Multilingual Domain-sensitive Temporal Tagging with HeidelTime

Jannik Strötgen, Max Planck Institute for Informatics, Saarland Informatics Campus, Saarbrücken, Germany

Temporal Tagging [1]
- extraction of temporal expressions
- normalization to standard format

Most Taggers
- focus on English
- focus on news documents

Main Challenge: ambiguities
- “May I march 2000 miles before I fall.”
- normalizing non-explicit expressions

Availability (github)
- as UIMA and GATE components
- Java standalone and online demo

easy to use / modify / extend

Challenges on Different Document Types [1]

News [1998-04-18]
... United States... he said... On May 22, 1995... Farkas was made a brigadier general... and was available for driving under the influence of alcohol...
- reference time often DCT
- relation to reference time

Narrative [2009-12-19]
... United States... Farkas was made a brigadier general... and was available for driving under the influence of alcohol...
- reference time often in text
- rich discourse structure

SMS [2010-01-10]
Y o! Rem to come for lab tmr
- relation to reference time
- noisy (word creations, ...)

Scientific [2009-12-19]
... Subjects consumed... Subjects were assessed... Clini-cal pathology analysis was performed... and six months...
- often no real reference time
- local time frame

HeidelTime's Architecture
- language-independent, domain-specific normalization strategies
- language-dependent resources

HeidelTime's Language Resources
- pattern files
  - reMonthLong
    - January
    - February
    - March
  - normMonth
    - "January","01"
    - "February","02"
- normalization files
- rules
  - normMonth...
  - Extr="%reMonthLong","%reDayTh","%reYear4D" %
  - Norm="%g-%normMonth(g)-%normDay(g)-%normYear(g)"
- Extraction regex, pos constraints, pattern files
- Normalization domain-sensitive strategies, linguistic clues (tense), normalization files

temporal expressions occur everywhere – and can be exploited in various applications

Literary Corpus Analysis
- when does literature take place?
- browsing “today in world literature”

Times in Street Names
- streets with dates commemorate important events in a region’s history

Time-centric Search
- time-aware search engine for Wikipedia
- handles explicit temporal constraints

What’s your domain?
What’s your language?
How temporal are your texts?
Test it!