Annotating Temporal Phenomena in Literary Text

Jannik Strötgen, Thomas Bögel, Michael Gertz

Database Systems Research Group, Heidelberg University, Im Neuenheimer Feld 348, 69120 Heidelberg, Germany

Context of the Work

- Natural language processing
- Information extraction
- Temporal annotation

Temporal information
- frequent in many texts
- can be normalized
- challenging to extract

Challenges for Temporal Tagging on Different Domains [1]

- reference time often DCT
- relation to reference time
- many underspecified and relative expressions

Stars of the Work

- TempEval-2 & TempEval-3 winner
- challenge: normalizing relative & underspecified expressions

Temporal Phenomena in Literary Text

- Literary: "Der Tod"
- Norman
- Heute morgen
- Vorzüglich Herbst
- Der 12. September
- Ichn bin mit der...

Temporal Tagging

- extraction & normalization of temporal expressions
- Challenge: normalizing relative & underspecified expressions

Most existing approaches
- focus on English
- focus on news documents

Temporal Expressions

- less frequent in literary text (usually)
- can be extracted
- normalization wrt local time frame

Tense information

- can be extracted (past, present, future)
- both help to detect more complex temporal phenomena:
  - as features for ML methods
  - as hints for manual annotations

Temporal narratological aspects

- relations in local time frame
- relations between discourse and history
- plot organizing sequences

The heureCLEA project [3]

HeidelTime: a Multilingual, Cross-domain Temporal Tagger [2]

Key Features
- rule-based system
- required: sentence, token, and POS information
- extraction: regular expressions & NLP features
- normalization: knowledge resources & linguistic clues
- TempEval-2 & TempEval-3 winner

Language-independent
- resource interpreter
- domain-dependent normalization strategies
- relation to reference time

Language-dependent
- pattern files
- month=(.../April/May/...)
- normalized files
- rule files

Normalizations to local time (year x)

<table>
<thead>
<tr>
<th>Der 10. September</th>
<th>09-10</th>
</tr>
</thead>
<tbody>
<tr>
<td>Norman</td>
<td>FA</td>
</tr>
<tr>
<td>Heute morgen</td>
<td>SU</td>
</tr>
<tr>
<td>Vorzüglich Herbst</td>
<td></td>
</tr>
<tr>
<td>Der 12. September</td>
<td></td>
</tr>
<tr>
<td>Ich bin mit der...</td>
<td></td>
</tr>
</tbody>
</table>

References


Temporal Tagging

HeidelTime: a Multilingual, Cross-domain Temporal Tagger [2]

Key Features
- rule-based system
- required: sentence, token, and POS information
- extraction: regular expressions & NLP features
- normalization: knowledge resources & linguistic clues
- TempEval-2 & TempEval-3 winner

Languages
- English, German, Dutch, Spanish, Italian, French, Arabic, Vietnamese
- more to come

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