ABSTRACT

Formal grammars play a vital role in various data and text processing applications, including the development of compilers for software languages and Natural Language Processing (NLP) systems. From a formal grammar, tools like scanners, parsers, compilers, and interpreters can be built. Moreover, these tools and grammars have a broader range of applications, encompassing areas such as speech recognition, information security, genetic sequencing, and more. Collectively, software applications which involve grammar knowledge are referred to as grammar-aware software.

Numerous tools have been proposed and used in the realm of software language development. However, software developers often find these tools difficult to integrate and use in development environments like NetBeans, Eclipse and IntelliJ IDEA, especially when working on applications that depend on grammars and associated tools. Many language implementation tools exist as standalone systems, in practice, they cannot be integrated with ease into the work flow of grammar-aware software development. Several approaches have been explored to simplify the usage and integration of language implementation tools into software development environments. However, while these approaches are very effective in the hands of experts, they typically have steep learning curves, which can be an obstacle for those with only a modest understanding of language implementation.

This thesis addresses the problem of integrating grammars and associated tools into the domain of grammar-aware software development by consciously applying component-based software development principles to develop a software component framework that integrates into a general-purpose programming environment. This is accomplished through pragmatic and solution-oriented research using the design science research methodology. The research proposed and constructed valuable artifacts in the form of a prescriptive architecture and an instantiation of the PAMOJA software component framework.

During the first phase, by analyzing literature and utilizing expert opinions, the study conceptualized the research problem and validated component-based software development principles as a potential solution. The second phase began by establishing a conceptual framework to guide the design and development of the PAMOJA component framework. Subsequently designing an architecture encompassing design-requirements, guiding principles, specific decisions, a structural model, and the development platform, followed by the instantiation of the PAMOJA component framework. Design and development was an iterative process that further helped to improve the component framework. In the third phase, the study demonstrated the value of the PAMOJA component framework in facilitating grammar-aware software development. First, by building and testing the framework the study showed that such an artifact is technically feasible, that it works under certain assumptions, and that the application of component-based software principles has the potential to solve or mitigate the stated problem. Demonstration cases, which involved utilizing PAMOJA to construct a generic hybrid text/structure editor adaptable to a specified programming language and extensible tools for language processor education, provided additional validation. Finally, expert evaluation and technical action research showed that potential users in the grammar-aware community consider PAMOJA component framework to be a useful artifact for improving and facilitating grammar-aware software development.