DEFESA DE DISSERTAÇÃO DE MESTRADO

Programa de Pós-Graduação em Ciência da Computação

ANALYSING SEMANTIC RESOURCES FOR COREFERENCE RESOLUTION

ALUNO: Thiago Machado Lima

ORIENTADORA: Dra. Renata Vieira

BANCA EXAMINADORA: Dr. Thiago Alexandre Salgueiro Pardo (ICMC/USP), Dr. Rafael Heitor Bordini (PPGCC/PUCRS)

DATA: 15 de março de 2019

LOCAL: Prédio 32, sala 506.2

HORÁRIO: 08:30

RESUMO:

Coreference Resolution is the task that consists of identifying mentions in a discourse that refer to the same entity. The task has the potential to improve other Natural Language Processing tasks such as sentiment analysis, information extraction, question answering, and others. Some coreferent relationships can be identified using lexical and syntactical rules, while others require semantic knowledge. However, few works focus on the possible improvements of using semantic knowledge. This work´s objective is to improve the coreference resolution task by using semantic knowledge. For that, we reviewed the semantic resources available for the Portuguese language, and selected ContoPT, Concept- Net and a word embedding model for our experiments. Experiments were performed using CORP, a coreference tool for the Portuguese language which already uses OntoPT as a semantic resource. The evaluation was composed of metrics MUC, B3 and CEAFe, using Corref-PT and Summ-it++ as corpora. Compared to OntoPT, we obtained better results in terms of F-Measure using ContoPT and ConceptNet. The experiments using the semantic similarity rule based in word embeddings was not able to surpass the results obtained with the structured semantic bases. Texts with more semantic relationships were selected for error analysis, and we were able to observe some difficulties involved in the detection of semantic relationships. To overcome these difficulties improvements are proposed. Besides the analysis of available semantic basis, this work brings as contribution a new CORP version, integrated with three new semantic resources, which obtained a higher F-Measure using semantics than the version that uses OntoPT.

Keywords: Coreference Resolution, Semantic Knowledge, Corpus Analysis.