

Université Lumière Lyon 2  
École doctorale : Sciences cognitives  
Institut de Psychologie  
*Équipe de recherche en Ingénierie des Connaissances*

# Couplage de l'analyse en ligne et de la fouille de données pour l'exploration, l'agrégation et l'explication des données complexes

**Par Riadh BEN MESSAOUD**

Thèse de doctorat de Sciences cognitives  
*Mention Informatique*

Dirigée par Nicolas NICOLOYANNIS

Présentée et soutenue publiquement le 30 novembre

Devant un jury composé de : Mme Rokia MISSAOUI Rapporteur (Professeur, Université du Québec en Outaouais) M. Gilles VENTURINI Rapporteur (Professeur, Université de Tours) M. Lotfi LAKHAL Examineur (Professeur, Université Aix-Marseille 2) M. Gilles ZURFLUH Examineur (Professeur, Université Toulouse 1) M. Nicolas NICOLOYANNIS Directeur de thèse (Professeur, Université Lyon 2) M. Omar BOUSSAID Examineur (Maître de conférences, Université Lyon 2) Mme Sabine LOUDCHER RABASÉDA Examineur (Maître de conférences, Université Lyon 2)



# Table des matières

<b>Résumé .</b>	<b>1</b>
<b>Abstract . .</b>	<b>3</b>
<b>Remerciements . .</b>	<b>5</b>
<b>Thèse au format PDF .</b>	<b>7</b>
Chapitre 1 Introduction générale . .	7
Chapitre 2 État de l'art .	7
Chapitre 3 Réorganisation des cubes de données par une approche factorielle .	7
Chapitre 4 Agrégation par classification dans les cubes de données . .	8
Chapitre 5 Explication dans les cubes de données par règles d'association .	8
Chapitre 6 Implémentation et cas d'application aux données complexes . .	8
Chapitre 7 Vers un cadre formel général .	8
Chapitre 8 Conclusion générale .	8
Bibliographie .	8
Listes des figures, tableaux et algorithmes . .	9
Table des matières .	9



## Résumé

Les entrepôts de données présentent une solution efficace à la gestion des grandes volumétries des données. L'analyse en ligne (OLAP) complète les entrepôts de données en proposant des outils pour la visualisation, la structuration et l'exploration des cubes de données afin d'y découvrir des informations pertinentes. D'un autre côté, la fouille de données emploie des techniques d'apprentissage afin d'induire des modèles de connaissances valides couvrant la description, la classification et l'explication.

L'idée de combiner l'analyse en ligne et la fouille de données est une solution prometteuse pour rehausser le processus d'aide à la décision, notamment dans le cas des données complexes. En effet, il s'agit de deux domaines qui peuvent se compléter dans le cadre d'un processus d'analyse unifié. L'objectif de cette thèse est d'avancer de nouvelles approches d'aide à la décision qui reposent sur le couplage de l'analyse en ligne et de la fouille de données.

Nous avons mis en place trois principales propositions pour ce problème. La première concerne la visualisation des données éparses. En se basant sur l'analyse des correspondances multiples, nous atténuons l'effet négatif de l'éparsité en réorganisant différemment les cellules d'un cube de données. Notre deuxième proposition fournit une nouvelle agrégation des faits d'un cube de données en se basant sur la classification ascendante hiérarchique. Les nouveaux agrégats obtenus sont sémantiquement plus riches que ceux fournis par l'OLAP classiques. Notre troisième proposition établit une démarche explicative en se basant sur les règles d'association. Nous avons élaboré un nouvel algorithme pour une recherche guidée des règles d'association dans les cubes de données. Nous avons également développé une plateforme logicielle pour concrétiser nos contributions théoriques et proposé un cas d'application à des données complexes afin de valider notre démarche de couplage. Enfin, en se basant sur une algèbre OLAP, nous avons mis en place les premières bases d'un cadre formel général dédié au couplage de l'analyse en ligne et de la fouille de données.

**Mots clés :** entrepôt de données, cubes de données, analyse en ligne, fouille de données, analyse des correspondances multiples, classification ascendante hiérarchique, extraction des règles d'association, données complexes, cadre formel général.



## Abstract

Data warehouses provide efficient solutions for the management of huge amounts of data. Online analytical processing (OLAP) is a key feature in data warehouses which enables users with visual tools to explore data cubes. Therefore, users are capable to extract relevant information for their decision-making. On the other hand, data mining offers automatic learning techniques in order to come out with comprehensive knowledge covering descriptions, clusterings and explanations.

The idea of combining online analytical processing and data mining is a promising solution to improve the decision-making process, especially in the case of complex data. In fact, OLAP and data mining could be two complementary fields that interact together within a unique analysis process. The aim of this thesis is to propose new approaches for decision support based on coupling online analytical processing and data mining.

In order to do so, we have established three main proposals. The first one concerns the visualization of sparse data. According to the multiple correspondence analysis, we have reduced the negative effect of sparsity by reorganizing the cells of a data cube. Our second proposal provides a new aggregation of facts in a data cube by using agglomerative hierarchical clustering. The obtained aggregates are semantically richer than those provided by traditional multidimensional structures. Our third proposal tries to explain possible relationships within multidimensional data by using association rules. We have designed a new algorithm for a guided-mining of association rules in data cubes. We have also developed a software platform which includes our theoretical contributions. In addition, we provided a case study on complex data in order to validate our approaches. Finally, based on an OLAP algebra, we have designed the first principles toward a general formal framework which models the problem of coupling online analytical processing and data mining.

**Keywords:** data warehouse, data cubes, on-line analytical processing, data mining, multiple correspondence analysis, agglomerative hierarchical clustering, extraction of association rules, complex data, general formal framework.





---

## Remerciements

Déjà trois ans depuis le tout débüt. Pendant ces trois ans, j'ai rencontré des hommes et des femmes qui ont contribué à ce projet et à qui j'adresse mes remerciements les plus sincères.

Tout d'abord, je remercie M. Nicolas NICOLOYANNIS, mon directeur de thèse, qui m'a accueilli à bras ouverts dans le laboratoire ERIC. Mes remerciements vont

aussi à la fondation VediorBis pour la Recherche et pour l'Emploi qui a financé mes travaux sous l'égide de la Fondation de France.

Je tiens à exprimer ma gratitude à mes encadrants de thèse : M. Omar BOUSSAID et Mme Sabine LOUDCHER RABASÉDA. Avec eux, j'ai apprécié une démarche organisée, une rigueur de travail et surtout un support scientifique inconditionné. Je les remercie également pour la précieuse confiance qu'ils m'ont donnée, pour la grande liberté d'idées et de travail qu'ils m'ont accordée et pour le temps qu'ils m'ont consacré. En dehors de leurs apports scientifiques, je n'oublierai pas aussi de les remercier pour leurs qualités humaines et leur soutien qui m'ont permis de mener à bien cet ouvrage.

Je remercie les membres du laboratoire ERIC, en particulier M. Stéphane LALLICH et M. Djamel A. ZIGHED. Je remercie aussi ceux que j'ai côtoyés et qui, de loin ou de près, ont apporté une plus-value à la qualité de mon travail par leurs conseils, leurs remarques et leurs observations. Je tiens spécialement à remercier mes anciens et actuels collègues de bureau : Kamel AOUICHE, Amandine DUFFOUX, Cécile FAVRE, Hadj MAHBOUBI, Nora MAIZ et Elie PRUDHOMME. C'était très agréable de travailler en leur compagnie.

Merci à mes parents et à mes soeurs qui, malgré l'éloignement, ont cru à ce projet, l'ont soutenu et l'ont défendu. Je leur suis reconnaissant pour les sacrifices qu'ils ont dû faire pendant mes longues années d'études et d'absence. Merci à mes amis de toujours : Asma, Anis, Boulbaba, Mehdi, Moncef et Sofiane. Merci à ma fiancée, Intissar, pour son soutien et le deuxième souffle qu'elle a toujours su apporter avec tendresse et sérénité dans les moments les plus difficiles. Merci à tous ceux qui y ont cru . . .

***“The White Rabbit put on his spectacles. ‘Where Shall I begin, Please your Majesty ?’ he asked. ‘Begin at the beginning,’ the King said gravely, ‘and go till you come to the end : then stop.’ ” Lewis Carroll, “Alice’s Adventures in Wonderland”***



# Thèse au format PDF

## Chapitre 1 Introduction générale

[benmessaoud\\_r\\_these\\_chap1.pdf](#)

## Chapitre 2 État de l'art

[benmessaoud\\_r\\_these\\_chap2.pdf](#)

## Chapitre 3 Réorganisation des cubes de données par une approche factorielle

[benmessaoud\\_r\\_these\\_chap3.pdf](#)

## **Chapitre 4 Agrégation par classification dans les cubes de données**

[benmessaoud\\_r\\_these\\_chap4.pdf](#)

## **Chapitre 5 Explication dans les cubes de données par règles d'association**

[benmessaoud\\_r\\_these\\_chap5.pdf](#)

## **Chapitre 6 Implémentation et cas d'application aux données complexes**

[benmessaoud\\_r\\_these\\_chap6.pdf](#)

## **Chapitre 7 Vers un cadre formel général**

[benmessaoud\\_r\\_these\\_chap7.pdf](#)

## **Chapitre 8 Conclusion générale**

[benmessaoud\\_r\\_these\\_chap8.pdf](#)

## **Bibliographie**

[benmessaoud\\_r\\_these\\_biblio.pdf](#)

## Listes des figures, tableaux et algorithmes

[benmessaoud\\_r\\_these\\_listes.pdf](#)

## Table des matières

[benmessaoud\\_r\\_these\\_tdm.pdf](#)