HeidelTime at TempEval-3
Tuning English and Developing Spanish Resources
Jannik Ströten, Julian Zell, Michael Gertz
Database Systems Research Group, Heidelberg University, Im Neuenheimer Feld 348, 69120 Heidelberg, Germany

Motivation
Temporal Tagging
● extraction & normalization of temporal expressions

Main Challenge
● normalizing relative and underspecified expressions

Different Domains
[1] pose different challenges
[2] require different strategies

Existing Approaches
[ existing focus on English
[ existing focus on news documents

News...[1998-04-18]
... for the United States, he said...[today]. On May 22, 1995, Farkas was made a brigadier general, and...[the following year]. However, cited by police in December for driving under the influence of alcohol...

Narrative...[2009-12-19]
[1979], Soviet invasion...[land in Kabul on December 25],...[they were complying with the Treaty of Friendship...[entered Afghanistan from the north on December 27]. In the morning, the 103rd...

HeidelTime: a multilingual, cross-domain temporal tagger

Key Features
[2] extraction: regular expressions & NLP features
[2] normalization: knowledge resources & linguistic clues

Language-independent
● resource interpreter
[2] domain-dependent normalization strategies
[2] reference time

Language-dependent
● pattern files
[2] month=[...|April|May|...]
[2] normalization files
[2] normMonth(April)=04
[2] rule files

TempEval-3
Temporal Tagging (Task A)
● English and Spanish news documents
● annotation according to TimeML

Evaluation
● strict and relaxed extraction
[2] type and value normalization
[2] ranking attribute: value F1 (relaxed)

From HeidelTime 1.2 to 1.3
● improved weekday normalization
[2] annotations closer to TimeML

English Adaptations
● X_REX values
[2] improved negative rules for ambiguous expressions (e.g., may, fall, march)

Developing Spanish Resources
Four Steps to Add a New Language:
(1) Preprocessing:
[2] sentence, token, PoS information
[2] HeidelTime uses TreeTagger
[2] Spanish TreeTagger module available

(2) Translation of Pattern Files:

(3) Translation of Normalization Files:

(4) Iterative Rule Development
[2] starting with (simple) English rules
[2] checking training corpus for errors (FP, FN, partial matches, incorrect values)
[2] adapting patterns, normalizations, and rules to improve results on training data

TempEval-3 Evaluation Results

English

<table>
<thead>
<tr>
<th></th>
<th>strict F1</th>
<th>relaxed F1</th>
<th>value F1</th>
</tr>
</thead>
<tbody>
<tr>
<td>HeidelTime 1.3</td>
<td>81.34</td>
<td>90.30</td>
<td>77.81</td>
</tr>
<tr>
<td>HeidelTime 1.2</td>
<td>78.07</td>
<td>86.99</td>
<td>72.12</td>
</tr>
<tr>
<td>Navy News</td>
<td>79.57</td>
<td>90.32</td>
<td>70.97</td>
</tr>
</tbody>
</table>

(next best system by value F1)

Spanish

<table>
<thead>
<tr>
<th></th>
<th>strict F1</th>
<th>relaxed F1</th>
<th>value F1</th>
</tr>
</thead>
<tbody>
<tr>
<td>HeidelTime 1.3</td>
<td>85.33</td>
<td>90.13</td>
<td>85.33</td>
</tr>
<tr>
<td>TipSemB</td>
<td>82.57</td>
<td>87.40</td>
<td>71.85</td>
</tr>
<tr>
<td>jrc</td>
<td>49.53</td>
<td>65.20</td>
<td>50.78</td>
</tr>
</tbody>
</table>

Summary
[2] English: 8 teams, 21 submissions
[2] Spanish: 3 teams, 3 submissions

Error Analysis & Conclusions
False Negatives
● expressions that cannot be normalized with high probability (some time)

False Negatives / False Positives
● trade-off due to X_REX expressions; annotated inconsistently (currently)

Incorrect Value Normalization
● due to partial matches
[2] incorrect relation to reference time

Spanish Resources
[2] benefit from high quality English resources as starting point
[2] contain many patterns and rules not in Spanish training data

Availability
HeidelTime’s Current Version
[2] as UIMA component
[2] as standalone version (Java)
[2] online demo
[2] @ Google code

Languages
[2] English, German, Dutch, Spanish, Italian, Arabic, Vietnamese

Ongoing Work
[2] further languages to come

References

This work was presented at SemEval 2013, the 7th International Workshop on Semantic Evaluation, June 14–15, 2013, Atlanta, Georgia, USA.

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

SemEval 2013