AGENT-BASED SOLUTION FOR IMPROVING ABSTRACTS

Adhikari Mudiyanselage Tharindu Buddhika Adhikari

University of Moratuwa, Sri Lanka.
Electronic Theses & Dissertations
www.lib.mrt.ac.lk

(139103T)

Degree of Master of Science in Artificial Intelligence

Department of Computational Mathematics

University of Moratuwa

Sri Lanka

December 2015

AGENT-BASED SOLUTION FOR IMPROVING ABSTRACTS

Adhikari Mudiyanselage Tharindu Buddhika Adhikari

University of Moratuwa, Sri Lanka.
Electronic Theses & Dissertations
www.lib.mrt.ac.lk

(139103T)

Thesis submitted in partial fulfillment of the requirements for the degree of MSc in Artificial Intelligence

Department of Computational Mathematics

University of Moratuwa

Sri Lanka

December 2015

Declaration

I declare that this dissertation does not incorporate, without acknowledgment, any material previously submitted for a Degree or a Diploma in any University and to the best of my knowledge and belief, it does not contain any material previously published or written by another person or myself except where due reference is made in the text. I also hereby give consent for my dissertation, if accepted, to be made available for photocopying and for interlibrary loans, and for the title and summary to be made available to outside organization.

A.M.T.B. Adhikari ……..

Date: 31st December 2015

University of Moratuwa, Sri Lanka.
Electronic Theses & Dissertations
www.lib.mrt.ac.lk

Supervised by

Prof. Asoka Karunananda ……..

Date: 31st December 2015

Acknowledgements

I would like to express my sincere gratitude to my supervisor Prof. Asoka Karunananda for guiding me throughout this research project. He has given his fullest corporation to me whenever I sought for advice. Also, Prof. Karunnanda's teaching techniques helped me a lot to consider problems in different avenues. I must also thank all the members of the lecturer panel. During the period of one year and three months, these valuable lectures helped me a lot to think differently.

I would like to place my gratitude to my loving parents for always encouraging for me on higher studies. Also, I would like to appreciate Mr. Dimuthu and Ms. Nadeeka from the non-academic staff for their support that they provided throughout the period of the course.

For this research, I had to refer to many books and research papers as reference. I would like to thank all the authors of those publications. Also, my batch mates helped me a lot in many different ways. I would like to pay my gratitude for them. They made my life enjoyable during the course period.

University of Moratuwa, Sri Lanka.
Electronic Theses & Dissertations
www.lib.mrt.ac.lk

Last but not least, I would like to thank all my colleagues and others who are not mentioned, for all the support extended to me. My co-workers helped me a lot. So, that I can focus much more on the research and the MSc. Without their dedication the project would not have been successful.

Abstract

Writing abstracts in a comprehensive and meaningful manner is a challenge for any researcher. However an abstract includes limited set of verbs and standard phrases and other good practices of structuring the contents. A research has been conducted to develop an Agent-based Solution for Improving Abstracts. This solution is based on multi agent systems technology and natural language processing together with commonly used verb phrases and other good practices. The system has been developed with nine agents, namely, coordination agent, parser agent, problem agent, solution agent, conclusion agent, content agent, synonym agent, improvement agent and restructure agent. The coordination agent coordinates entire process. The parser agent identifies syntactic information of each sentence and prepares the contents of the abstract for further analysis. The problem agent ensures whether the research problem has been stated in the early part of the abstract and it's proportion within the abstract. The solution agent checks for the contents in terms of concepts such as hypothesis, methodology, approach, design, implementation, methods, theoretical framework, technology, hardware, software, and sampling based on the key words. The conclusion agent searches for concepts such as testing, evaluation, data analysis and statistical significance based on the key words. The content agent, improvement agent, synonym agent, and restructure agent are responsible to offer guidelines to modify and improving of the abstract. More importantly, these agents interact with each other and deliberate to reach consensus regarding a solution. For instance, problem agent and solution agent may agree on the proportion of respective contents within the abstract. Each agent has its own Ontology for deliberating with other agents. The Stanford CoreNLP Natural Language Processing Toolkit has been used to develop parser and JADE has been used for development of the entire multi agent system. The system has been developed with JAVA to run on Windows. It has been incrementally tested, and shown interesting results related to checking for completeness of the abstract in terms required materials and suggestion for improvements.

University of Moratuwa, Sri Lanka.
Electronic Theses & Dissertations
www.lib.mrt.ac.lk

Contents

Chapter 1 Introduction 14

1.1 Prolegomena 14

1.2 State of the art Artificial Intelligence field 15

1.3 Aim and Objectives 19

1.4 Background and Motivation 19

1.5 Problem in brief 19

1.6 Approach 19

1.7 Structure of the thesis 20

1.8 Summary 20

Chapter 2 Review of the State of the Art 21

2.1 Introduction 21

2.2 An Overview of Natural Language Processing by Computers 21

2.3 Natural Language Processing using Machine Learning Algorithms 23

2.4 Most Famous Research Areas in Natural Language Processing 24

2.4.1 Automatic summarization 24

2.4.2 Coreference resolution 24

2.4.3 Discourse analysis 24

2.4.4 Machine translation 25

2.4.5 Morphological segmentation 25

2.4.6 Named Entity Recognition (NER) 25

2.4.7 Natural language generation 25

2.4.8 Natural language understanding 25

2.4.9 Optical Character Recognition (OCR) 26

2.4.10 Part-of-speech tagging 26

2.4.11 Parsing 26

2.4.12 Question Answering 26

2.4.13 Relationship Extraction 26

2.4.14 Sentence Breaking 26

2.4.15 Sentiment Analysis 26

2.4.16 Speech Recognition 27

University of Moratuwa, Sri Lanka.
Electronic Theses & Dissertations
www.lib.mrt.ac.lk

2.4.17 Speech Segmentation 27
2.4.18 Topic Segmentation and Recognition 27
2.4.19 Word Segmentation 27
2.4.20 Word Sense Disambiguation 27
2.4.21 Information retrieval (IR) 27
2.4.22 Information extraction (IE) 28
2.4.23 Speech processing 28
2.5 Statistical Methods in Natural Language Processing Algorithms 28
2.6 Evaluation Methods of Natural Language Processing Algorithms 28
2.6.1 Intrinsic vs. Extrinsic Evaluation 29
2.6.2 Black-Box vs. Glass-Box Evaluation 30
2.6.3 Automatic vs. Manual Evaluation 30
2.7 An Overview of Multi Agent Systems (MAS) 32
2.8 Concept of Multi Agent Systems 32
2.9 Characteristics of Multi Agent Systems 33
2.10 Self-Organization and Self-Steering of Multi Agent Systems 33
2.11 Systems Paradigms of Multi Agent Systems 33
2.12 Properties of Multi Agent Systems 34
2.13 Research Areas of Multi Agent Systems 34
2.14 Frameworks of Multi Agent Systems 34
2.15 Real World Applications of Multi Agent Systems 35
2.16 Agent-based Modeling (ABM) Software of Multi Agent Systems 35
2.17 Advantages of a Multi-Agent Approach 36
2.18 An Overview of Automated Evaluation of Texts 36
2.19 Overview of Language Assessment 37
2.20 An Overview of Automated Assessment 38
2.21 Machine Learning in Automated Assessment 40
2.22 Other Approaches to Automated Assessment 43
2.23 Summary 44

Chapter 3 Cutting Edge Technologies Integrated 45

3.1 Introduction 45
3.2 Natural Language Processing 45
3.3 Bayesian approach 46

University of Moratuwa, Sri Lanka.
Electronic Theses & Dissertations
www.lib.mrt.ac.lk

3.4 Latent Semantic Analysis 47
3.5 Multi Agent Systems 48
3.6 Ontology 50
3.7 Summary 51

Chapter 4 Top-notch Approach to Improve the Abstracts 53
4.1 Introduction 53
4.2 Hypothesis 53
4.3 Inputs to System 54
4.4 Outputs from the System 54
4.5 Process of the System 54
4.6 Features of the System 55
4.7 Functional Requirements of the Proposed System 55
4.8 Non Functional Requirements of the Proposed System 55
4.9 Users of the Proposed System 55
4.10 Summary 56

Chapter 5 Analysis and Design of the Proposed Solution 57
5.1 Introduction 57
5.2 Analysis of the Proposed Design 57
5.3 Design Architecture Diagram 58
5.4 Multi Agent Swarm Intelligence 59
5.5 Checking the Overall Match between the Title and the Abstract 59
5.6 Check whether the Problem, Solution & Conclusion is mentioned in the Abstract 59
5.7 Check the Balance between the Problem, Solution & Conclusion in the Abstract 59
5.8 Check the Accuracy of the Content of the Abstract 60
5.9 Suggestions for Improvements 60
5.10 Ontology 60
5.11 Summary 66

Chapter 6 Implementation of the Novel Solution 67
6.1 Introduction 67
6.2 Overview of the Implementation 67

University of Moratuwa, Sri Lanka.
Electronic Theses & Dissertations
www.lib.mrt.ac.lk

6.3 Coordination Agent and Parser Agent 68
6.4 Problem Agent 68
6.5 Solution Agent 68
6.6 Conclusion Agent 68
6.7 Content Agent, Synonym Agent, Improvement Agent and Restructure Agent 69
6.8 Multi Agent System 70
6.9 Testing 70
6.10 The Stanford CoreNLP Natural Language Processing Toolkit 71
6.11 JADE Framework 72
6.12 Summary 73

Chapter 7 Evaluation of the Proposed Solution 74
7.1 Introduction 74
7.2 Evaluating the Overall System 74
7.3 Evaluating the Title and Abstract Overall Match 75
7.4 Evaluating the Checking whether the Problem is stated 75
7.5 Evaluating the Checking whether the Solution is stated 76
7.6 Evaluating the Checking whether the Conclusion is stated 77
7.7 Evaluating the Checking the Balance between Problem, Solution & Conclusion 78
7.8 Evaluating the Checking the Content 79
7.9 Evaluating the Suggesting the Synonyms 80
7.10 Evaluating the Suggesting the Improvements 81
7.11 Evaluating the Suggesting the Restructurings 82
7.12 Summary 93

Chapter 8 Conclusion and Further Work of the Research 94
8.1 Introduction 94
8.2 Conclusion 94
8.3 Further Work 94
8.4 Improving the Multi Agent System 95
8.5 Improving the Ontologies 95
8.6 Improving the Natural Language Processing Capabilities 96
8.7 Improving the User Friendliness 96

University of Moratuwa, Sri Lanka.
Electronic Theses & Dissertations
www.lib.mrt.ac.lk

8.8 Summary 96

Chapter 9 References 97

Appendix A: Software Code 100

A.1 Introduction 100
A.2 Coordination Agent 100
A.3 Parser Agent 100
A.4 Problem Agent 100
A.5 Solution Agent 100
A.6 Conclusion Agent 100
A.7 Content Agent 100
A.8 Synonym Agent 101
A.9 Improvement Agent 101
A.10 Restructure Agent 101
A.11 Abstract Concept 101
A.12 Message Concept 101
A.13 Parser Ontology 101
A.14 Problem Ontology 101
A.15 Solution Ontology 102
A.16 Conclusion Ontology 102
A.17 Content Ontology 102
A.18 Synonym Ontology 102
A.19 Improvement Ontology 102
A.20 Restructure Ontology 102
A.21 Abstract GUI 102
A.22 Text Area Output Stream 102

Appendix B: User Interfaces of the Proposed System 103

B.1 Introduction 103
B.2 Main User Interface of ASIA 103
B.3 JADE Remote Agent Management GUI 104
B.4 Open Abstract Dialog GUI 104
B.5 Abstract Open in the Text Area 105
B.6 Enter the Title of the Abstract in the Text Area 106

University of Moratuwa, Sri Lanka.
Electronic Theses & Dissertations
www.lib.mrt.ac.lk

B.7 Submit the Abstract to the System 107
B.8 Save the Suggested Improvements 108
B.9 Save the Agent Message Space 109
B.10 Clear the Agent Message Space 110
B.11 Clear the Suggested Improvements 111
B.12 Clear the Abstract and Title 112
B.13 Project Workspace in NetBeans IDE 113
B.14 Project Run Output in NetBeans IDE 114

University of Moratuwa, Sri Lanka.
Electronic Theses & Dissertations
www.lib.mrt.ac.lk

List of Figures

Figure 1.1: Acting humanly: The Turing Test approach 15

Figure 1.2: Thinking humanly: The cognitive modeling approach 16

Figure 1.3: Acting rationally: The rational agent approach 17

Figure 1.4: Thinking rationally: The "laws of thought" approach 18

Figure 2.1: Major research areas in Natural Language Processing 35

Figure 3.1: System Architecture of Stanford CoreNLP Toolkit 54

Figure 4.1: High Level Process Diagram 59

Figure 5.1: Design Architecture Diagram 64

Figure 6.1: System Architecture of Stanford CoreNLP Toolkit 72

Figure 7.1: Evaluation process of the proposed solution 75

Figure 7.2: Evaluation process of the Title and Abstract Overall Match 76

Figure 7.3: Evaluation process of the Checking whether the Problem is stated 77

Figure 7.4: Evaluation process of the Checking whether the Solution is stated 78

Figure 7.5: Evaluation process of the Checking whether the Conclusion is stated 79

Figure 7.6: Evaluation process of the Checking the Balance between Problem, Solution & Conclusion 80

Figure 7.7: Evaluation process of the Checking the Content 81

Figure 7.8: Evaluation process of the Suggesting the Synonyms 82

Figure 7.9: Evaluation process of the Suggesting the Improvements 83

Figure 7.10: Evaluation process of the Suggesting the Restructurings 84

Figure 8.1: Further Work of the Research 86

Figure B.1: Main User Interface of ASIA 95

Figure B.2: JADE Remote Agent Management GUI 95

Figure B.3: Open Abstract Dialog GUI 96

Figure B.4: Abstract Open in the Text Area 97

Figure B.5: Enter the Title of the Abstract in the Text Area 98

Figure B.6: Submit the Abstract to the System 99

Figure B.7: Save the Suggested Improvements 100

Figure B.8: Save the Agent Message Space 101

Figure B.9: Clear the Agent Message Space 102

Figure B.10: Clear the Suggested Improvements 103

University of Moratuwa, Sri Lanka.
Electronic Theses & Dissertations
www.lib.mrt.ac.lk

Figure B.11: Clear the Abstract and Title 104
Figure B.12: Project Workspace in NetBeans IDE 105
Figure B.13: Project Run Output in NetBeans IDE 105

University of Moratuwa, Sri Lanka.
Electronic Theses & Dissertations
www.lib.mrt.ac.lk

List of Tables

Table 1.1: Applications of Artificial Intelligence Technologies 18
Table 2.1: Applications of Multi Agent Systems 40
Table 3.1: Adapted Cutting Edge Technologies 56
Table 4.1: Inputs and Outputs of the System 60
Table 5.1: Process Description in the Proposed Design 67

University of Moratuwa, Sri Lanka.
Electronic Theses & Dissertations
www.lib.mrt.ac.lk

Chapter 1

Introduction

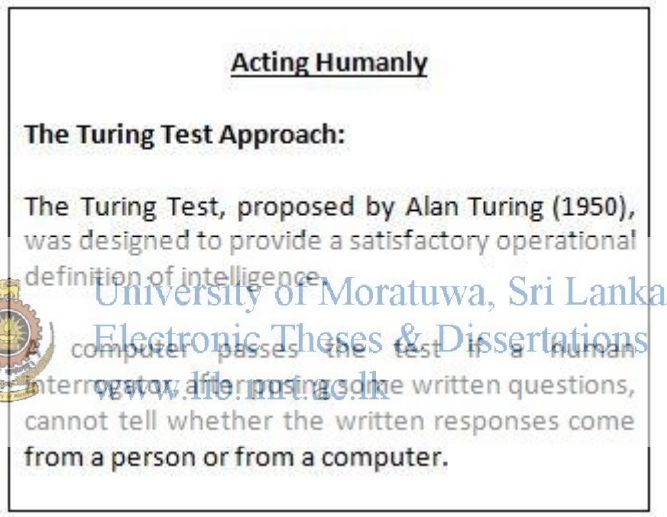
1.1 Prolegomena

In the course of the most recent six decades of time period Artificial Intelligence (AI) methods have demonstrated interesting ability to tackle diverse sorts of different complex genuine issues which couldn't be explained generally utilizing some other procedures. Specifically these present realities to a great degree of exceptionally complex frameworks are included in huge number of interconnected elements which are working in a circulated domain under flighty instabilities. With the rapidly increasing popularity in Artificial Intelligence, numerous complex intelligent techniques has been developed including Artificial Neural Networks (ANN), Genetic Algorithms (GA), Expert Systems (ES), Recommender Systems (RS), Natural Language Processing (NLP), Multi Agent Systems (MAS), etc… Considerably very large volume of literature on AI has shown the power of AI including Natural Language Processing, Multi Agent Systems, Artificial Neural Networks, Expert Systems, Genetic Algorithms, Ontology, etc… Quickly expanding prominence and the infiltration of AI innovations into wide range of branches of knowledge into the intricate genuine has made AI as an unmistakable innovation in the last century. Among other AI strategies multi agent systems and natural language processing has given powerful answers for critical problem solving. Energizing utilization of natural language processing and multi agent systems innovations have been accounted for in the mind boggling certifiable issue, for example, web search engines, speech recognition, machine translation, etc… Thus, having recognized creating answer for the improving abstracts as an intrinsically complex issue, this task has been directed to build up a multi agent based solution for improving the abstracts. In this connection, this chapter presents state of the art AI field, aim and objectives, background and motivation, problem in brief, novel approach to Agent-based Solution for Improving Abstracts and overall thesis.

University of Moratuwa, Sri Lanka.
Electronic Theses & Dissertations
www.lib.mrt.ac.lk

1.2 State of the art Artificial Intelligence field

Artificial Intelligence field right now covers an immense assortment of subfields, going from the general, for example, learning and perception up to the particular, for example, playing chess, proving complex mathematical theorems, writing poetry, storytelling, driving a vehicle on a very crowded street, and diagnosing cancers. Artificial Intelligence is genuinely general field which is important to performing any smart assignment [1]. Most of the famous definitions of Artificial Intelligence can be structured into four categories as indicated in the Figure 1.1, Figure 1.2, Figure 1.3 and Figure 1.4



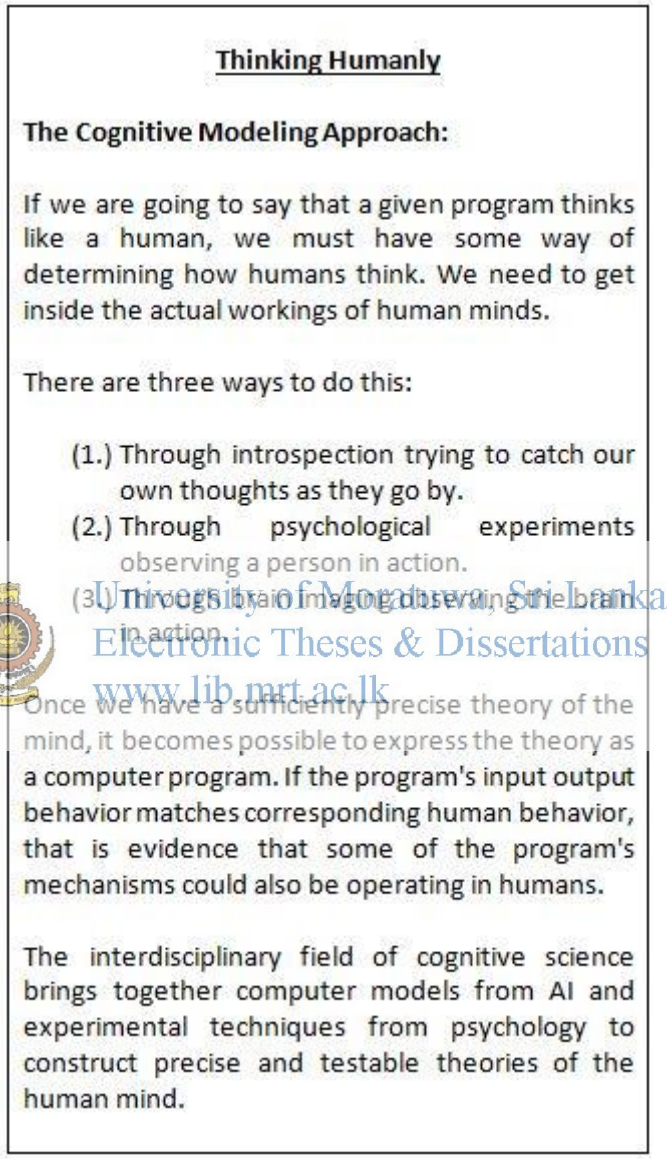
University of Moratuwa, Sri Lanka.
Electronic Theses & Dissertations
www.lib.mrt.ac.lk

Figure 1.1: Acting humanly: The Turing Test approach

The definitions on Thinking Humanly and Thinking Rationally are worried with points of view and thinking, while the ones on the Acting Humanly and Acting Rationally address conduct. The definitions on the Thinking Humanly and Acting Humanly measure accomplishment regarding loyalty to human execution, while the ones on the Thinking Rationally and Acting Rationally measure against a perfect execution measure, called rationality. A system can be ordered as rational if the system does the "proper right thing," given that what the system knows [1].

Every one of the four here said ways to deal with construct Artificial Intelligent systems have been taken after, each by distinctive researchers with various types of methods. The human focused methodology which is thinking humanly and acting

humanly must be in the part experimental sciences, which are including perceptions and speculations on human conduct. The realists approach which is on thinking rationally and acting rationally includes a mix of arithmetic and engineering. The different research groups from different parts of the world have both helped and also disparaged one another on this [1].

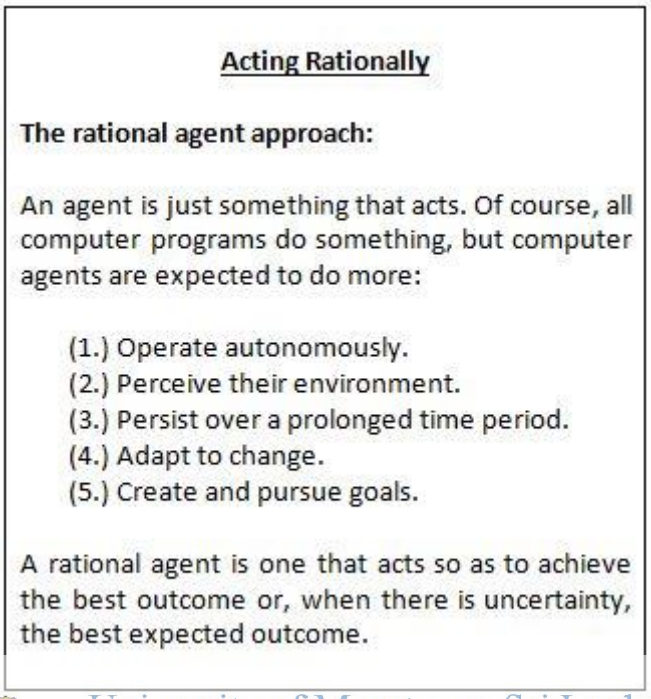


University of Moratuwa, Sri Lanka.
Electronic Theses & Dissertations
www.lib.mrt.ac.lk

Figure 1.2: Thinking humanly: The cognitive modeling approach

Deep Blue computer program from IBM turned into the first computer program to overcome the best on the planet Garry Kasparov in a chess match when it bested the world chess champion by a score of 3.5 to 2.5 in a show match. The world chess champion Garry Kasparov expressed that he felt "new kind of intelligence" in all cases from him while this match in the middle of him and the IBM Deep Blue

computer program. The estimation of IBM's stock in the share market has been expanded by $18 billion after this noteworthy defining moment [1].



University of Moratuwa, Sri Lanka.
Electronic Theses & Dissertations
www.lib.mrt.ac.lk

Figure 1.3: Acting rationally: The rational agent approach

Consistently, artificial intelligence learning algorithms group more than a billion email messages as spam, sparing the valuable time of the recipient from wasting time erasing what, for some email clients, could involve 80% or 90% of all email messages, if not characterized away by these artificial intelligence learning algorithms. The spammers are continuously upgrading their methodologies. Along these lines, it is troublesome for a static modified technique to keep this up, and artificial intelligence learning algorithms work fine in this sort of circumstances [1].

Amid the time of Persian Gulf War in 1991, United State security forces deployed a Dynamic Analysis and Replanning Tool (DART), to do mechanized logistics arranging and booking for transportation. This included up to around 50,000 automobiles, cargo, and individuals at a given purpose of time, furthermore this DART device needed to represent starting points, destinations, routes, and conflict resolution among all the considered parameters. The Artificial Intelligence planning techniques created in hours an impeccable arrangement that would have been bring numerous weeks with more seasoned utilized systems [1].

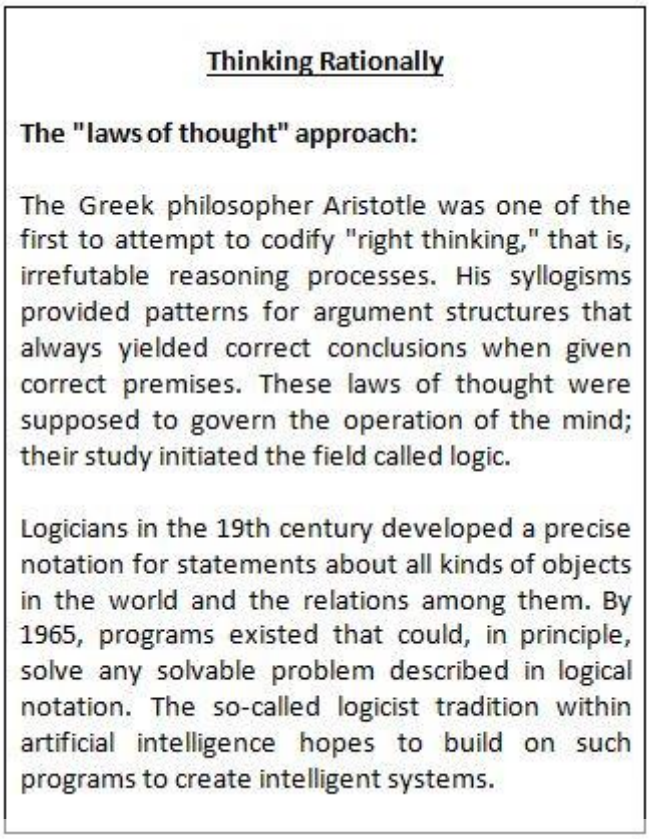


Figure 1.4: Thinking rationally: The "laws of thought" approach

University of Moratuwa, Sri Lanka
Electronic Theses & Dissertations
www.lib.mrt.ac.lk

The Defense Advanced Research Project Agency (DARPA) expressed this single artificial intelligence application paid back more than DARPA's around 30 year downright interest of total investment in artificial intelligence research projects [1].

Application Industry	Artificial Intelligence Technology
Computer Science	Machine Learning Algorithms
Finance	Artificial Neural Networks
Hospitals and Medicine	Artificial Neural Networks
Heavy Industry	Robotics
Online and Telephone Customer Service	Natural Language Processing
Transportation	Fuzzy Logic
Telecommunications Maintenance	Heuristic Search Algorithms
Toys and Games	Search Algorithms
Music	Genetic Algorithms
Aviation	Expert Systems
News, Publishing and Writing	Natural Language Processing

Table 1.1: Applications of Artificial Intelligence Technologies

1.3 Aim and Objectives

The aim of this project is to develop an Agent-based Solution for Improving Abstracts. In order to reach this aim the following objectives are defined.

1. To critically study the improving abstracts with a view to identify current practices and issues.
2. Critically analyze the existing solution for improving abstracts with the view to define the research problem and possible technology.
3. In depth study about improving abstracts and its applications
4. Design and implement Agent-based Solution for Improving Abstracts.
5. Evaluate the improving the abstracts using the real world scenario.

1.4 Background and Motivation

Abstract gives an overview of the whole research. So, writing abstracts in a comprehensive and meaningful manner is a challenge task for any researcher. However an abstract includes limited set of verbs and standard phrases and other good practices of structuring the contents. This research has been conducted to develop an Agent-based Solution for Improving Abstracts. This solution is based on natural language processing together with commonly used verb phrases and other good practices. The system suggests the improvement with relevant comments for the user to finalize the abstract. So, this system helps the writers to improve their abstracts in a comprehensive manner.

University of Moratuwa, Sri Lanka.
Electronic Theses & Dissertations
www.lib.mrt.ac.lk

1.5 Problem in brief

Improving abstracts has been seriously affected by the topic and the content of the abstract. Lack of proper expertise in automated assessment solutions for improving the abstracts have resulted in malfunctioning of whole improving the abstracts process leading to dissatisfaction of both the writers and the readers.

1.6 Approach

Improving abstracts has been seriously affected by the limitation of domain expertise. However an abstract includes limited set of verbs and standard phrases and other good

practices of structuring the contents. This suggested solution is based on natural language processing together with commonly used verb phrases and other good practices to assist the writer to improve the abstracts.

1.7 Structure of the thesis

Rest of the thesis is organized as follows. Chapter 2 critically reviews the domain of Agent-based Solution for Improving Abstracts by highlighting current solutions, practices, technologies, limitations defining the research problem. Chapter 3 describes the essentials of Agent-based Solution for Improving Abstracts technology showing its relevance to Agent-based Solution for Improving Abstracts domain. Chapter 4 presents our novel approach to Agent-based Solution for Improving Abstracts with NLP technology. Chapter 5 is on the design of NLP for Agent-based Solution for Improving Abstracts. Chapter 6 contains details of implementation of the NLP solution for Agent-based Solution for Improving Abstracts. Chapter 7 illustrates a real world application of the novel approach. Chapter 8 reports on the evaluation of the new solution by explaining the evaluation strategy, participants, data collection, representation and analysis. Chapter 9 concludes the outcome of the research with the note on further work.

University of Moratuwa, Sri Lanka.
Electronic Theses & Dissertations
www.lib.mrt.ac.lk

1.8 Summary

This chapter describes the full picture of the whole research project showing research problem objectives, hypothesis and novel solution. Next chapter will be on literature review of solution for improving abstracts, practices, technologies and issues with a view to define the research problem.

Chapter 2

Review of the State of the Art

2.1 Introduction

Recent increasing popularity and the penetration of NLP technologies into wide spectrum of subject areas has speed up the developments of automated solutions which uses NLP for the assessments.

2.2 An Overview of Natural Language Processing by Computers

Natural Language Processing (NLP) is a field of computer engineering, computerized reasoning, and computational phonetics worried with the cooperation in the middle of computers and human natural languages. In that capacity, NLP is identified with the zone of human–computer cooperation. Numerous difficulties in NLP include natural language understanding, that is, empowering computers to get significance from human or common language data, and others include characteristic language era [2].

University of Moratuwa, Sri Lanka.
Electronic Theses & Dissertations
www.lib.mrt.ac.lk

The historical backdrop of NLP by and large begins in the 1950s, in spite of the fact that work can be found from before periods. In 1950, Alan Turing distributed an article titled "Computing Machinery and Intelligence" which proposed what is presently called the Turing test as a foundation of insight [4].

Utilizing no data about human thought or feeling, ELIZA some of the time gave a startlingly human-like association [6]. Whenever the "patient" surpassed the little information base, ELIZA may give a nonexclusive reaction, for instance, reacting to "My head harms" with "Why do you say your head harms?"

Amid the 1970s numerous computer engineers started to compose 'applied ontologies', which organized true data into computer reasonable information.

A percentage of the soonest utilized machine learning algorithm, for example, choice trees, created frameworks of hard if-then principles like existing transcribed standards. On the other hand, Grammatical form labeling presented the utilization of Shrouded

Markov Models to NLP, and progressively, research has concentrated on measurable models, which make delicate, probabilistic choices in light of joining genuine esteemed weights to the elements making up the information. The store language models whereupon numerous discourse acknowledgment frameworks now depend are samples of such measurable models. Such models are by and large more vigorous when given new info, particularly include that contains blunders (as is exceptionally regular for certifiable information), and deliver more dependable results when incorporated into a bigger framework involving various subtasks [9].

A hefty portion of the striking early triumphs happened in the field of machine interpretation, because of work at IBM Research, where progressively more entangled factual models were produced. Be that as it may, most different frameworks relied on upon corpora particularly produced for the undertakings executed by these frameworks, which was (and frequently keeps on being) a noteworthy confinement in the achievement of these frameworks. Accordingly, a lot of examination has gone into strategies for all the more viably gaining from constrained measures of information [10].

University of Moratuwa, Sri Lanka.
Electronic Theses & Dissertations
www.lib.mrt.ac.lk

Late research has progressively centered on unsupervised and semi-regulated learning algorithm. Such algorithm can gain from information that has not been hand-commented with the wanted replies, or utilizing a mix of expounded and non-explained information. For the most part, this undertaking is substantially more troublesome than managed learning, and normally delivers less exact results for a given measure of information. In any case, there is a colossal measure of non-clarified information accessible (counting, in addition to other things, the whole substance of the Internet), which can regularly compensate for the mediocre results [11].

An ISO subcommittee is working with a specific end goal to straightforwardness. Some ISO gauges are as of now distributed yet the majority of them are under development, for the most part on dictionary representation (see LMF), annotation, and information classification registry [12].

Human-level characteristic language preparing, be that as it may, is an AI-complete issue. That is, it is proportionate to taking care of the focal computerized reasoning

issue—production computers as shrewd as individuals, or solid AI. NLP's future is along these lines fixing intently to the improvement of AI when all is said in don [13].

2.3 Natural Language Processing using Machine Learning Algorithms

Advanced NLP algorithms depend on machine adapting, particularly factual machine learning. The worldview of machine taking in is not the same as that of earlier endeavors at language preparing [14]. Earlier executions of language preparing errands commonly included the immediate hand coding of huge arrangements of standards [15]. The machine-learning worldview calls rather to use general learning algorithm — frequently, despite the fact that not generally, grounded in measurable deduction — to consequently learn such guidelines through the investigation of expansive corpora of common true cases. A corpus (plural, "corpora") is an arrangement of records (or some of the time, individual sentences) that have been hand-commented with the right values to be learned [16].

University of Moratuwa, Sri Lanka.
Electronic Theses & Dissertations
www.lib.mrt.ac.lk

A percentage of the most punctual utilized algorithm, for example, choice trees, delivered frameworks of hard if-then guidelines like the frameworks of manually written principles that were then normal. Progressively, then again, research has concentrated on factual models, which make delicate, probabilistic choices taking into account connecting genuine esteemed weights to every data highlight. These have a wide range of conceivable answers as opposed to one and only, delivering more solid results when such a model is incorporated as a segment of a bigger framework.

Frameworks in light of machine-learning algorithm have numerous focal points over hand-created rules:

The learning methods utilized amid machine adapting naturally concentrate on the most widely recognized cases, while when composing tenets by hand it is frequently not evident at all where the exertion ought to be coordinated.

For the most part, taking care of such data smoothly with transcribed principles — or all the more for the most part, making frameworks of written by hand decides that settles on delicate choices — is amazingly troublesome, blunder inclined and tedious.

Frameworks in light of consequently taking in the guidelines can be made more precise basically by supplying more information. In any case, frameworks in light of manually written principles must be made more exact by expanding the many-sided quality of the tenets, which is an a great deal more troublesome errand. Specifically, there is a point of confinement to the multifaceted nature of frameworks in view of hand-made guidelines, past which the frameworks turn out to be more unmanageable.

2.4 Most Famous Research Areas in Natural Language Processing

The accompanying is a rundown of the absolute most ordinarily scrutinized undertakings in NLP.

2.4.1 Automatic summarization

Produce an intelligible outline of a piece of content. Regularly used to give outlines of content of a referred to sort, For example, articles in the budgetary area of a daily paper.

University of Moratuwa, Sri Lanka.
Electronic Theses & Dissertations
www.lib.mrt.ac.lk

2.4.2 Coreference resolution

Given a sentence or bigger piece of content, figure out which words ("notice") allude to the same articles ("elements"). Anaphora determination is a particular sample of this undertaking, and is particularly worried with coordinating up pronouns with the things or names that they allude to. The broader errand of coreference determination likewise incorporates distinguishing purported "spanning connections" including alluding expressions. For instance, in a sentence, for example, "He went into John's home through the front entryway", "the front entryway".

2.4.3 Discourse analysis

This rubric incorporates various related assignments. One undertaking is distinguishing the talk structure of associated content, i.e. the way of the talk connections between sentences (e.g. elaboration, clarification, contrast).

2.4.4 Machine translation

Naturally decipher content starting with one human language then onto the next. This is a standout amongst the most troublesome issues, and is an individual from a class of issues conversationally termed "AI-complete", i.e. requiring the majority of the distinctive sorts of information that people have (linguistic use, semantics, actualities about this present reality, and so forth.) keeping in mind the end goal to settle legitimately.

2.4.5 Morphological segmentation

English has genuinely basic morphology, particularly inflectional morphology, and accordingly it is regularly conceivable to overlook this undertaking altogether and essentially display every single conceivable type of a word (e.g. "open, opens, opened, opening") as discrete words. In languages, for example, Turkish or Manipuri [4] a very agglutinated Indian language, then again, such a methodology is impractical, as every lexicon section has a large number of conceivable word frames.

University of Moratuwa, Sri Lanka.
Electronic Theses & Dissertations
www.lib.mrt.ac.lk

2.4.6 Named Entity Recognition (NER)

Note that, in spite of the fact that upper casing can help in perceiving named elements in languages, for example, English, this data can't help in deciding the kind of named element, and regardless is frequently off base or inadequate. For instance, the first expression of a sentence is likewise promoted, and named substances frequently compass a few words, just some of which are promoted.

2.4.7 Natural language generation

Change over data from computer databases into coherent human language.

2.4.8 Natural language understanding

Proselyte pieces of content into more formal representations, for example, first-arrange rationale structures that are simpler for computer projects to control.

2.4.9 Optical Character Recognition (OCR)

Determine the corresponding text, given printed text representing in an image.

2.4.10 Part-of-speech tagging

Given a sentence, decide the grammatical form for every word. Numerous words, particularly regular ones, can serve as different parts of discourse. Such intonation is not promptly passed on through the elements utilized inside of the orthography to pass on expected importance.

2.4.11 Parsing

Decide the parse tree (syntactic examination) of a given sentence.

2.4.12 Question Answering

Given a human-language inquiry, decide its answer. Run of the mill inquiries have a particular right reply, (for example, "What is the capital of Canada?"), however some of the time open-finished inquiries are likewise viewed as, (for example, "What is the significance of life?"). Late works have taken a gander at considerably more mind boggling inquiries [6].

University of Moratuwa, Sri Lanka.
Electronic Theses & Dissertations
www.lib.mrt.ac.lk

2.4.13 Relationship Extraction

Given a piece of content, recognize the connections among named substances.

2.4.14 Sentence Breaking

Sentence breaking is otherwise called sentence limit disambiguation (e.g. stamping contractions).

2.4.15 Sentiment Analysis

Remove subjective data for the most part from an arrangement of archives, regularly utilizing online audits to decide "extremity" about particular items. It is particularly

helpful for distinguishing patterns of popular sentiment in the online networking, with the end goal of showcasing.

2.4.16 Speech Recognition

Given a sound clasp of a man or individuals talking, decide the printed representation of the discourse. This is the inverse of content to discourse and is one of the greatly troublesome issues informally termed "AI-complete". In regular discourse there are not really any stops between progressive words, and in this manner discourse division is a fundamental subtask of discourse acknowledgment.

2.4.17 Speech Segmentation

Given a sound clasp of a man or individuals talking, separate it into words. A subtask of discourse acknowledgment and regularly gathered with it.

2.4.18 Topic Segmentation and Recognition

University of Moratuwa, Sri Lanka.
Electronic Theses & Dissertations
www.lib.mrt.ac.lk

Given a piece of content, separate it into portions each of which is dedicated to a point, and distinguish the subject of the fragment.

2.4.19 Word Segmentation

Separate a lump of persistent content into partitioned words. For a language like English, this is genuinely unimportant, since words are normally isolated by spaces.

2.4.20 Word Sense Disambiguation

Numerous words have more than one significance; we need to choose the importance which bodes well in setting.

2.4.21 Information retrieval (IR)

This is worried with putting away, looking and recovering data. It is a different field inside of software engineering (closer to databases), however IR depends on some

NLP strategies (for instance, stemming). Some ebb and flow examination and applications look to cross over any barrier in the middle of IR and NLP.

2.4.22 Information extraction (IE)

This is worried when all is said in done with the extraction of semantic data from content. This spreads undertakings, for example, named substance acknowledgment, Coreference determination, relationship extraction, and so forth.

2.4.23 Speech processing

This spreads discourse acknowledgment, content to-discourse and related undertakings.

Different undertakings include: Local Language Distinguishing proof, Stemming, Content disentanglement, Content to-discourse, Content sealing, Regular language hunt, Inquiry development, Robotized exposition scoring, Truecasing.

University of Moratuwa, Sri Lanka.
Electronic Theses & Dissertations
www.lib.mrt.ac.lk

2.5 Statistical Methods in Natural Language Processing Algorithms

Measurable natural language preparing uses stochastic, probabilistic, and factual systems to determine a percentage of the challenges talked about above, particularly those which emerge on the grounds that more drawn out sentences are exceedingly vague when handled with sensible punctuations, yielding thousands or a large number of conceivable examinations. Strategies for disambiguation frequently include the utilization of corpora and Markov models. The ESPRIT Venture P26 (1984 - 1988), drove by CSELT, investigated the issue of discourse acknowledgment looking at learning based methodology and measurable ones: the picked result was a totally factual model [7]. The innovation for factual NLP comes for the most part from machine learning and information mining, both of which are fields of counterfeit consciousness that include gaining from information.

2.6 Evaluation Methods of Natural Language Processing Algorithms

The objective of NLP assessment is to quantify one or more characteristics of an algorithm or a framework, so as to figure out if (or to what degree) the framework

answers the objectives of its creators, or addresses the issues of its clients. Research in NLP assessment has gotten extensive consideration, on the grounds that the meaning of appropriate assessment criteria is one approach to determine correctly a NLP issue, going subsequently past the unclearness of assignments characterized just as language comprehension or language era. An exact arrangement of assessment criteria, which incorporates primarily assessment information and assessment measurements, empowers a few groups to contrast their answers with a given NLP issue.

The main assessment crusade on composed writings is by all accounts a battle devoted to message understanding in 1987 (Bed 1998). At that point, the Parseval/GEIG task looked at expression structure sentence structures (Dark 1991). A progression of battles inside of Tipster undertaking were acknowledged on assignments like rundown, interpretation and looking (Hirschman 1998). In 1994, in Germany, the Morpholympics analyzed German taggers. At that point, the Senseval and Romanseval battles were directed with the targets of semantic disambiguation. In 1996, the Radiance battle analyzed syntactic parsers in four unique languages (English, French, German and Italian). In France, the Elegance undertaking looked at an arrangement of 21 taggers for French in 1997 (Adda 1999). In Italy, the EVALITA battle was directed in 2007, 2009, 2011, and 2014 [11] to look at different NLP and discourse apparatuses for Italian - EVALITA site. In France, inside of the ANR-Entry venture (end of 2007), 10 parsers for French were thought about - section site.

University of Moratuwa, Sri Lanka.
Electronic Theses & Dissertations
www.lib.mrt.ac.lk

Contingent upon the assessment systems, various refinements are customarily made in NLP assessment.

2.6.1 Intrinsic vs. Extrinsic Evaluation

Inherent assessment considers a disconnected NLP framework and portrays its execution chiefly regarding a best quality level result, pre-characterized by the evaluators. Extraneous assessment, likewise called assessment being used considers the NLP framework in a more intricate setting, either as an inserted framework or serving an exact capacity for a human client. The extraneous execution of the framework is then described regarding its utility as for the general undertaking of the intricate framework or the human client. An inborn assessment would run the POS

tagger on some named information, and think about the framework yield of the POS tagger to the best quality level (right) yield. An extraneous assessment would run the parser with some different POS tagger, and afterward with the new POS tagger, and think about the parsing exactness.

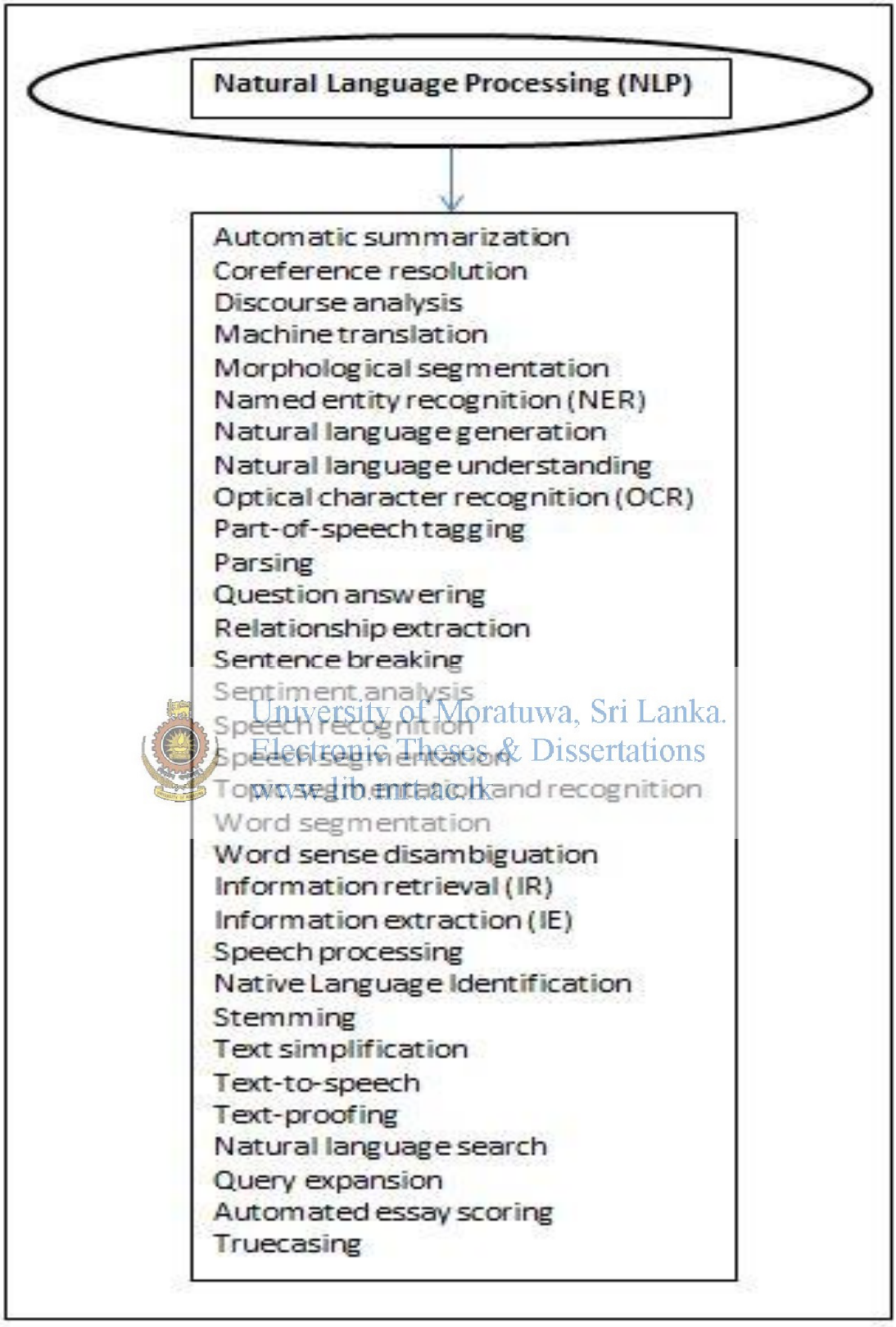
2.6.2 Black-Box vs. Glass-Box Evaluation

Discovery assessment obliges one to run a NLP framework on a given information set and to gauge various parameters identified with the nature of the procedure (speed, unwavering quality, asset utilization) and, above all, to the nature of the outcome (e.g. the exactness of information annotation or the loyalty of an interpretation). Glass-box assessment takes a gander at the outline of the framework, the algorithm that are actualized, the semantic assets it utilizes (e.g. vocabulary size), and so forth. Given the many-sided quality of NLP issues, it is regularly hard to foresee execution just on the premise of glass-box assessment; however this sort of assessment is more instructive as for mistake investigation or future improvements of a framework.

University of Moratuwa, Sri Lanka.
Electronic Theses & Dissertations
www.lib.mrt.ac.lk

2.6.3 Automatic vs. Manual Evaluation

As a rule, programmed strategies can be characterized to assess a NLP framework by contrasting its yield and the best quality level (or craved) one. Despite the fact that the expense of delivering the best quality level can be very high, programmed assessment can be rehashed as frequently as required without much extra expenses (on the same information). On the other hand, for some NLP issues, the meaning of a best quality level is a perplexing assignment, and can demonstrate unimaginable when between annotator understanding is deficient. Manual assessment is performed by human judges, which are told to appraise the nature of a framework, or frequently of an example of its yield, taking into account various criteria. Despite the fact that, because of their phonetic ability, human judges can be considered as the reference for various language preparing assignments, there is likewise extensive variety over their evaluations. This is the reason programmed assessment is now and then alluded to as target assessment, while the mankind gives off an impression of being more "subjective."



University of Moratuwa, Sri Lanka.
Electronic Theses & Dissertations
www.lib.mrt.ac.lk

Figure 2.1: Major research areas in Natural Language Processing

2.7 An Overview of Multi Agent Systems (MAS)

A multi-agent system (MAS) is a mechanized framework made out of different interfacing keen agents inside of a situation. Multi-agents systems can be utilized to take care of issues that are troublesome or unthinkable for an individual agents or a solid framework to tackle. Knowledge may incorporate some methodic, utilitarian, procedural methodology, algorithmic hunt or support learning. Despite the fact that there is extensive cover; a multi-agents system is not generally the same as an agents based model (ABM). The objective of an ABM is to hunt down illustrative understanding into the aggregate conduct of agents (which don't as a matter of course should be "savvy") obeying straightforward guidelines, normally in common frameworks, instead of in taking care of particular commonsense or building issues. The phrasing of ABM has a tendency to be utilized all the more frequently as a part of the sciences, and MAS in designing and technology [1].

In manmade brainpower research, agents based frameworks innovation has been hailed as another worldview for conceptualizing, outlining, and actualizing programming frameworks. Agents are refined computer programs that demonstration self-rulingly in the interest of their clients, crosswise over open and dispersed situations, to unravel a developing number of complex issues. Progressively, in any case, applications require numerous agents that can cooperate. A multi-agents system (MAS) is an inexactly coupled system of programming agents that interface to take care of issues that are past the individual limits or information of every issue solver.

University of Moratuwa, Sri Lanka.
Electronic Theses & Dissertations
www.lib.mrt.ac.lk

2.8 Concept of Multi Agent Systems

Multi-agents systems comprise of agents and their surroundings. Regularly multi-agents systems examination alludes to programming agents.

Agents situations can likewise be sorted out as indicated by different properties like: availability (in the event that it is conceivable to accumulate complete data about nature), determinism (if an activity performed in the earth causes a distinct impact), elements (what number of elements impact the earth in the occasion), discreteness (whether the quantity of conceivable activities in the earth is limited), episodicity (whether agents activities in certain time periods impact other periods) [8] and

dimensionality (whether spatial attributes are essential components of the earth and the agents considers space in its choice making) [9]. Agent's activities in the earth are commonly intervened by means of a fitting middleware. This middleware offers a top of the line outline reflection for multi-agents systems, giving intends to oversee asset access and agents coordination [10].

2.9 Characteristics of Multi Agent Systems

The agents in a multi-agent system have a few imperative characteristics [11].

Autonomy: the agents are at any rate mostly free, mindful, self-governing.

Local views: no agent has a full worldwide perspective of the framework, or the framework is excessively unpredictable for an agent, making it impossible to make commonsense utilization of such information.

Decentralization: there is no assigned controlling agents (or the framework is successfully decreased to a solid system) [12].

University of Moratuwa, Sri Lanka.
Electronic Theses & Dissertations
www.lib.mrt.ac.lk

2.10 Self-Organization and Self-Steering of Multi Agent Systems

Multi-agents systems can show self-association and also self-controlling and other control ideal models and related complex practices notwithstanding when the individual procedures of every one of their agents are straightforward. At the point when agents can share learning utilizing any concurred language, inside of the requirements of the framework's correspondence convention, the methodology may prompt a typical change. Case languages are Learning Inquiry Control Language (KQML) or FIPA's Agents Correspondence Language (ACL).

2.11 Systems Paradigms of Multi Agent Systems

Numerous Multi Agent Systems are actualized in computer simulations, venturing the framework through discrete "time steps". The MAS segments impart ordinarily utilizing a weighted solicitation network and a weighted reaction grid. A test reaction

contract plan is normal in MAS frameworks, where likewise considering different parts, developing "contracts", and the limitation sets of the segment algorithm.

Another worldview normally utilized with MAS frameworks is the pheromone, where parts "leave" data for different segments "next in line" or "in the region".

2.12 Properties of Multi Agent Systems

MAS, likewise alluded to as "self-composed frameworks", tend to locate the best answer for their issues "without intercession". There is high closeness here to physical wonders, for example, vitality minimizing, where physical articles tend to achieve the least vitality conceivable inside of the physically obliged world. For instance: huge numbers of the autos entering a city in the morning will be accessible for leaving that same city at night.

The principle highlight which is accomplished when creating multi-agents systems is adaptability, since a multi-agents system can be added to, adjusted and recreated, without the requirement for nitty gritty revamping of the application [13]. The frameworks likewise have a tendency to avoid engendering of issues, self-recuperate and be blame tolerant, for the most part because of the repetition of segments.

University of Moratuwa, Sri Lanka.
Electronic Theses & Dissertations
www.lib.mrt.ac.lk

2.13 Research Areas of Multi Agent Systems

The investigation of multi-agents systems is "worried with the advancement and examination of refined AI critical thinking and control architectures for both single-agents and numerous agents systems"[14].

2.14 Frameworks of Multi Agent Systems

While impromptu multi-agent systems are regularly made sans preparation by agents and engineers, a few structures have emerged that actualize normal gauges, (for example, the FIPA agents framework stages and correspondence languages). These systems spare designers time furthermore help in the institutionalization of MAS improvement. One such formative system for mechanical autonomy is given in [18]. See additionally Correlation of agents based displaying programming.

2.15 Real World Applications of Multi Agent Systems

Multi-agent systems are connected in this present reality to graphical applications, for example, computer amusements. Agents systems have been utilized as a part of films [19]. They are additionally utilized for facilitated safeguard frameworks. Different applications incorporate transportation, logistic, design, GIS and in addition in numerous different fields. It is generally being supported for use in systems administration and versatile advances, to accomplish programmed and element burden adjusting, high adaptability, and self-recuperating systems.

Application Area	Application Details
Computer Games	Graphics Processing
Defense Systems	Coordinate Systems
Transportation	Logistics Planning
Networking	Dynamic Load Balancing
Mobile Technologies	Achieve High Scalability

University of Moratuwa, Sri Lanka.
Electronic Theses & Dissertations
www.lib.mrt.ac.lk

Table 2.1: Applications of Multi Agent Systems

2.16 Agent-based Modeling (ABM) Software of Multi Agent Systems

In the most recent couple of years, the agent-based modeling (ABM) group has added to a few handy agents based demonstrating toolboxes that empower people to create agents based applications. More such toolboxes are appearing, and each tool stash has an assortment of attributes. A few people have made endeavors to contrast toolboxes with one another.

2.17 Advantages of a Multi-Agent Approach

A MAS has the accompanying favorable circumstances over a solitary agents or brought together approach:

A MAS conveys computational assets and abilities over a system of interconnected agents. Though an incorporated framework may be tormented by asset restrictions, execution bottlenecks, or basic disappointments, MAS is decentralized and consequently does not experience the ill effects of the "single purpose of disappointment" issue connected with brought together frameworks.

A MAS takes into consideration the interconnection and interoperation of numerous current legacy frameworks. By building an agents wrapper around such frameworks, they can be incorporated into an agent's society.

A MAS models issues as far as self-sufficient collaborating segment agents, which is turned out to be a more common method for speaking to assignment portion, group arranging, client inclinations, open situations, et cetera.

University of Moratuwa, Sri Lanka.
Electronic Theses & Dissertations
www.lib.mrt.ac.lk

A MAS productively recovers, channels, and comprehensively organizes data from sources that are spatially dispersed.

A MAS gives arrangements in circumstances where mastery is spatially and transiently conveyed.

A MAS upgrades general framework execution, particularly along the measurements of computational productivity, unwavering quality, extensibility, strength, practicality, responsiveness, adaptability and reuse.

2.18 An Overview of Automated Evaluation of Texts

Automated Essay Scoring (AES) is characterized as the technology innovation that assesses and scores the written composition. AES frameworks are created to help educators in low-stakes classroom appraisal and in addition testing organizations and states in extensive scale high-stakes evaluation. They are for the most part used to

overcome time, expense, unwavering quality, and generalizability issues in composing evaluation [1].

Various studies have been directed to survey the exactness and dependability of the AES frameworks concerning composing appraisal. The aftereffects of a few AES studies reported high understanding rates between AES frameworks and human raters [2].

AES frameworks have been condemned for lacking human collaboration, defenselessness to tricking, and their requirement for a huge corpus of test content to prepare the framework. In spite of its shortcomings, AES keeps drawing in the consideration of state funded schools, colleges, testing organizations, analysts and instructors [3].

2.19 Overview of Language Assessment

University of Moratuwa, Sri Lanka.
Electronic Theses & Dissertations
www.lib.mrt.ac.lk

English is utilized as a most widely used language, and English capability is a vital aptitude throughout today's worldwide business market. It is thusly imperative, in an inexorably globalized environment, to have the capacity to show one's English-language abilities by means of equitably surveyed capabilities [5].

Language evaluation gives the way to recognizing and measuring a singular's language aptitudes, capacities, and capability level. There is an extensive variety of appraisals accessible changing in arrangement, thoroughness, and prerequisites, managed on paper or on computer. Addressing is a standout amongst the most well-known appraisal instruments and may utilize various evaluation systems; for instance, certain sorts of inquiries require a particular foreordained answer, for example, different decision inquiries, genuine or-false inquiries, in with no reservations the-clear inquiries and built short reactions. Others may concentrate on developed composed reactions, including prompts inspiring free-content answers, for example, expositions and reports. Each is intended to respond different learning focuses on; these may be low-arrange psychological aptitudes, for example, remembrance, or high-arrange ones, for example, thinking, sorting out thoughts, combination and contention abilities, and investigative intuition [6].

Appraisal instruments, utilized as a part of mix with institutionalized estimations of fluctuating execution levels, give solid confirmation of somebody's language capacities. Evaluations and scores are essential estimations embraced and utilized for purposes, for example, affirmation and self-appraisal. They are appointed on the premise of particular stamping criteria that serve as formats for appraisal, concocted to portray logically key elements in one's capacities [7].

2.20 An Overview of Automated Assessment

Robotized evaluation concentrates on naturally investigating and surveying somebody's ability. The field of robotized appraisal can be followed back to the mid 1960s and developed as a way to overcome issues emerging with institutionalized evaluation. For instance, it underpins a quicker appraisal and conveyance of results, favorable position for a few reasons, for example, moment input at the level of a person, as well as to foundations wishing to address instructive shortages expeditiously. Further points of interest turn out to be more claimed with regards to checking broadened writings, an undertaking inclined to a component of subjectivity. Mechanized frameworks ensure the use of steady checking criteria, in this way lessening irregularity, which may emerge specifically when more than one human analyst is utilized. Regularly, usage incorporate more point by point criticism on the journalists' written work capacities, along these lines encouraging self-evaluation and self-mentoring. In addition, the capability of a lessened workload is turning out to be more appealing, particularly in vast scale evaluations. Institutionalized appraisal involves a costly and major logistical exertion; mechanized evaluation can possibly definitely lessen time and expenses for preparing and utilizing human scorers [8].

University of Moratuwa, Sri Lanka.
Electronic Theses & Dissertations
www.lib.mrt.ac.lk

In spite of the fact that it is genuinely simple to develop a model that surveys shut class sorts of inquiries precisely, computerized content evaluation confronts numerous difficulties. A standout amongst the most imperative contemplations is the likelihood of building a framework that copies human conduct in perusing and making worth judgments about somebody's composition. This is generally managed by the capacity to assess vocabulary, sentence structure and linguistic structure, as well as different perspectives; diverse written work kinds, for example, expositions, stories, letters,

verse, fiction, thus on and additionally subjective viewpoints, for example, language development, scholarly substance, the rationale behind a contention, talk structure, clarity and familiarity are just a little piece of the range that should be considered. Moreover, it is just as essential to have the capacity to distinguish and naturally remove from writings measures of composing quality that are additionally a genuine response of the characteristic qualities that shape the premise of human judgments. The technique and appraisal criteria received by such frameworks ought to be straightforward, justifiable and significant. As the commonsense utility of mechanized frameworks depends unequivocally on their vigor to subversion, dangers to their legitimacy ought to additionally be distinguished and tended to. For instance, authors who comprehend something of a framework's workings may endeavor to abuse this to amplify their scores, autonomously of their basic capacity. A few different difficulties emerge, for example, their further improvement to work as learning apparatuses, giving input on somebody's written work abilities and advancement in comparative courses and as conveniently as people normally do [9].

University of Moratuwa, Sri Lanka.
Electronic Theses & Dissertations
www.lib.mrt.ac.lk

Mechanized content evaluation frameworks abuse printed components picked trying to adjust proof of composing ability against confirmation of execution mistakes keeping in mind the end goal to quantify the general quality and appoint a score to a content. The soonest frameworks utilized super uncommon elements, for example, word and sentence length, as intermediaries for comprehension the content. Later frameworks have utilized more refined robotized content handling systems to quantify grammaticality, printed lucidness, pre indicated blunders, et cetera. In the following area, we give an outline of this theory, trailed by our examination objectives [10].

Certainly or unequivocally, past work has for the most part regarded computerized appraisal as a regulated content characterization assignment that may be, anticipating a name for a content that is illustrative of its quality (e.g., an evaluation), taking into account an arrangement of cases named with the classes or evaluations the framework is attempting to foresee. Diverse methods have been utilized, for case cosine closeness of vectors speaking to message in different ways, regularly consolidated with dimensionality lessening strategies, for example, Latent Semantic Analysis (LSA), generative and discriminative machine learning models, space particular element extraction, and altered syntactic parsers [11].

We approach computerized appraisal as a managed discriminative machine learning issue, which empowers us to exploit clarified information. Our work researches routines for evaluating diverse parts of composing exposition, investigates the significance of an assortment of composing quality components, and locations legitimacy issues identified with their arrangement. Further, we distinguish new methods that beat beforehand created ones, and location speculation issues [12].

Procedures, for example, LSA can be utilized to quantify, notwithstanding composing capability, the semantic importance of a content written in light of a given brief. Rather than past work, we contend that a methodology which does not depend on (physically created) errand subordinate parts or information, and straightforwardly evaluates learner English, can deliver results in the same class as brief particular models. Further, it has the extra favorable position that it may not require re-preparing or tuning for new prompts or appraisal assignments. Frameworks that measure English capability specifically are less demanding and quicker to convey, since they will probably be re-usable and sum up better crosswise over distinctive sorts contrasted with subject particular ones; the last turns into a problem that needs to be addressed while endeavoring new assignments, since the model can't be connected until a considerable measure of physically expounded reaction writings are gathered for a particular brief. A non specific methodology has the upside of requiring littler example sizes, while its definition speaks to genuinely predictable `marking criteria' paying little respect to the brief conveyed. We ought to, nonetheless, take note of that human scoring rubrics additionally assume a vital part in the advancement of computerized frameworks [13].

University of Moratuwa, Sri Lanka.
Electronic Theses & Dissertations
www.lib.mrt.ac.lk

2.21 Machine Learning in Automated Assessment

There is a great collection of writing with respect to the advancement, execution, ease of use and assessment of mechanized content appraisal and scoring frameworks. Most as of late, report a thorough examination of the capacities of eight existing business paper scoring frameworks, assessed as a feature of the ASAP challenge [14].

Surviving ways to deal with Automated Assessment (AA) convey an extensive variety of methods from dimensionality decrease over frameworks of terms through to extraction of phonetically more profound elements, for example, sorts of syntactic developments and particular blunder sorts (e.g., non-assention of subject and primary verb). In this area, we talk about some of the more compelling and/or better portrayed methodologies since the early stages of computerized content appraisal and give a review of the different procedures received; further frameworks will be thoroughly analyzed to our work in the accompanying sections. Nitty gritty diagrams of existing AA frameworks have been distributed in different studies [15].

Project Essay Grade (PEG) is one of the most punctual frameworks, to a great extent spurred by the possibility to decrease work escalated stamping exercises. The framework utilizes various physically recognized generally shallow literary elements, which are thought to be intermediaries for inborn characteristics of composing skill. Samples of such elements incorporate the article length, number of pronouns and different POS labels, number of accentuation denote, the vicinity of a title, number of sections, etc. Direct relapse is utilized to dole out ideal component weights to amplify the connection with the inspector scores. The primary issue with this framework is that components, for example, word length and script length are anything but difficult to control freely of authentic written work capacity, possibly undermining the legitimacy of the framework. Later forms were adjusted to incorporate more advanced modules, for example, ones consolidating the utilization of parsers [9].

University of Moratuwa, Sri Lanka.
Electronic Theses & Dissertations
www.lib.mrt.ac.lk

e-Rater, a mechanized exposition scoring framework created by Educational Testing Service (ETS), was the first to be conveyed for operational scoring of high-stakes appraisals. In e-Rater writings are spoken to utilizing vectors of weighted components. Every component compares to an alternate property of writings, for example, a part of linguistic use (e.g., pronoun blunders, missing words, and subject verb assention), style (e.g., word redundancy, aloof voice, and sentence length), mechanics (e.g., upper casing of formal people, places or things, missing accentuation, spelling), association and talk (e.g., number of talk components, subordinating provisions), semantic rationality and theme likeness (e.g., similitude between words in a content and those found in physically evaluated samples for every evaluation). A few elements speaking

to cliché syntactic blunders, for instance, are removed utilizing physically coded undertaking particular indicators based, to some degree, on ordinary stamping criteria.

An unmarked content is scored in view of the cosine similitude between its weighted component vector and the ones got from the preparation set. Highlight weights and/or scores can be admitted to a stamping plan by direct relapse to create an all encompassing score. On the other hand, the framework contains some physically created errand particular segments and may require re-preparing or tuning for new prompts and appraisal undertakings. The previous shows high results on five distinctive datasets utilizing Naive Bayes prepared on vectors of stemmed words.

Later, depict the Bayesian Essay Test Scoring sYstem (BETSY), a framework which is uninhibitedly accessible for examination purposes. BETSY utilizes multinomial or Bernoulli Naive Bayes models to arrange writings into diverse classes (e.g., pass and come up short, or evaluations in the middle of An and F) in view of substance and style elements, for example, word unigrams and bigrams, sentence length, number of verbs, thing verb sets, et cetera. Grouping choices depend on the contingent likelihood of a class given an arrangement of elements, which is figured under the supposition that every component is autonomous of the others. Relapse is utilized to upgrade between the classifier's concordance and the evaluation point scales utilized. These frameworks demonstrate that regarding AA as a content order issue is feasible; then again, the component sorts utilized are all genuinely shallow, and the methodology does not make proficient utilization of the preparation information, as a different classifier is prepared for every evaluation point [10].

University of Moratuwa, Sri Lanka.
Electronic Theses & Dissertations
www.lib.mrt.ac.lk

Unsupervised clustering way to deal with AA of writings tending to the same theme, in view of a voting algorithm. The basic thought behind the algorithm is like e-Rater's speculation: great writings ought to look like other great ones. Writings are grouped by evaluation and given a starting Z-score. A model is prepared where the introductory score of a content changes iteratively in light of its similitude with whatever remains of the writings and additionally their Z-scores.

The methodology may be better portrayed as feebly directed as the appropriation of content evaluations in the preparation information is utilized to at the last Z-scores to

reviews. The framework utilizes a sack of-words representation of content, which is inclined to subversion and can possibly undermine its legitimacy. In any case, investigation of the exchange offs between the level of supervision required in preparing and evaluating precision is an imperative territory for future examination [12].

Lack of studies examining the use of discriminative machine figuring out how to AA. Generative models frequently utilize mistaken presumptions about the fundamental properties of writings, for instance, that the likelihood of an element given a class is restrictively autonomous of the remaining elements. Discriminative learning procedures make weaker suppositions, straightforwardly enhance execution on the preparation information, and frequently outflank non-discriminative ones in the connection of content order.

Novel discriminative model, a variation of the group perceptron algorithm and report better results thought about than probabilistic classifiers, for example, Naive Bayes and Maximum Entropy, and also to dimensionality decrease strategies that have been effectively utilized as a part of prior AA thinks about. They tentatively demonstrate that their model, prepared on CLC messages and utilizing an assortment of lexical and linguistic elements (e.g., POS ngrams and expression structure tenets) performs near the upper bound as depended by the understanding between human inspectors [1].

University of Moratuwa, Sri Lanka.
Electronic Theses & Dissertations
www.lib.mrt.ac.lk

2.22 Other Approaches to Automated Assessment

Intelligent Essay Assessor (IEA) utilizes Latent Semantic Analysis (LSA) to figure the semantic comparability between writings, at a particular evaluation point, and a test content. In spite of different systems, LSA can be understood as both a model of human learning representation and obtaining and as a strategy for catching semantic substance in writings. In LSA, content is spoken to by a network, where lines compare to words and sections to setting (writings).

Singular Value Decomposition (SVD) is utilized to acquire a decreased measurement framework grouping words and settings. The framework is prepared on theme and/or brief particular writings while test writings are alloted a score in view of the ones in

the preparation set that are generally comparable. The general score, which is ascertained utilizing relapse systems, depends on the substance score and additionally on different properties of writings, for example, style, syntax, et cetera, however the routines used to evaluate these are not depicted in any point of interest in distributed work [6].

Nonetheless, the framework requires re-preparing or tuning for new prompts and appraisal errands. A fairly distinctive approach is received to utilize an adjusted syntactic parser to investigate and score writings. This system depends on an altered adaptation of the Link Grammar parser where the general score of a content is computed as the normal of the scores alloted to every sentence.

Sentences are scored on a five-point scale taking into account the parser's expense metric, which generally measures the many-sided quality and deviation of a sentence from the parser's linguistic model. This methodology bears a few likenesses to the representation of our element space; on the other hand, syntactic elements portray stand out segment of our general framework and of the assignment [10].

University of Moratuwa, Sri Lanka.
Electronic Theses & Dissertations
www.lib.mrt.ac.lk

2.23 Summary

This chapter describes the recent increasing popularity and the penetration of NLP technologies into wide spectrum of subject areas which has speed up the developments of automated solutions which uses NLP for the assessments. Next chapter will be on adapted AI technologies for developing the proposed Agent-based Solution for Improving Abstracts.

Chapter 3

Cutting Edge Technologies Integrated

3.1 Introduction

Increasing popularity and the penetration of AI technologies into wide spectrum of subject areas into the complex real world has made AI as rapidly expanding area of research.

3.2 Natural Language Processing

Natural Language Processing (NLP) is thought to be a standout amongst the most difficult regions of AI. The exploration in NLP contains an assortment of fields including corpus-based routines, talk systems, formal models, machine interpretation, normal language era, and talked language understanding. There have been a few experimental systems utilized as a part of NLP. Past strategies (e.g., realist routines) required manual encoding of phonetic learning, which has turned out to be troublesome because of the perplexing way of human language. Late routines (e.g., observational strategies), in any case, utilize methods that naturally separate semantic learning from vast content corpora. As such, observational strategies utilize factual or machine learning methods to prepare the framework on a lot of real language information [1].

University of Moratuwa, Sri Lanka.
Electronic Theses & Dissertations
www.lib.mrt.ac.lk

NLP is asserted to be an intricate errand to understand in light of the fact that it contains a few levels of handling and subtasks. It has four classifications of language assignments including discourse acknowledgment, syntactic investigation, talk examination, data extraction, and machine interpretation. Discourse acknowledgment concentrates on charting a constant discourse signal into a succession of known words. Syntactic investigation, then again, decides the ways words are grouped into parts like thing and verb-phrases. Semantic investigation includes charting a sentence to a kind of significance representation, for example, an intelligent expression. Though talk examination concentrates on how connection effects sentence translation and data [2].

3.3 Bayesian approach

Bayesian systems have a few applications, for example, distinguishing spam and other undesirable messages in view of their similitude with already arranged email, and sorting the resumes of occupation candidates into different employment classifications as indicated by their likeness to beforehand characterized resumes. A few Microsoft items, for example, Answer Wizard of Office 95®, the Office Assistant of Office 97®, and various specialized troubleshooters are different uses of the Bayesian approach [1].

There are two Bayesian models broadly utilized as a part of content grouping: the Multivariate Bernoulli Model and the Multinominal Model. While the previous perspectives every article as an exceptional instance of aligned elements, the last perspectives every exposition as a specimen of adjusted components. In the Bernoulli display, the restrictive likelihood of vicinity of a particular element is evaluated by the extent of articles inside of every classification that incorporate the element [6].

University of Moratuwa, Sri Lanka.
Electronic Theses & Dissertations
www.lib.mrt.ac.lk

In Multinomial model, then again, the likelihood of every score for a given exposition is processed as the result of the probabilities of the elements incorporated into the article. To outline, the Bernoulli model explores whether a particular component exists in an exposition or not, though the Multinominal model checks the various utilization of a particular element in a paper. The Bernoulli model registers moderately gradually contrasted with the Multinominal model [8].

The Bayesian methodology incorporates key ideas, for example, stemming, stop words, and highlight determination. Stemming indicates the procedure of dispensing with additions to get stems. For instance, getting "educ" as a stem for teach, training, instructs, instructive, and taught. Stop words allude to different articles, pronouns, descriptors, and relational words. Web indexes don't list these sorts of words in light of the fact that they can bring about extensive number of unimportant results. One way to deal with highlight choice is the lessening in entropy. By minimizing entropy, it is conceivable to pick the things with most extreme potential data pick up [1].

3.4 Latent Semantic Analysis

Latent Semantic Analysis (LSA) is characterized as "a factual model of word utilization that allows examinations of the semantic closeness between bits of literary data". LSA first procedures a corpus of machine-coherent language and afterward speaks to the words that are incorporated into a sentence, section, or exposition through measurable algorithm. LSA measures of closeness are considered exceedingly corresponded with human importance likenesses among words and messages. Besides, it effectively emulates human word choice and class judgments. The basic thought is that the importance of an entry is all that much reliant on its words and changing even one and only word can bring about significance contrasts in the section. Then again, two entries with diverse words may have a fundamentally the same significance. The hidden thought can be condensed as: "which means of word 1 + importance of word 2 + … + significance of word k = significance of entry" [1].

The instructive uses of LSA incorporate picking the most suitable content for understudies with diverse levels of foundation learning, programmed scoring of exposition substance, and helping understudies in condensing messages effectively. Keeping in mind the end goal to assess the general nature of a paper, LSA should be prepared on space delegate writings (messages that best speak to the composition brief). At that point the exposition should be portrayed by LSA vectors (a scientific representation of the paper). At long last, the reasonable significance and the substance of the exposition are contrasted with different writings [5].

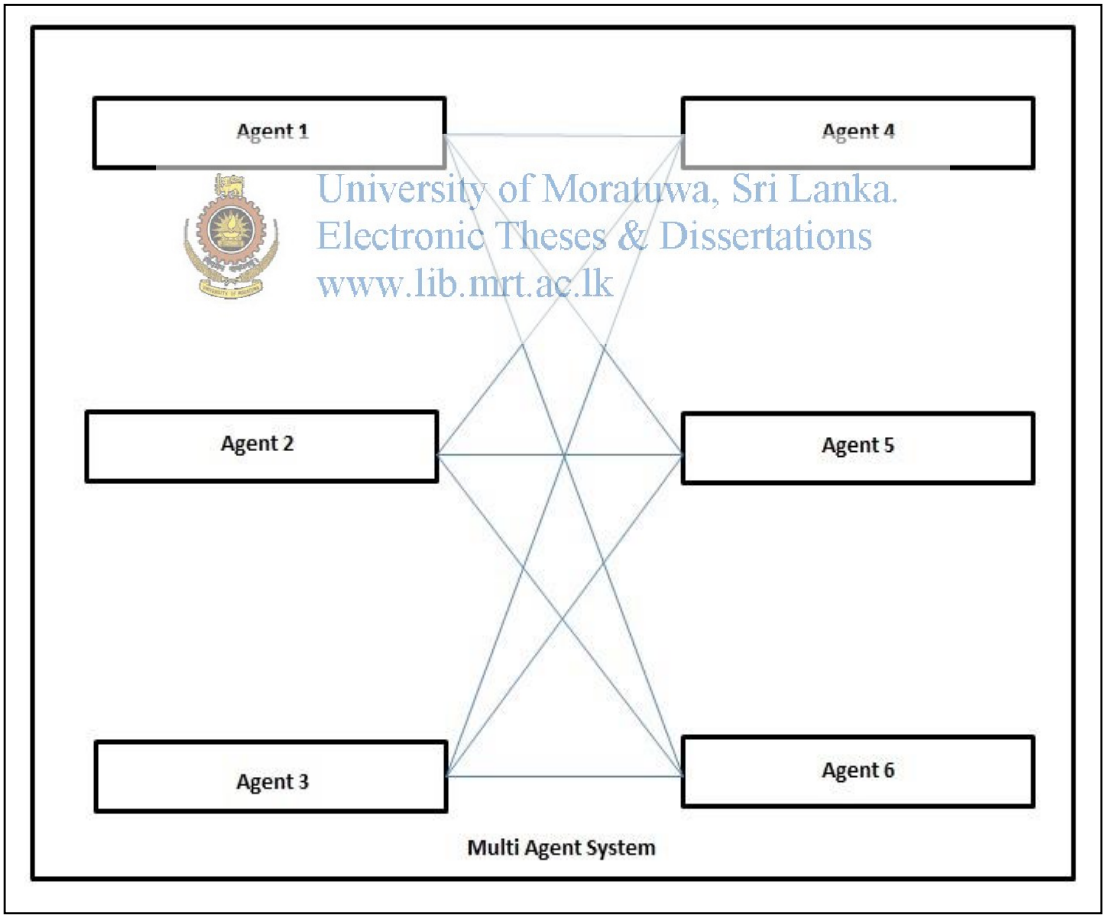
University of Moratuwa, Sri Lanka.
Electronic Theses & Dissertations
www.lib.mrt.ac.lk

In the LSA based methodology, the content is spoken to as a network. Every line in the grid speaks to an extraordinary word, while every section speaks to setting. Every cell includes the recurrence of the word. At that point, every cell recurrence is considered by an element that indicates not just the significance of the word in that connection additionally the extent to which the word sort conveys data in the area talk. The semantics of a word are confirmed through every one of the settings in which the word happens. The quantity of events of every word in a content decides its semantic space. For instance, 300 sections and 2000 words give a 300x2000 matrix. Here, while every word is spoken to by a 300-dimentional vector, every passage is spoken to by a 2000-dimentional vector. By diminishing these measurements, LSA incites

semantic similitudes between words. This diminishment is basic since it allows the representation of word implications through the connection in which they happen. The quantity of measurements is additionally significant. That is, if the number is too little, a great part of the data will be lost. In actuality, if the number is too huge, restricted conditions will be drawn between vectors. As indicated by this strategy, the semantic data is resolved just through the co-event of words in an extensive corpus of writings [1].

3.5 Multi Agent Systems

Agent software is a quickly creating zone of examination. In any case, the abuse of 'agent' has tended to cover the way that, in actuality, there is a genuinely heterogeneous assortment of examination being done under this standard [16].



University of Moratuwa, Sri Lanka.
Electronic Theses & Dissertations
www.lib.mrt.ac.lk

Figure 3.1: Overview of Multi Agent System

Software agents have advanced from multi-agent frameworks, which thus shape one of three wide zones which fall under disseminated computerized reasoning, the other

two being Distributed Problem Solving and Parallel AI. Consequently, as with multi-agents systems, they acquire a considerable lot of appropriated counterfeit consciousness' inspirations, objectives and potential advantages. For instance, because of dispersed processing, programming agents acquire conveyed computerized reasoning's potential advantages including seclusion, speed (because of parallelism) and dependability (because of repetition). It likewise acquires those because of AI, for example, operation at the information level, less demanding upkeep, reusability and stage freedom [16], [17].

There are no less than two reasons why it is so hard to characterize unequivocally what agents are. Firstly, agents don't" "possess" this term in the same path as fluffy rationalists/AI analysts, for instance, claim the term 'fluffy rationale' - it is one that is utilized broadly as a part of regular speech as in travel agents, domain agents, and so forth. Besides, even inside of the product crew, "agents" is truly an umbrella term for a heterogeneous group of innovative work. The reaction of some agents scientists to this absence of definition has been to create yet some more equivalent words, and it is doubtful if these comprehend anything or simply further add to the disarray. So we now have equivalent words including knowbots (i.e. information based robots), softbots (programming robot), taskbots (task based robots), userbots, robots, individual agents, self-ruling agents and individual colleagues. To be reasonable, there are some great explanations behind having such equivalent words [16], [17], [18].

University of Moratuwa, Sri Lanka.
Electronic Theses & Dissertations
www.lib.mrt.ac.lk

The speculation/objective of multi-agent frameworks is sufficiently clear and has been demonstrated in numerous multi-agent models over the globe: making a framework that interconnects independently created agents, in this way empowering the gathering to work past the abilities of any solitary agents in the set-up. Much critical ground in such frameworks had been secured certainly before 1994, and it is genuinely a matter of verbal confrontation the amount of genuine advancement has been made subsequent to. Obviously, there has been much solidification of pre-1994 work. Be that as it may, let us return to the guarantees all the more nearly whilst at the same time adding with a percentage of the truth [18], [19], [20].

Co-ordination is a focal issue in software agent frameworks specifically, and in conveyed manmade brainpower all in all. On the other hand, it has additionally been considered by scientists in assorted controls in the sociologies, including association hypothesis, political science, social brain research, human sciences, law and humanism [20], [21], [22].

Adapted Technology	Description
Multi Agent Systems	Coordinate and control the entire process
Natural Language Processing	Process the content of the abstract to suggest the improvements
Ontology	Keep the domain knowledge required to process the content of the abstract

University of Moratuwa, Sri Lanka.
Electronic Theses & Dissertations
www.lib.mrt.ac.lk

Table 3.1: Adapted Cutting Edge Technologies

3.6 Ontology

Speculations in AI fall into two general classes: mechanism theories and content theories. Ontologies are content speculations about the sorts of articles, properties of items, and relations between items that are conceivable in a predefined space of information. They give potential terms to portraying our insight about the space [23].

Most research on ontologies concentrates on what one may portray as area verifiable information, on the grounds that learning of that sort is especially helpful in characteristic language understanding. There is another class of ontologies that are

essential in Knowledge Based Systems one that aides in sharing learning about thinking techniques or critical thinking routines [24].

In AI, learning in frameworks is considered as something that is unequivocally spoken to and worked on by surmising procedures. In any case, that is an excessively tight view. All data frameworks movement in information. Any product that does anything valuable can't be composed without a guarantee to a model of the applicable world to substances, properties, and relations in that world. Information structures and methodology verifiably or unequivocally make responsibilities to a space metaphysics. It is regular to ask whether a finance framework "knows" about the new assessment law, or whether a database framework "knows" about representative compensations. Data recovery frameworks, computerized libraries, mix of heterogeneous data sources, and Internet web crawlers need area ontologies to arrange data and direct the pursuit forms. For instance, a web crawler has classes and subcategories that arrange the pursuit. The web crawler group generally alludes to these classifications and subcategories as ontologies [25].

University of Moratuwa, Sri Lanka.
Electronic Theses & Dissertations
www.lib.mrt.ac.lk

Article situated configuration of programming frameworks also relies on upon a suitable area cosmology. Objects, their properties, and their methods pretty much reflect parts of the space that are significant to the application. Object frameworks speaking to a valuable examination of a space can frequently be reused for an alternate application program. Object frameworks and ontologies accentuate distinctive viewpoints, however we suspect that after some time joining between these advances will increment. As data frameworks demonstrate expansive information spaces, area ontologies will get to be as critical when all is said in done programming frameworks as in numerous regions of AI [26].

3.7 Summary

This chapter describes the adapted AI technologies for the development of proposed Agent-based Solution for Improving Abstracts and explained the increasing popularity and the penetration of AI technologies into wide spectrum of subject areas. Next chapter will be on approach by describing the hypothesis input, output, process,

features and users for novel solution for natural language processing based Agent-based Solution for Improving Abstracts.

University of Moratuwa, Sri Lanka.
Electronic Theses & Dissertations
www.lib.mrt.ac.lk

Chapter 4

Top-notch Approach to Improve the Abstracts

4.1 Introduction

In the previous two chapters we define the research problem as the inefficient in improving abstracts and the existing solutions by describing why NLP technology could be a potential technology to develop novel solution for improving the abstracts. This chapter presents our approach by describing the hypothesis input, output, process, features and users for novel solution for Agent-based Solution for Improving Abstracts. The new solution has been named as ASIA, an acronym for Agent-based Solution for Improving Abstracts.

4.2 Hypothesis

Preliminary editing of a document in a specified domain can be automated with NLP techniques.

University of Moratuwa, Sri Lanka.
Electronic Theses & Dissertations
www.lib.mrt.ac.lk

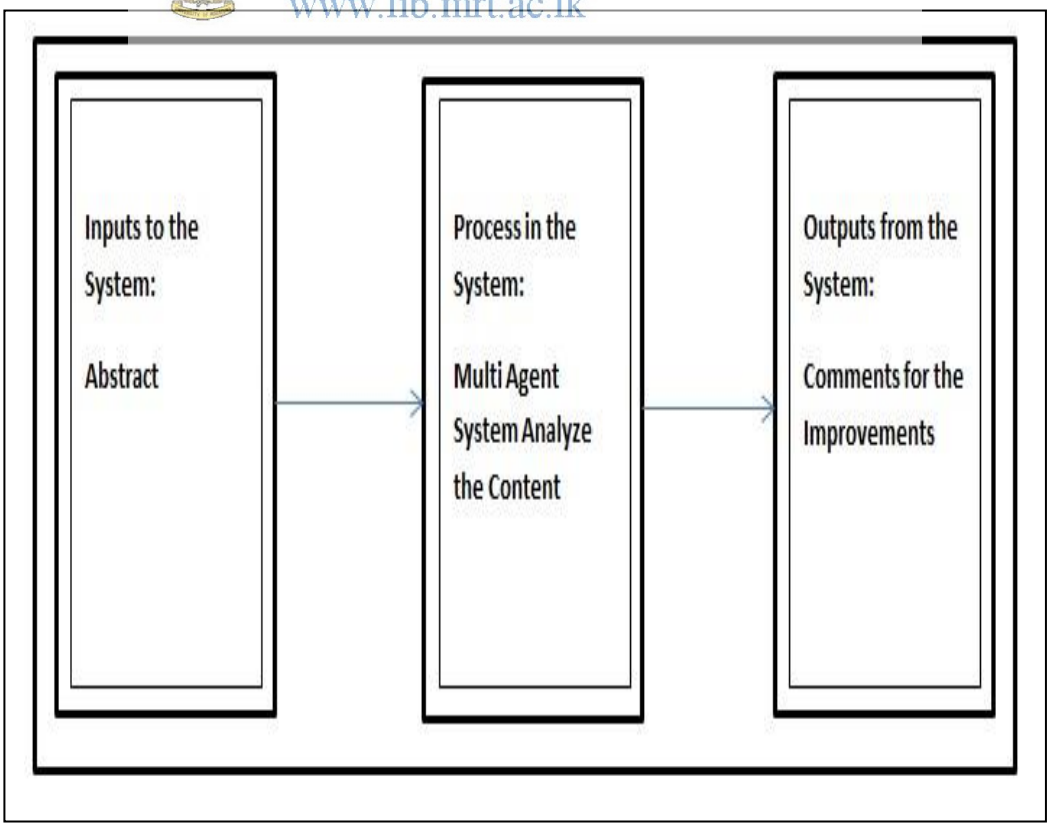


Figure 4.1: High Level Process Diagram

4.3 Inputs to System

NLP for Agent-based Solution for Improving Abstracts has been designed to accept multiple inputs coming from different entities of the improving the abstracts.

Agent-based Solution for Improving Abstracts has been designed to accept the abstracts as the inputs to the system.

4.4 Outputs from the System

These outputs are coming as user related and the improving related aspects. For example

Pre edited abstract with comments for the improvement.

Input / Output	Description
Inputs to the System	Title and Abstract of the Research Paper
Outputs from the System	Pre edited abstract with comments for the improvement.

University of Moratuwa, Sri Lanka.
Electronic Theses & Dissertations
www.lib.mrt.ac.lk

Table 4.1: Inputs and Outputs of the System

4.5 Process of the System

Having entered the inputs the system will uses NLP to generate the output. In this process two major types of processes are defined. Knowledge required to operate these processes are stored in a common domain. This has knowledge required for the editing of the documents together with the context specified knowledge. For example morphology, syntax and pragmatic are included from the language perspective. Further context based solution such as problem, solution, conclusion are also available.

4.6 Features of the System

Following features are available in the system:

Minimal resource usage,
Online available,
Development cost is marginally.

4.7 Functional Requirements of the Proposed System

Following functional requirements are available in the system:

Take the Title and the abstract of the research paper as the inputs,
Process the content taken as the input to find out the areas that can be improved,
Suggest the improvements for the abstract as the outputs.

University of Moratuwa, Sri Lanka.
Electronic Theses & Dissertations
www.lib.mrt.ac.lk

4.8 Non Functional Requirements of the Proposed System

Following non-functional requirements are available in the system:

Parallel processing,
Reasonable response time,
User friendliness,
Ability to update the Ontologies,
Ability to improve the remaining Agents in the Multi Agent System,
Ability to add new Agents to the Multi Agent System.

4.9 Users of the Proposed System

This proposed system is mainly focus on assisting the researchers to improve the abstract part of their research papers. This system would be also useful for the technical writers who are improving the research papers to improve the abstract part of the research papers. This system would be useful to undergraduate and postgraduate students who are not that much familiar with writing research papers to

take assistance on improving the abstract part of the research paper. This proposed system would assist them with suggestions for the improvements in their abstract part of the research papers.

4.10 Summary

This chapter describes the approach by describing the hypothesis input, output, process, features and users for novel solution for natural language processing based Agent-based Solution for Improving Abstracts. Next chapter will be on high level analysis and the design of the Agent-based Solution for Improving Abstracts.

University of Moratuwa, Sri Lanka.
Electronic Theses & Dissertations
www.lib.mrt.ac.lk

Chapter 5

Analysis and Design of the Proposed Solution

5.1 Introduction

This chapter discusses the high level analysis and the design of the Agent-based Solution for Improving Abstracts using multi agent systems and natural language processing techniques.

5.2 Analysis of the Proposed Design

This section critically discusses about the overview of the proposed design. When an abstract has been inputted to the proposed system, parser agent would act on it. That is the first step in the system.

After it has gone through the parser agent, other agents of the multi agent system would act on it to analyze the content & balance and suggest restructuring & improvements. This multi agent system would argue about the problem, solution & conclusion.

University of Moratuwa, Sri Lanka.
Electronic Theses & Dissertations
www.lib.mrt.ac.lk

In the second step title of the research paper is checked against the abstract and look for the overall match of the abstract to the title. In the third step proposed system checks whether the Problem, Solution & Conclusion is there in the abstract.

Then in the fourth step proposed system checks the content for accuracy. The proposed system got ontology of keywords which should and shouldn't be used in an abstract. The proposed system automatically checks whether the content have those keywords which should be there in an abstract and the content haven't those keywords which shouldn't be there in an abstract. So, the system checks whether the necessity is there in the abstract.

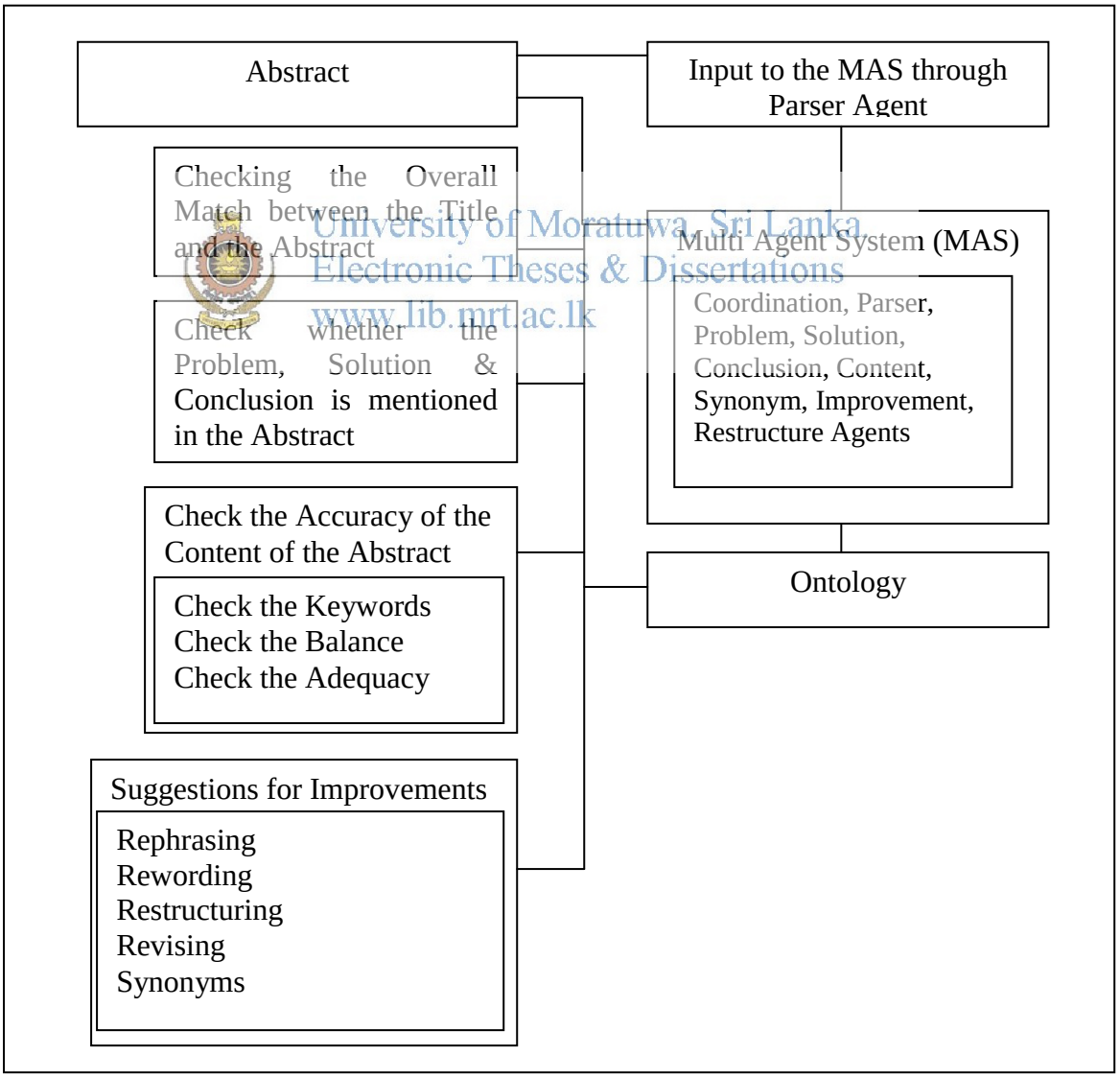
Then the system checks whether the balance between the problem, solution & conclusion is there. As example the abstract should discuss about the problem less than the solution and the solution should be explained much more than the problem.

Also the system assures that the references, citation and further work is not there in the abstract.

Then in the fifth step the proposed system makes suggestions for the improvements and the editing of the document using its ontology. This includes rephrasing, rewording, restructuring and revising for improving the abstract. The system would suggest synonyms which are much more suitable to be used in the abstract.

5.3 Design Architecture Diagram

In below Figure 5.1, we have shown the high level design architecture diagram for an Agent-based Solution for Improving Abstracts



University of Moratuwa, Sri Lanka.
Electronic Theses & Dissertations
www.lib.mrt.ac.lk

Figure 5.1: Design Architecture Diagram

5.4 Multi Agent Swarm Intelligence

When an abstract has been inputted to this system it would be feed into a parser agent which is the first step in the system.

After that has gone through the parser agent, multi agent swarm of agents would act on it to analyze the content & balance and suggest restructuring & improvements.

The agents in this multi agent swarm discuss about the problem, solution & conclusion and provide suggestions to improve the abstract.

5.5 Checking the Overall Match between the Title and the Abstract

In this step title of the research paper is matched against the abstract and look for the overall match of the abstract to the title.

Here the system analyzes the title of the research paper against the content of the abstract and check whether the abstract discuss about the topic mentioned in the title and gives the feedback.

University of Moratuwa, Sri Lanka.
Electronic Theses & Dissertations
www.lib.mrt.ac.lk

5.6 Check whether the Problem, Solution & Conclusion is mentioned in the Abstract

In this step proposed system checks whether the Problem, Solution & Conclusion is there in the abstract.

System looks for the words like issue, address, proposes, approach, methodology, etc… and identify whether the abstract consists of the problem, solution & conclusion and gives the feedback whether those sections are included in the abstract.

5.7 Check the Balance between the Problem, Solution & Conclusion in the Abstract

Then the system checks whether the balance between the problem, solution & conclusion is there. As example the abstract should discuss about the problem less than the solution and the solution should be explained much more than the problem.

5.8 Check the Accuracy of the Content of the Abstract

Then in this step proposed system checks the content for accuracy. The proposed system got ontology of keywords which should and shouldn't be used in an abstract.

The proposed system automatically checks whether the content have those keywords which should be there in an abstract and the content haven't those keywords which shouldn't be there in an abstract. So, the system checks whether the necessity is there in the abstract.

Also the system assures that the references, citation and further work is not there in the abstract.

5.9 Suggestions for Improvements

As the last step the proposed system makes suggestions for the improvements and the editing of the document using its ontology.

University of Moratuwa, Sri Lanka.
Electronic Theses & Dissertations
www.lib.mrt.ac.lk

This includes rephrasing, rewording, restructuring and revising for improving the abstract. The system would suggest synonyms which are much more suitable to be used in the abstract.

5.10 Ontology

Ontology which consists of keywords that should and shouldn't be there in an abstract would be maintained to be accessed by the multi agent swarm to check whether the content of the abstract have those keywords that should be used in the abstracts and the content of the abstract haven't those keywords that shouldn't be used in abstracts.

So, necessity of the content of the abstract can be checked and assured. Also, this ontology would consist of the synonyms for suggesting rewording, verbs, standard phrases and other good practices for structuring the contents of the abstracts.

Process	Description
Input the Title and Abstract into the Multi Agent System (MAS)	When an abstract has been inputted to this system it would be feed into a parser agent which is the first step in the system.
Check the Overall Match between the Title and the Abstract by MAS	In this step title of the research paper is matched against the abstract and look for the overall match of the abstract to the title.
Check whether the Problem, Solution & Conclusion is mentioned in the Abstract by MAS	In this step proposed system checks whether the Problem, Solution & Conclusion is there in the abstract.
Check the Balance between the Problem, Solution & Conclusion in the Abstract by MAS	Then the system checks whether the balance between the Problem, Solution & Conclusion is there.
Check the Accuracy of the Content of the Abstract by MAS	Then in this step proposed system checks the content for accuracy. The proposed system got ontology of keywords which should and shouldn't be used in an abstract.
Suggestions for Improvements by MAS	As the last step the proposed system makes suggestions for the improvements and the editing of the document using its ontology.

University of Moratuwa, Sri Lanka.
Electronic Theses & Dissertations
www.lib.mrt.ac.lk

Table 5.1: Process Description in the Proposed Design

5.11 Analysis of the Proposed Design

This section critically discusses about the overview of the proposed design. When an abstract has been inputted to the proposed system, parser agent would act on it. That is the first step in the system.

After it has gone through the parser agent, other agents of the multi agent system would act on it to analyze the content & balance and suggest restructuring & improvements. This multi agent system would argue about the problem, solution & conclusion.

In the second step title of the research paper is checked against the abstract and look for the overall match of the abstract to the title. In the third step proposed system checks whether the Problem, Solution & Conclusion is there in the abstract.

University of Moratuwa, Sri Lanka.
Electronic Theses & Dissertations
www.lib.mrt.ac.lk

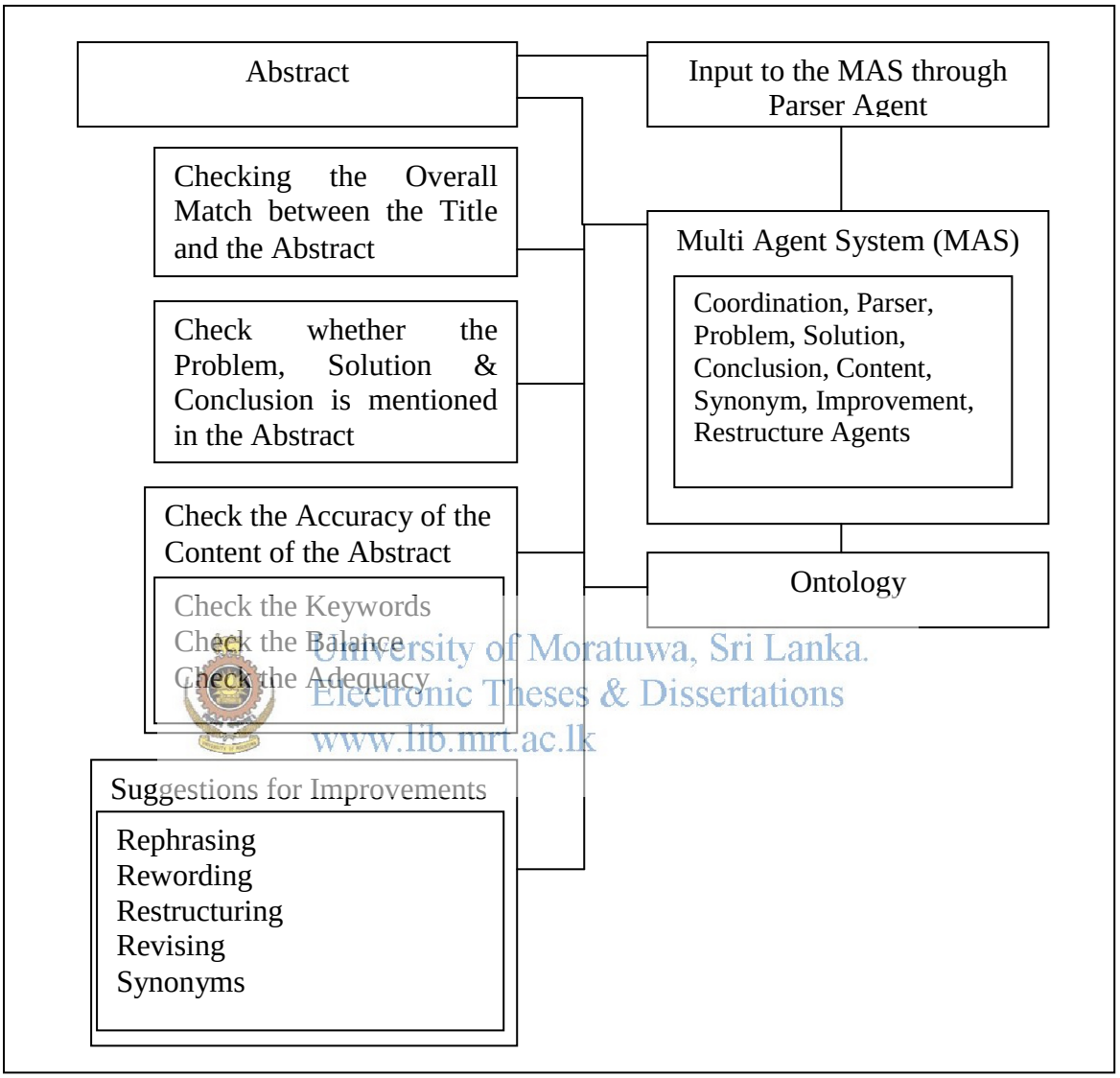
Then in the fourth step proposed system checks the content for accuracy. The proposed system got ontology of keywords which should and shouldn't be used in an abstract. The proposed system automatically checks whether the content have those keywords which should be there in an abstract and the content haven't those keywords which shouldn't be there in an abstract. So, the system checks whether the necessity is there in the abstract.

Then the system checks whether the balance between the problem, solution & conclusion is there. As example the abstract should discuss about the problem less than the solution and the solution should be explained much more than the problem. Also the system assures that the references, citation and further work is not there in the abstract.

Then in the fifth step the proposed system makes suggestions for the improvements and the editing of the document using its ontology. This includes rephrasing, rewording, restructuring and revising for improving the abstract. The system would suggest synonyms which are much more suitable to be used in the abstract.

5.12 Design Architecture Diagram

In below Figure 5.1, we have shown the high level design architecture diagram for an Agent-based Solution for Improving Abstracts



University of Moratuwa, Sri Lanka.
Electronic Theses & Dissertations
www.lib.mrt.ac.lk

Figure 5.2: Design Architecture Diagram

5.13 Multi Agent Swarm Intelligence

When an abstract has been inputted to this system it would be feed into a parser agent which is the first step in the system.

After that has gone through the parser agent, multi agent swarm of agents would act on it to analyze the content & balance and suggest restructuring & improvements.

The agents in this multi agent swarm discuss about the problem, solution & conclusion and provide suggestions to improve the abstract.

5.14 Checking the Overall Match between the Title and the Abstract

In this step title of the research paper is matched against the abstract and look for the overall match of the abstract to the title.

Here the system analyzes the title of the research paper against the content of the abstract and check whether the abstract discuss about the topic mentioned in the title and gives the feedback.

5.15 Check whether the Problem, Solution & Conclusion is mentioned in the Abstract

In this step proposed system checks whether the Problem, Solution & Conclusion is there in the abstract.

University of Moratuwa, Sri Lanka.
Electronic Theses & Dissertations
www.lib.mrt.ac.lk

System looks for the words like issue, address, proposes, approach, methodology, etc… and identify whether the abstract consists of the problem, solution & conclusion and gives the feedback whether those sections are included in the abstract.

5.16 Check the Balance between the Problem, Solution & Conclusion in the Abstract

Then the system checks whether the balance between the problem, solution & conclusion is there. As example the abstract should discuss about the problem less than the solution and the solution should be explained much more than the problem.

5.17 Check the Accuracy of the Content of the Abstract

Then in this step proposed system checks the content for accuracy. The proposed system got ontology of keywords which should and shouldn't be used in an abstract.

The proposed system automatically checks whether the content have those keywords which should be there in an abstract and the content haven't those keywords which

shouldn't be there in an abstract. So, the system checks whether the necessity is there in the abstract.

Also the system assures that the references, citation and further work is not there in the abstract.

5.18 Suggestions for Improvements

As the last step the proposed system makes suggestions for the improvements and the editing of the document using its ontology.

This includes rephrasing, rewording, restructuring and revising for improving the abstract. The system would suggest synonyms which are much more suitable to be used in the abstract.

5.19 Ontology

University of Moratuwa, Sri Lanka.
Electronic Theses & Dissertations
www.lib.mrt.ac.lk

Ontology which consists of keywords that should and shouldn't be there in an abstract would be maintained to be accessed by the multi agent swarm to check whether the content of the abstract have those keywords that should be used in the abstracts and the content of the abstract haven't those keywords that shouldn't be used in abstracts.

So, necessity of the content of the abstract can be checked and assured. Also, this ontology would consist of the synonyms for suggesting rewording, verbs, standard phrases and other good practices for structuring the contents of the abstracts.

Process	Description
Input the Title and Abstract into the Multi Agent System (MAS)	When an abstract has been inputted to this system it would be feed into a parser agent which is the first step in the system.

Check the Overall Match between the Title and the Abstract by MAS	In this step title of the research paper is matched against the abstract and look for the overall match of the abstract to the title.
Check whether the Problem, Solution & Conclusion is mentioned in the Abstract by MAS	In this step proposed system checks whether the Problem, Solution & Conclusion is there in the abstract.
Check the Balance between the Problem, Solution & Conclusion in the Abstract by MAS	Then the system checks whether the balance between the Problem, Solution & Conclusion is there.
Check the Accuracy of the Content of the Abstract by MAS	Then in this step proposed system checks the content for accuracy. The proposed system got ontology of keywords which should and shouldn't be used in an abstract.
Suggestions for Improvements by MAS	As the last step the proposed system makes suggestions for the improvements and the editing of the document using its ontology.

University of Moratuwa, Sri Lanka.
Electronic Theses & Dissertations
www.lib.mrt.ac.lk

Table 5.2: Process Description in the Proposed Design

5.20 Summary

This chapter discussed about the high level analysis and the design of the Agent-based Solution for Improving Abstracts using natural language processing techniques, multi agent systems and ontology. Next chapter will be on the implementation of the Agent-based Solution for Improving Abstracts.

Chapter 6

Implementation of the Novel Solution

6.1 Introduction

This chapter provides the implementation details that are stated in the design and explains the consistency between the design and the implementation. This chapter describes the implementation of the Agent-based Solution for Improving Abstracts using natural language processing techniques in detail.

6.2 Overview of the Implementation

This section critically discusses about the overview of the proposed implementation of the Agent-based Solution for Improving Abstracts. This implementation of the proposed Agent-based Solution for Improving Abstracts is based on the natural language processing and multi agent systems technology together with commonly used verb phrases and other good practices as ontology.

University of Moratuwa, Sri Lanka.
Electronic Theses & Dissertations
www.lib.mrt.ac.lk

The system has been developed with a multi agent system of six agents, namely, parser agent, problem agent, solution agent, conclusion agent, synonym agent and rephrase agent. This multi agent system of six agents interacts with each other and deliberates to reach consensus regarding a solution.

The system has been developed with JAVA to run on Windows platform. This proposed Agent-based Solution for Improving Abstracts has been incrementally tested for results and has been evaluated on Windows environment. However as JAVA is platform independent, this proposed implementation of Agent-based Solution for Improving Abstracts would be capable of easily be ported to other platforms as well.

The Stanford CoreNLP Natural Language Processing Toolkit which is a very powerful API for Natural Language Processing has been mainly used to develop the parser agent and JADE which is a very popular framework for agent development has been used for development of the entire multi agent system.

6.3 Coordination Agent and Parser Agent

The coordination agent initiates and coordinates the entire process. The parser agent identifies syntactic information of each sentence and prepares the contents of the abstract for further analysis. These agents would interact with other agents of the multi agents system developed in this proposed implementation to reach consensus regarding the solution. The coordination agent and parser agent would use it's ontologies for deliberating with other agents.

6.4 Problem Agent

The problem agent ensures whether the research problem has been stated in the early part of the abstract and its proportion within the abstract. This agent would interact with other agents of the multi agents system developed in this proposed implementation to reach consensus regarding the solution in terms of the proportion of problem section within the abstract. The problem agent would use it's ontology for deliberating with other agents.

University of Moratuwa, Sri Lanka.
Electronic Theses & Dissertations
www.lib.mrt.ac.lk

6.5 Solution Agent

The solution agent checks for the contents in terms of concepts such as hypothesis, methodology, approach, design, implementation, methods, theoretical framework, technology, hardware, software, and sampling. This agent would interact with other agents of the multi agents system developed in this proposed implementation to reach consensus regarding the solution in terms of the proportion of solution section within the abstract. The solution agent would use it's ontology for deliberating with other agents.

6.6 Conclusion Agent

The conclusion agent searches for concepts such as testing, evaluation, data analysis and statistical significance. This agent would interact with other agents of the multi agents system developed in this proposed implementation to reach consensus regarding the solution in terms of the proportion of conclusion section within the abstract. The conclusion agent would use it's ontology for deliberating with other agents.

6.7 Content Agent, Synonym Agent, Improvement Agent and Restructure Agent

The content, synonym, improvement and restructure agents are responsible to offer guidelines to modify and improving of the abstract. These agents would interact with other agents of the multi agents system developed in this proposed implementation to reach consensus regarding the solution in terms of the synonyms and rephrasing the abstract. The content, synonym, improvement and restructure agents would use it's ontologies for deliberating with other agents.

6.8 Coordination Agent and Parser Agent

The coordination agent initiates and coordinates the entire process. The parser agent identifies syntactic information of each sentence and prepares the contents of the abstract for further analysis. These agents would interact with other agents of the multi agents system developed in this proposed implementation to reach consensus regarding the solution. The coordination agent and parser agent would use it's ontologies for deliberating with other agents.

University of Moratuwa, Sri Lanka.
Electronic Theses & Dissertations
www.lib.mrt.ac.lk

6.9 Problem Agent

The problem agent ensures whether the research problem has been stated in the early part of the abstract and its proportion within the abstract. This agent would interact with other agents of the multi agents system developed in this proposed implementation to reach consensus regarding the solution in terms of the proportion of problem section within the abstract. The problem agent would use it's ontology for deliberating with other agents.

6.10 Solution Agent

The solution agent checks for the contents in terms of concepts such as hypothesis, methodology, approach, design, implementation, methods, theoretical framework, technology, hardware, software, and sampling. This agent would interact with other agents of the multi agents system developed in this proposed implementation to reach consensus regarding the solution in terms of the proportion of solution section within

the abstract. The solution agent would use it's ontology for deliberating with other agents.

6.11 Conclusion Agent

The conclusion agent searches for concepts such as testing, evaluation, data analysis and statistical significance. This agent would interact with other agents of the multi agents system developed in this proposed implementation to reach consensus regarding the solution in terms of the proportion of conclusion section within the abstract. The conclusion agent would use it's ontology for deliberating with other agents.

6.12 Content Agent, Synonym Agent, Improvement Agent and Restructure Agent

The content, synonym, improvement and restructure agents are responsible to offer guidelines to modify and improving of the abstract. These agents would interact with other agents of the multi agents system developed in this proposed implementation to reach consensus regarding the solution in terms of the synonyms and rephrasing the abstract. The content, synonym, improvement and restructure agents would use it's ontologies for deliberating with other agents.

University of Moratuwa, Sri Lanka.
Electronic Theses & Dissertations
www.lib.mrt.ac.lk

6.13 Multi Agent System

The agents in this multi agent system of the proposed implementation of Agent-based Solution for Improving Abstracts would interact with each other and deliberate to reach consensus regarding a solution. As example, problem agent, solution agent and conclusion agent may agree on the proportion of respective contents within the abstract. Each agent has its own ontology for deliberating with other agents in this multi agent system.

6.14 Testing

The proposed implementation of Agent-based Solution for Improving Abstracts has been incrementally tested on Windows environment, and shown interesting results

related to checking for completeness of the abstract in terms required materials and suggestion for improvements.

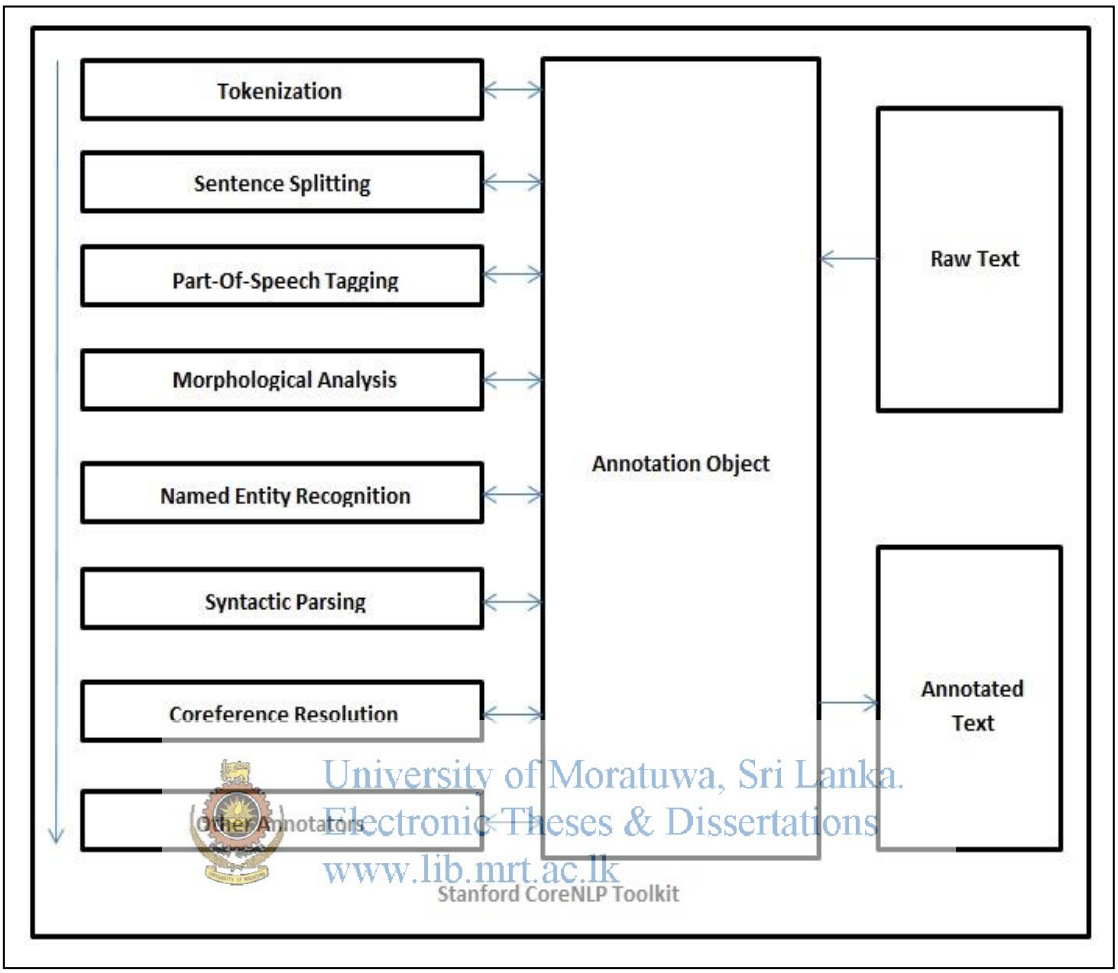
6.15 The Stanford CoreNLP Natural Language Processing Toolkit

Stanford CoreNLP gives an arrangement of natural language examination apparatuses which can take crude content data and give the base types of words, their parts of discourse, whether they are names of organizations, individuals, and so on., standardize dates, times, and numeric amounts, and checkup the structure of sentences as far as expressions and word conditions, demonstrate which thing expressions allude to the same substances, show assessment, and so on.

Stanford CoreNLP is an incorporated structure. It's will probably make it simple to apply a bundle of semantic investigation devices to a bit of content. Beginning from plain content, designer can run every one of the instruments on it with only two lines of code. It is intended to be exceptionally adaptable and extensible. With a solitary alternative engineer can change which apparatuses ought to be empowered and which ought to be crippled. Its investigations give the foundational building squares to larger amount and area particular content comprehension applications.

University of Moratuwa, Sri Lanka.
Electronic Theses & Dissertations
www.lib.mrt.ac.lk

Stanford CoreNLP coordinates large portions of their NLP apparatuses, including the grammatical feature (POS) tagger, the named substance recognizer (NER), the parser, the coreference determination framework, the feeling investigation, and the bootstrapped example learning devices. The fundamental appropriation gives model records to the investigation of English, however the engine is good with models for different languages.



University of Moratuwa, Sri Lanka.
Electronic Theses & Dissertations
www.lib.mrt.ac.lk

Figure 6.1: System Architecture of Stanford CoreNLP Toolkit

6.16 JADE Framework

JAVA Agent DEvelopment Framework (JADE) is an open source stage for distributed agent based applications. JADE is a product Framework completely executed in the Java language. It streamlines the usage of multi-agents systems through a center product that consents to the FIPA determinations and through an arrangement of graphical instruments that bolster the investigating and organization stages.

A JADE-based framework can be disseminated crosswise over machines (which not even need to have the same OS) and the arrangement can be controlled by means of a remote GUI. The setup can be even changed at run-time by moving agents starting

with one machine then onto the next, as and when required. JADE is totally executed in Java language and the negligible framework prerequisite is the adaptation 5 of JAVA (the run time environment or the JDK).

Other than the agents deliberation, JADE gives a basic yet effective undertaking execution and piece model, shared agents correspondence in view of the nonconcurrent message passing worldview, a business catalog administration supporting distribute subscribe disclosure instrument and numerous other propelled highlights that encourages the improvement of a disseminated framework.

On account of the commitment of the LEAP venture, specially appointed renditions of JADE exist intended to convey JADE agents straightforwardly on diverse Java-situated situations, for example, Android gadgets and J2ME-CLDC MIDP 1.0 gadgets.

Besides suitable setups can be determined to run JADE agents in systems described by incomplete availability including NAT and firewalls and additionally irregular scope and IP-location changes.

University of Moratuwa, Sri Lanka.
Electronic Theses & Dissertations
www.lib.mrt.ac.lk

JADE is free programming and is circulated by Telecom Italia, the copyright holder, in open source under the terms and states of the LGPL (Lesser General Public License Version 2) permit. Other than the JADE Team, be that as it may, a genuinely huge Community of designers assembled around the JADE Framework in these years.

6.17 Summary

This chapter discussed about the high level implementation details of the Agent-based Solution for Improving Abstracts which has been implemented using natural language processing techniques, multi agent systems and ontology. Next chapter will be on the evaluation of the Agent-based Solution for Improving Abstracts.

Chapter 7

Evaluation of the Proposed Solution

7.1 Introduction

This chapter reports on how evaluated the proposed Agent-based Solution for Improving Abstracts to see whether the objectives have been achieved. Experimental design, selection of inputs, outputs, etc. Here also present the results from the evaluation. This chapter analyzes the input and output from the proposed solution.

7.2 Evaluating the Overall System

Overall system has been evaluated by using real world scenarios to verify the accuracy and the performances of the Agent-based Solution for Improving Abstracts system with the help of the human experts.

University of Moratuwa, Sri Lanka.
Electronic Theses & Dissertations
www.lib.mrt.ac.lk

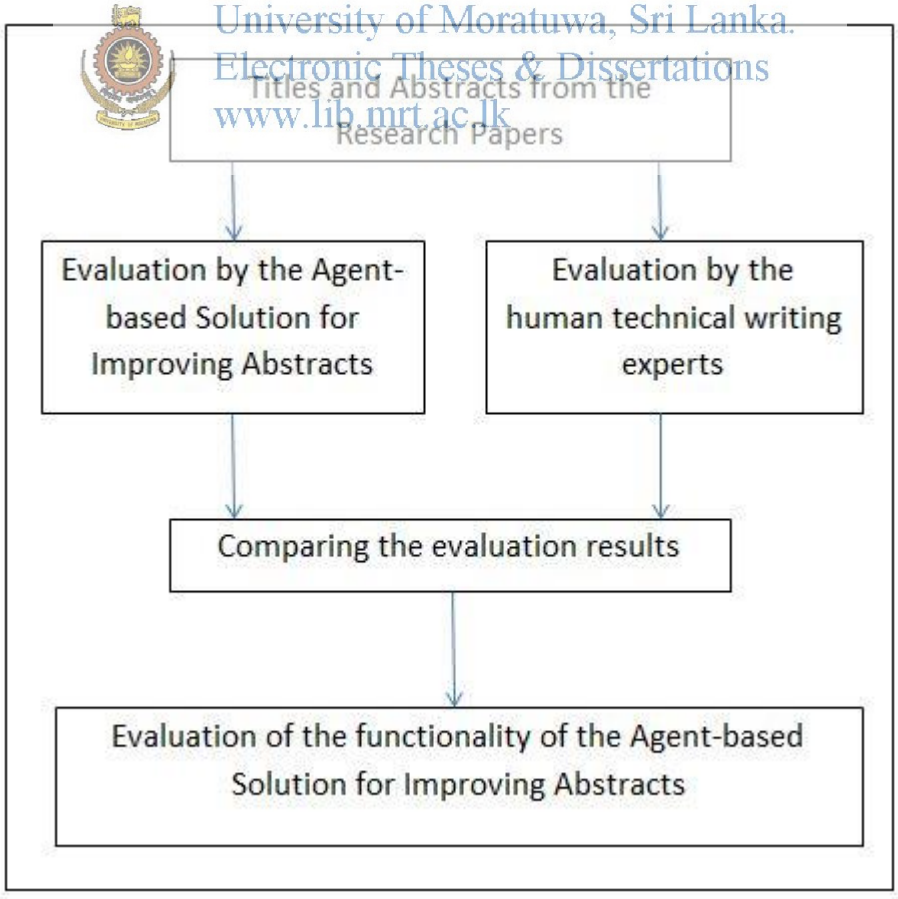
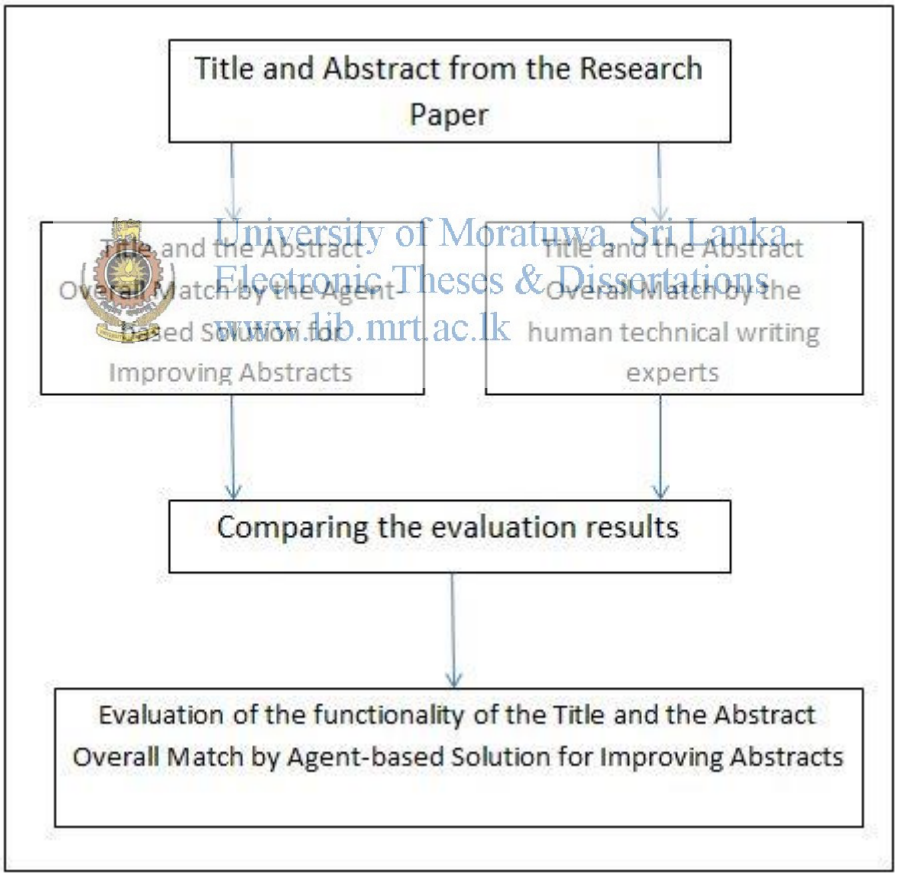


Figure 7.1: Evaluation process of the proposed solution

As this proposed solution deals with more subjective areas, evaluation is much more needed to verify the solutions provided in this system. Many evaluations were carried out on the various parts of the system with the help of the human domain experts.

7.3 Evaluating the Title and Abstract Overall Match

Overall match between the title and the abstract has been evaluated by entering sample abstracts and their titles which have been taken from real world well recognized research papers to the system. System has been evaluated by entering the titles of it's and title of other research papers against a particular abstract. Overall matching scores have been compared to evaluate the title abstract matching accuracy.



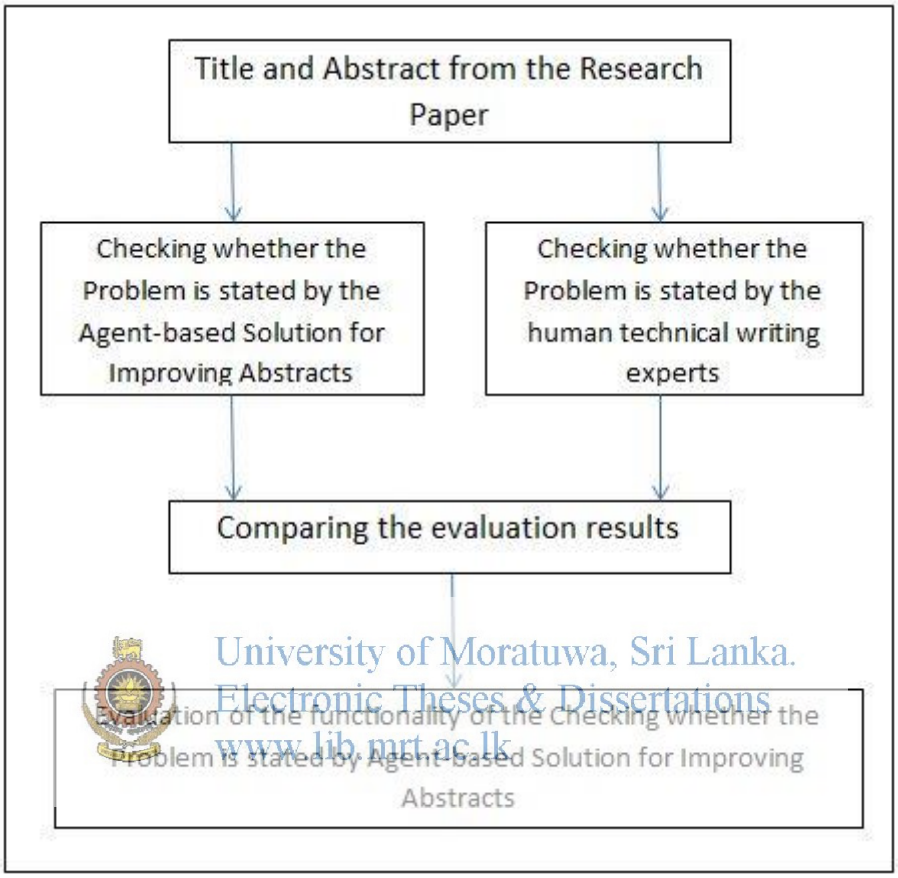
University of Moratuwa, Sri Lanka.
Electronic Theses & Dissertations
www.lib.mrt.ac.lk

Figure 7.2: Evaluation process of the Title and Abstract Overall Match

7.4 Evaluating the Checking whether the Problem is stated

The Agent-based Solution for Improving Abstracts system has been evaluated for its accuracy in checking whether the problem is stated in the abstract by inserting

abstracts from the real world well recognized research papers with and without the problem statement parts. The system has been able to correctly recognize whether the problem is stated in the abstract or not.



University of Moratuwa, Sri Lanka.
Electronic Theses & Dissertations
www.lib.mrt.ac.lk

Figure 7.3: Evaluation process of the Checking whether the Problem is stated

7.5 Evaluating the Checking whether the Solution is stated

The Agent-based Solution for Improving Abstracts system has been evaluated for its accuracy in checking whether the solution is stated in the abstract by inserting abstracts from the real world well recognized research papers with and without the solution statement parts. The system has been able to correctly recognize whether the solution is stated in the abstract or not.

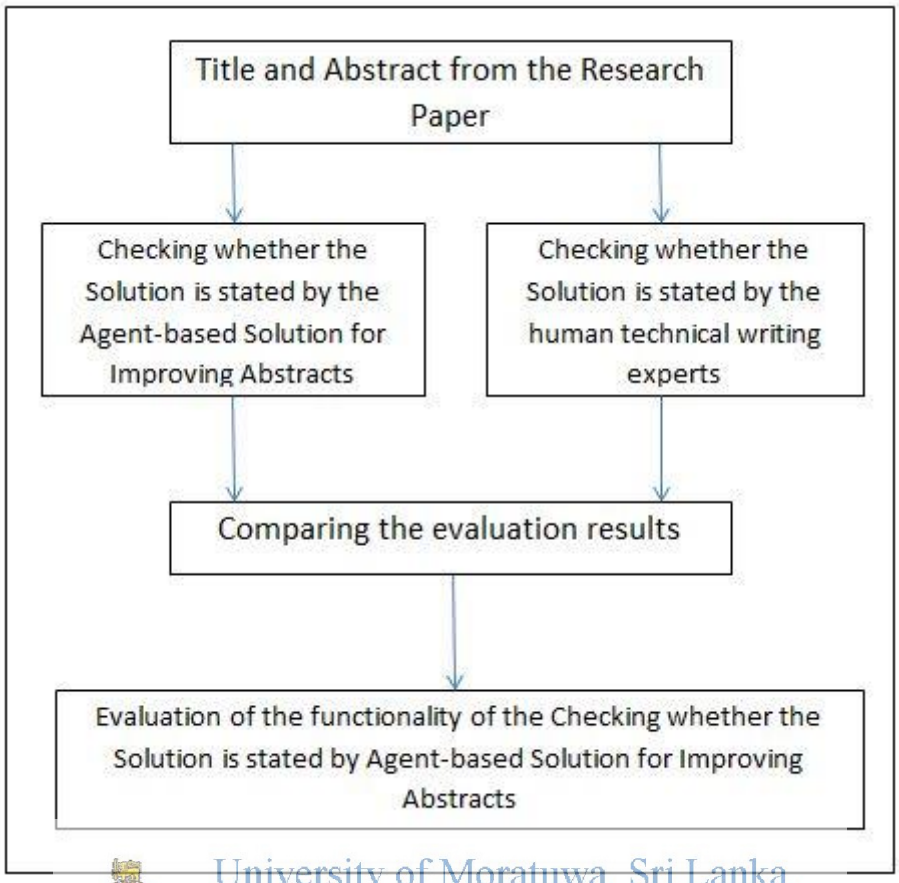
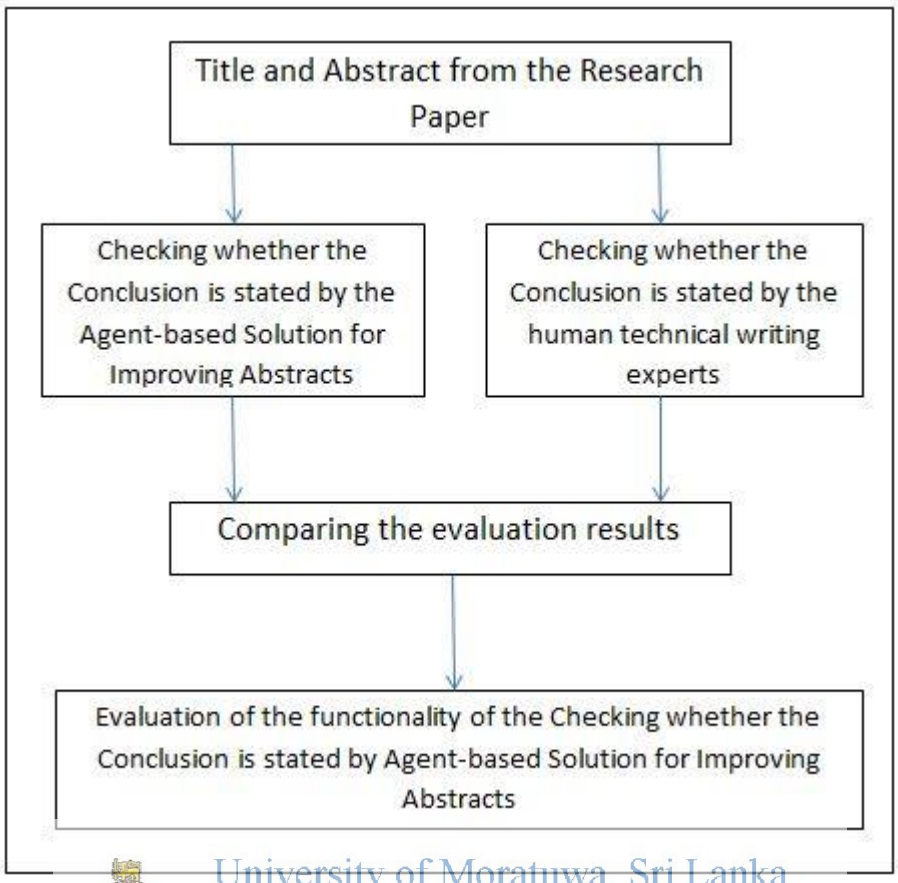


Figure 7.4: Evaluation process of the Checking whether the Solution is stated

University of Moratuwa, Sri Lanka.
Electronic Theses & Dissertations
www.lib.mrt.ac.lk

7.6 Evaluating the Checking whether the Conclusion is stated

The Agent-based Solution for Improving Abstracts system has been evaluated for its accuracy in checking whether the conclusion is stated in the abstract by inserting abstracts from the real world well recognized research papers with and without the conclusion statement parts. The system has been able to correctly recognize whether the conclusion is stated in the abstract or not.

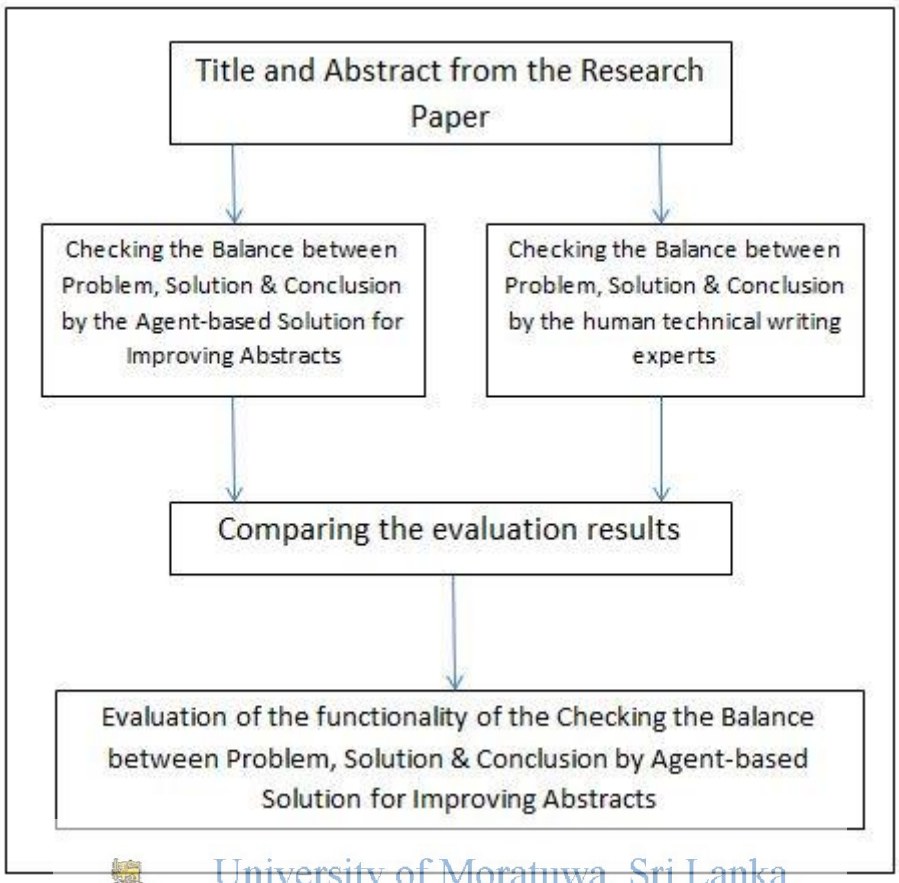


University of Moratuwa, Sri Lanka.
Electronic Theses & Dissertations
www.lib.mrt.ac.lk

Figure 7.5: Evaluation process of the Checking whether the Conclusion is stated

7.7 Evaluating the Checking the Balance between Problem, Solution & Conclusion

The Agent-based Solution for Improving Abstracts system has been evaluated for its accuracy in checking the balance between the problem, solution and conclusion parts stated in the abstract by inserting abstracts from the real world well recognized research papers with and without the correct balance in between the problem, solution & conclusion parts. The system has been able to correctly recognize whether the correct balance is there in the abstract in between the problem, solution & conclusion or not.

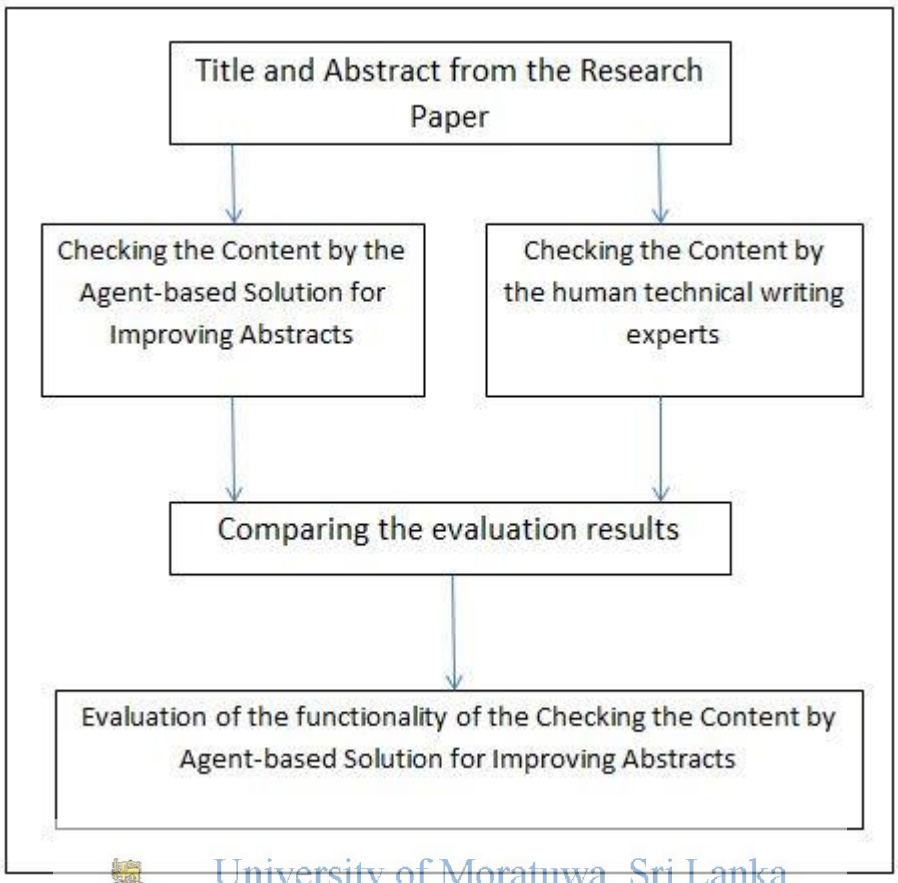


University of Moratuwa, Sri Lanka.
Electronic Theses & Dissertations
www.lib.mrt.ac.lk

Figure 7.6: Evaluation process of the Checking the Balance between Problem, Solution & Conclusion

7.8 Evaluating the Checking the Content

The Agent-based Solution for Improving Abstracts system has been evaluated for its accuracy in checking the content in the abstract by inserting the abstracts from the real world well recognized research papers with and without the correct usage of suitable words in the content. The system has been able to correctly recognize whether the correct usage of suitable words is there in the abstract content or not.

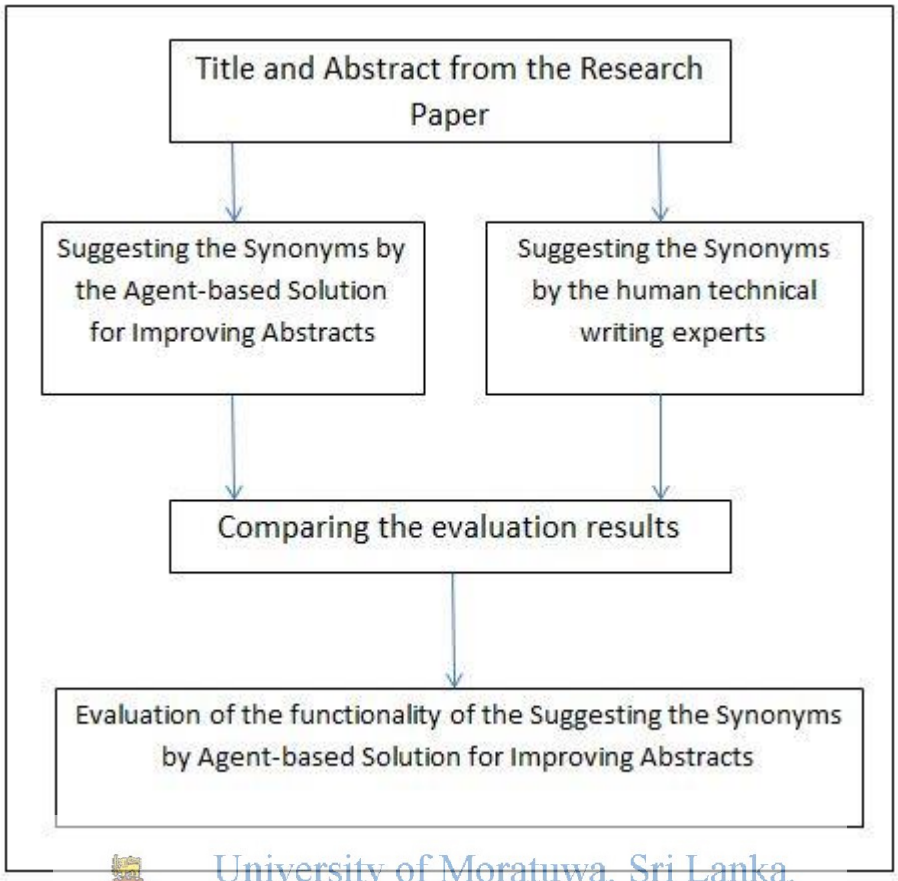


University of Moratuwa, Sri Lanka.
Electronic Theses & Dissertations
www.lib.mrt.ac.lk

Figure 7.7: Evaluation process of the Checking the Content

7.9 Evaluating the Suggesting the Synonyms

The Agent-based Solution for Improving Abstracts system has been evaluated for its accuracy in suggesting the synonyms for the words used in the abstract by inserting the abstracts from the real world well recognized research papers. The system has been able to correctly suggest the synonyms for the words used in the abstract.



University of Moratuwa, Sri Lanka.
Electronic Theses & Dissertations
www.lib.mrt.ac.lk

Figure 7.8: Evaluation process of the Suggesting the Synonyms

7.10 Evaluating the Suggesting the Improvements

The Agent-based Solution for Improving Abstracts system has been evaluated for its accuracy in suggesting the improvements for the content in the abstract by inserting the abstracts from the real world well recognized research papers. The system has been able to correctly suggest the improvements for the content in the abstract in terms of spellings and grammar.

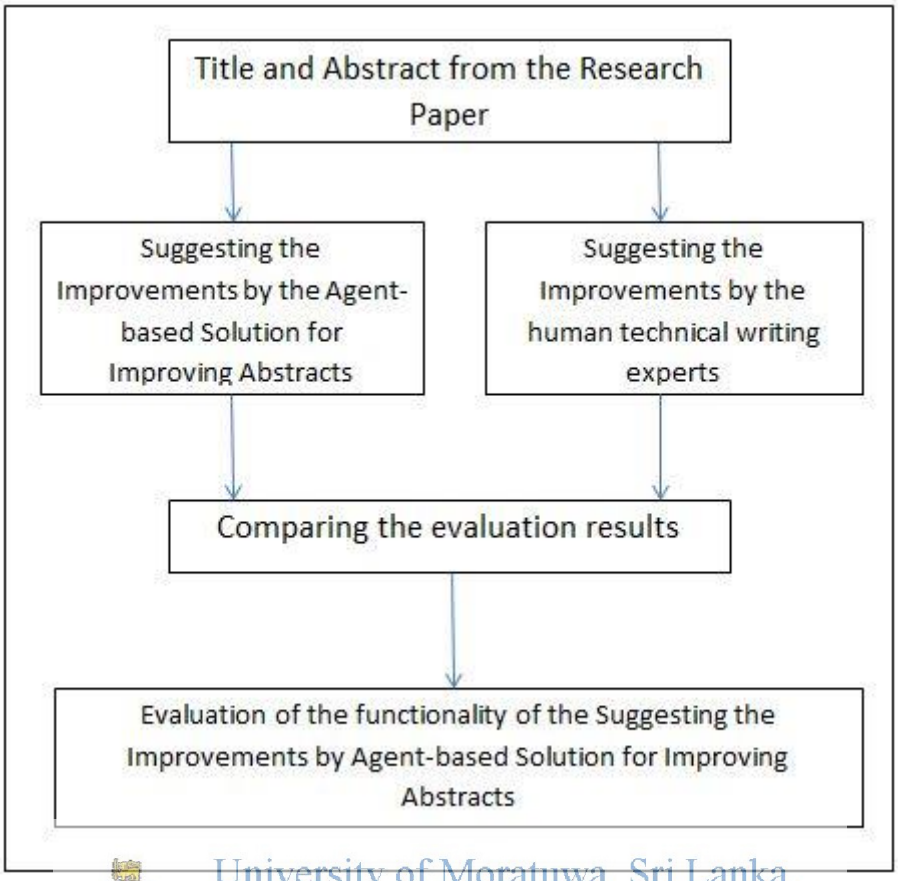
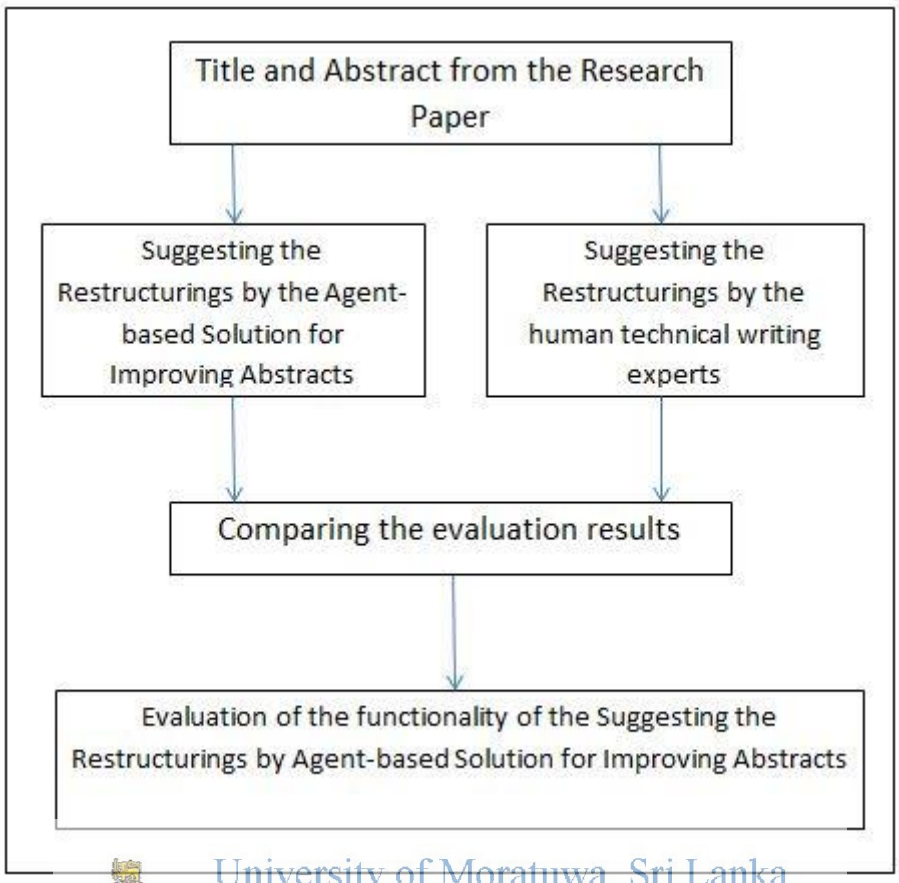


Figure 7.9: Evaluation process of the Suggesting the Improvements

University of Moratuwa, Sri Lanka.
Electronic Theses & Dissertations
www.lib.mrt.ac.lk

7.11 Evaluating the Suggesting the Restructurings

The Agent-based Solution for Improving Abstracts system has been evaluated for its accuracy in suggesting the restructurings for the content in the abstract by inserting the abstracts from the real world well recognized research papers. The system has been able to correctly suggest the restructurings for the content in the abstract in terms of balance and the richness of the problem, solution and conclusion statements.



University of Moratuwa, Sri Lanka.
Electronic Theses & Dissertations
www.lib.mrt.ac.lk

Figure 7.10: Evaluation process of the Suggesting the Restructurings

7.12 Evaluating the Overall System

Overall system has been evaluated by using real world scenarios to verify the accuracy and the performances of the Agent-based Solution for Improving Abstracts system with the help of the human experts.

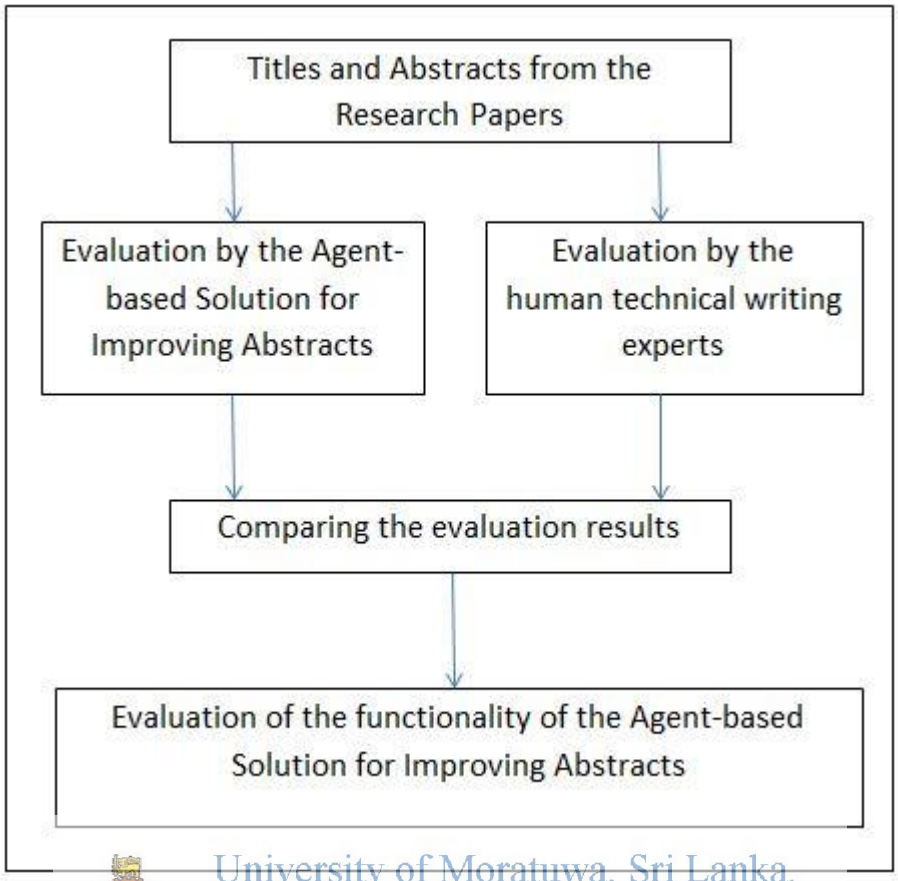


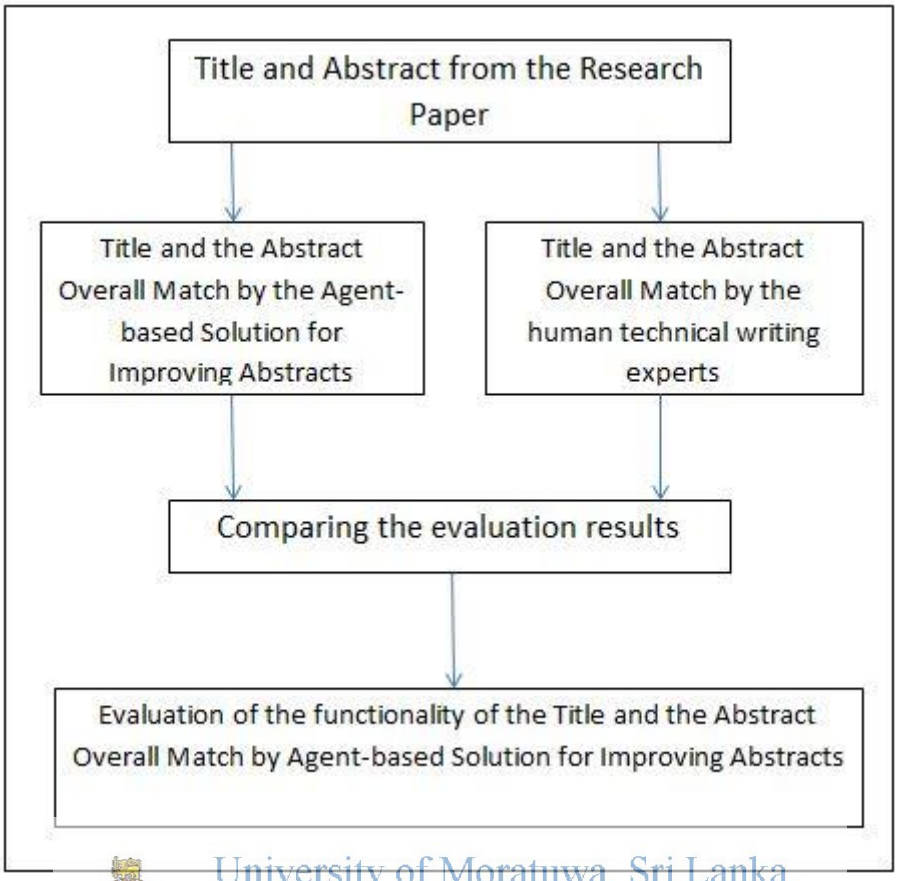
Figure 7.11: Evaluation process of the proposed solution

University of Moratuwa, Sri Lanka.
Electronic Theses & Dissertations
www.lib.mrt.ac.lk

As this proposed solution deals with more subjective areas, evaluation is much more needed to verify the solutions provided in this system. Many evaluations were carried out on the various parts of the system with the help of the human domain experts.

7.13 Evaluating the Title and Abstract Overall Match

Overall match between the title and the abstract has been evaluated by entering sample abstracts and their titles which have been taken from real world well recognized research papers to the system. System has been evaluated by entering the titles of it's and title of other research papers against a particular abstract. Overall matching scores have been compared to evaluate the title abstract matching accuracy.

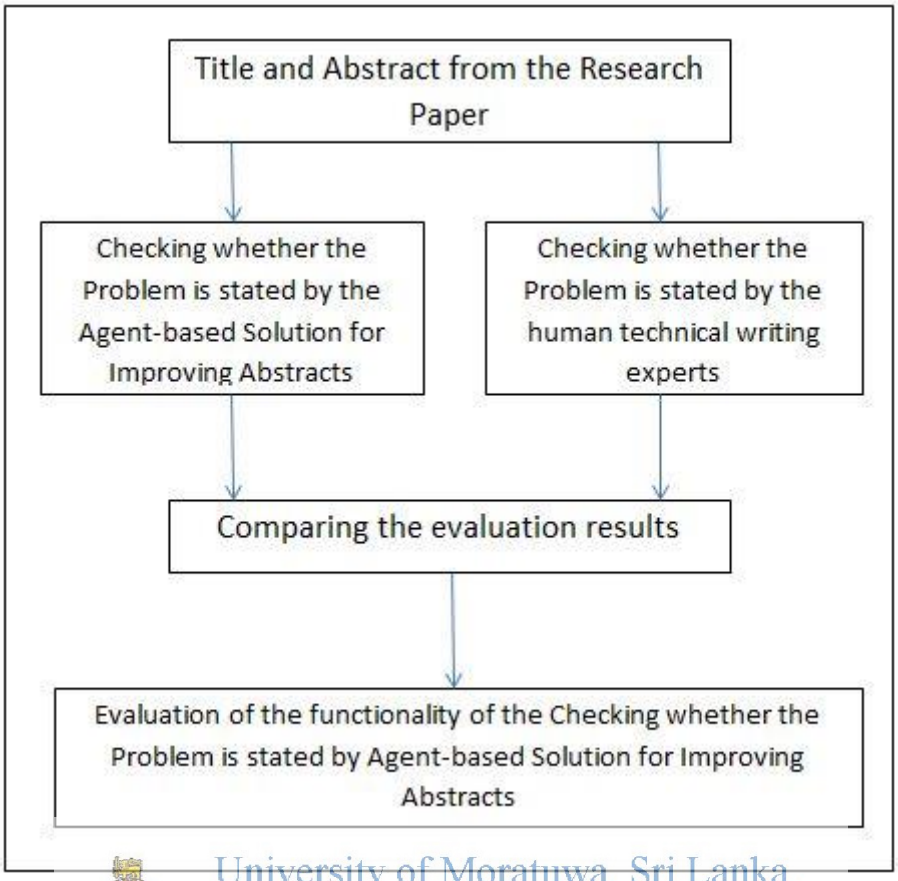


University of Moratuwa, Sri Lanka.
Electronic Theses & Dissertations
www.lib.mrt.ac.lk

Figure 7.12: Evaluation process of the Title and Abstract Overall Match

7.14 Evaluating the Checking whether the Problem is stated

The Agent-based Solution for Improving Abstracts system has been evaluated for its accuracy in checking whether the problem is stated in the abstract by inserting abstracts from the real world well recognized research papers with and without the problem statement parts. The system has been able to correctly recognize whether the problem is stated in the abstract or not.

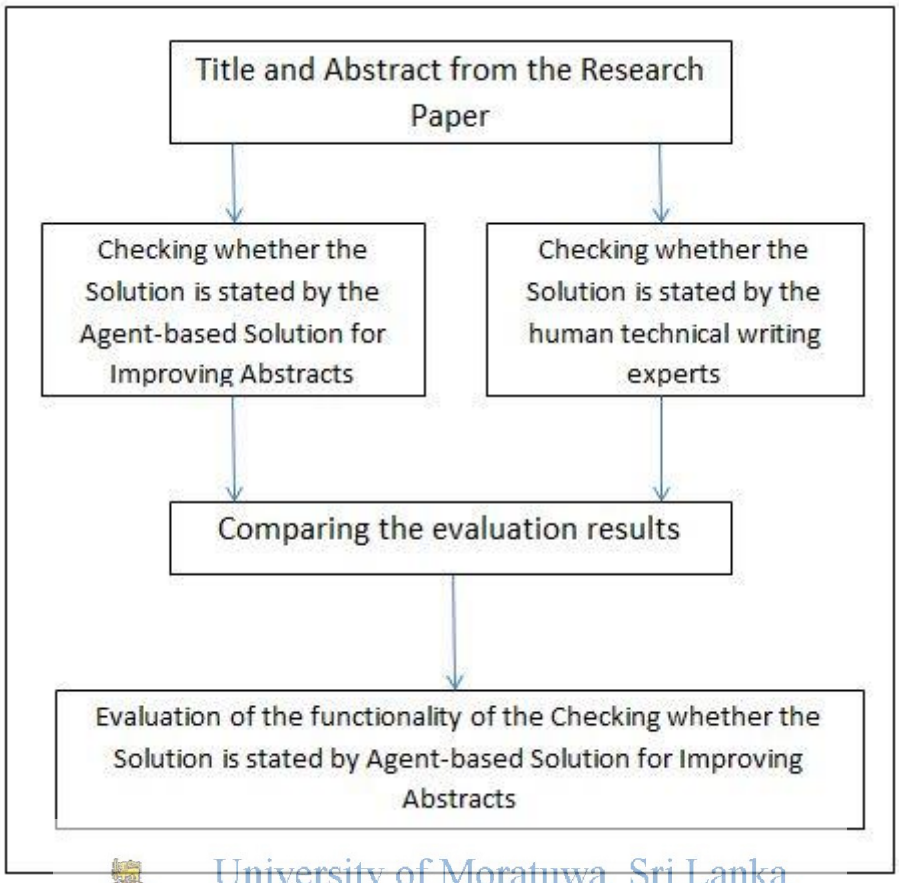


University of Moratuwa, Sri Lanka.
Electronic Theses & Dissertations
www.lib.mrt.ac.lk

Figure 7.13: Evaluation process of the Checking whether the Problem is stated

7.15 Evaluating the Checking whether the Solution is stated

The Agent-based Solution for Improving Abstracts system has been evaluated for its accuracy in checking whether the solution is stated in the abstract by inserting abstracts from the real world well recognized research papers with and without the solution statement parts. The system has been able to correctly recognize whether the solution is stated in the abstract or not.

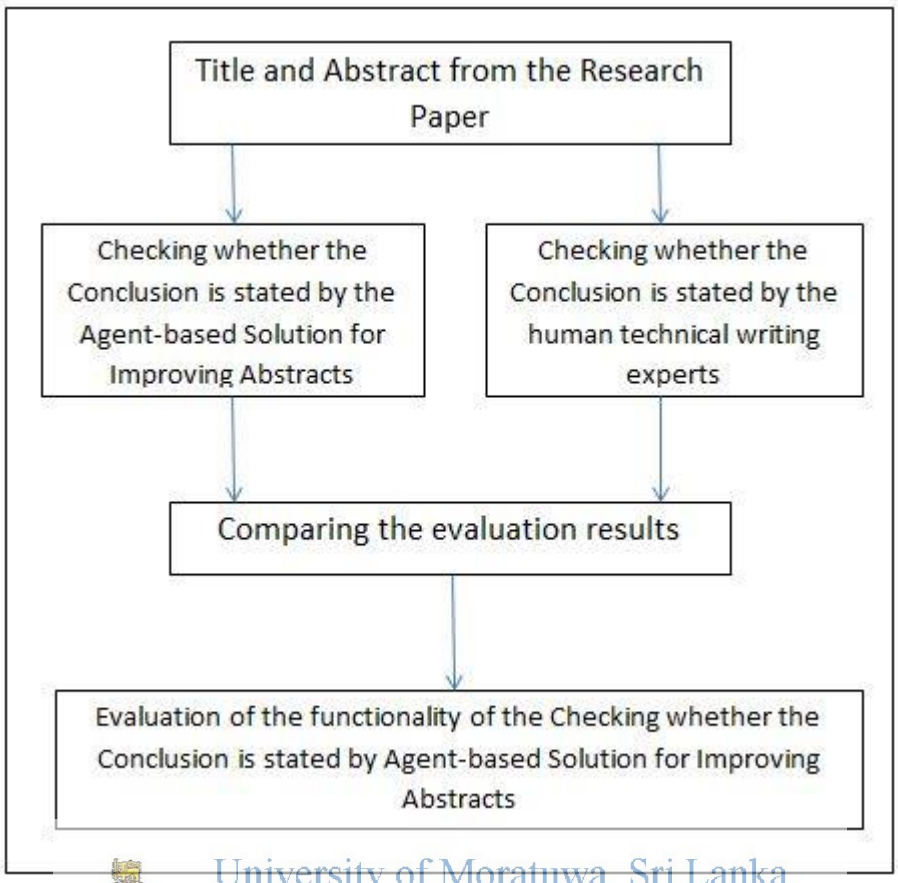


University of Moratuwa, Sri Lanka.
Electronic Theses & Dissertations
www.lib.mrt.ac.lk

Figure 7.14: Evaluation process of the Checking whether the Solution is stated

7.16 Evaluating the Checking whether the Conclusion is stated

The Agent-based Solution for Improving Abstracts system has been evaluated for its accuracy in checking whether the conclusion is stated in the abstract by inserting abstracts from the real world well recognized research papers with and without the conclusion statement parts. The system has been able to correctly recognize whether the conclusion is stated in the abstract or not.

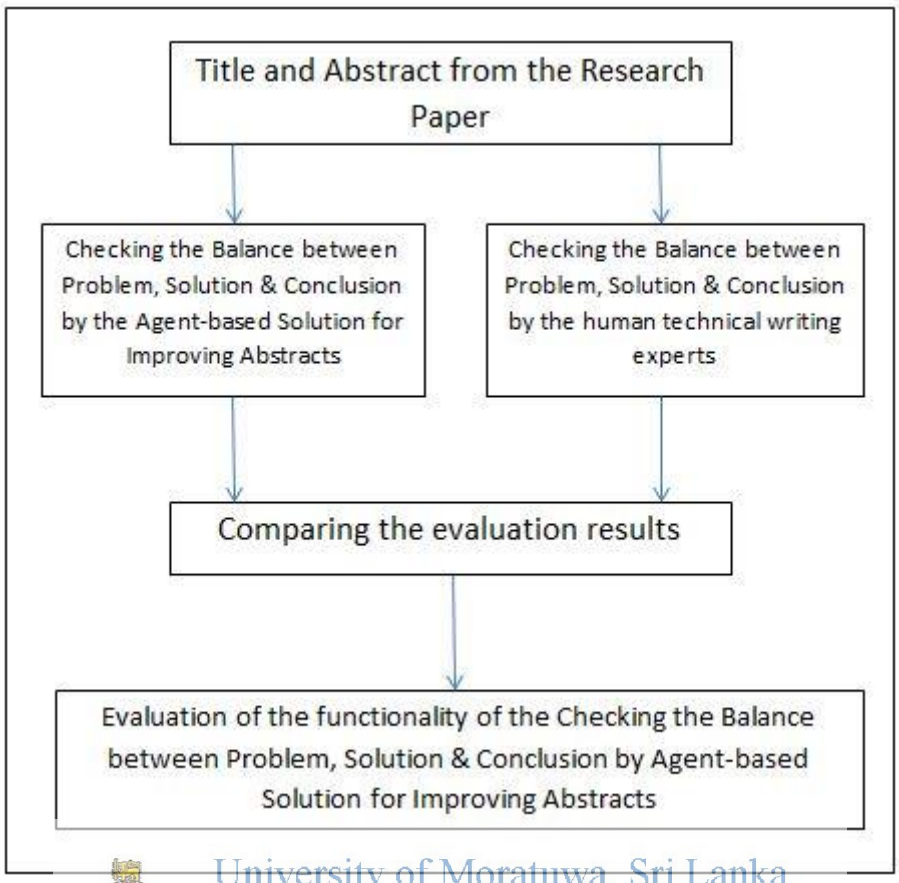


University of Moratuwa, Sri Lanka.
Electronic Theses & Dissertations
www.lib.mrt.ac.lk

Figure 7.15: Evaluation process of the Checking whether the Conclusion is stated

7.17 Evaluating the Checking the Balance between Problem, Solution & Conclusion

The Agent-based Solution for Improving Abstracts system has been evaluated for its accuracy in checking the balance between the problem, solution and conclusion parts stated in the abstract by inserting abstracts from the real world well recognized research papers with and without the correct balance in between the problem, solution & conclusion parts. The system has been able to correctly recognize whether the correct balance is there in the abstract in between the problem, solution & conclusion or not.

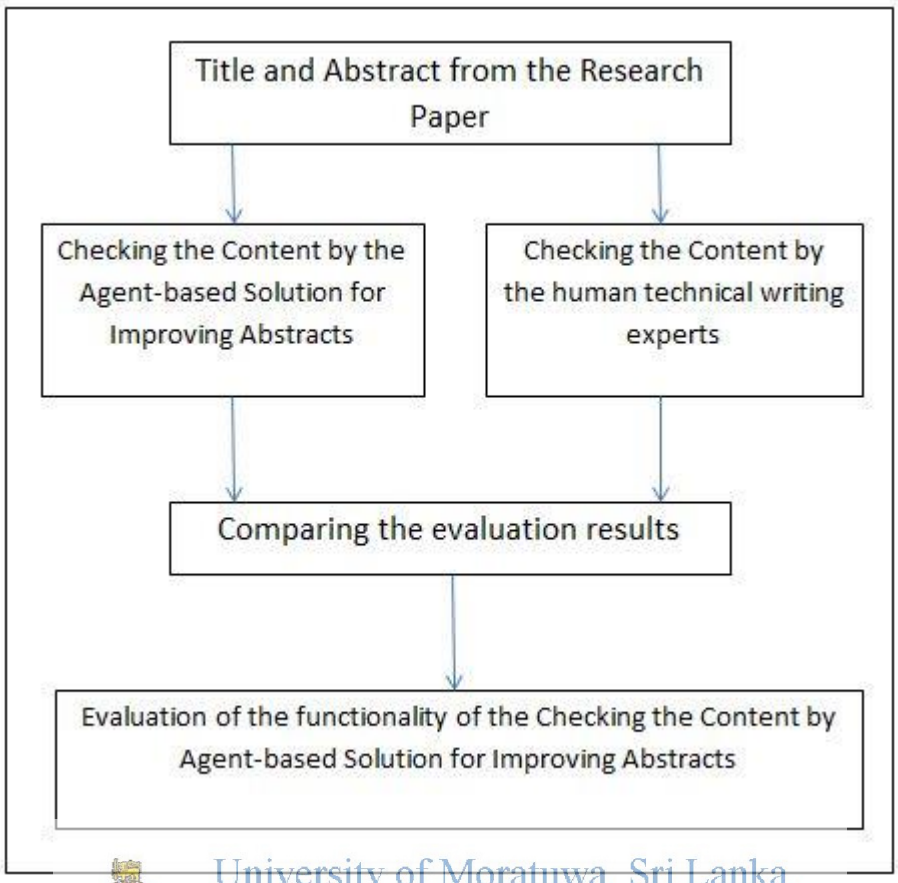


University of Moratuwa, Sri Lanka.
Electronic Theses & Dissertations
www.lib.mrt.ac.lk

Figure 7.16: Evaluation process of the Checking the Balance between Problem, Solution & Conclusion

7.18 Evaluating the Checking the Content

The Agent-based Solution for Improving Abstracts system has been evaluated for its accuracy in checking the content in the abstract by inserting the abstracts from the real world well recognized research papers with and without the correct usage of suitable words in the content. The system has been able to correctly recognize whether the correct usage of suitable words is there in the abstract content or not.

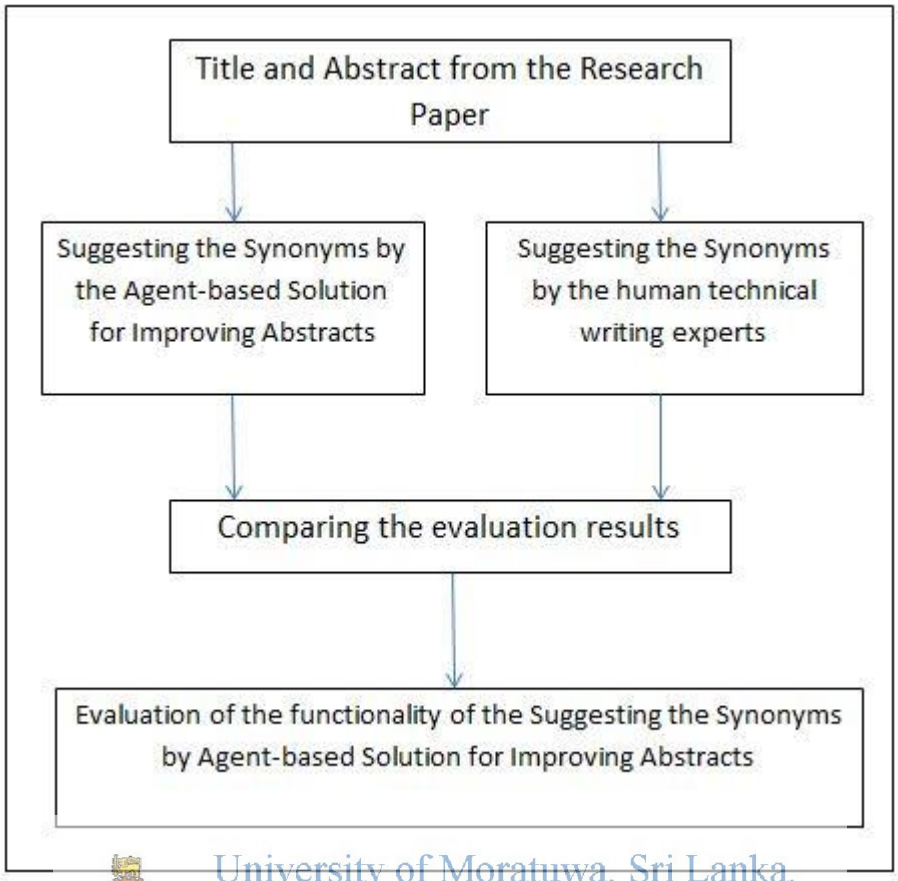


University of Moratuwa, Sri Lanka.
Electronic Theses & Dissertations
www.lib.mrt.ac.lk

Figure 7.17: Evaluation process of the Checking the Content

7.19 Evaluating the Suggesting the Synonyms

The Agent-based Solution for Improving Abstracts system has been evaluated for its accuracy in suggesting the synonyms for the words used in the abstract by inserting the abstracts from the real world well recognized research papers. The system has been able to correctly suggest the synonyms for the words used in the abstract.

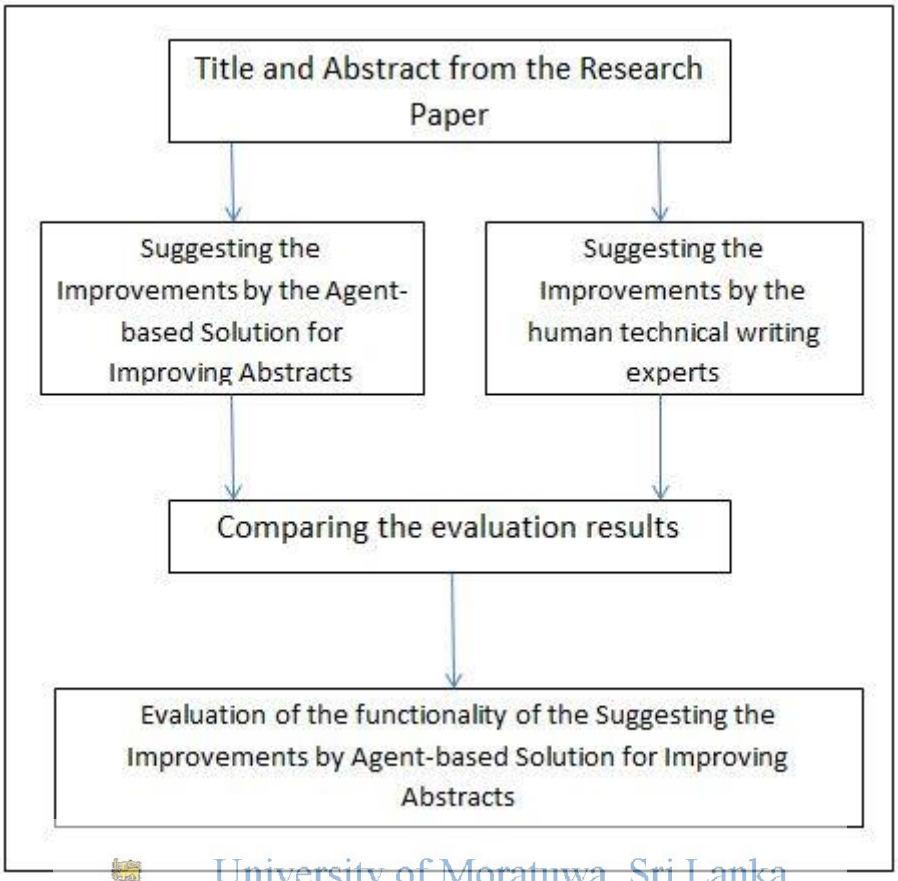


University of Moratuwa, Sri Lanka.
Electronic Theses & Dissertations
www.lib.mrt.ac.lk

Figure 7.18: Evaluation process of the Suggesting the Synonyms

7.20 Evaluating the Suggesting the Improvements

The Agent-based Solution for Improving Abstracts system has been evaluated for its accuracy in suggesting the improvements for the content in the abstract by inserting the abstracts from the real world well recognized research papers. The system has been able to correctly suggest the improvements for the content in the abstract in terms of spellings and grammar.

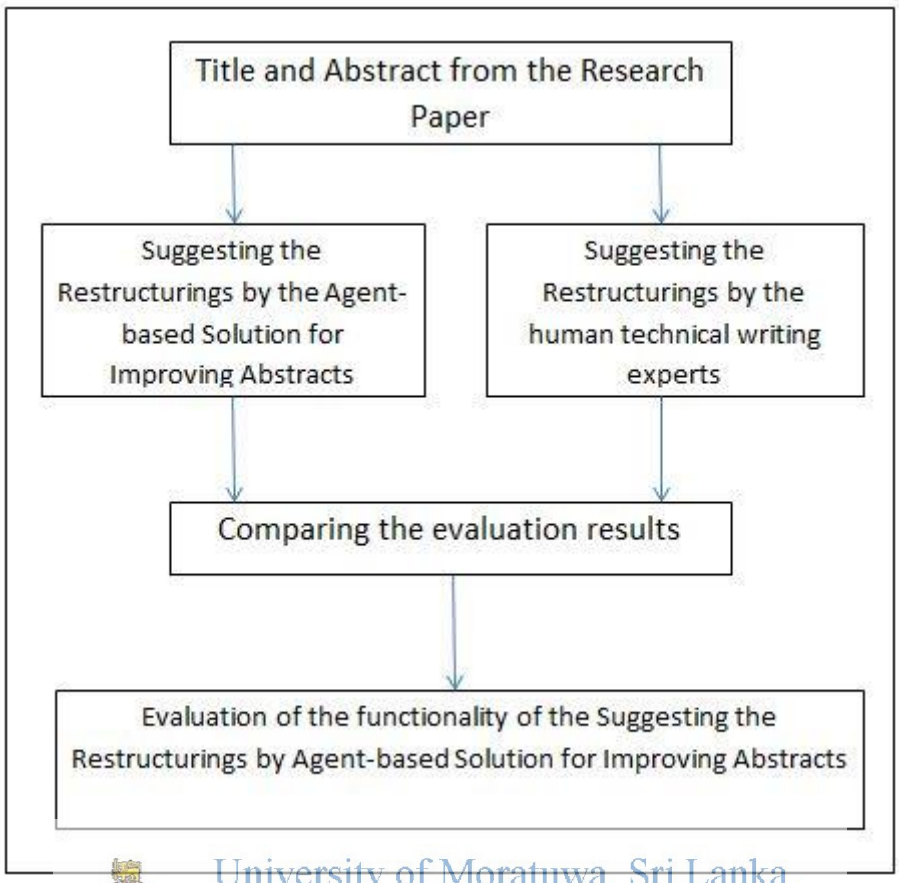


University of Moratuwa, Sri Lanka.
Electronic Theses & Dissertations
www.lib.mrt.ac.lk

Figure 7.19: Evaluation process of the Suggesting the Improvements

7.21 Evaluating the Suggesting the Restructurings

The Agent-based Solution for Improving Abstracts system has been evaluated for its accuracy in suggesting the restructurings for the content in the abstract by inserting the abstracts from the real world well recognized research papers. The system has been able to correctly suggest the restructurings for the content in the abstract in terms of balance and the richness of the problem, solution and conclusion statements.



University of Moratuwa, Sri Lanka.
Electronic Theses & Dissertations
www.lib.mrt.ac.lk

Figure 7.20: Evaluation process of the Suggesting the Restructurings

7.22 Summary

This chapter discussed about the evaluation details of the proposed Agent-based Solution for Improving Abstracts which has been implemented using natural language processing techniques, multi agent systems and ontology, to see whether the objectives have been achieved and analyzed the input and output data from the proposed solution. After evaluating the proposed system with different aspects, it can be concluded that the proposed system can be used at acceptable level with respect to features and performance. Next chapter will be on the conclusion and the further work.

Chapter 8

Conclusion and Further Work of the Research

8.1 Introduction

Here concludes the overall achievements quantitatively in the first place. Then state about the achievement of each objective. Also mention about problem encountered, limitations of solution, and some further work.

8.2 Conclusion

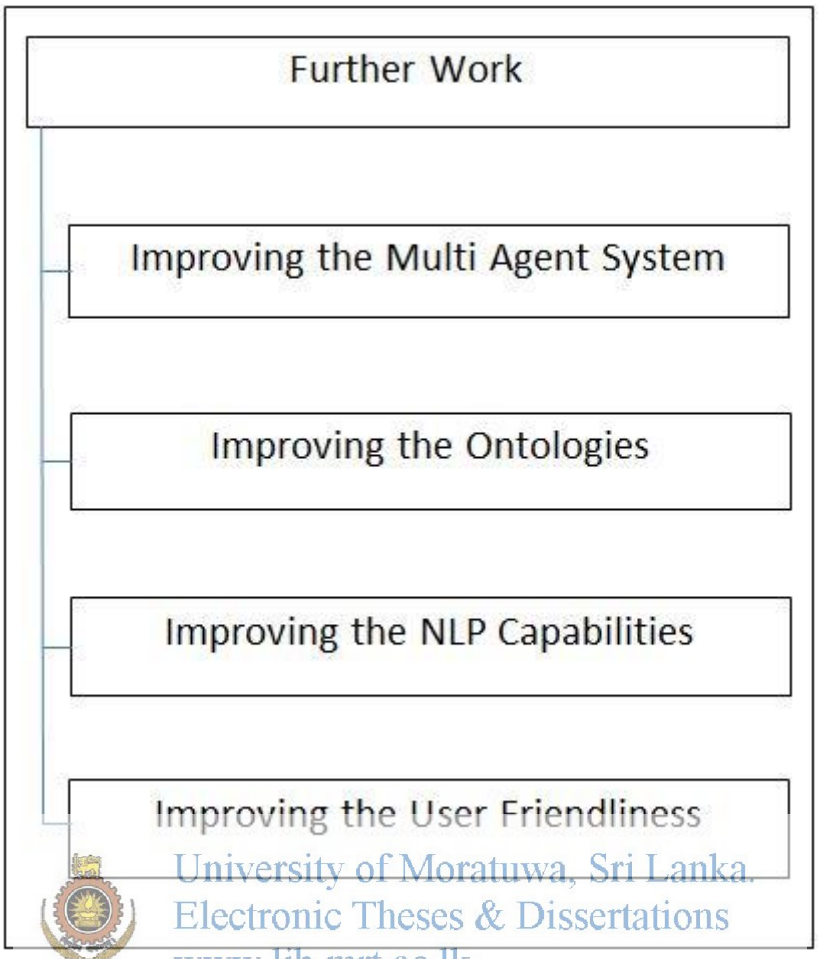
Research has been breakdown into multiple objectives. Main objective has been achieved by critically studying the improving abstracts with a view to identify current practices and issues. Next challenge was to critically analyze the existing solutions in Agent-based Solution for Improving Abstracts with the view to define the research problem and possible technology. After it was achieved then next step was in depth study about improving abstracts and its applications. Then moved to design and implement the Agent-based Solution for Improving Abstracts.

University of Moratuwa, Sri Lanka.
Electronic Theses & Dissertations
www.lib.mrt.ac.lk

Finally evaluated the improving the abstracts using the real world scenario to test and verify the accuracy and the performance of the proposed system. Every research open the paths and leads to few more other new researches and this research is also no difference. Since Artificial Intelligence techniques are new to improving the abstracts, this research will open lot of paths for new other researches. Even though this research has delivered all of its main objectives, there are a lot of areas can be identified to improve by accuracy and as well as performance wise.

8.3 Further Work

Agents in the Multi Agent System can be improved and the respective ontologies can be improved for increasing the accuracy and deriving the better performance. Also the user friendliness of the system can be improved to increase the end user satisfaction.



University of Moratuwa, Sri Lanka.
Electronic Theses & Dissertations
www.lib.mrt.ac.lk
Figure 8.1: Further Work of the Research

8.4 Improving the Multi Agent System

Multi Agent System in the proposed solution for improving the abstracts can be improved by increasing the capabilities of the remaining Agents in the Multi Agents System as well as introducing new Agents to the Multi Agents System. Agents' features could be sharpened by adding much more logic and increasing the communication among the Agents.

8.5 Improving the Ontologies

Proposed solution for improving the abstracts can be improved by improving the ontologies of the respective agents. Much more knowledge can be added to the remaining ontologies to improve the capabilities of those agents. So, those agents can make better decisions and can make better contributions for the improving abstracts.

8.6 Improving the Natural Language Processing Capabilities

Proposed solution for improving the abstracts can be improved by improving the natural language processing capabilities in the system. Much more efficient and capable natural language processing algorithms can be added to the proposed system to get a better output from the system. So, that would help agents to make better decisions and would be able to make better contributions for the improving abstracts.

8.7 Improving the User Friendliness

User friendliness of the proposed solution for improving the abstracts can be improved by adding much more user friendly features to the current system. By improving the user friendliness of the system, proposed solution can be improved significantly. Graphical User Interface of the current system can be improved for a better communication with the end users and can provide much more attractiveness to the end users.

University of Moratuwa, Sri Lanka.
Electronic Theses & Dissertations
www.lib.mrt.ac.lk

8.8 Summary

This chapter discussed about the conclusion and the further work details of the proposed Agent-based Solution for Improving Abstracts which has been developed using the natural language processing techniques, multi agent systems and ontology. By analyzing all the aims and objectives, it can be said the research has achieved its objectives effectively. Next chapter will list down the references used during this research of proposing an Agent-based Solution for Improving Abstracts.

Chapter 9

References

[1] Stuart Russell, Peter Norvig, “Artificial Intelligence: A Modern Approach”, 3rd ed. Prentice Hall, 2009.

[2] Helen Yannakoudakis, “Automated assessment of English-learner writing”, Technical Report Number 842, University of Cambridge, Computer Laboratory, October 2013

[3] Semire Dikli, “An Overview of Automated Scoring of Essays”, The Journal of Technology, Learning, and Assessment Volume 5, Number 1, August 2006

[4] Salvatore Valenti, Francesca Neri and Alessandro Cucchiarelli, “An Overview of Current Research on Automated Essay Grading”, Journal of Information Technology Education, Volume 2, 2003

[5] Felice Dell’Orletta, Martijn Wieling, Andrea Cimino, Giulia Venturi and Simonetta Montemagni, “Assessing the Readability of Sentences: Which Corpora and Features?”, Proceedings of the Ninth Workshop on Innovative Use of NLP for Building Educational Applications, pages 163–173, Baltimore, Maryland USA, June 26, 2014, Association for Computational Linguistics

University of Moratuwa, Sri Lanka.
Electronic Theses & Dissertations
www.lib.mrt.ac.lk

[6] Jill Burstein, Martin Chodorow and Claudia Leacock, “Automated Essay Evaluation: The Criterion Online Writing Service”, AI Magazine Volume 25 Number 3 (2004), American Association for Artificial Intelligence

[7] Rachele De Felice, Stephen G. Pulman, “Automatic detection of preposition errors in learner writing”, Oxford University Computing Laboratory, 19 March 2008

[8] Ryu Iida, Takenobu Tokunaga, “Building a Corpus of Manually Revised Texts from Discourse Perspective”, Graduate School of Information Science and Engineering, Tokyo Institute of Technology

[9] Siddhartha Ghosh, "Online Automated Essay Grading System as a Web Based Learning (WBL) Tool in Engineering Education", Chapter 5, G. Narayanamma Institute of Technology and Science, India

[10] Elizabeth Salesky, Wade Shen "Exploiting Morphological, Grammatical, and Semantic Correlates for Improved Text Difficulty Assessment", Proceedings of the Ninth Workshop on Innovative Use of NLP for Building Educational Applications,

pages 155–162, Baltimore, Maryland USA, June 26, 2014. Association for Computational Linguistics

[11] Attali, Y. and Burstein, J. (2006). Automated essay scoring with e-Rater v.2.0. Journal of Technology, Learning, and Assessment, 4(3):1 – 30.

[12] Landauer, T. K., Laham, D., and Foltz, P. W. (2003). Automated scoring and annotation of essays with the Intelligent Essay Assessor. In Shermis, M. and Burstein, J. C., editors, Automated essay scoring: A cross-disciplinary perspective, pages 87 - 112.

[13] Briscoe, T., Medlock, B., and Andersen, . E. (2010). Automated assessment of ESOL free text examinations. Technical Report UCAM-CL-TR-790, University of Cambridge, Computer Laboratory, http://www.cl.cam.ac.uk/techreports/UCAM-CL-TR-790.pdf.

[14] Rudner, L. and Liang, T. (2002). Automated essay scoring using Bayes' theorem. The Journal of Technology, Learning and Assessment, 1(2):3 - 21.

[15] Higgins, D., Burstein, J., and Attali, Y. (2006). Identifying off-topic student essays without topic-specific training data. Natural Language Engineering, 12(2):145 - 159.

University of Moratuwa, Sri Lanka.
Electronic Theses & Dissertations
www.lib.mrt.ac.lk

[16] Lonsdale, D. and Strong-Krause, D. (2003). Automated rating of ESL essays. In Proceedings of the HLT-NAACL 2003 workshop on Building Educational Applications Using Natural Language Processing, pages 61 - 67. ACL.

[17] Nwana, H. (1996), “Software Agents: An Overview”, The Knowledge Engineering Review, 11 (3), 205-244.

[18] Nwana, H & Azarmi, N. (eds) (1997), “Software Agent and Soft Computing: towards enhancing machine intelligence”, Lecture Notes in Artificial Intelligence 1198, Springer-Verlag, New York.

[19] Hyacinth S. Nwana and Divine T. Ndumu, "A Perspective on Software Agents Research".

[20] Nwana, H. S. (1990), “Intelligent Tutoring Systems: An Overview”, Artificial Intelligence Review 4 (4).

[21] Nwana, H. S., Ndumu, D. T., Lee, L. C. & Collis, J. C. (1999), “ZEUS: A Toolkit for Building Distributed Multi-Agent Systems”, Applied Artificial Intelligence Journal 13 (1/2), 129-185.

[22] Nwana, H.S., Lee, L. Jennings, N.R. (1996), “Coordination in software agent systems”, BT Technology Journal, 14(4) (1996) 79–88.

[23] Nwana. H. S. & Ndumu, D. T. (1999), (eds)., Special Issue of Applied Artificial Intelligence Journal 13 (1/2).

[24] B. Chandrasekaran and John R. Josephson, Ohio State University, V. Richard Benjamins, University of Amsterdam, "What Are Ontologies, and Why Do We Need Them?", IEEE INTELLIGENT SYSTEMS, JANUARY/FEBRUARY 1999

[25] N. Guarino and R. Poli, “The Role of Ontology in the Information Technology,” Int’l J. Human-Computer Studies, Vol. 43, Nos. 5/6, Nov.-Dec. 1995, pp. 623–965.

[26] G. Van Heijst, A.T. Schreiber, and B.J. Wielinga, “Using Explicit Ontologies in KBS Development,” Int’l J. Human-Computer Studies, Vol. 46, Nos. 2/3, Feb.-Mar. 1997, pp. 183–292.

[27] M. Uschold and A. Tate, “Putting Ontologies to Use,” Knowledge Eng. Rev., Vol. 13, No. 1, Mar. 1998, pp. 1–3.

University of Moratuwa, Sri Lanka.
Electronic Theses & Dissertations
www.lib.mrt.ac.lk

Appendix A:

Software Code

A.1 Introduction

Software code used to implement the Agent-based Solution for Improving Abstracts is provided in a CD along with this thesis and the description of the source code is presented here.

A.2 Coordination Agent

Software code used to implement the coordination agent of the Agent-based Solution for Improving Abstracts is presented here.

A.3 Parser Agent

Software code used to implement the parser agent of the Agent-based Solution for Improving Abstracts is presented here.

University of Moratuwa, Sri Lanka.
Electronic Theses & Dissertations
www.lib.mrt.ac.lk

A.4 Problem Agent

Software code used to implement the problem agent of the Agent-based Solution for Improving Abstracts is presented here.

A.5 Solution Agent

Software code used to implement the solution agent of the Agent-based Solution for Improving Abstracts is presented here.

A.6 Conclusion Agent

Software code used to implement the conclusion agent of the Agent-based Solution for Improving Abstracts is presented here.

A.7 Content Agent

Software code used to implement the content agent of the Agent-based Solution for Improving Abstracts is presented here.

A.8 Synonym Agent

Software code used to implement the synonym agent of the Agent-based Solution for Improving Abstracts is presented here.

A.9 Improvement Agent

Software code used to implement the improvement agent of the Agent-based Solution for Improving Abstracts is presented here.

A.10 Restructure Agent

Software code used to implement the restructure agent of the Agent-based Solution for Improving Abstracts is presented here.

A.11 Abstract Concept

Software code used to implement the abstract concept of the Agent-based Solution for Improving Abstracts is presented here.

University of Moratuwa, Sri Lanka.
Electronic Theses & Dissertations
www.lib.mrt.ac.lk

A.12 Message Concept

Software code used to implement the message concept of the Agent-based Solution for Improving Abstracts is presented here.

A.13 Parser Ontology

Software code used to implement the parser ontology of the Agent-based Solution for Improving Abstracts is presented here.

A.14 Problem Ontology

Software code used to implement the problem ontology of the Agent-based Solution for Improving Abstracts is presented here.

A.15 Solution Ontology

Software code used to implement the solution ontology of the Agent-based Solution for Improving Abstracts is presented here.

A.16 Conclusion Ontology

Software code used to implement the conclusion ontology of the Agent-based Solution for Improving Abstracts is presented here.

A.17 Content Ontology

Software code used to implement the content ontology of the Agent-based Solution for Improving Abstracts is presented here.

A.18 Synonym Ontology

Software code used to implement the synonym ontology of the Agent-based Solution for Improving Abstracts is presented here.

University of Moratuwa, Sri Lanka.
Electronic Theses & Dissertations
www.lib.mrt.ac.lk

A.19 Improvement Ontology

Software code used to implement the improvement ontology of the Agent-based Solution for Improving Abstracts is presented here.

A.20 Restructure Ontology

Software code used to implement the restructure ontology of the Agent-based Solution for Improving Abstracts is presented here.

A.21 Abstract GUI

Software code used to implement the graphical user interface of the Agent-based Solution for Improving Abstracts is presented here.

A.22 Text Area Output Stream

Software code used to implement the redirecting the output stream to text area of the Agent-based Solution for Improving Abstracts is presented here.

Appendix B:

User Interfaces of the Proposed System

B.1 Introduction

Screen shots of the user interfaces of the proposed system, Agent-based Solution for Improving Abstracts are presented here.

B.2 Main User Interface of ASIA

Screen shot of the Main User Interface of the proposed system, Agent-based Solution for Improving Abstracts is presented here.

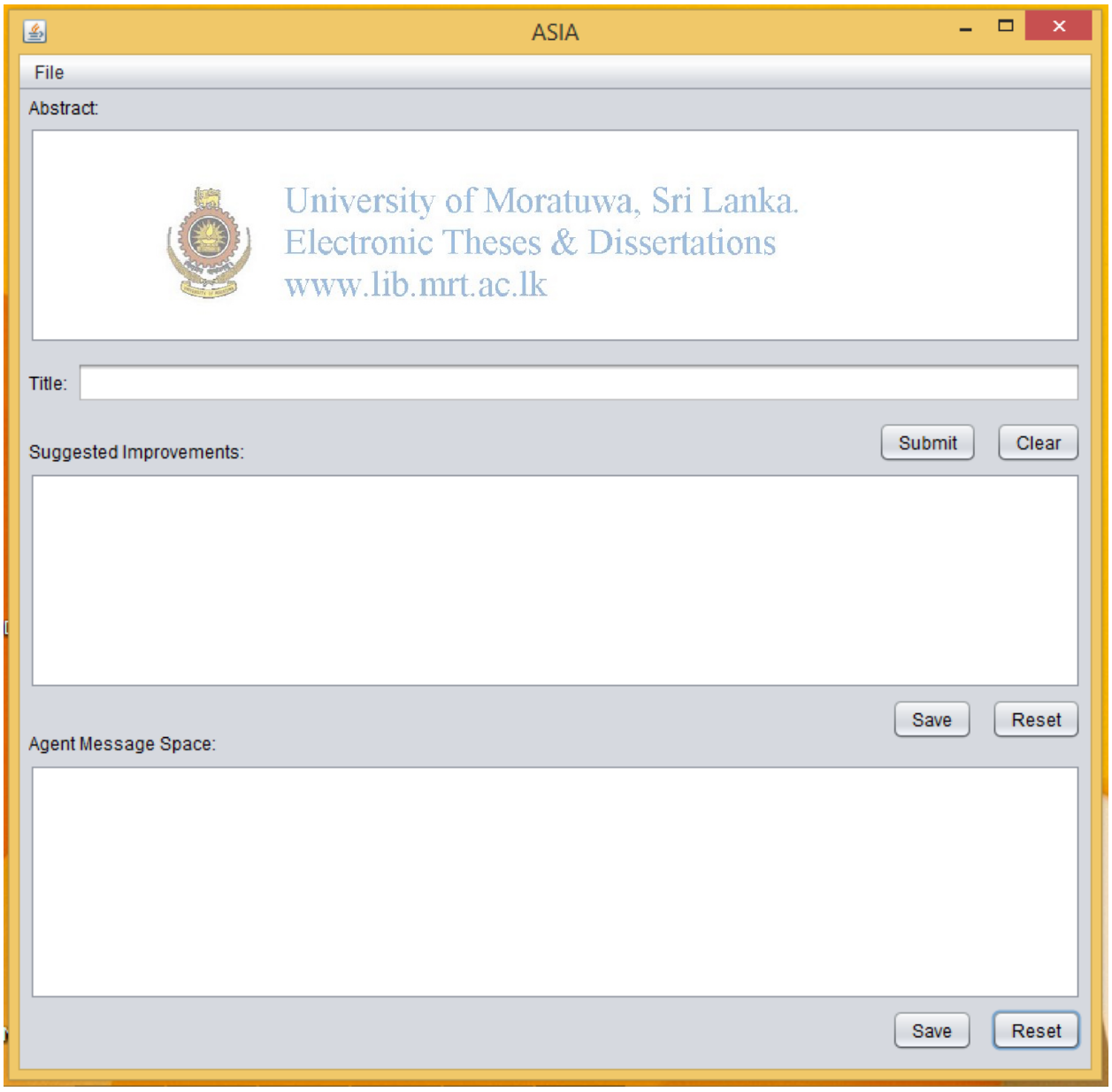
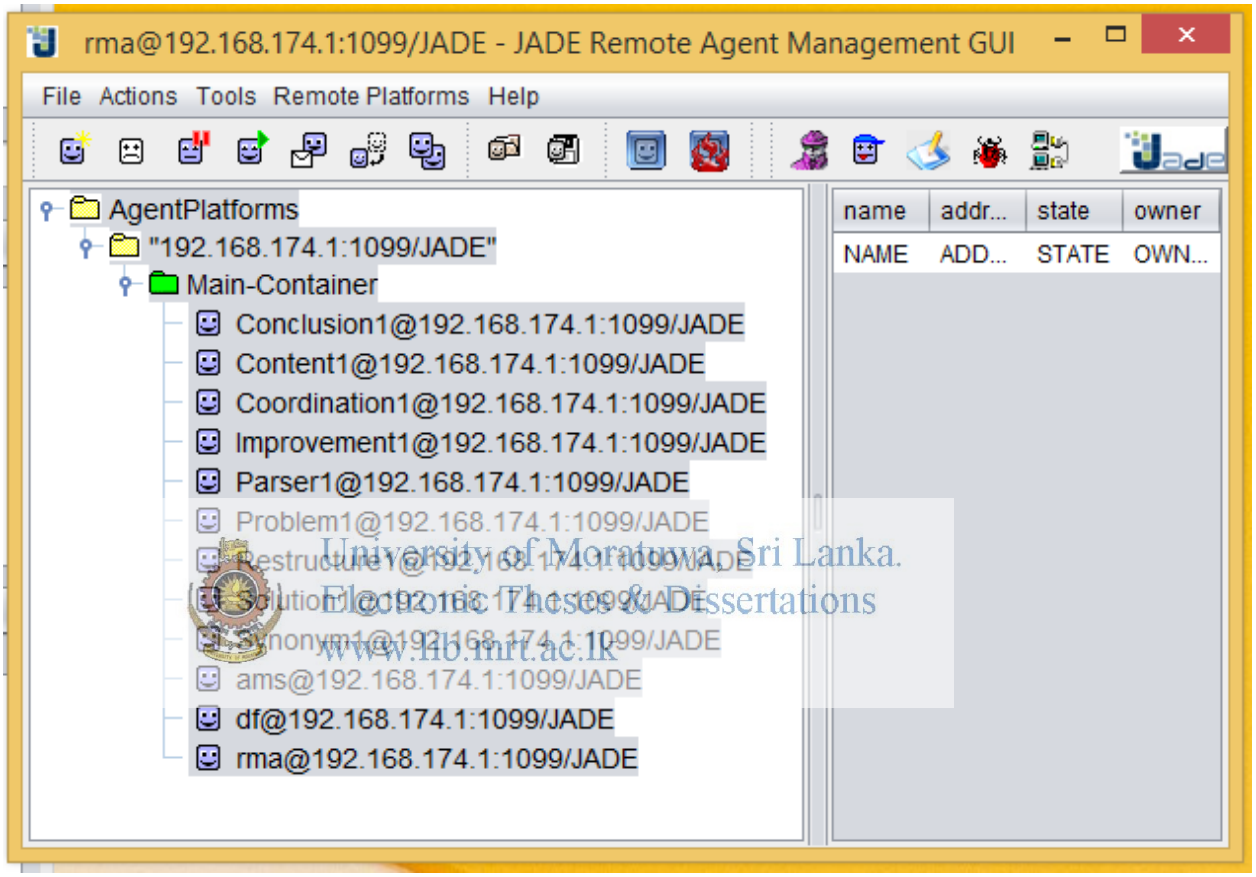


Figure B.1: Main User Interface of ASIA

B.3 JADE Remote Agent Management GUI

Screen shot of the user interface, JADE Remote Agent Management GUI of the proposed system, Agent-based Solution for Improving Abstracts is presented here.

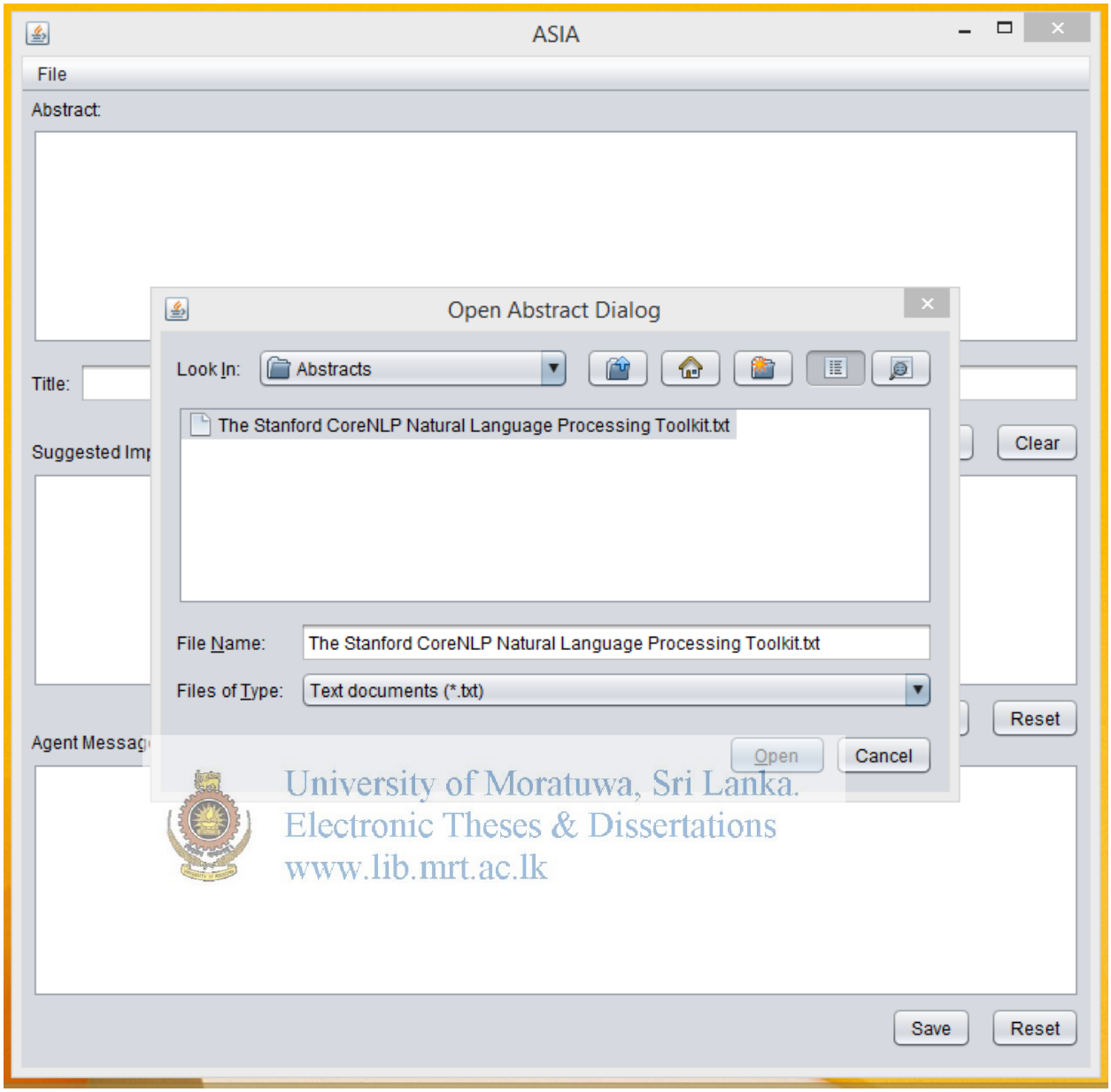


University of Moratuwa, Sri Lanka.
Electronic Theses & Dissertations
www.lib.mrt.ac.lk

Figure B.2: JADE Remote Agent Management GUI

B.4 Open Abstract Dialog GUI

Screen shot of the user interface, Open Abstract Dialog of the proposed system, Agent-based Solution for Improving Abstracts is presented here.

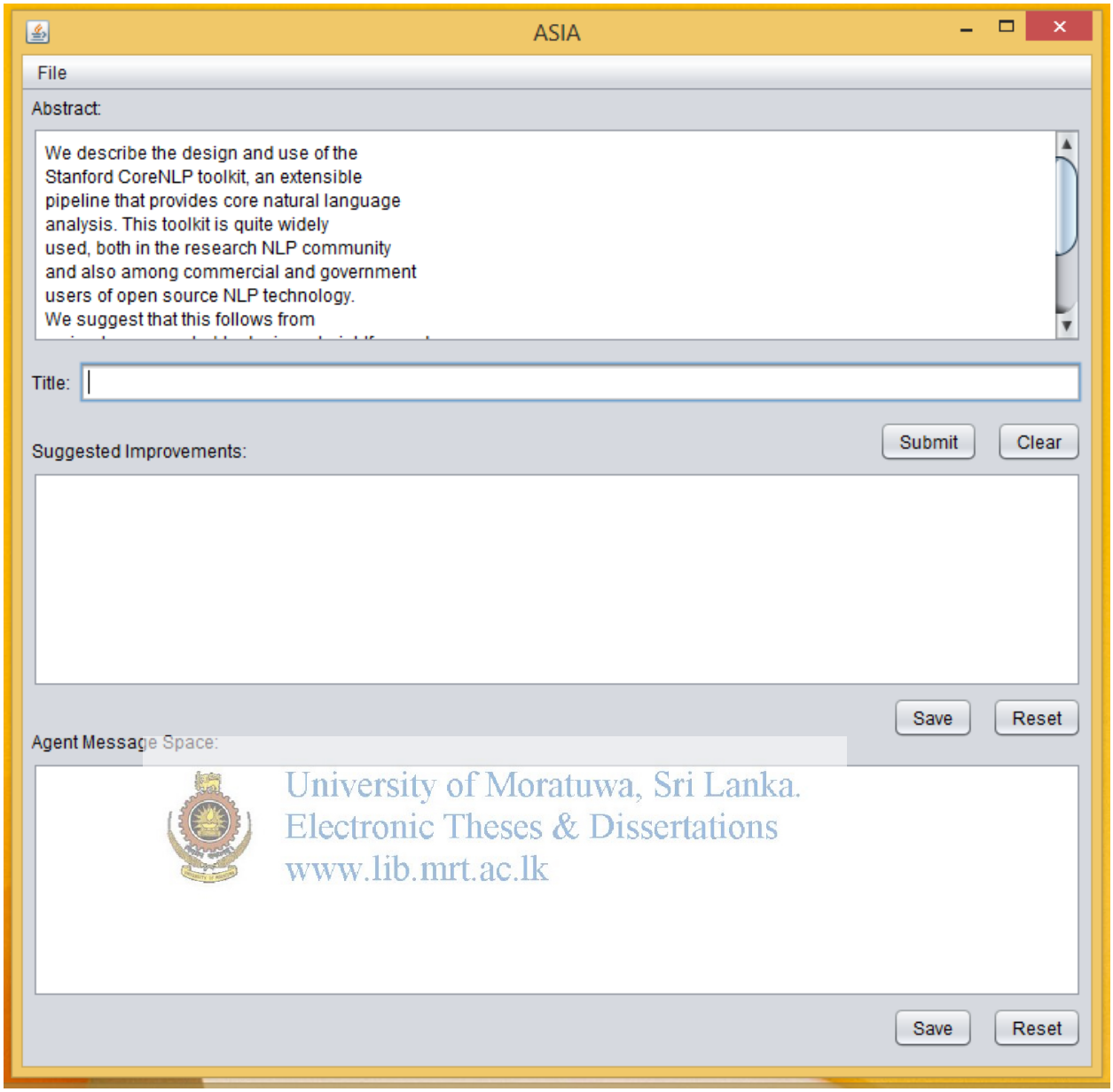


University of Moratuwa, Sri Lanka.
Electronic Theses & Dissertations
www.lib.mrt.ac.lk

Figure B.3: Open Abstract Dialog GUI

B.5 Abstract Open in the Text Area

Screen shot of the user interface; Abstract Open in the Text Area of the proposed system, Agent-based Solution for Improving Abstracts is presented here.

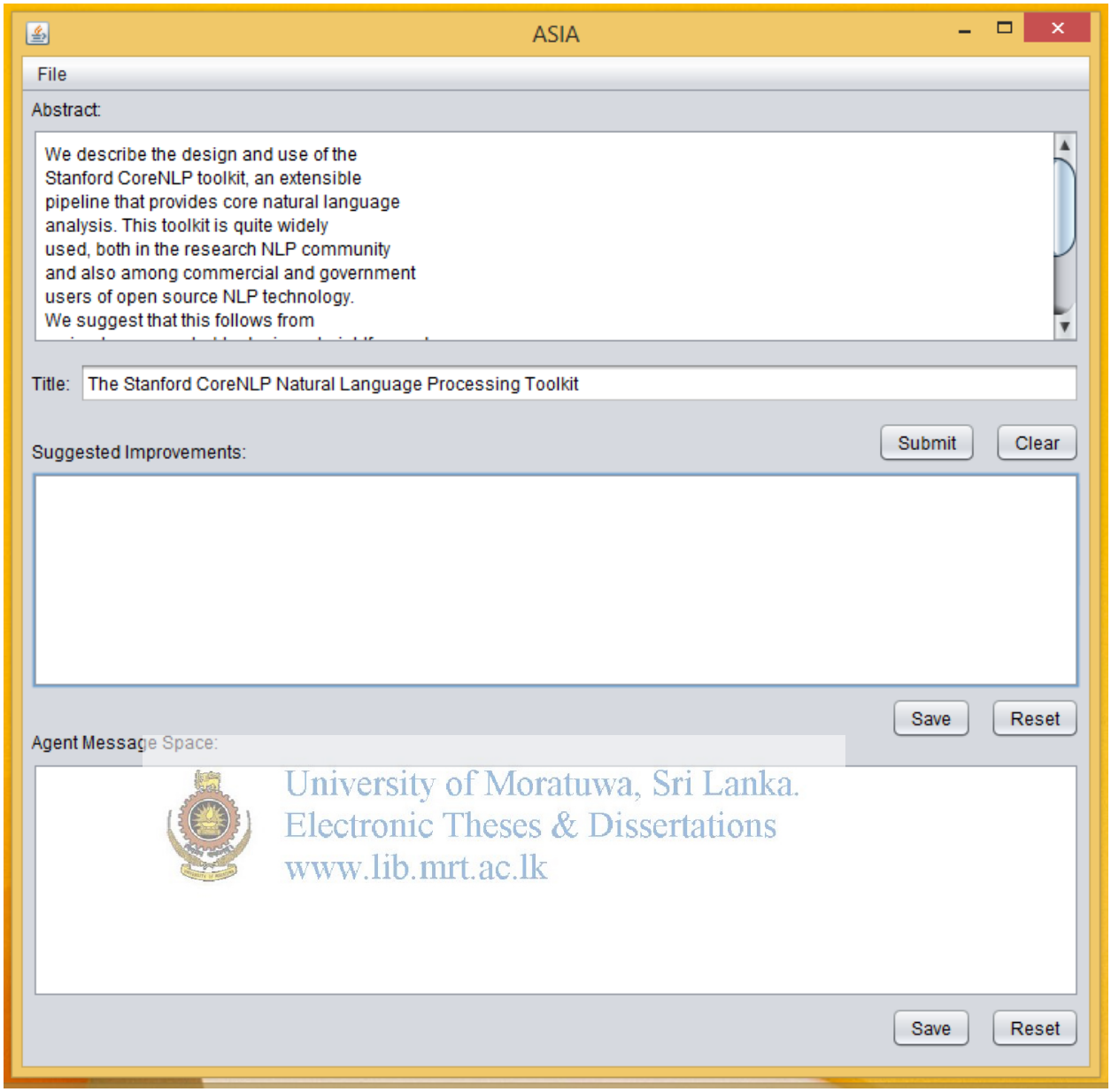


University of Moratuwa, Sri Lanka.
Electronic Theses & Dissertations
www.lib.mrt.ac.lk

Figure B.4: Abstract Open in the Text Area

B.6 Enter the Title of the Abstract in the Text Area

Screen shot of the user interface; enter the Title of the Abstract in the Text Area of the proposed system, Agent-based Solution for Improving Abstracts is presented here.



University of Moratuwa, Sri Lanka.
Electronic Theses & Dissertations
www.lib.mrt.ac.lk

Figure B.5: Enter the Title of the Abstract in the Text Area

B.7 Submit the Abstract to the System

Screen shot of the user interface; Submit the Abstract to the System of the proposed system, Agent-based Solution for Improving Abstracts is presented here.

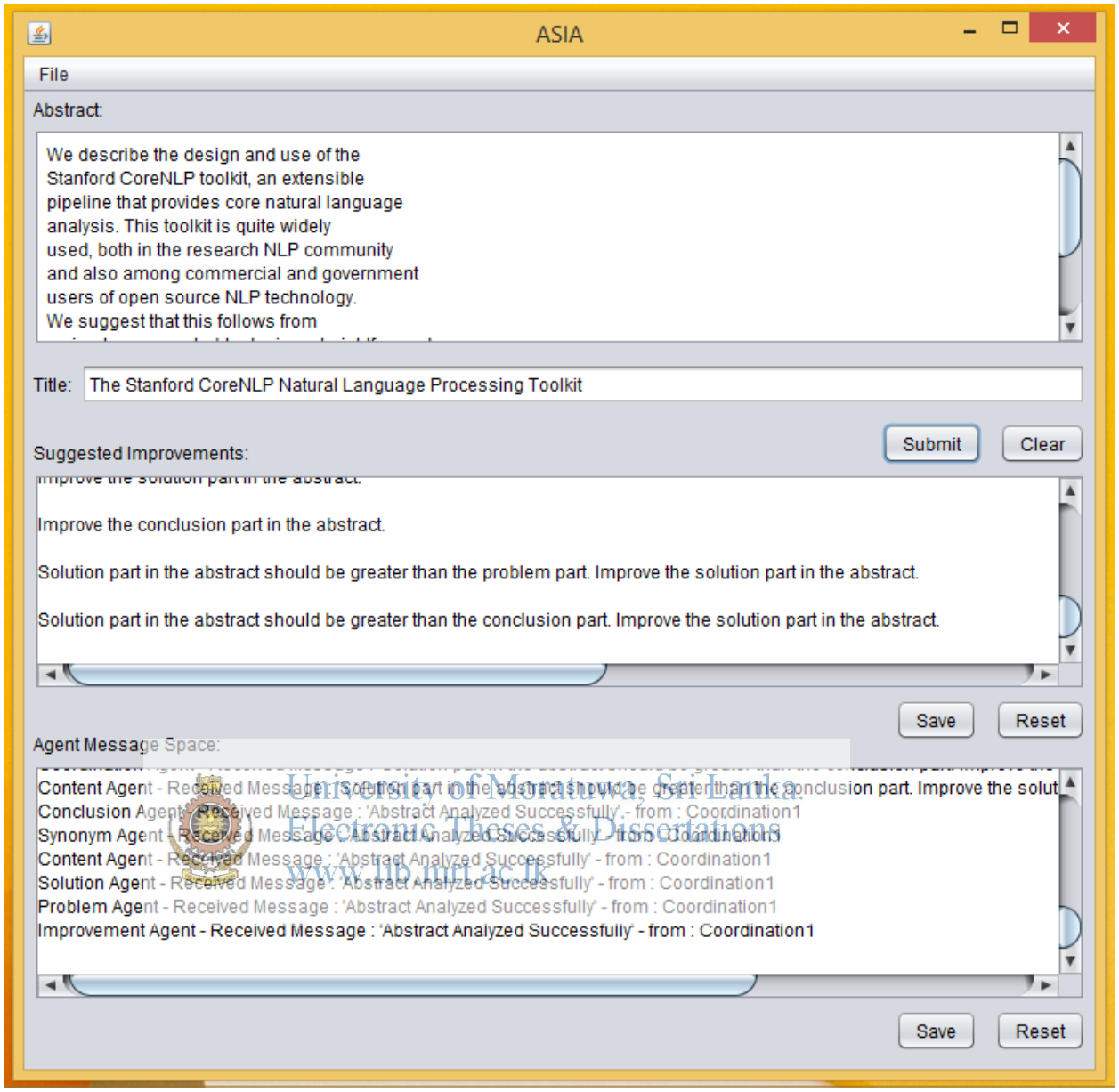


Figure B.6: Submit the Abstract to the System

B.8 Save the Suggested Improvements

Screen shot of the user interface; Save the Suggested Improvements of the proposed system, Agent-based Solution for Improving Abstracts is presented here.

University of Moratuwa, Sri Lanka.
Electronic Theses & Dissertations
www.lib.mrt.ac.lk

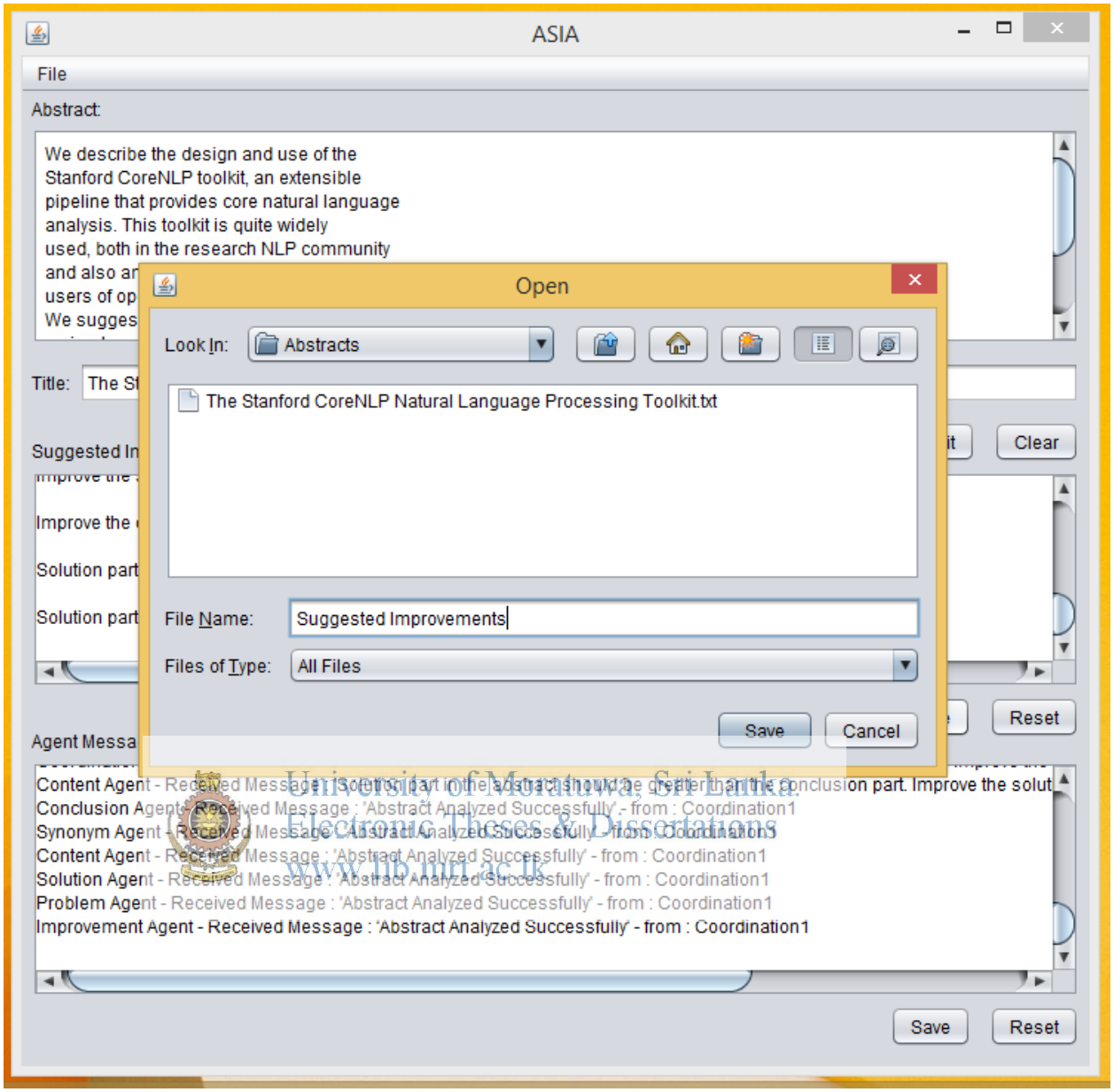
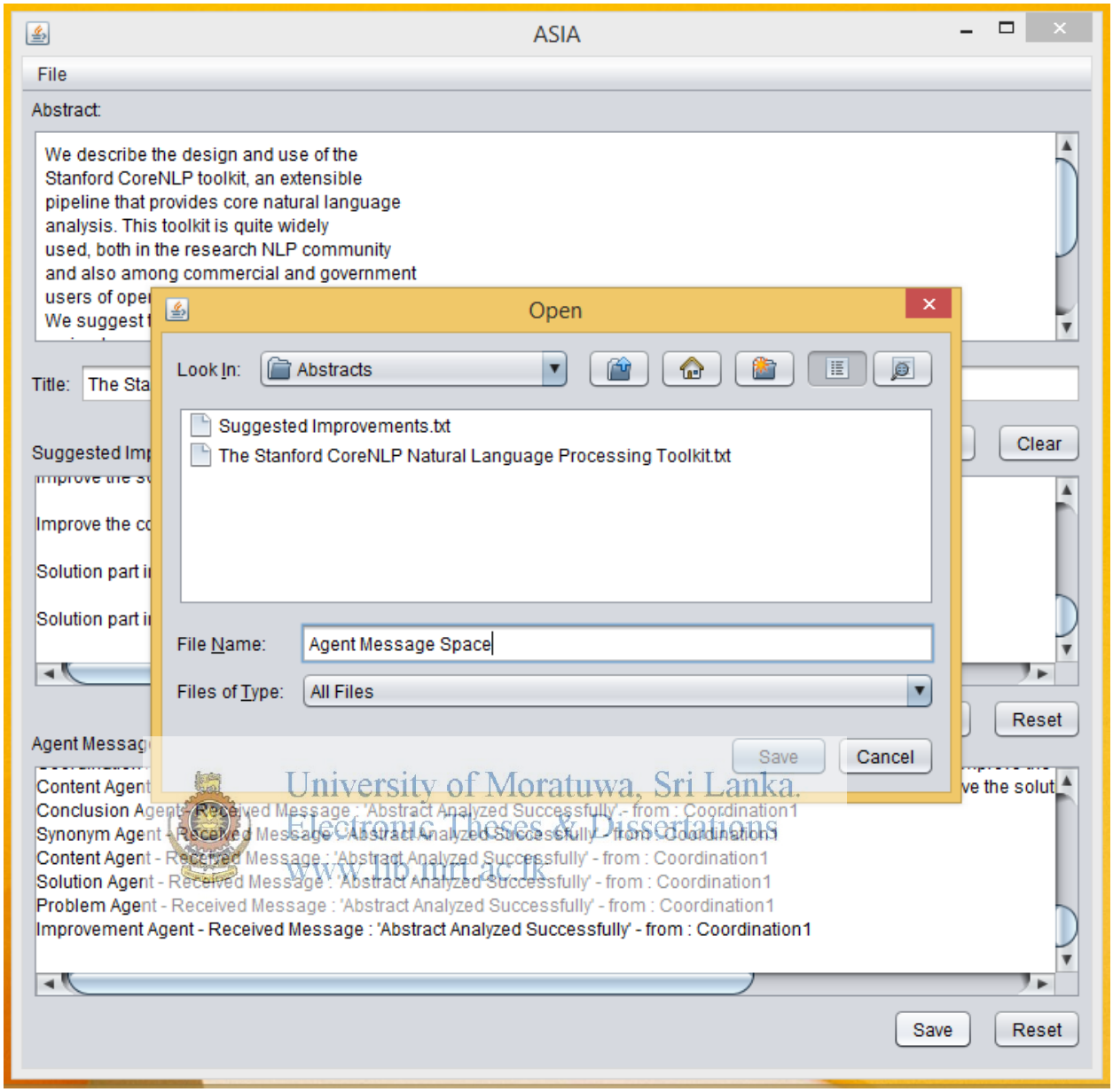


Figure B.7: Save the Suggested Improvements

B.9 Save the Agent Message Space

Screen shot of the user interface; Save the Agent Message Space of the proposed system, Agent-based Solution for Improving Abstracts is presented here.

University of Moratuwa, Sri Lanka.
Electronic Theses & Dissertations
www.lib.mrt.ac.lk

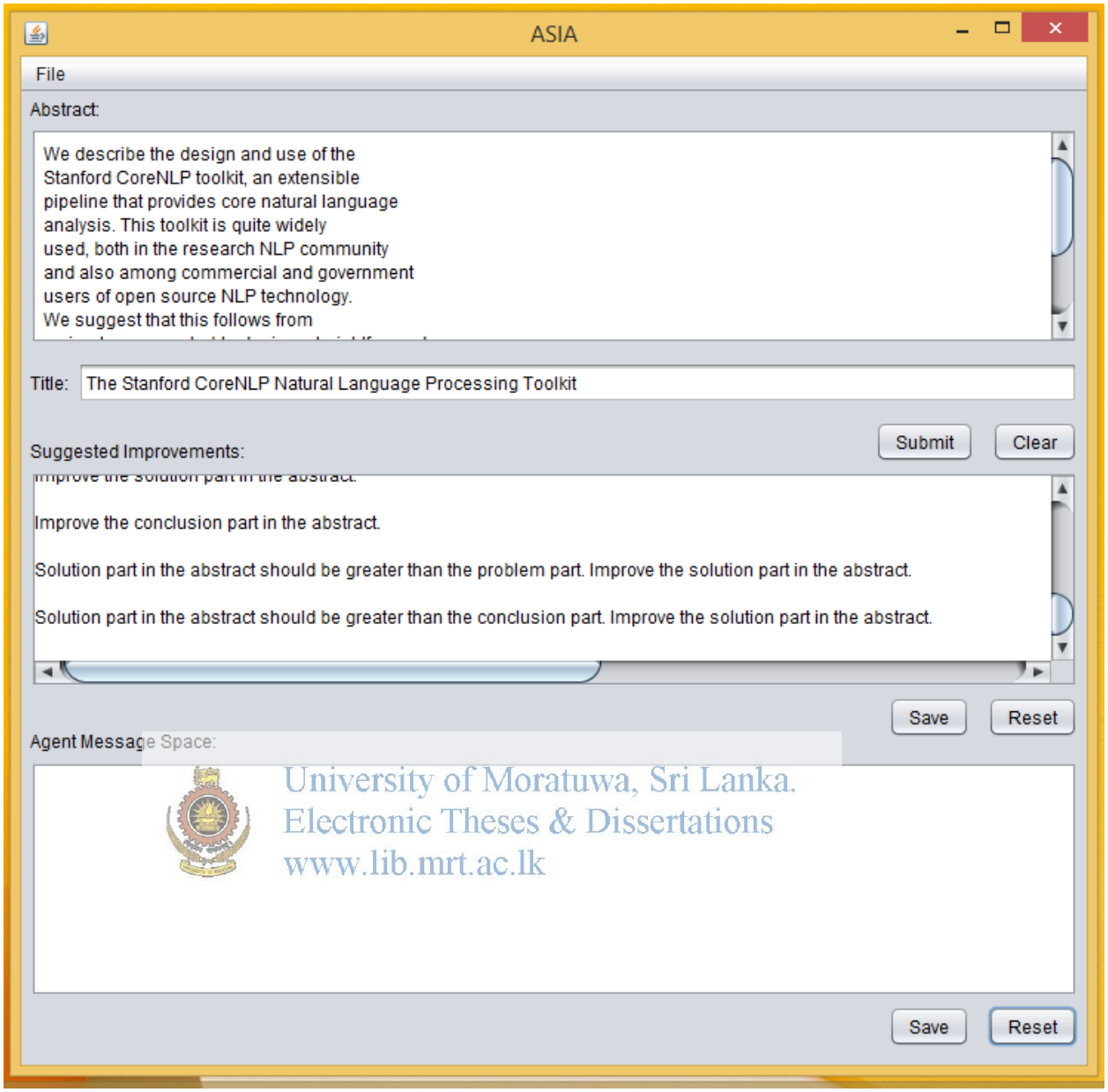


University of Moratuwa, Sri Lanka.
Electronic Theses & Dissertations
www.lib.mrt.ac.lk

Figure B.8: Save the Agent Message Space

B.10 Clear the Agent Message Space

Screen shot of the user interface; Clear the Agent Message Space of the proposed system, Agent-based Solution for Improving Abstracts is presented here.

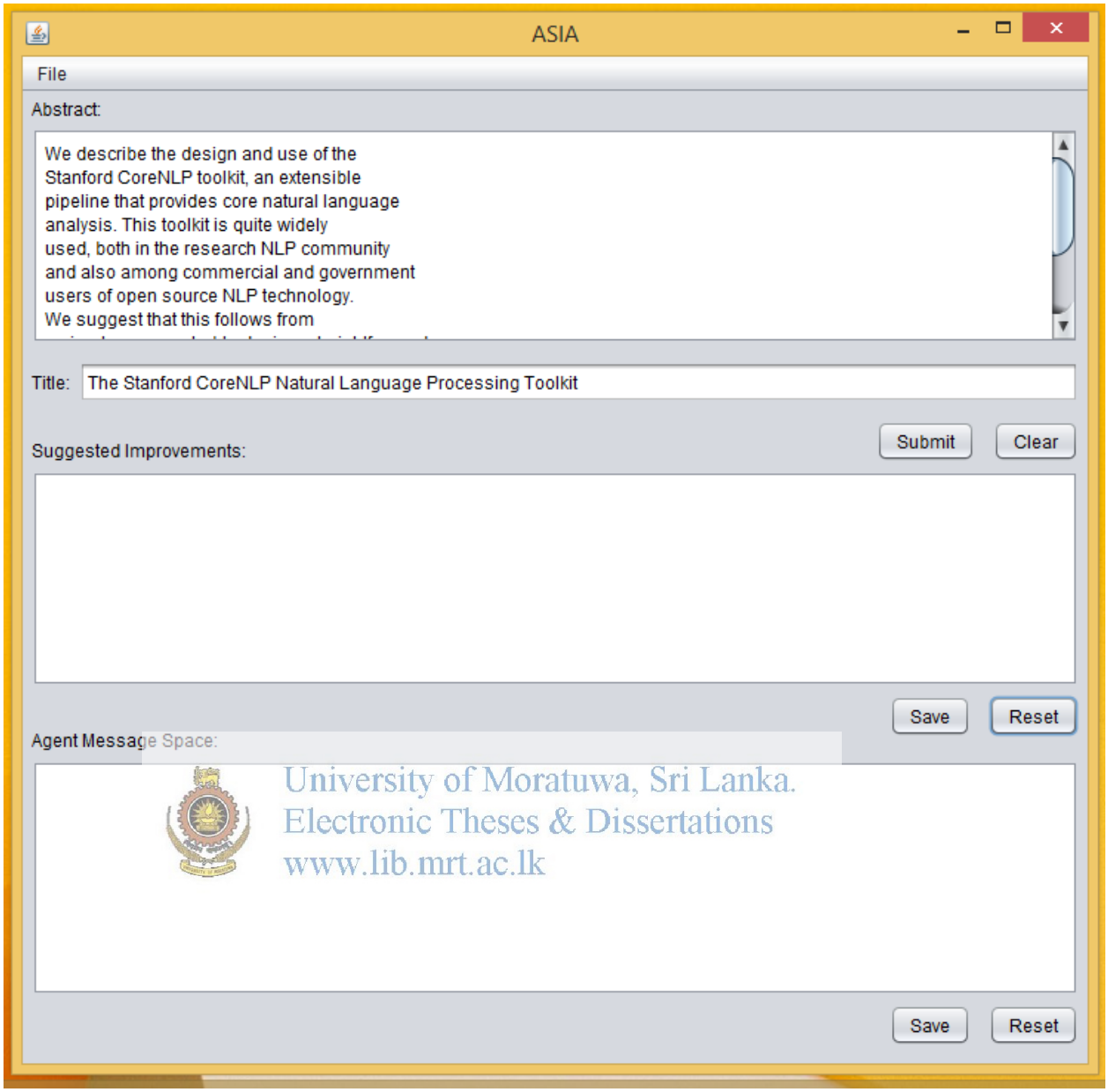


University of Moratuwa, Sri Lanka.
Electronic Theses & Dissertations
www.lib.mrt.ac.lk

Figure B.9: Clear the Agent Message Space

B.11 Clear the Suggested Improvements

Screen shot of the user interface; Clear the Suggested Improvements of the proposed system, Agent-based Solution for Improving Abstracts is presented here.

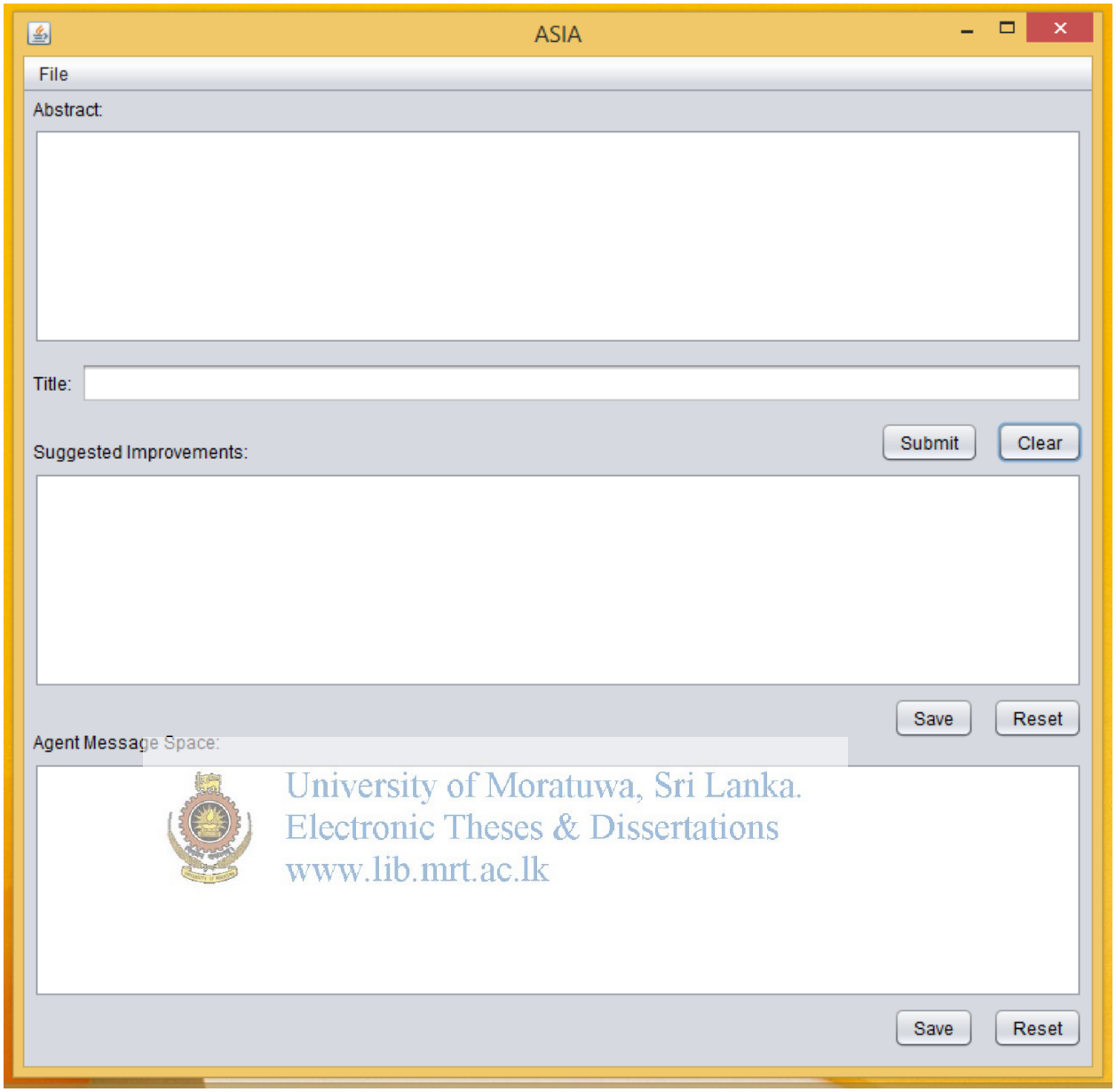


University of Moratuwa, Sri Lanka.
Electronic Theses & Dissertations
www.lib.mrt.ac.lk

Figure B.10: Clear the Suggested Improvements

B.12 Clear the Abstract and Title

Screen shot of the user interface Clear the Abstract and Title of the proposed system, Agent-based Solution for Improving Abstracts is presented here.



University of Moratuwa, Sri Lanka.
Electronic Theses & Dissertations
www.lib.mrt.ac.lk

Figure B.11: Clear the Abstract and Title

B.13 Project Workspace in NetBeans IDE

Screen shot of the Project Workspace in NetBeans IDE of the proposed system, Agent-based Solution for Improving Abstracts is presented here.

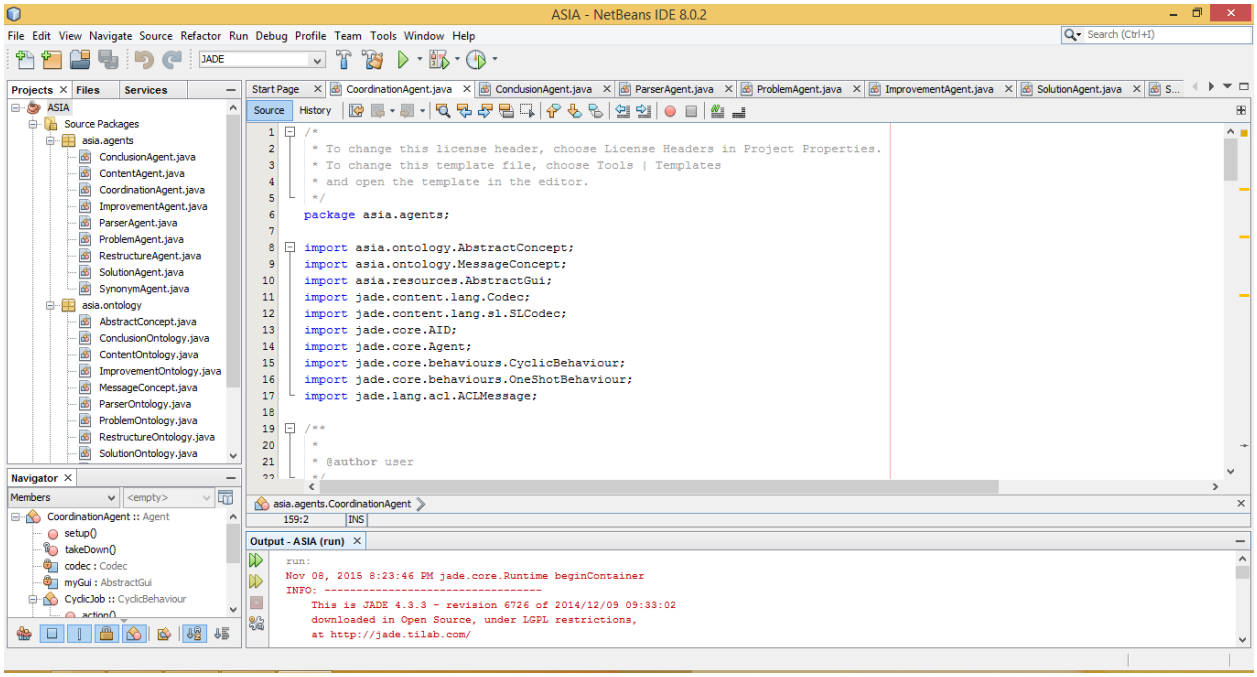


Figure B.12: Project Workspace in NetBeans IDE

B.14 Project Run Output in NetBeans IDE

University of Moratuwa, Sri Lanka.
Electronic Theses & Dissertations
www.lib.mrt.ac.lk

Screen shot of the Project Run Output in NetBeans IDE of the proposed system, Agent-based Solution for Improving Abstracts is presented here.

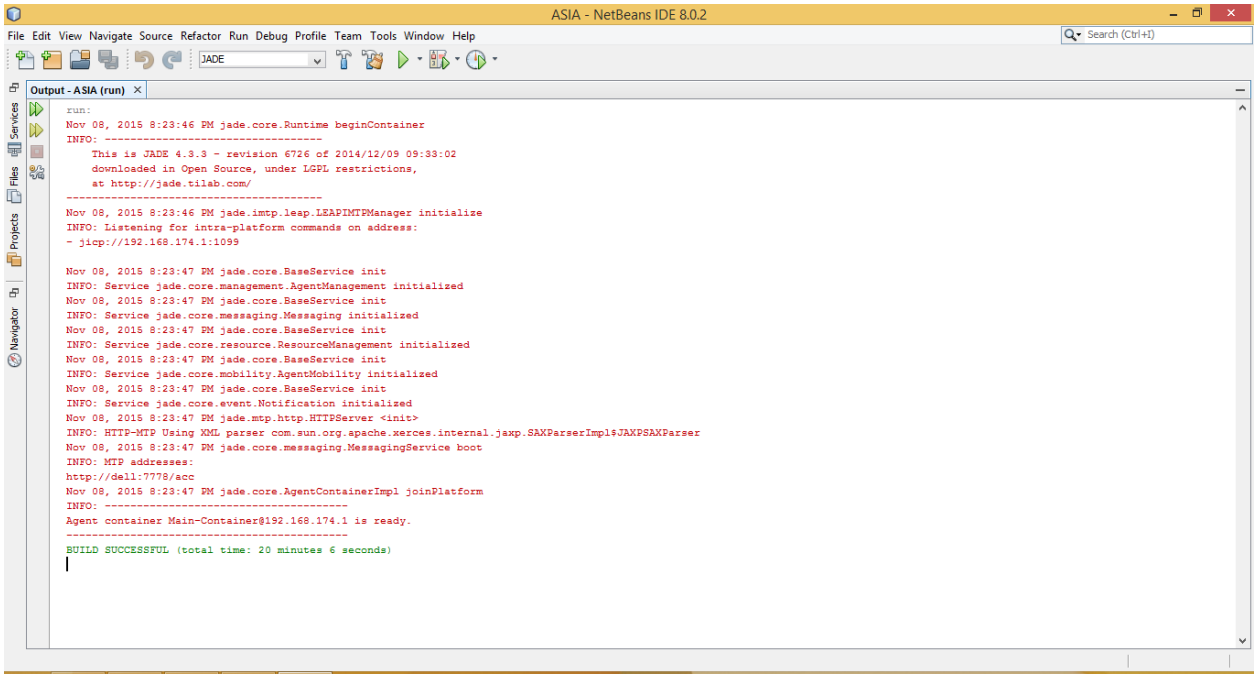


Figure B.13: Project Run Output in NetBeans IDE