiques ;
- Les contraintes sur les mesures **ne sont pas exprimées**.



Exemple : Limites du schéma en étoile

- Des relations **structurelles** entre la **table de faits** et les **tables des dimensions** ;
- Il n'y a pas de différence entre les mesures additives et semi-additives.

Fouille des règles multidimensionnelles

Le manque de soutien des utilisateurs

???



Fouille des règles multidimensionnelles

Le manque de soutien des utilisateurs

- Comment choisir l'outil d'extraction approprié ?

???



Fouille des règles multidimensionnelles

Le manque de soutien des utilisateurs

- Comment **choisir** l'outil d'extraction approprié ?
- Comment **sélectionner le sous-ensemble de données approprié** ?

???



Fouille des règles multidimensionnelles

Le manque de soutien des utilisateurs

- Comment **choisir** l'outil d'extraction approprié ?
- Comment **sélectionner** le sous-ensemble de données approprié ?
- **Comment formuler une requête appropriée ?**

???

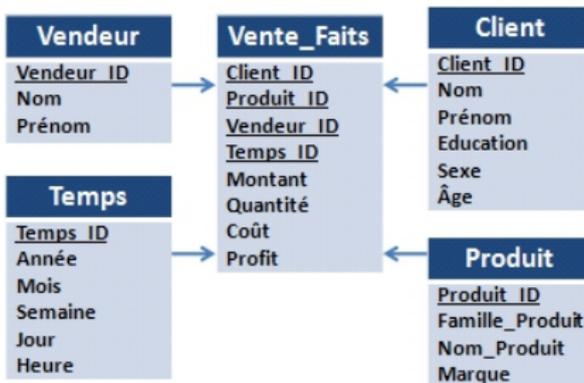


Fouille des règles multidimensionnelles

???

Le manque de soutien des utilisateurs

- Comment **choisir** l'outil d'extraction approprié ?
- Comment **sélectionner** le sous-ensemble de données approprié ?
- Comment **formuler** une requête appropriée ?



Requête

```

Grouping ID Client_ID
Mining Set Temps.Mois,
                Produit.Nom.
FROM "Vente_faits"
With minsup=10%,
        minconf=80%.
  
```

Fouille des règles multidimensionnelles

Des difficultés dans la compréhension des intentions des utilisateurs

Fouille des règles multidimensionnelles

Des difficultés dans la compréhension des intentions des utilisateurs

- Les paramètres fixés lors du lancement des requêtes ;

Fouille des règles multidimensionnelles

Des difficultés dans la compréhension des intentions des utilisateurs

- Les **paramètres** fixés lors du lancement des requêtes ;
- **OLAP Mining offre peu de mécanismes permettant de capturer et d'analyser les anciennes requêtes ;**

Fouille des règles multidimensionnelles

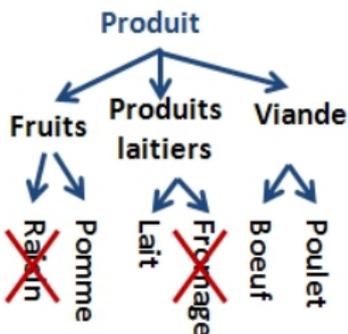
Des difficultés dans la compréhension des intentions des utilisateurs

- Les **paramètres** fixés lors du lancement des requêtes ;
- OLAP Mining offre **peu de mécanismes** permettant de capturer et d'analyser les anciennes requêtes ;
- **Les résultats peuvent être analysés afin de répondre aux futur recommandations des utilisateurs.**

Fouille des règles multidimensionnelles

Des difficultés dans la compréhension des intentions des utilisateurs

- Les **paramètres** fixés lors du lancement des requêtes ;
- OLAP Mining offre **peu de mécanismes** permettant de capturer et d'analyser les anciennes requêtes ;
- Les résultats peuvent être analysés afin de répondre aux **futur recommandations** des utilisateurs.

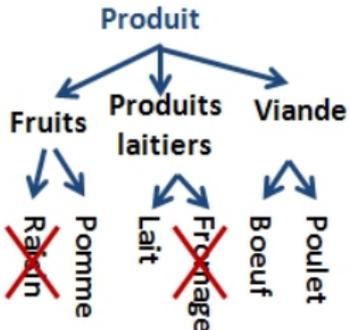


Le manque d'une mise à jour active et efficace des connaissances

Fouille des règles multidimensionnelles

Des difficultés dans la compréhension des intentions des utilisateurs

- Les **paramètres** fixés lors du lancement des requêtes ;
- OLAP Mining offre **peu de mécanismes** permettant de capturer et d'analyser les anciennes requêtes ;
- Les résultats peuvent être analysés afin de répondre aux **futur recommandations** des utilisateurs.



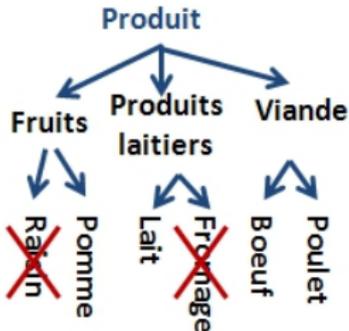
Le manque d'une mise à jour active et efficace des connaissances

- **Un environnement dynamique ;**

Fouille des règles multidimensionnelles

Des difficultés dans la compréhension des intentions des utilisateurs

- Les **paramètres** fixés lors du lancement des requêtes ;
- OLAP Mining offre **peu de mécanismes** permettant de capturer et d'analyser les anciennes requêtes ;
- Les résultats peuvent être analysés afin de répondre aux **futur recommandations** des utilisateurs.



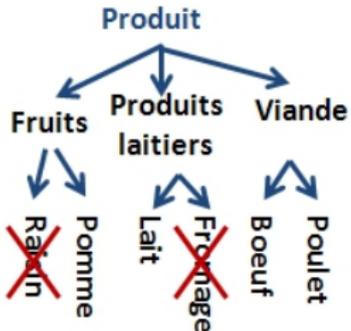
Le manque d'une mise à jour active et efficace des connaissances

- Un environnement **dynamique** ;
- **Processus de fouille interactif, itératif et évolutif** ;

Fouille des règles multidimensionnelles

Des difficultés dans la compréhension des intentions des utilisateurs

- Les **paramètres** fixés lors du lancement des requêtes ;
- OLAP Mining offre **peu de mécanismes** permettant de capturer et d'analyser les anciennes requêtes ;
- Les résultats peuvent être analysés afin de répondre aux **futur recommandations** des utilisateurs.



Le manque d'une mise à jour active et efficace des connaissances

- Un environnement **dynamique** ;
- Processus de fouille **interactif**, **itératif** et **évolutif** ;
- **Comment mettre à jour efficacement les connaissances découvertes devient un défis majeur.**

Conclusion

- 1 OLAP, les ontologies et la FD : des évolutions parallèles et indépendantes ;

Conclusion

- 1 OLAP, les ontologies et la FD : des évolutions parallèles et indépendantes ;
- 2 L'importance et l'intérêt de l'association entre ces trois domaines scientifiques ;

Conclusion

- 1 OLAP, les ontologies et la FD : des évolutions parallèles et indépendantes ;
- 2 L'importance et l'intérêt de l'association entre ces trois domaines scientifiques ;
- 3 Un état de l'art que nous espérons le plus exhaustif possible ;

Conclusion

- 1 OLAP, les ontologies et la FD : des évolutions parallèles et indépendantes ;
- 2 L'importance et l'intérêt de l'association entre ces trois domaines scientifiques ;
- 3 Un état de l'art que nous espérons le plus exhaustif possible ;
- 4 Les ontologies dans les EDs en vue de l'extraction des règles d'association multidimensionnelles.

Perspectives

- 1 Opérer sur des cubes de données **MOLAP** (Multi-Dimensional eXpression) ;

Perspectives

- 1 Opérer sur des cubes de données **MOLAP** (Multi-Dimensional eXpression) ;
- 2 Gérer des **relations autres** que celles de **compositions** et de **classifications** ;

Perspectives

- 1 Opérer sur des cubes de données **MOLAP** (Multi-Dimensional eXpression) ;
- 2 Gérer des **relations autres** que celles de **compositions** et de **classifications** ;
- 3 Utiliser des **formes condensées** (fermés, fermés non-dérivables, etc) ;

Perspectives

- 1 Opérer sur des cubes de données **MOLAP** (Multi-Dimensional eXpression) ;
- 2 Gérer des **relations autres** que celles de **compositions** et de **classifications** ;
- 3 Utiliser des **formes condensées** (fermés, fermés non-dérivables, etc) ;
- 4 Employer les **méta-règles inter-dimensionnelles** afin de **piloter** le processus de recherche des règles.



Cláudia Antunes.

ONTO4AR : A Framework for Mining Association Rules.

In Proceedings of the International Workshop on Constraint-Based Mining and Learning (CMILE-PKDD), Warsaw, Poland, pages 37–48, 2007.



Andrea Bellandi, Barbara Furletti, Valerio Grossi, and Andrea Romei.

Ontological Support for Association Rule Mining.

In Proceedings of the 26th IASTED International Conference on Artificial Intelligence and Applications, Innsbruck, Austria, pages 110–115, 2008.



Laurent Brisson and Martine Collard.

How to Semantically Enhance a Data Mining Process ?, volume 19, chapter Enterprise Information Systems, pages 103–116.

Springer Berlin Heidelberg, 2009.



Eduardo L. G. Escovar, Cristiane A. Yaguinuma, and Mauro Biajiz.

Using Fuzzy Ontologies to Extend Semantically Similar Data Mining.

In Proceedings of the 21st Brazilian Symposium of Databases, Florianópolis, Brazil, pages 16–30, 2006.



Chimène Fankam, Ladjel Bellatreche, Dehainsala Hondjack, Yamine Aït Ameer, and Guy Pierra.

SISRO, Conception de Bases de Données à partir d'Ontologies de Domaine.

Technique et Science Informatiques, 28 :1233–1261, 2009.



Inhaúma Neves Ferraz and Ana Cristina Bicharra Garcia.

Ontology in Association Rules Pre-Processing and post-processing.

In Proceedings of the Third International Conference on Information Technology and Applications, Washington, USA, pages 87–91, 2008.



Xiangdan Hou, Junhua Gu, Xueqin Shen, and Weili Yan.
Application of Data Mining in Fault Diagnosis Based on Ontology.

In Proceedings of the Third International Conference on Information Technology and Applications, Washington, USA, pages 260–263, 2005.



Gunjan Mansingh, Kweku-Muata Osei-Bryson, and Han Reichgelt.

Using Ontologies to Facilitate Post-processing of Association Rules by Domain Experts.

Information Sciences, 181 :419–434, 2011.



Claudia Marinica and Fabrice Guillet.

Knowledge-Based Interactive Postmining of Association Rules Using Ontologies.

IEEE Transactions on Knowledge and Data Engineering, 22 :784–797, 2010.



Rafael Garcia Miani, Cristiane A. Yaguinuma, Marilde Terezinha Prado Santos, and Mauro Biajiz.

NARFO Algorithm : Mining Non-redundant and Generalized Association Rules Based on Fuzzy Ontologies.

In Proceedings of the 11th International Conference on Enterprise Information Systems, ICEIS 2009, Milan, Italy, pages 415–426, 2009.



Ahlem Nabli, Jamel Feki, and Faîez Gargouri.

Advanced Internet Based Systems and Applications.

chapter An Ontology Based Method for Normalisation of Multidimensional Terminology, pages 235–246. 2009.



T. Priebe and G. Pernul.

Ontology-based Integration of OLAP and Information Retrieval.

In Proceedings of the 14th International Workshop on Database and Expert Systems Applications, Washington, DC, USA, DEXA '03, pages 610–614, 2003.



K. Stoffel, J. Saltz, J. Hendler, J. Dick, W. Merz, and R. Miller.

Semantic Indexing For Complex Patient Grouping.

In In Proceedings of the Annual Conference of the American Medical Informatics Association, 1997.



Ming-Cheng Tseng and Wen-Yang Lin.

Incremental Mining of Ontological Association Rules in Evolving Environments.

In *Proceedings of the 22nd International Conference on Industrial, Engineering and Other Applications of Applied Intelligent Systems, Tainan*, pages 142–151, 2009.



Ming-Cheng Tseng, Wen-Yang Lin, and Rong Jeng.

Mining Association Rules with Ontological Information.

In *Proceedings of the Second International Conference on Innovative Computing, Informatio and Control, ICICIC '07, Washington, DC, USA*, pages 300–303, 2007.



Chin-Ang Wu, Wen-Yang Lin, Chang-Long Jiang, and Chuan-Chun Wu.

Toward Intelligent Data Warehouse Mining : An Ontology-Integrated Approach for Multi-dimensional Association Mining.

Expert System Application., 38 :11011–11023, 2011.



Martin Zeman, Martin Ralbovský, Vojtech Svatek, and Jan Rauch.

Ontology-Driven Data Preparation for Association Mining.
In *Proceedings of the 8th Znalosti Conference, Brno.*,
pages 1–12, 2009.

Merci pour votre attention