**HeidelTime:**
Temporal Tagging on Different Domains

Jannik Strötgen and Michael Gertz

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

---

**Motivation**

Temporal Tagging
- focus mainly on news documents
- only recently on narratives

Existing Annotated Corpora
- news: TimeBank
- narratives: WikiWars

Different Domains
- pose different challenges
- require different strategies

---

**Challenges on Different Corpora**

**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...

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

**Narrative**

2009-12-19
1979: Soviet invasion and landing on Kabul on December 25. ... they were complying with the 1978 Treaty of Friendship. ... entered Afghanistan from the north on December 27. In the morning, the 103rd...

- reference time often in text
- relation to reference time
- long documents, rich discourse structure

**SMS**

2010-01-10
10:51:19
Whats u wanted 2 say last night?

2010-09-23
09:50:50
Yo! Rem 2 come for lab tmr...

2011-02-16
12:42:42
... and is available at 10 am in his office

- relation to reference time
- non-standard language (errors, word creations, ...)
- missing context information

**Scientific**

2009-12-19
... Subjects consumed one tablet each day containing... Subjects were assessed at baseline, three and six months...

- often no real reference time
- local semantics (document time frame)
- many durations and sets

---

**HeidelTime’s Strategy**

- Reference Time Detection
  → domain-dependent

- Relation to Reference Time
  → domain-dependent

- Broad Coverage of Expression Types
  → domain-dependent rules

- Non-standard Language
  → domain-dependent pattern
  e.g., colloquial synonyms

- Local Semantics
  → based on “time point zero”

---

**Domain-sensitive Evaluation [1]**

<table>
<thead>
<tr>
<th>Corpus</th>
<th>Extraction</th>
<th>Normalization</th>
</tr>
</thead>
<tbody>
<tr>
<td>news</td>
<td>91.1</td>
<td>84.1</td>
</tr>
<tr>
<td>TimeBank</td>
<td>91.1</td>
<td>84.1</td>
</tr>
<tr>
<td>(news)</td>
<td>91.1</td>
<td>84.1</td>
</tr>
<tr>
<td>colloq</td>
<td>81.1</td>
<td>74.0</td>
</tr>
<tr>
<td>scient</td>
<td>80.0</td>
<td>80.1</td>
</tr>
<tr>
<td>WikiWars</td>
<td>91.7</td>
<td>91.3</td>
</tr>
<tr>
<td>(narrative)</td>
<td>89.7</td>
<td>80.5</td>
</tr>
<tr>
<td>colloq</td>
<td>81.6</td>
<td>79.6</td>
</tr>
<tr>
<td>scient</td>
<td>80.8</td>
<td>80.1</td>
</tr>
<tr>
<td>Time4SMS</td>
<td>91.7</td>
<td>91.3</td>
</tr>
<tr>
<td>(colloquial)</td>
<td>95.1</td>
<td>93.9</td>
</tr>
<tr>
<td>colloq</td>
<td>91.8</td>
<td>91.3</td>
</tr>
<tr>
<td>scient</td>
<td>89.7</td>
<td>89.8</td>
</tr>
<tr>
<td>Time4SCI</td>
<td>69.7</td>
<td>55.8</td>
</tr>
<tr>
<td>(scientific)</td>
<td>72.8</td>
<td>58.2</td>
</tr>
<tr>
<td>colloq</td>
<td>78.3</td>
<td>72.4</td>
</tr>
<tr>
<td>scientist</td>
<td>64.7</td>
<td>69.5</td>
</tr>
</tbody>
</table>

- HeidelTime’s F-scores
- correct strategy outperforms others
- especially for normalization task

---

**HeidelTime’s Features [2]**

- multilingual (English, German, Dutch)
- easy extensible to further languages
- cross-domain
  (news, narrative, colloquial, scientific)

---

**Availability**

HeidelTime
- as UIMA component
- as standalone version (Java)
- online demo
- @ Google code

Corpora
- Time4SMS
- Time4SCI
- WikiWarsDE

---

**References**


---

Contact Information:

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

This work was presented at LREC 2012, the 8th Conference on Language Resources and Evaluation, May 21-27, 2012, Istanbul, Turkey.