



وزارة التعليم العالي والبحث العلمي
جامعة الموصل
كلية علوم الحاسوب والرياضيات
قسم علوم الحاسوب

التصنيف والوصف المرئي للأنشطة الأكاديمية في قاعدة البيانات المتدفقة لقسم علوم الحاسوب

رسالة مقدمة
الى مجلس كلية علوم الحاسوب والرياضيات في جامعة الموصل
كجزء من متطلبات نيل شهادة ماجستير علوم في
علوم الحاسوب

من قبل

أزهار محمد صالح يوسف محمد

بإشراف
أ.م.د. عمار ظاهر ياسين طه

المُلخَص

تُطبق شركات الإنترنت وغيرها من المؤسسات تقنيات لغات NoSQL بدلا من SQL؛ لسرعتها في إدخال، وإسترجاع، ومعالجة البيانات الضخمة. إن سهولة التصميم والقدرة على التوسع أفقيا وبساطة أنموذج البيانات الخالي من المخططات جعلت الشركات تبدأ في إدراك أن قواعد البيانات NoSQL هي طريقة مميزة لإدارة الكميات الكبيرة من البيانات تنشئها مصادر مختلفة وغير متجانسة وباستمرار ولها أشكال وأحجام مختلفة وتسدعي معالجة في الوقت الفعلي وتدعى "البيانات المتدفقة Data Stream DS".

تهدف الرسالة إلى إنشاء قاعدة بيانات متدفقة لإدارة معلومات الأساتذة والطلبة والموظفين في قسم علوم الحاسوب في كلية علوم الحاسوب والرياضيات في جامعة الموصل. فضلا عن تصنيف أنشطة الأساتذة المتدفقة في قاعدة البيانات بتحديد نوع النشاط الدقيق للأستاذ من حيث وقوعه ضمن التخصص ام مجرد اهتمام بحثي. تم ذلك من خلال استخدام قاعدة البيانات MongoDB والتي هي إحدى أنواع قواعد البيانات NoSQL التي تتميز بالمرونة وقابلية التوسع الأفقي ومناسبتها للبيانات المتدفقة. تم قراءة ومعالجة البيانات المتدفقة وحقنها في قاعدة البيانات MongoDB من خلال استخدام أداة أباتشي سبارك التي تعد من الأدوات المتميزة للتعامل مع البيانات المتدفقة. إذ تم تمثيل أنموذج محاكاة افتراضي لعمل أداة سبارك لعرض أنشطة الاساتذة بشكل آني لتغيير الانشطة حسب وقت ونوع النشاط باستخدام مجموعة من الفقاعات كل فقاعة تمثل استاذ، يتغير لون الفقاعة بالاعتماد على نوع النشاط. تم تطبيق ثلاث خوارزميات للتعلم الآلي (التصنيف) باستخدام لغة بايثون3 متمثلة بخوارزمية شجرة القرار Decision Tree DT، وخوارزمية آلة المتجه الداعمة Support Vector Machine SVM، وخوارزمية الإدراك الحسي متعدد الطبقات Multilayer Perceptron MLP على قاعدة البيانات المتدفقة لأنشطة 45 أستاذا والتي تضم 3881 قيدا. تم استخدام 70% من البيانات للتدريب، و30% للاختبار.

أظهرت النتائج أن خوارزمية MLP هي أفضل خوارزمية لتصنيف أنشطة الأساتذة؛ لحصولها على أعلى دقة 0.9795%، بينما دقة خوارزمية DT 0.9792%، ودقة خوارزمية SVM كانت 0.9283%. نتيجة التصنيف هذه أثبتت مدى فاعلية النظام المقترح لمساعدة القسم وأساتذته والكلية على حد سواء في إدارة النشاطات العلمية وتبويبها بشكل صحيح، فضلا عن إمكانية استخدام النظام كوسيلة مساعدة في تقييم ضمان الجودة لأساتذة القسم وموظفيه.

**Ministry of Higher Education and
Scientific Research
University of Mosul
College of Computer Science and
Mathematics
Department of Computer Science**



The Classification and Visualization of the Academic Activities in the Streaming Database of the Computer Science Department

**A Thesis Submitted to the Council of the College of
Computer Science and Mathematics
University of Mosul
as a Partial Fulfillment of Requirements
for the Degree of Master of Science**

Computer Science

By

Azhar Mohammed Saleh Yousef Mohammed

Supervised by

Assist Prof. Dr. Ammar Thaher Yaseen Taha

2025 A.D.

1446 A.H

Abstract

Internet companies and other organizations are implementing NoSQL languages technologies instead of SQL because of their speed in inserting, retrieving, and processing big data. The ease of design, the ability to scale horizontally, and the simplicity of the schema-free data model let companies begin to realize that NoSQL databases are a great way to manage large amounts of data. Large amounts of data that are generated from different, heterogeneous sources on a continuous basis and of different formats and sizes and that require real-time processing are called “streaming data.”

The aim of the thesis is to create a streaming database to manage information of professors, students and employees in the Department of Computer Science in the College of Computer Science and Mathematics at the University of Mosul. In addition to classifying the streaming activities of professors in the database by specifying the exact type of activity of the professor in terms of whether it falls within the specialization or just a research interest. This was done by using the MongoDB database, which is one of the types of NoSQL databases that are characterized by flexibility, horizontal scalability and suitability for streaming data. The streaming data was read, processed and injected into the MongoDB database using the Apache Spark tool, which is one of the distinguished tools for dealing with streaming data. A virtual simulation model was represented for the Spark tool to display professors' activities instantly to change activities according to the time and type of activity using a set of bubbles, each bubble represents a professor, the color of the bubble changes depending on the type of activity. Three machine learning (classification) algorithms were applied using Python3 language represented by Decision Tree DT, Support Vector Machine SVM, and Multilayer Perceptron MLP on the streaming database of 45 professors' activities which includes 3881 entries. 70% of the data was used for training, and 30% for testing..

The results showed that MLP is the best algorithm for classifying professors' activities, as it obtained the highest accuracy of 0.9795%, while the accuracy of Decision Tree is 0.9792% and the accuracy of Support Vector Machine is 0.9283%. This classification result proved the effectiveness of the proposed system in helping professors, department and college alike in managing and classifying scientific activities correctly in addition to that the ability of using the suggested system for helping checking professors and employee quality.