

Improvement of TA-ARM Algorithm for Concept Maps Based on Text Analysis

**By
Sara Saleh Ahmad Alomari**

**A thesis submitted for the requirements of the master's degree of Science in Information
Systems**

**Supervised by
Dr. Salha Binti Abdullah**

تحسين خوارزمية TA-ARM لخرائط المفاهيم بناءً على تحليل النص

سارة صالح العمري

المستخلص

لقد أصبحت خرائط المفاهيم مؤخرًا واحدة من وسائل تمثيل المعرفة التي أثبتت فاعليتها كوسيلة تعليمية. قد سهلت عملية الدراسة للمتعلمين خصوصًا عندما تكون مصادر المادة الدراسية غير منظمة وغنية بالمعلومات عن مفاهيم مختلفة مرتبطة ببعضها بطريقة ما. تساعد خرائط المفاهيم هؤلاء المتعلمين لتحديد العلاقات والروابط بين هذه المفاهيم.

بالرغم من إيجابيات خرائط المفاهيم فإن إنشاء خريطة مفاهيم واحدة يعتبر عملية غير سهلة بحيث تتطلب مجهودًا خاصًا من شخص متخصص يمتلك المعرفة الكافية في المجال بجميع المفاهيم والعلاقات الممكنة بينها. لهذا السبب اجتهد الباحثون في تطوير خوارزميات تساعد في أتمتة عملية إنشاء خرائط المفاهيم تضمن أقل تدخل من الانسان وتوفير وقته وجهده وذلك باستخدام أساليب تحليتي النص.

في هذا البحث قام الباحث باقتراح تحسين مرحلة تحليل النص في خوارزمية TA-ARM التي تنشئ خريطة مفاهيم باستخدام مصنف K-NN لتصنيف الأسئلة إلى مفاهيم و Apriori principle لإيجاد العلاقات بين هذه المفاهيم. قام الباحث بتطبيق المصنف distance-weighted K-NN بدلا من مصنف K-NN التقليدي بهدف تحسين دقة التصنيف إلى مفاهيم وقد أظهرت النتائج تفوق مصنف distance-weighted K-NN في الدقة ومقياس درجة ف 1 وجميع مقاييس التقييم الأخرى على مصنف K-NN التقليدي في تصنيف البيانات المختارة لهذا البحث مما يدل على فاعليته في التعامل مع مجموعة البيانات الغير متوازنة.

Improvement of TA-ARM Algorithm for Concept Maps Based on Text Analysis

**By
Sara Saleh Ahmad Alomari**

**A thesis submitted for the requirements of the master's degree of Science in Information
Systems**

**Supervised by
Dr. Salha Binti Abdullah**

ABSTRACT

Text analysis is one of the emerged methods to deal with unstructured text data that allows to automatically extract and classify information from text. It has many applications in Knowledge Visualization (KV) field as it serves as the base of transforming textual data into visual presentations such as sketches and conceptual diagrams.

Concept map is a form of conceptual diagrams that shows the main concepts of a given text and shows the relationships among these concepts. However, constructing a manual concept map from a given text is deemed to be a tedious task. It is time-consuming, and it demands an extensive effort in reading the textual content and reasoning the relationships among concepts. For this reason, many studies have utilized text analysis methods to develop intelligent algorithms in order to generate concept maps automatically with a high accuracy at a minimal effort.

This thesis aims at improving the Text Analysis phase in Text Analysis and Association Rules Mining (TA-ARM) algorithm that automatically generates concept maps from test questions and answer records. The improvement of test questions classification in Text Analysis phase is supposed to lead to more accurately classified test questions into concepts.

The distance-weighted K-Nearest Neighbors classifier is applied to classify test questions instead of the traditional K-NN. This approach is expected to handle the imbalanced dataset and increases classification accuracy.

A comparison between the distance-weighted K-NN classifier and the traditional K-NN were undertaken. The results have shown better classification performance by the distance-weighted K-NN than by the traditional K-NN in terms of accuracy, precision, recall, and F1-score metrics. the distance-weighted K-NN achieved an accuracy of 95% and F1-score of 96% where the traditional K-NN achieved an accuracy of 71% and F1-score of 55%. These results have confirmed the effectiveness of the proposed improvement.