

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

Newly Developed Corpora

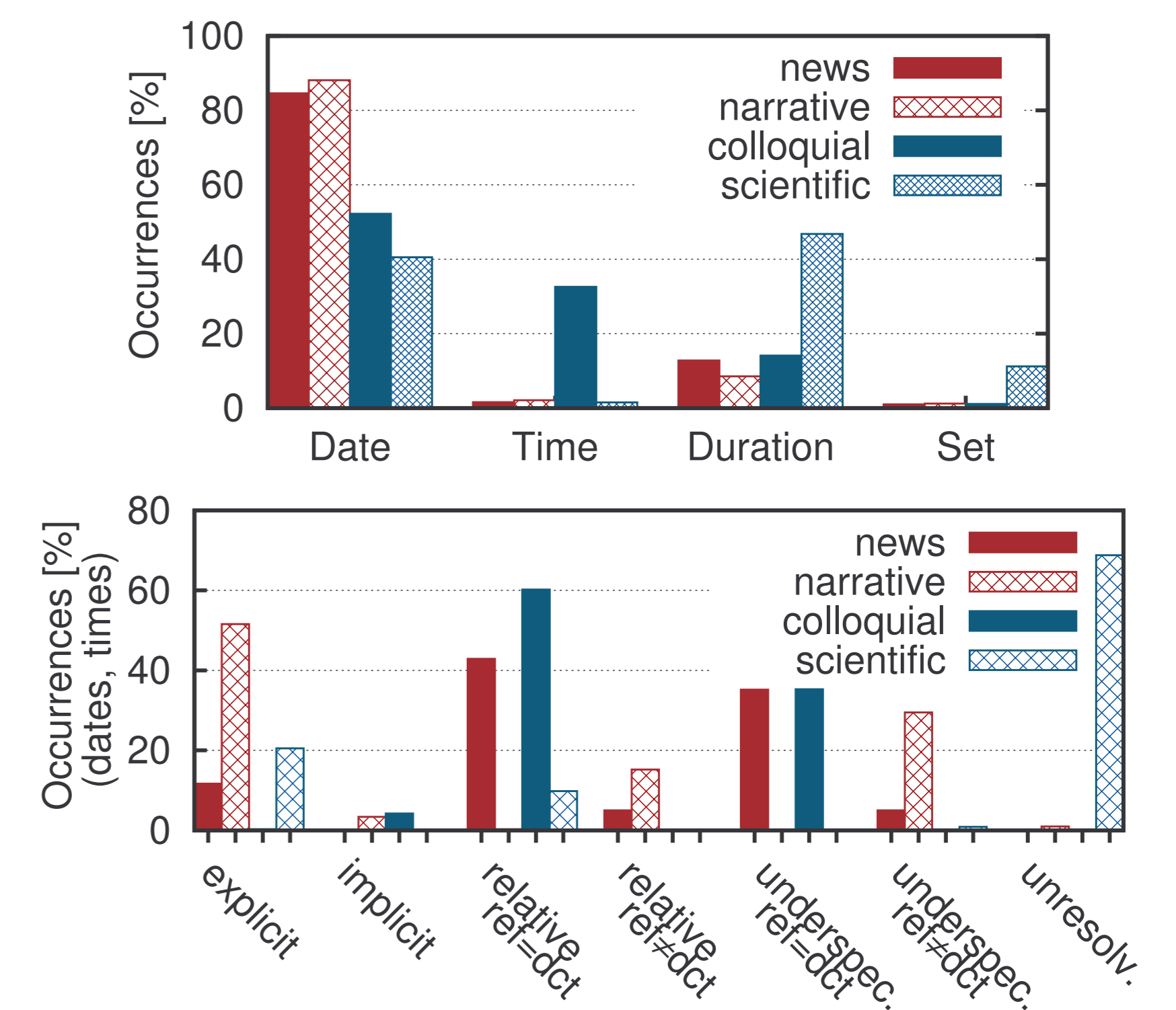
Time4SMS

- 1,000 documents of NUS SMS corpus
- manually annotated
- total of 1,341 temporal expressions

Time4SCI

- 50 PubMed articles on clinical trails
- manually annotated
- total of 317 temporal expressions

Different Domains – Different Types



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 ... land in 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

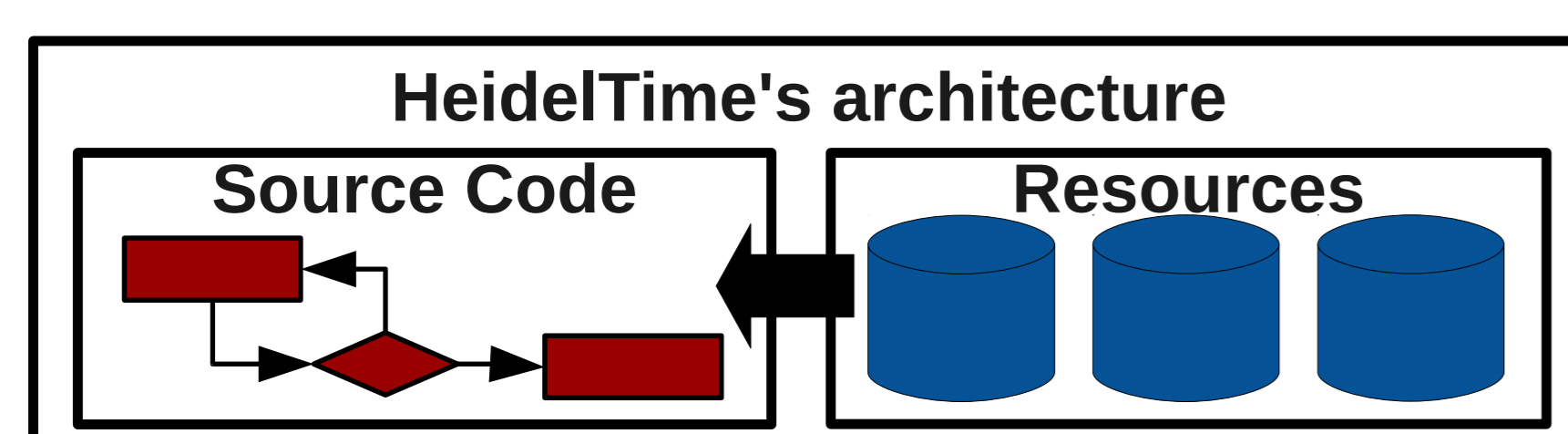
SMS 2010-01-10T05:19
Whats it u wanted 2 say **last nite**?
SMS 2010-09-23T09:50
Yo! Rem to come for lab **tmr** :) ...
SMS 2011-02-16T12:42
... andy is available at **10 am** in his office

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

Scientific 2009-12-19
... Subjects consumed one tablet **per day** containing ... Subjects were assessed **at baseline**, **three** and **six months** ... Clinical pathology analysis was performed **at baseline** 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

→ via tense, heuristics

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]

corpus	strategy	extraction		normalization	
		lenient	strict	lenient	strict
TimeBank (news)	news	91.1	84.1	78.6	73.8
	narrat	91.1	84.1	61.5	57.7
	colloq	91.1	83.4	78.4	72.9
WikiWars (narrative)	news	87.9	80.5	56.9	53.9
	narrat	87.9	80.5	78.7	74.5
	colloq	88.1	79.6	56.5	52.8
Time4SMS (colloquial)	news	91.7	91.3	89.8	89.7
	narrat	91.7	91.3	88.4	88.3
	colloq	95.1	93.9	92.3	91.9
Time4SCI (scientific)	news	69.7	55.8	51.9	49.5
	narrat	69.7	55.8	51.9	49.5
	colloq	72.8	58.2	55.1	52.0
	scient	78.3	72.4	69.5	64.7

- 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

- [1] J. Strötgen and M. Gertz: **Temporal Tagging on Different Domains: Challenges, Strategies, and Gold Standards**. *LREC*, 2012.
- [2] J. Strötgen and M. Gertz: **Multilingual and Cross-domain Temporal Tagging**. *Language Resources and Evaluation*, Online First, 2012.

Contact Information:

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