



Named Entities Annotator

OVERVIEW

The Named entities annotator module is intended for annotation of textual news items with named entity instances from the Papyrus news ontology, such as Organisation, Location and Person. First, the ontology is parsed to extract the list of named entities with their related keywords, and then the extracted grammar is used for identification of entities in the input document.

INNOVATION

The module improves the current State-of-the-Art by using fast and scalable compiler-style techniques for implementing matching rules for named entities, i.e. TXL transformation framework.

BUSINESS IMPACT

This module can be integrated into any textual analysis component that needs named entities annotation given an existing ontology that contains a predefined set of entities.

The annotation provided consists of a list of named entities, where each of output named entities is represented by its identifier and a set of its alternative names. Thus, given a textual input file, the module returns a list of named entities identified in this input. Alternatively, annotations can be represented as embedded XML-tags in the text.

INTEROPERABILITY

The core application is a TXL program that can be easily run using a command line interface and requires a TXL interpreter for x86-based architectures (32/64bit) using MS Windows or Linux. The module can also be wrapped in a web service or called from a program written in other programming language command line calls.

