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

تطوير أداة نمذجة مفاهيمية ومنطقية لقواعد البيانات

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

من قبل

عمر بشار علي احسان اسماعيل

بإشراف

أ.م.د ندى نعمت سليم داؤد

١٤٤٧ هـ ٢٠٢٥ م

المستخلص

تعد هندسة البرمجيات من فروع علوم الحاسوب الأساسية، وتركّز على تطوير أنظمة برمجية عالية الجودة من خلال جمع وتحليل المتطلبات بشكل منظم. وتُعد نمذجة المتطلبات خطوة مهمة تترجم احتياجات المستخدم إلى تمثيل واضح يسهل تنفيذه.

لكن صياغة المتطلبات بلغة طبيعية قد تسبب غموضًا وسوء فهم. ولتجاوز ذلك تُستخدم تقنيات معالجة اللغات الطبيعية (Natural Language Processing–NLP) لتحليل النصوص لاستخلاص المكونات الاساسية لنمذجة البيانات في مستوييها المفاهيمي والمنطقي، مما يعزز دقة النمذجة ويقلل الأخطاء، ويساعد على بناء أنظمة بيانات أكثر وضوحًا وموثوقية.

تم في هذه الرسالة تصميم وبناء اداة مقترحة (Database Conceptual and Logical Model Tool– DBCLM) التي تقوم بقراءة وتحليل متطلبات العمل النصية المنتظمة (المكتوبة باللغة الانكليزية) لنمذجة البيانات بمستويين المفاهيمي والمنطقي، وبشكل آلي من خلال استخدام تقنيات معالجة اللغات الطبيعية، وبالاعتماد على تطبيق مجموعة من التعابير المنتظمة (Regular Expressions) والقواعد الارشادية (Heuristics Rules) لاستخلاص المعلومات والمكونات المطلوبة لتصميم وانشاء النموذج المفاهيمي (Conceptual Model) ورسمه وتمثيله بمخطط الكيان والعلاقة (Entity Relationship Diagram). كما ان الاداة لها امكانية بناء النموذج المنطقي من خلال تحويل النموذج المفاهيمي الى النموذج المنطقي (Logical Model) وتمثيله من خلال انشاء النموذج العلائقي (Relational Model)، مما يسهل على المطورين بناء قواعد بيانات الانظمة وفقاً للمعايير الدولية وبشكل سليم.

تم تطوير الأداة باستخدام لغة البرمجة بايثون(Python) ، بالاعتماد على مجموعة من القواعد الارشادية والتعابير المنتظمة بالإضافة الى مكتبات متخصصة في تقنيات معالجة اللغات الطبيعية، مثل مكتبة (Stanza وNLTK)، وذلك لتنفيذ مهام التحليل اللغوي واستخراج مكونات النموذج. كما تم توظيف المعجم الدلالي الإلكتروني (WordNet) باللغة الإنجليزية، لدعم تفسير المفردات وتقليل الغموض في نصوص متطلبات العمل، مما ساهم في تحسين دقة الفهم والتحليل اللغوي للنصوص المدخلة.

تم تطبيق وتنفيذ الأداة المقترحة (DBCLM) على مجموعة من حالات الدراسة في مجالات مختلفة من خلال إدخال متطلبات العمل النصية للمستخدم. وأظهرت النتائج كفاءة عمل الاداة في نمذجة البيانات، إذ بلغ معدل الدقة الإجمالي(98.18%) لخمسة حالات دراسية في مجالات مختلفة. وقدمت الأداة مخرجات واضحة ومتكاملة من خلال استخراج النماذج المفاهيمية والمنطقية لبناء قواعد البيانات وبشكل متزامن مما يسهم في دعم المطورين وأصحاب المصلحة لتطوير انظمة قواعد البيانات بكفاءة أعلى، وجودة أفضل وتقليل الاخطاء البشرية.

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

Developing Conceptual and Logical Database Modeling Tool

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 in Software

By

Omer Bashar Ali Ehsan Ismael

Supervised by

Dr. Nada Nimat Saleem Daood
Assistant Professor

2025 A.D. **1447 A.H.**

Abstract

Software engineering is considered one of the fundamental branches of computer science, focusing on the development of high-quality software systems through the systematic collection and analysis of requirements. Requirements modeling plays a vital role in translating user needs into a clear and implementable representation. However, expressing requirements in natural language often leads to ambiguity and misinterpretation. To overcome this, Natural Language Processing (NLP) techniques are employed to analyze textual requirements and extract the core components necessary for both conceptual and logical data modeling. This enhances modeling accuracy, reduces errors, and contributes to the development of more reliable and comprehensible data systems.

In this thesis, a proposed tool named Database Conceptual and Logical Model Tool (DBCLM) was designed and implemented to automatically read and analyze structured textual business requirements (written in English) for the purpose of data modeling at both the conceptual and logical levels. The tool utilizes NLP techniques, supported by a set of regular expressions and heuristic rules, to extract the required information and components for designing the Conceptual Model, and then generates and visualizes it using Entity-Relationship Diagrams (ERDs). Furthermore, the tool is capable of transforming the conceptual model into a Logical Model by constructing its corresponding Relational Model, thus facilitating the development of database systems in accordance with international standards.

The tool was developed using the Python programming language, relying on heuristic rules, regular expressions, and specialized NLP libraries such as Stanza and NLTK to perform linguistic analysis and extract model components. Additionally, the WordNet lexical database was integrated to support word interpretation and reduce ambiguity in requirement texts, which significantly enhanced the precision of linguistic understanding and analysis.

The proposed (DBCLM) tool was tested on a variety of case studies from different domains using actual user requirement texts. The results demonstrated the effectiveness of the tool in data modeling, achieving an overall accuracy rate of 98.18% across five different case studies. The tool produced clear and comprehensive outputs by generating both conceptual and logical models simultaneously, which supports developers and stakeholders in building database systems with higher efficiency, improved quality, and reduced human errors.