EventMiner

Michal Starý

EventMiner

- extract dates from text
- interpret (= normalize) dates to absolute values
- connect dates to relevant events
- deliver results

- Both EN, CS supported

Motivation

Temporal entity extraction
- quick recap

Temporal entity extraction process

Entity recognition

Entity normalization

- absolute
- relative (needs correct reference date)

Temporal entity extraction process example

Kate,

please, can you confirm the meeting at 6 PM tomorrow? It's about time to prepare our talk at the NLP conference, it's on the 27th of November, remember? We need to submit the presentation 2 weeks before the conference itself.

PS. I've talked to John on Wed. and the vacation won't be a problem.

See you,

Michal

Send: 02-11-2019

03-11-2019T18:00:00

27-11-2019

13-11-2019

30-10-2019

Existing tools

SUTime

Heideltime

Dateparser

Duckling

(Natty)

SUTime

Rule based temporal entity extractor

Great performance on more formal and shorter texts (legal records), the less formal and the longer the worse performance was observed

Uses rules for matching and interpreting, POS tags for past/future distinction and filtering

Problems/limitations:

- Not handling shortcuts, typos
- Only one reference date for the whole document
- Only EN!

SUTime result

Kate, please, can you confirm the meeting at 6 PM tomorrow? It's about time to prepare our talk at the NLP conference, it's on the 27th of November, remember? We need to submit the presentation 2 weeks before the conference itself. PS. I've talked to John on Wed. and the vacation won't be a problem. See you, Michal

6 PM tomorrow **2019-11-03T18:00** <TIMEX3 tid="t1" type="TIME" value="2019-11-03T18:00">6 PM tomorrow</TIMEX3>

the 27th of November **2019-11-27** <TIMEX3 tid="t2" type="DATE" value="2019-11-27">the 27th of November</TIMEX3>

2 weeks **P2W** <TIMEX3 tid="t3" type="DURATION" value="P2W">2 weeks</TIMEX3>

HeidelTime

Similar to SUTime, but keeps evolving

Looks for the reference date inside the texts.

For EN, DE, ES, IT, NE, ...(~10 languages) have high F1 (manual rules)

Extended for all languages using rules automatic translation (not so useful)

- Good baseline
- *"Note that recall is much worse than precision for all languages and that the normalization of extracted expressions works quite well (value acc.)."*

Strötgen, J., & Gertz, M. (2015, September). A baseline temporal tagger for all languages. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing* (pp. 541-547).

HeidelTime result

Kate, please, can you confirm the meeting at 6 PM tomorrow? It's about time
to prepare our talk at the NLP conference, it's on the 27th of November,
remember? We need to submit the presentation 2 weeks before the conference
itself. PS. I've talked to John on Wed. and the vacation won't be a problem.
See you, Michal

6 PM	XXXX-XX-XXT18:00
tomorrow	2019-11-03
the 27th of November	2019-11-27
2 weeks	P2W

	F1	P	R	strict F1	value F1
HeidelTime-t	90.30	93.08	87.68	81.34	**77.61**
HeidelTime-bf	87.31	90.00	84.78	78.36	72.39
HeidelTime-1.2	86.99	89.31	84.78	78.07	72.12
NavyTime-1,2	**90.32**	89.36	**91.30**	79.57	70.97
ManTIME-4	89.66	95.12	84.78	74.33	68.97
ManTIME-6	87.55	98.20	78.99	73.09	68.27
ManTIME-3	87.06	94.87	80.43	69.80	67.45
SUTime	**90.32**	89.36	**91.30**	79.57	67.38
ManTIME-1	87.20	97.32	78.99	70.40	67.20
ManTIME-5	87.20	97.32	78.99	69.60	67.20
ManTIME-2	88.10	97.37	80.43	72.22	66.67
ATT-2	85.25	98.11	75.36	78.69	65.57
ATT-1	85.60	**99.05**	75.36	79.01	65.02
ClearTK-1,2	90.23	93.75	86.96	**82.71**	64.66
JU-CSE	86.38	93.28	80.43	75.49	63.81
KUL	83.67	92.92	76.09	69.32	62.95
KUL-TE3RunABC	82.87	92.04	75.36	73.31	62.15
ClearTK-3,4	87.94	94.96	81.88	77.04	61.48
ATT-3	80.85	97.94	68.84	72.34	60.43
FSS-TimEx	85.06	90.24	80.43	49.04	58.24
TIPSem (TE2)	84.90	97.20	75.36	81.63	65.31

Table 3: Task A - Temporal Expression Performance.

UzZaman, N., Llorens, H., Derczynski, L., Allen, J., Verhagen, M., & Pustejovsky, J. (2013, June). Semeval-2013 task 1: Tempeval-3: Evaluating time expressions, events, and temporal relations. In *Second Joint Conference on Lexical and Computational Semantics (* SEM), Volume 2: Proceedings of the Seventh International Workshop on Semantic Evaluation (SemEval 2013)* (pp. 1-9).

Duckling

Haskell library for generic entities interpretation

Does not support entity recognition

Temporal entity interpretation for 20+ languages

Common problem - Reference date acquisition

Relation extraction task

The **hardest task** in whole temporal entity extraction process

Solutions:

- Create own model to get a correct reference date for each relative entity
  - Either heuristics, machine learning or combination of both could be used
  - UzZaman, N., Llorens, H., Derczynski, L., Allen, J., Verhagen, M., & Pustejovsky, J. (2013, June). Semeval-2013 task 1: Tempeval-3: Evaluating time expressions, events, and temporal relations. In *Second Joint Conference on Lexical and Computational Semantics (* SEM), Volume 2: Proceedings of the Seventh International Workshop on Semantic Evaluation (SemEval 2013)* (pp. 1-9).
- Be satisfied with HeidelTime reference date acquisition

Reference date acquisition example

I'm looking forward to `Saturday`'s tournament. It all starts at `8 AM` and ends before `noon`. Make sure to be there!

Saturday 2019-11-23   **8 AM 2019-11-26T08:00**   **noon 2019-11-26T12:00**

I'm looking forward to Saturday's tournament. It all starts at 8 AM and ends before noon. Make sure to be there!

Saturday's 23-11-2019   **8 AM 23-11-2019T08:00**

Temporal entity extraction - progress

Work done

1) Explored EN, CS datasets
2) Evaluated different NER tools/models on related datasets
3) Tried several aproaches
4) Built Albert-biLSTM-CRF NER model for English
5) Prepared Iterative normalization scheme

Findings

1) Datasets - temporal entities

a) CS
   i) None

b) EN
   i) TBAQ
   ii) TimeBank and Aquaint corpora
   iii) well annotated
   iv) very good dataset
   v) ~1800 datetime entities

1) Datasets - general NE including date/time

a) CS
   i) CNEC
   ii) not so good
   iii) date/time annotation is focused only on more obvious absolute values
   iv) ~3k datetime entities
   v) not suitable for training

b) EN
   i) Ontonotes
   ii) surprisingly good
   iii) very well annotated
   iv) ~15k datetime entities
   v) suitable for training

2) Tools / models

Evaluation of recognition on TBAQ

Manual observation

2) Tools - evaluation on TBAQ

Metrics - soft:

a) preprocess
b) BOW approach
c) remove entity after match
d) add prefix "the"
e) check if reference ent is subword in predicted entity
f) erase errors caused by "now" keyword - easy to match manually

2) Tools - evaluation score

Sutime

soft_precision: 0.805
soft_recall: 0.754
soft_F1: 0.778

HeidelTime

soft_precision: 0.887
soft_recall: 0.741
soft_F1: 0.808

Multilingual BERT NER

soft_precision: 0.860
soft_recall: 0.752
soft_F1: 0.803

BERT NER

soft_precision: 0.725
soft_recall: 0.695
soft_F1: 0.710

2) Tools - observation

multilingual BERT based model recover also surrounding with temporal function

multilingual BERT based model is surprisingly good for CS

HeidelTime and Sutime are much faster and lighter

Bonus - general NER for CS (simplified boolean matching):

Model	soft F1
multiNER (NER currently used at FI)	0.580
multilingual BERT NER	0.675

Albert based NER

Why Albert based NER?

BERT ~ 5GB RAM

Albert > BERT (performance)

Albert < BERT (memory need)

(Hopefully) soon - multilingual Albert

Properties

Trained on Ontonotes dataset (finalized on TBAQ)

~ 20 entities (including date, time)

Useful for following tasks in the pipeline as well

Outcome

EN

1) HeidelTime + Sutime [+Albert based] (recognition and basic normalization)
2) Iterative normalization

CS

1) Deeppavlov multilingual BERT NER (recognition)
2) Feature engineering on top of duckling (basic normalization)
3) Iterative normalization

Iterative normalization

Issues to solve

Need of domain knowledge - catch and interpret the specific entities with high accuracy; past vs. future focus

- "due XXX", "within XXX", ...

Way to properly distinguish the absolute values of several orders

- totally absolute (specific date): DD/MM/YYYY, DD/MM
- partially absolute (specific interval): MM/YYYY, YYYY

Model choice

Draft

Finetuned by hand/ ML model

Features:

- distance
- anaphora
- syntactic tree relationship
- POS
- modifying keyword in surrounding
- paragraph?
- topics?
- ...
- ...any suggestions?