Effective hybrid preprocessing algorithms

György Orosz

Faculty of Information Technology and Bionics

Pázmány Péter Catholic University

This dissertation is submitted for the degree of

Doctor of Philosophy

July 2014

Table of contents

Chapter 1

Introduction

1. basic concepts (Pos tagging, morphological analyzer/tagging, lemmatizing)
2. importance of the problems
3. historical overview

1.1 Research Aims

1. research aims and questions

1.2 Methodology

1. commonly used algorithms
2. how to evaluate
3. used resources)

Chapter 2

Algorithms for Morphological Disambiguation

2.1 Introduction

1. Focused research aim: effective algorithms that can be easily domain adapted
2. importance of the problem
3. historical notes (?)

2.1.1 Background

- Commonly used methods: Rules, Brill, Tnt, HunPos, Maxent modells, CRF modells, SVM modells
- Uses cases for other languages

2.2 Experiments

2.2.1 Context Dependent Statistically Driven Lemmatization

Methods

Evaluation

2.2.2 Hybrid Morphological Disambiguation

Methods

Evaluation

2.2.3 Combining Morphological Disambiguation Tools

Methods

Evaluation

2.3 Summary

Chapter 3

Preprocessing Algorithms for Clinical Hungarian

Chapter 4

Parsing Speech Transcripts of Children

Chapter 5

Summary – New Scientific Results