



Pôle des Etudes Doctorales
Centre des Etudes Doctorales
Sciences et Techniques et Sciences Médicales

AVIS DE SOUTENANCE DE THESE DE DOCTORAT

Madame EL HADDADI Oumaima
Présentera ses travaux de recherche en vue de l'obtention du
Doctorat



**Formation Doctorale : Sciences et Techniques de
l'Ingénieur(STI)**
Discipline : Informatique
Spécialité : Gestion des données, Big Data et Intelligence Artificielle

**Le 05/07/2025 à 14H00 à la salle de conférence F de la Faculté
des Sciences et Techniques de Tanger, UAE**

Sous le thème

**Alignement incrémental de données hétérogènes et
dynamiques pour l'aide à la décision**

Devant le jury composé de :

Nom et Prénom	Etablissement	Qualité
Pr. AKNIN Noura	FS de Tétouan, UAE	Présidente
Pr. BENATTOU Mohammed	FS de Kénitra, UIT	Rapporteur
Pr. FENNAN Abdelhadi	FST de Tanger, UAE	Rapporteur
Pr. MONTICOLO Davy	Université de Lorraine, Nancy, France	Rapporteur
Pr. FARHAOUI Youssef	FST d'Errachidia, UMI	Examinateur
Pr. KOULALI Sara	FSJES de Oujda, UMP	Examinatrice
Pr. TESTE Olivier	Université de Toulouse, France	Examinateur
Pr. EL ALLAOUI Ahmad	FST Errachidia, UMI	Co-Directeur
Pr. BAHRI Abdelkhalek	ENSAH, d'Al-Hoceima, UAE	Directeur
Pr. CHEVALIER Max	Université de Toulouse, France	Directeur

Structure de recherche : Science des données et intelligence Compétitive – SDIC (ENSAH – Université Abdelmalek Essaâdi, Maroc) /
Systèmes d'Information Généralisés – SIG (IRIT - Université de Toulouse, France)

Résumé

A l'ère du Big Data, où les données sont massives et leur complexité croissante, le besoin en informations reste crucial pour chaque organisation. En effet, qu'il s'agisse d'informations collectées sur l'environnement direct (clients, fournisseurs, concurrents...) ou des informations qu'elle a elle-même générées, celles-ci représentent un capital que l'organisation doit exploiter pour rester performante. C'est pourquoi les décideurs doivent prendre, au moment opportun, les meilleures décisions. Dans ce cadre, les systèmes d'aide à la décision visent à les aider à analyser l'ensemble de ces données afin de faciliter et de guider leurs choix stratégiques.

Du point de vue du système d'information, deux perspectives peuvent être identifiées : la première concerne l'intégration des données (entrepôts de données, silos...) qui vise à rassembler toutes les données nécessaires à la prise de décision dans un même espace homogène. La seconde vise à stocker les données de manière brute (lacs de données ou data lakes en anglais) dans un même espace qui lui sera, par nature, hétérogène.

Quelle que soit la perspective choisie, un des grands enjeux reste l'alignement des données. En effet, l'alignement permet d'établir des correspondances entre les données disponibles dans l'organisation afin de mieux les intégrer, les cataloguer et plus globalement mieux les exploiter pour l'aide à la décision. Dans les environnements dynamiques dans lesquels les organisations se retrouvent, une double problématique impacte les performances des approches d'alignement de données. En effet, dans ces environnements les données exploitées par l'organisation peuvent être à la fois hétérogènes (la diversité des données et de leur structure) et dynamiques (évolution dans le temps du contenu et des structures des données, apparition de nouvelles sources de données).

De multiples approches d'alignement de données ont été proposées dans la littérature. Toutefois ces travaux s'intéressent exclusivement, à notre connaissance, à la problématique de l'hétérogénéité de données. Cependant, ces travaux ne prennent pas en compte la problématique liée à la dynamique des données. En réponse à cela, l'objectif de nos travaux dans cette thèse est de considérer la dynamique des données pour améliorer leurs alignements dans les environnements dynamiques.

Or, la problématique de la dynamique des données a été explorée dans des domaines autres que l'alignement. Parmi eux, on trouve la gestion des flux de données et les réseaux dynamiques. Diverses solutions ont émergé pour répondre à ces défis. Par exemple, les graphes dynamiques permettent de représenter et de maintenir les relations évolutives entre les entités. Ils offrent une approche efficace pour modéliser des systèmes en constante mutation. Parallèlement, les mises à jour en flux (stream processing en anglais) proposent des mécanismes pour traiter les données en temps réel. Elles éliminent ainsi la nécessité de recalculs globaux. Plus récemment, les techniques de plongement (embeddings en anglais) dynamiques se sont distinguées. Elles capturent des relations complexes tout en permettant une mise à jour progressive. Ces approches répondent aux besoins croissants des environnements où les données évoluent rapidement.

Ces travaux nous ont inspiré pour concevoir et développer l'approche incrémentale d'alignement de données, nommée 'IDAGEmb', que nous proposons. Cette contribution originale dans le contexte de l'aide à la décision est basée sur des plongements dynamiques de graphe. Cette contribution considère la nature évolutive des données en actualisant progressivement les alignements tout en minimisant les ressources nécessaires au calcul de ces derniers.

Les résultats obtenus dans nos expérimentations soulignent l'intérêt de l'alignement incrémental pour l'aide à la décision dans des environnements fortement dynamiques. Nous proposons une solution concrète qui vise à améliorer le calcul des alignements de données et par voie de conséquence l'efficacité de l'analyse des données et du processus globale de prise de décision.

Mots clés : Gestion de Données, Aide à la décision, Environnement Dynamique, Hétérogénéité, Alignement de données, Plongement Dynamique, Méthode Incrémentale.