

Avis de Soutenance

Madame Ikram NEKKACHE

Informatique et applications

Soutiendra publiquement ses travaux de thèse intitulés

Extraction de motifs fréquents sous contraintes

dirigés par Monsieur Lakhdar SAÏS et Nadjat KAMEL
Co-tutelle avec l'Université Ferhat-Abbas de Sétif - ALGÉRIE

Soutenance prévue le **dimanche 05 décembre 2021** à 10h00

Lieu : Université de Ferhat Abbas - Sétif 1, Campus el Bez 19000 Sétif, Algérie
Salle de visio conférence

Composition du jury proposé

M. Lakhdar SAÏS	Université d'Artois	Directeur de thèse
Mme Nadjat KAMEL	Université Ferhat Abbas Sétif 1	Co-directrice de thèse
M. Belaid BENHAMOU	Aix-Marseille Université	Rapporteur
Mme Habiba DRIAS	Université des Sciences et de la Technologie Houari Boumedienne	Rapporteuse
Mme Zibouda ALIOUAT	Université Ferhat Abbas Sétif1	Examinatrice
M. Saïd JABBOUR	Université d'Artois	Examinateur

Résumé :

L'extraction de motifs fréquents est l'une des tâches fondamentales de la fouille de données. Elle consiste à découvrir des motifs intéressants pour l'utilisateur à partir des bases de données. Différents types des motifs peuvent être trouvés à partir de divers types de données, tels que les données transactionnelles, les séquences, les graphes et les arbres. Cette thèse s'intéresse aux deux tâches bien connues de la fouille de motifs à savoir la fouille des itemsets fréquents et la fouille des motifs séquentiels. La première contribution de cette thèse concerne l'utilisation des approches déclaratives de type SAT pour l'extraction des itemsets fréquents fermés. Ces approches sont connues pour être flexibles en permettant d'ajouter des contraintes pour extraire des motifs particuliers. L'un des problèmes majeurs de ces approches est le passage à l'échelle dû à la taille de l'encodage des grandes bases transactionnelles. Pour réduire la taille de ces encodages, nous avons proposé des représentations plus compactes. Plus précisément, nous montrons que notre approche peut être vue comme un problème de compression de matrices booléennes. Nos résultats montrent une réduction significative de la taille de l'encodage. Notre seconde contribution consiste à exploiter les symétries dans le cadre de la fouille des motifs séquentiels. La recherche de ces symétries est effectuée en encodant la table des séquences sous forme d'un graphe orienté coloré. Les symétries sont alors énumérées en calculant les automorphismes du graphe. Finalement, nous proposons deux approches pour exploiter les symétries détectées. Dans la première, nous montrons comment nous pouvons améliorer l'extraction de motifs séquentiels en intégrant les symétries découvertes dans un algorithme de type Apriori. Dans la seconde approche nous démontrons comment ces symétries peuvent être utilisées en prétraitement. Cela est réalisé en modifiant la base de séquences originale.