Using MADA+TOKAN to Generate Use Case Models from Arabic User Requirements in a Semi-Automated Approach

Nabil Arman Department of Computer Science and Engineering Palestine Polytechnic University Hebron, Palestine narman@ppu.edu

Abstract—Automated software engineering has attracted a large amount of research efforts. The need for new approaches that reduces the cost of developing software systems within project schedule has made it necessary to develop approaches that aid in the construction of different UML models in a semi-automated approach from Arabic textual user requirements. UML use case models represent an essentialartifact that provide a perspective of the system under analysis or development. Thedevelopment of such use case models is very crucial in an object-oriented development methodology. In this paper, MADA+TOKAN is used to parse different statements of the user requirements written in Arabic to obtain different components of a sentence like lists of nouns, noun phrases, verbs, verb phrases, etc. that aid in finding potential actors and use cases. A set of steps that represent our approach for constructing a use case model is presented. Finally, the proposed approach is to be validated and implemented at a later stage of the research project.

Keywords—Arabic User Requirements, Use Case Model, MADA+TOKAN tool.

I. INTRODUCTION

Object-oriented methodologies are used for software systems development for the many benefits they provide like software reuse, reducing software development costs, to name just a few. Therefore, there is a need for development of automated tools that can help in constructing different components of an object-oriented software system.

A use case diagram shows a set of use cases and actors and their relationships. Use case diagrams address the static use case view of a system. These diagrams are especially important in organizing and modeling the behaviors of a system. This paper addresses the problem of generating a use case model from user requirements, written in Arabic, in a semi-automated approach. An Arabic natural language processing tool/software, namely MADA+TOKAN, is used to parse different statements of the user requirements, written in Arabic, to obtain lists of nouns, noun phrases, verbs, verb phrases, etc. that aid in finding potential actors and use cases. A set of steps that represent our approach for constructing a use case model is presented.

The rest of the paper is organized as follows: the section about related works presents the literature review and any related approaches; the section about constructing use cases describes the process of constructing use case models from Arabic user requirements; the section about validationpresents the validation and implementation of our proposed approach, and finally, the section about conclusion presents the main issues related to the proposed approach.

II. RELATED WORKS

Recently there is a great interest in automating software engineering activities. Many tools were developed to automate different activities of software systems development like normalizing relational database schemas, reverse engineering of relational database and generating the corresponding entityrelationship data model, ...etc. [1, 2]. In addition, many CASE tools were developed to aid in drawing different diagrams of UML. For example, Rational Rose is an object-oriented Unified Modeling Language (UML) software design tool intended for visual modeling and component construction of enterprise-level software applications [3]. Rational Unified Process (RUP) is an object-oriented Web-enabled program development methodology. [4].

More advanced tools were developed to automate software engineering activities that are more complicated than just aiding in drawing a UML diagram or checking its overall structure. Arman and Daghameen proposed a systematic approach that generates class diagrams from textual software requirements. They presented some steps to build a matrix that was used to obtain classes and their associations to generate class diagrams [5]. The same authors later developed a CASE tool, called, SDLCCASE tool that implemented their approach [6]. Kothari proposed an approach that can extract the basic elements for generating a class diagram from user requirements written in a clear way. The Natural Language Processing for Class (NLPC) can extract classes, data members and member functions from the given user requirements [7]. This approach