

Universitatea POLITEHNICA din București

Facultatea de Automatică și Calculatoare,

Departamentul de Calculatoare



# LUCRARE DE DIPLOMĂ

Agent conversațional pentru  
interacțiunea cu personalități istorice

**Conducător Științific:**

Ș.l. Dr. ing. Traian REBEDEA

**Autor:**

Adrian BOGATU

București, 2015

University POLITEHNICA of Bucharest  
Faculty of Automatic Control and Computers,  
Computer Science and Engineering Department



## Bachelor Thesis

# Conversational Agent for Interacting with Historical Figures

**Scientific Adviser:**

Ș.l. Dr. ing. Traian REBEDEA

**Author:**

Adrian BOGATU

Bucharest, 2015

# Abstract

This paper describes a conversational agent that answers questions pertaining to a given historical figure. The answers are selected from a set of sentences extracted from the biographical text of that personality using syntactic and semantic analysis on both the question and the sentences at hand. This is accomplished with a top-down approach, starting from the entire biographical text and working our way down, in several steps, to the

# Contents

<b>Abstract</b>	<b>i</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Question Answering . . . . .	1
1.2 Conversational Agent . . . . .	1
1.3 Project Objectives and Motivation . . . . .	1
1.3.1 Project Scope . . . . .	1
1.3.2 Project Objectives . . . . .	1
1.3.3 Related Work . . . . .	1
1.3.4 Tables . . . . .	1
<b>2 Related Work</b>	<b>2</b>
<b>3 Tools</b>	<b>3</b>
3.1 DBpedia . . . . .	3
3.2 ChatScript . . . . .	3
3.3 WordNet . . . . .	3
3.3.1 MIT Java Wordnet Interface (JWI) . . . . .	3
3.4 Apache Lucene . . . . .	3
3.5 Stanford CoreNLP . . . . .	3
3.5.1 Stanford Log-linear Part-Of-Speech Tagger . . . . .	4
3.5.2 Stanford Tokenizer . . . . .	4
3.5.3 Stanford Named Entity Recognizer (NER) . . . . .	4
3.5.4 Stanford Deterministic Coreference Resolution System . . . . .	4
<b>4 Implementation</b>	<b>5</b>
4.1 Testing . . . . .	5
<b>5 Conclusions</b>	<b>6</b>
<b>6 Future Work</b>	<b>7</b>
<b>7 Results</b>	<b>8</b>

# Chapter 1

## Introduction

More than 60 years ago, Alan Turing raised the question "Can machines think?" [1]. Can people believe that a machine thinks.

### 1.1 Question Answering

### 1.2 Conversational Agent

### 1.3 Project Objectives and Motivation

#### 1.3.1 Project Scope

#### 1.3.2 Project Objectives

#### 1.3.3 Related Work

#### 1.3.4 Tables

## Chapter 2

## Related Work

# Chapter 3

## Tools

The next sections describe the most important programming tools used in the implementation of the conversational agent.

### 3.1 DBpedia<sup>1</sup>

DBpedia [2]

### 3.2 ChatScript<sup>2</sup>

ChatScript [3]

### 3.3 WordNet<sup>3</sup>

WordNet [4]

#### 3.3.1 MIT Java Wordnet Interface (JWI)

JWI [5]

### 3.4 Apache Lucene

### 3.5 Stanford CoreNLP

Stanford CoreNLP is one of the most exhaustive tools for natural language analysis. [6]

---

<sup>1</sup>DBpedia

<sup>2</sup>ChatScript

<sup>3</sup>WordNet

- 3.5.1 Stanford Log-linear Part-Of-Speech Tagger
- 3.5.2 Stanford Tokenizer
- 3.5.3 Stanford Named Entity Recognizer (NER)
- 3.5.4 Stanford Deterministic Coreference Resolution System



## Chapter 4

# Implementation

### 4.1 Testing

## Chapter 5

## Conclusions

## Chapter 6

# Future Work

## Chapter 7

# Results

# Bibliography

- [1] A. Turing, “Computing machinery and intelligence,” *Mind*, vol. 59, no. 236, pp. 433–460, 1950.
- [2] S. Auer, C. Bizer, G. Kobilarov, J. Lehmann, R. Cyganiak, and Z. Ives, “DBpedia: A nucleus for a Web of open data,” in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 4825 LNCS, pp. 722–735, 2007.
- [3] B. Wilcox, “Beyond Façade: Pattern Matching for Natural Language Applications,” *Gamasutra*, pp. 1–5, 2011.
- [4] G. a. Miller, “WordNet: a lexical database for English,” *Communications of the ACM*, vol. 38, no. 11, pp. 39–41, 1995.
- [5] M. A. Finlayson, “Code for Java Libraries for Accessing the Princeton Wordnet: Comparison and Evaluation,” *the 7th Global Wordnet Conference*, 2013.
- [6] C. Manning, M. Surdeanu, J. Bauer, J. Finkel, S. Bethard, and D. McClosky, “The Stanford CoreNLP Natural Language Processing Toolkit,” *Proceedings of 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, pp. 55–60, 2014.