**Motivation**

Temporal Tagging
- extraction & normalization of temporal expressions

Main Challenge
- normalizing relative and underspecified expressions

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

Existing Approaches
- focus on English
- focus on news documents

**HeidelTime: Multilingual, Cross-domain Temporal Tagger**

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

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

Language-dependent
- pattern files
- month=.../April/...)
- normalization files
- normMonth(April)=04
- rule files

**EVENTI at EVALITA 2014**

Temporal Tagging (Subtask A)
- Italian news documents
- TimeML's TIMEX3 annotations
- task guidelines and Ita-TimeBank corpus as reference

Evaluation
- strict and relaxed extraction
- type and value normalization
- ranking criterion: value F1 (strict)

**From HeidelTime 1.7 to 1.8**

- collection reader & CAS consumer to process EVENTI data
- empty TIMEX3 tags
- anchored durations
- range expressions
- improved Italian resources

**Developing Italian Resources**

Four Steps to Add a New Language:
(1) Preprocessing:
- TreeTagger for sentence splitting, tokenization, and POS tagging

(2) Translation of Pattern Files:
- extracted as separate tags

(3) Translation of Normalization Files:
- only current version

(4) Iterative Rule Development
- writing rules from scratch or taking an existing language as reference
- checking training corpus for errors (FP, FN, partial matches, incorrect values)
- adapting patterns, normalizations, and rules to improve results on training data

**False Negatives**
- expressions that cannot be normalized with high probability (some time)
- allora - therefore or at that time?

**Incorrect Value Normalization**
- underspecified DATE expressions (XXXX-XX-XX)
- TIME expressions without date
- wrong date for TIME expressions

**Partial Matches with Correct Value**
- un lasso di tempo di 14 giorni - only 14 giorni extracted
- ieri verso le 11 - ieri and verso le 11 extracted as separate tags

**Error Analysis**

False Positives
- 26 anni - 26 years or 26 year old?

**EVENTI Task Evaluation Results**

<table>
<thead>
<tr>
<th></th>
<th>strict F1</th>
<th>relaxed F1</th>
<th>value F1</th>
</tr>
</thead>
<tbody>
<tr>
<td>HeidelTime 1.7</td>
<td>66.2</td>
<td>78.0</td>
<td>57.1</td>
</tr>
<tr>
<td>HeidelTime 1.8</td>
<td>80.4</td>
<td>87.8</td>
<td>68.0</td>
</tr>
<tr>
<td>(without empty tags)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HeidelTime 1.8</td>
<td>82.1</td>
<td>89.3</td>
<td>70.9</td>
</tr>
</tbody>
</table>

Team B-1  
82.7  88.6  66.5
Team C-1  
66.2  76.8  56.6
Team C-2  
65.9  77.1  56.3

**False Positives**
- 26 anni - 26 years or 26 year old?

**Availability**

HeidelTime's Current Version
- as UIMA component
- as standalone version (Java)
- online demo
- @ Google code

11 Languages
- English, German, Dutch, Spanish, Italian, French, Arabic, Vietnamese, Russian, Chinese, Croatian

Ongoing Work
- further languages to come

**References**


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

This work was presented at EVALITA 2014, the 4th Evaluation Campaign of NLP and Speech Tools for Italian, December 11, 2014, Pisa, Italy.