

الجمهورية الشعبية الديمقراطية الجزائرية
République Algérienne Démocratique et Populaire
وزارة التعليم العالي والبحث العلمي
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique
المدرسة العليا للإعلام الآلي ٠80 ماي 5491. بسيدي بلعباس
École Supérieure en Informatique
-08 Mai 1945- Sidi Bel Abbès



THESIS

To obtain the diploma of Masters
Field: **Computer Science**
Specialty: **Intelligence Artificielle et Sciences de Données (IASD)**
Theme

**Querying Probabilistic OLAP data cubes with
hierarchies: models and algorithms**

Presented by:
TALBI Youcef
BENDOU Ahmed El Mahdi

Submission Date: **June, 2026**
In front of the jury composed of:

Prof. Nabil Keskes
Dr. Belkacem Khaldi
Prof. Alfredo Cuzzocrea
Dr. Mohamed Salih Bendella

President
Supervisor
Co-Supervisor
Examiner

Academic Year : 2025/2026

Abstract

On-Line Analytical Processing assumes that every fact occupies a single, fully determined position in the multidimensional space, but large-scale data increasingly violates this premise. This thesis develops the theory of OLAP over data with imprecise dimensions and uncertain measures, relaxing the point-fact assumption so that a fact becomes a region of cells. Two formal accounts of the resulting ambiguity are developed: a probabilistic account via possible-worlds semantics induced by allocation, and a fuzzy account via graded membership and hesitation margins. The treatment covers consistency, faithfulness, and correlation-preservation, together with domain integrity constraints and hierarchy irregularities. A final part analyses the algorithms and their complexity: EM-based allocation, top- k query processing, incremental OLAP, and probabilistic inverse ranking, all unified by replacing exponential world sets with compact linear summaries of weighted claims.

Keywords: Probabilistic OLAP, fuzzy OLAP, uncertain data, hierarchical dimensions, allocation algorithms, intuitionistic fuzzy sets, top-k queries, incremental OLAP

Résumé

Le Traitement Analytique En Ligne (OLAP) suppose que chaque fait occupe une position unique et entièrement déterminée dans l'espace multidimensionnel, mais les données à grande échelle violent de plus en plus cette hypothèse. Cette thèse développe la théorie de l'OLAP sur des données aux dimensions imprécises et aux mesures incertaines, en relâchant l'hypothèse de fait ponctuel. Deux comptes formels de l'ambiguïté résultante sont développés : un compte probabiliste via la sémantique des mondes possibles induite par l'allocation, et un compte flou via l'adhésion graduée et les marges d'hésitation. Le traitement couvre la cohérence, la fidélité et la préservation des corrélations, ainsi que les contraintes d'intégrité de domaine et les irrégularités hiérarchiques. Une dernière partie analyse les algorithmes et leur complexité, notamment l'allocation EM, requêtes Top- k , OLAP incrémental et classement inverse probabiliste, tous unifiés par le remplacement des ensembles exponentiels de mondes par des résumés linéaires compacts.

Mots-clés : OLAP probabiliste, OLAP flou, données incertaines, dimensions hiérarchiques, algorithmes d'allocation, ensembles flous intuitionnistes, requêtes Top- k , OLAP incrémental

ملخص

تفترض المعالجة التحليلية الفورية (On-Line Analytical Processing - OLAP) أن كل واقعة (fact) تشغل موضعاً مفرداً ومحددًا بالكامل في الفضاء متعدد الأبعاد، إلا أن البيانات واسعة النطاق تتجاوز هذه الفرضية على نحو متزايد. وتعمل هذه الأطروحة على تطوير نظرية ال OLAP على البيانات ذات الأبعاد غير الدقيقة والمقاييس غير اليقينية، مما يوجب التخلي عن فرضية "الواقعة النقطية" لتصبح الواقعة حيناً من الخلايا. وتُقدّم الأطروحة صياغتين رسميتين للغموض الناشئ عن ذلك: صياغة احتمالية عبر دلالات العوالم الممكنة (possible-worlds semantics) المستحثة بالتخصيص، وصياغة ضبابية عبر درجات العضوية وهوامش التردد. وتغطي هذه الدراسة مفاهيم الاتساق، والأمانة الدلالية، والحفاظ على الارتباط، بالإضافة إلى قيود سلامة النطاق والتشوهات الهرمية. وفي الجزء الأخير، تجري الأطروحة تحليلاً للخوارزميات وتعقيدها الحسابي، بما في ذلك التخصيص القائم على تعظيم التوقع (EM-based allocation)، ومعالجة استعلامات أفضل- k (query processing k top-)، وال OLAP التزايد (incremental OLAP)، والترتيب العكسي الاحتمالي، حيث تُوحّد هذه الأدوات كافة عبر استبدال مجموعات العوالم الأسية بملخصات خطية مدمجة للادعاءات الموزونة.

الكلمات المفتاحية:

ال OLAP الاحتمالي، ال OLAP الضبابي، البيانات غير اليقينية، الأبعاد الهرمية، خوارزميات التخصيص، المجموعات الضبابية الحدسية (Intuitionistic fuzzy sets)، استعلامات Top- k ، ال OLAP التزايد.