



Bahria University Islamabad
Department of Software Engineering

Dated: 14th May, 2018

CERTIFICATE

We accept the work contained in the report titled (Automatic Intent Classification using Parts of Speech Tags and Sentence Dependency Structure) as a confirmation to the required standard for the partial fulfillment the degree of BSE -8.

Ms. Sadia Ashraf
Project Coordinator
Date: _____

Dr. Raja M. Suleman
Supervisor
Date: _____

Dr. Tamim Ahmed Khan
Internal Examiner
Date: _____

Dr. Iftikhar Azim Niaz
External Examiner
Date: 14-05-2018

Dr. Awais Majeed
Head of Department (SE)
Date: 01-06-2018

ABSTRACT

The year 2016-17 was termed as “The Year of the Chatbots”. Chatbots are becoming part of our everyday life fabric. From intelligent personal assistants like Apple’s Siri to psychology therapist like WoeBot, they can be seen in every domain and have support from almost all major tech companies. The research on conversational agents or Chatbots has received a lot of attention in the recent years. One of the biggest challenges in developing such agents is to enable them to understand natural language. Natural language is very hard to understand because it is ambiguous and words have multiple meanings. In natural language, the purpose of a sentence is called its “intent”. In intelligent systems, an intent represents the mapping between what a user “wants the system to do” and “what action should be taken by the system” in response. When developing Chatbots, developers have to train these agents on large fully-annotated corpuses, using Supervised Machine Learning to automatically classify user intents. These corpuses contain sentences that are tagged or “labelled” to a specific intent. Compiling a ‘complete’ corpus with all possible sentence-intent mapping is almost impossible and a complete dataset of all sentences and their intents is not available. This is the major drawback and explains why Chatbots have not been able to achieve the success envisioned for such agents. Almost all tech giants like Google, Facebook, IBM, and Microsoft have also introduced frameworks to develop Chatbot but they, again, are dependent on data and require developers to manually enter all possible sentences along with their intended intent. Entering data in this way is a very hectic process and inefficient. Keeping in mind the above problems, we have proposed an algorithm which can classify intents of short sentences automatically without any dependence on a corpus. The algorithm uses “Parts-of-Speech” tags along with “Sentence Dependency Structure” to identify the intents from given input text. The results of our approach are very encouraging as the system is able to generate intents for given inputs automatically and is also able to annotate similar sentence with the same intent. This algorithm is a step towards improving Natural Language Understanding of computers while reducing the dependence on data.

Table of Contents

DECLARATION	i
CERTIFICATE	ii
ACKNOWLEDGMENT.....	iii
ABSTRACT.....	iv
CHAPTER – 1 INTRODUCTION	1
1.1 Motivation.....	2
1.2 Problem Statement	2
1.3 Goals /Objectives	2
1.4 Main Contributions	3
1.5 Thesis Organization	3
CHAPTER – 2 BACKGROUND/ LITERATURE REVIEW	4
2.1 Literature Review.....	5
CHAPTER – 3 SYSTEM REQUIREMENTS.....	7
3.1 Interface Requirements	8
3.2 Functional Requirements	8
3.3 Use Cases.....	9
3.4 Non-Functional Requirements.....	12
CHAPTER – 4 SYSTEM DESIGN	13
4.1 Design Approach	14
4.2 Design Constraints.....	14
4.3 Interface Design.....	14
4.4 Entity-Relational Diagram (ERD)	20
4.5 Logical Data Model and Functional Flow	21
CHAPTER – 5 SYSTEM IMPLEMENTATION	23
5.1 Strategy	24
5.2 Tools Used	24
5.3 Algorithm.....	25
5.4 System Architecture.....	26
CHAPTER – 6 SYSTEM TESTING.....	27
6.1 Test Strategy	28
6.2 Test Cases	28
CHAPTER – 7 CONCLUSION.....	30

7.1 Conclusion	31
BIBLIOGRAPHY / REFERENCES.....	32
APPENDIX-I	33