HeidelTime: High Quality Rule-based Extraction and Normalization of Temporal Expressions

Jannik Strötgen and Michael Gertz
Institute of Computer Science, University of Heidelberg
Im Neuenheimer Feld 348, 69120 Heidelberg, Germany

Temporal Expressions (TEs)
Different types
• Date: On May 22, 1995, Farkas was ...
• Time: ... in Brownsville around 7:15 p.m.
• Duration: He spent six days abroad ...
• Set: ... for liver transplants each year ...
Different occurrences in documents
• explicit: easy to normalize
• implicit: knowledge is needed
• relative: reference time is needed
(under additional information)
Annotation scheme
• TimeML: ISO standard for temporal annotation (Timex3) [2]

Main Challenges
Disambiguation of:
• the reference time of relative TEs
• underspecified relative TEs
• relations between underspecified TEs
Document Creation Time: 1990-08-15
Also today, King Hussein of Jordan arrived in Washington ...
he had rejected in peace talks following the August 1988 cease-fire ...
which had condemned Iraq ’s invasion of Kuwait on Aug 2 ...
The monarch will meet Bush on Thursday ...
... Independence Day 1989 ...

Task Description
• Identify extents of TEs
• Assign type information to TEs
• Normalize values according to Timex3

HeidelTime - Overview
• Rule-based system
• Realized as UIMA component [1]
• Methods for Extraction: regular expressions & NLP features
• Methods for Normalization: knowledge resources & linguistic clues
• Two rule sets: high precision (HeidelTime-1) and high recall (HeidelTime-2)

HeidelTime - System Architecture

UIMA Text Mining Pipeline

Iterative Rule Development Process
• Training data (sentence, token, Timex annotations → gold standard)
• Evaluator compares gold standard with HeidelTime Timex annotations
• Evaluator creates lists with FN, FP, TP
• These lists are used to improve rules

HeidelTime - Resources

Extraction Results

Normalization Results

Ongoing Work
Other languages
• new pattern resources (e.g., names of months, weekdays, ...)
• new normalization resources
• new rules

Other types of corpora
• TempEval-2 corpus: news documents
• adaptations for and evaluation on other types of corpora

Contact information:
Jannik Strötgen
stroetgen@uni-hd.de
http://dbs.ifi.uni-heidelberg.de/

References