Contents

Introduction

1.1 Motivation

I was motivated to write a Phd thesis because I did not want to work directly after finishing my study

1.2 Organization

This thesis is organized as follows, ...

Background

The aim of this work is to create a model that predicts if an article comment or a forum post can be classified as "persuasive" or "non-persuasive". We'll first give general definitions about persuasion and argumentation, and how the corpus was annotated. Then, the tool that was used to perform the classfication, DKPro Text Classification Framework, will be introduced and its functionnalities will be explained. Finally we'll discuss about the algorithm used and the metrics to evaluate our model.

2.1   Argumentation

2.1.1   General Definitions

2.1.2   Persuasion

2.1.3   Persuasion 2.0

2.2   NLP and the DKPro Framework

2.2.1   Natural Language Processing

2.2.2   UIMA

DKPro stands for *Darmstadt Knowledge Processing* [3] and it's a software suite for NLP based on the Apache UIMA Framework. UIMA are software systems that analyse large volumes of unstructured information in order to discover knowledge that is relevant to an end user. An example UIMA application might ingest plain text and identify entities, such as persons, places, organizations; or relations, such as works-for or located-at:

The real power of UIMA is its Analysis Engines (AE) which basically analyse a document and record descriptive attributes. Those descriptive attributes will form the docu-
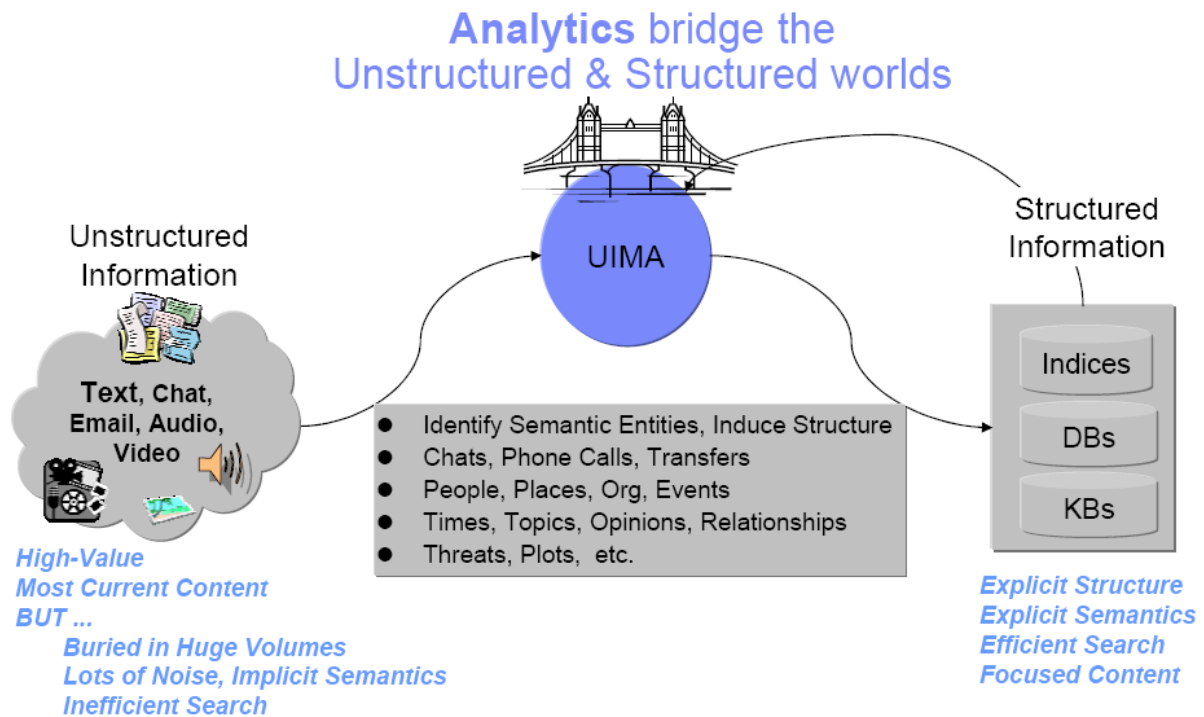
Figure 2.1: UIMA: Order unstructured data, from UIMA website [1]

ment's metadata that will be used for further analysis (such as Text Classification in our case).

GIVE EXAMPLES OF AE HERE AND DKPRO PROC PIPILINE AND CAS ?

Thus, the DKPRo software use UIMA's AE in order to collect structured information about textual data.

2.2.3 DKPro Core

Many NLP tools are already freely available in the NLP research community. DKPro Core [2] provides UIMA components wrapping these tools (and some original tools) so they can be used interchangeably in UIMA processing pipelines. The provided components wrap a constantly growing set of stand-of-the-art NLP tools and also include several original components covering a wide range of tasks including: tokenization/segmentation, compound splitting, stemming, part-of-speech tagging, lemmatization, constituency parsing, dependency parsing, named entity recognition, coreference resolution, language identification, spelling correction, grammar checking, and support for reading and writing various file and corpus formats.
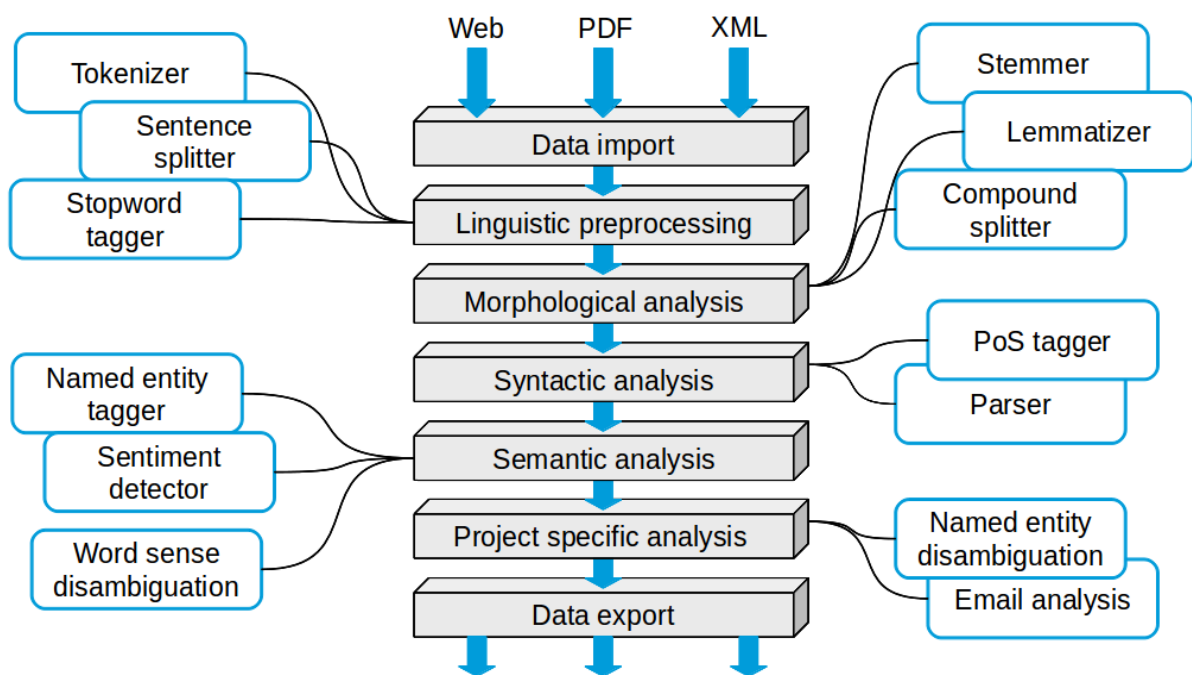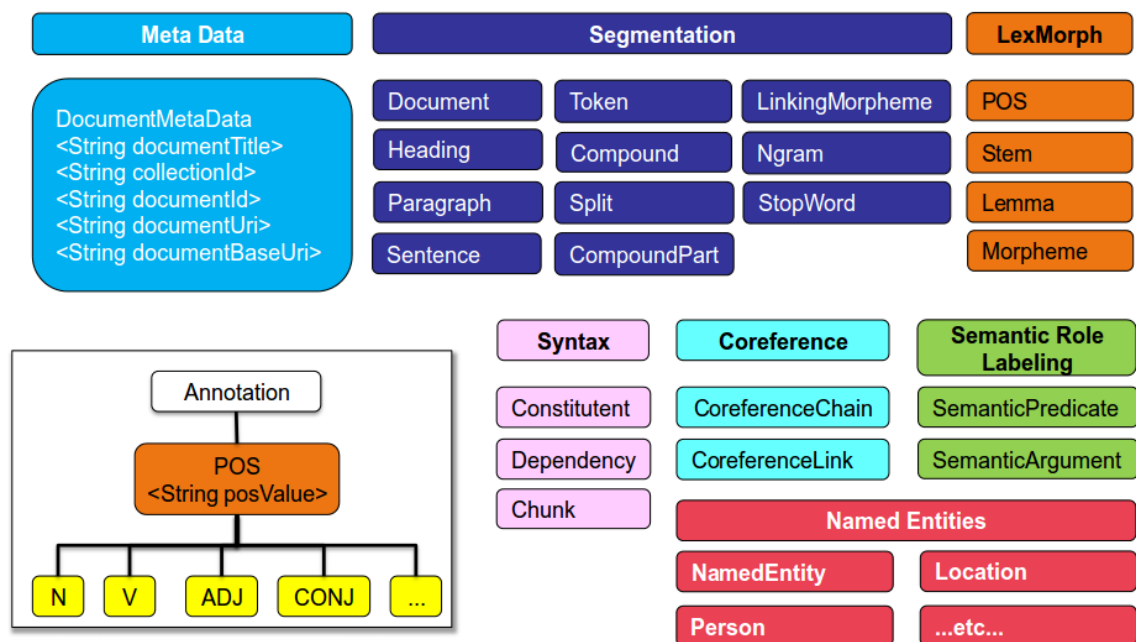
Figure 2.2: DKPro Core Pipeline

DKPro Core Overview

Figure A.1: DKPro Type System

Glossary

mathematics Mathematics is what mathematicians do. 3

Bibliography

[1] Overview and setup, uima.apache.org, 2006.

[2] Richard Eckart de Castilho and Iryna Gurevych. A broad-coverage collection of portable nlp components for building shareable analysis pipelines. In *Proceedings of the Workshop on Open Infrastructures and Analysis Frameworks for HLT (OIAF4HLT) at COLING 2014*, pages 1–11, August 2014.

[3] Iryna Gurevych, Max Mühlhäuser, Christof Müller, Jürgen Steimle, Markus Weimer, and Torsten Zesch. Darmstadt knowledge processing repository based on uima. In *Proceedings of the First Workshop on Unstructured Information Management Architecture at Biannual Conference of the Society for Computational Linguistics and Language Technology*, Tübingen, Germany, April 2007.