APPLICATION OF MACHINE LEARNING FOR EXTRACTING PROGRAMMING LANGUAGE CONSTRUCTS FROM 4GL LEGACY CODE

W. S. A. Ilakshini C. Subasinghe

138237A

Degree of Master of Science

Department of Computer Science and Engineering

University of Moratuwa

Sri Lanka

May 2015

APPLICATION OF MACHINE LEARNING FOR EXTRACTING PROGRAMMING LANGUAGE CONSTRUCTS FROM 4GL LEGACY CODE

W. S. A. Ilakshini C. Subasinghe

138237A

Dissertation submitted in partial fulfilment of the requirements for the degree Master of Science in Computer Science specializing in Software Architecture

Department of Computer Science and Engineering

University of Moratuwa

Sri Lanka

May 2015

DECLARATION OF THE CANDIDATE & SUPERVISOR

I declare that this is my own work and this dissertation does not incorporate without acknowledgement any material previously submitted for a Degree or Diploma in any other University or institute of higher learning and to the best of my knowledge and belief it does not contain any material previously published or written by another person except where the acknowledgement is made in the text.

Also, I hereby grant to University of Moratuwa the non-exclusive right to reproduce and distribute my thesis/dissertation, in whole or in part in print, electronic or other medium. I retain the right to use this content in whole or part in future works (such as articles or books).

Signature: ..... Date: .....

Name: W. S. A. Ilakshini C. Subasinghe

The above candidate has carried out research for the Masters Dissertation under my supervision.

Signature: ..... Date: .....

Name of Supervisor: Dr. Amal Shehan Perera

ACKNOWLEDGEMENTS

I most gratefully acknowledge my appreciation to Dr. Shehan Perera for his acceptance, guidance and encouragement as my supervisor for this research, without which this work would not have been successful. I further extend my appreciation to Dr. Malaka Walpola and all my lecturers at the department for their encouragement and support to complete the research accurately in a timely manner.

I sincerely acknowledge the support my colleagues and seniors at work have provided, especially Mr. Mifraz Marzoon for his invaluable expertise as well as my friends who motivated me to complete the research.

Last but not least, I extend my deepest gratitude towards my parents and family for their tolerance and undivided love and support throughout the program.

ABSTRACT

With the progression and innovations of the Information Technology industry, computer systems have become not only a part of an organization but the heart of it that drives their daily routines and manages and tracks the entire business process for most enterprises and for decades Advanced Business Languages (ABL) have been evolving to provide successful economic solutions to drive these businesses. Progress 4GL (Fourth Generation Language) is one such Advanced Business Language where organizations have developed entire business process on for 30 years. However, with the advancement of Free and Open Sourced Software providing business solutions, some organizations using these legacy systems are looking for means of migration. Even though proprietary service providers exists for the migration process, organizations with decades old data are reluctant to use them for both cost and security reasons. Yet, in house development is also costly since ABL experts are very few and would require much time and effort to complete the process.

This research project is focused on a solution to develop such expert system that can interpret progress 4GL code to aid not only enterprises with migration but also engineers to learn and understand the language logic with ease. With the use of the Machine Learning technologies where research concerning modelling human thinking into machines are popular, this thesis provides a Proof of Concept for a methodology in which, an expert system can be created to read 4GL code, analyse the code, understand and infer the code logic and output the workflow in a graphical Flow Chart format. The prototype is run through several training 4GL programs to evaluate the implementation of the proposed theory. Current application proves to be successful for code with simple syntax and leaves room for further improvements to the system that can be enhanced to process 4GL's many complex and evolving constructs and also the possibility of translating to a different language.

Keywords: Expert Systems, Natural Language Processing, CLIPSJNI, Progress 4GL, mxGraph, Java-ML, Proparse

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

Abbreviation	Description
4GL	Fourth Generation Languages
ABL	Advanced Business Language
AI	Artificial Intelligence
API	Application Program Interface
ANTLR	Another Tool for Language Recognition
CHUI	Character User Interface
CLIPS	C Language Integrated Production System – Expert System Dev Tool
CRUD	Create, Read, Update and Delete Operations
EBMT	Example Based Machine Translation
EGL	Enterprise Generation Language
GUI	Graphical User Interface
NLP	Natural Language Processing
PSC	Progress Software Corporation
RBMT	Rule Based Machine Translation
SDL	Specification and Description Language
WEKA	Weikato Environment for Knowledge Analysis

LIST OF APPENDICES