i2b2
The Sixth i2b2 Workshop on Challenges in NLP for Clinical Data , Chicago, 2nd November 2012

Extraction of temporal expressions, events and relations from clinical narratives using rules and machine-learning

Aleksandar Kovačević[1], Azad Dehghan[2], Michele Filannino[2]*, John A. Keane[2,3] and Goran Nenadic[2,3]

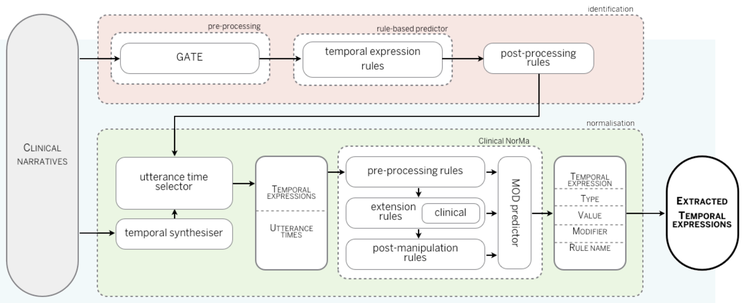
[1] Faculty of Technical Sciences, University of Novi Sad, Serbia
[2] School of Computer Science, University of Manchester, United Kingdom
[3] Manchester Institute of Biotechnology, United Kingdom
* Presenting author, e-mail: filannim@cs.man.ac.uk

TASK 1: TEMPORAL EXPRESSIONS

The **identification** phase consists of a pre-processing step, using GATE[2], followed by a **rule-based** pipeline. For the **normalisation** phase we have developed Clinical NorMA, an **open-source** dictionary-driven rule-based system.

- **Run 1** uses rules optimised for F-measure
- **Run 2** uses recall-optimised rule-based predictions
- **Run 3** uses rules optimised according to the precision measure

Normalisation settings have been kept constant through the different runs.



# Run	Strict matching			Lenient matching			Normalisation		
	P	R	F	P	R	F	Type	Value	Modifier
Run 1	**83.79**	81.37	**82.56**	**91.61**	89.05	90.31	81.37	68.40	80.67
Run 2	76.81	**83.00**	79.78	86.04	**93.17**	89.46	**84.86**	**72.05**	**83.93**
Run 3	82.58	82.22	82.40	90.59	90.30	**90.44**	82.38	69.25	81.68

training set - micro-averaged results

# Run	Strict matching			Lenient matching			Normalisation		
	P	R	F	P	R	F	Type	Value	Modifier
Run 1	78.03	78.41	78.22	89.23	89.62	89.42	76.47*	39.95*	79.89*
Run 2	77.03	**79.62**	78.30	88.68	**91.54**	**90.08**	**84.73**	**70.44**	**82.75**
Run 3	**79.85**	77.09	**78.45**	**90.38**	87.25	88.79	65.49*	39.89*	77.80*

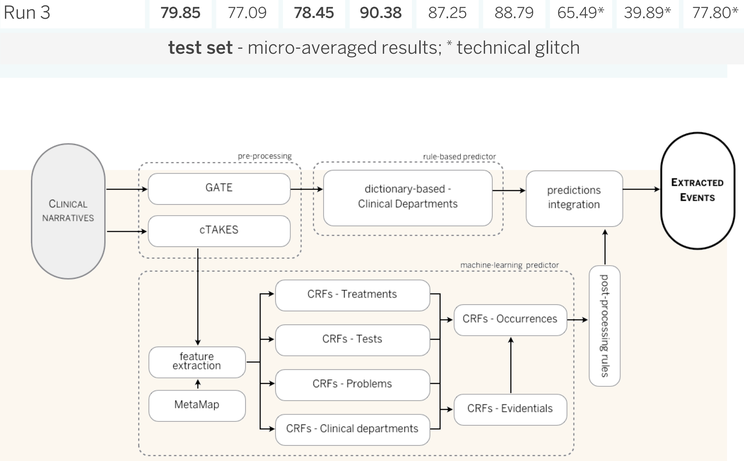
test set - micro-averaged results; * technical glitch

TASK 2: EVENTS

The event recognition task relies on a **hybrid** architecture (rules and machine learning). The pre-processing module uses cTAKES[4]. For each type of event, a CRF[3]-based component has been trained, whereas the recognition of clinical departments has been tackled using a dictionary-based component trained on **i2b2 2010**[5] and **2012 training sets**.

Runs differ for the clinical department predictions:

- **Run 1** uses the dictionary-based component for clinical departments
- **Run 2** uses the union of both the components for clinical departments



# Run	Strict matching			Lenient matching			Attributes	
	P	R	F	P	R	F	Polarity	Modality
Run 1	82.54	**74.71**	**78.43**	89.95	**81.42**	**85.47**	**75.22**	**78.42**
Run 2	**82.56**	74.68	78.42	**89.96**	81.39	85.46	75.19	78.39

training set - micro-averaged results

# Run	Strict matching			Lenient matching			Attributes	
	P	R	F	P	R	F	Polarity	Modality
Run 1	**82.05**	77.05	79.71	**89.64**	84.66	87.08	78.81	80.08
Run 2	81.74	**78.05**	**79.85**	89.35	**85.32**	**87.29**	**79.45**	**81.53**

test set - micro-averaged results

TASK 3: TEMPORAL RELATIONS

Recognition of temporal links is divided into **two steps**:

- **identify** pairs of entities (events and temporal expressions) to be linked together

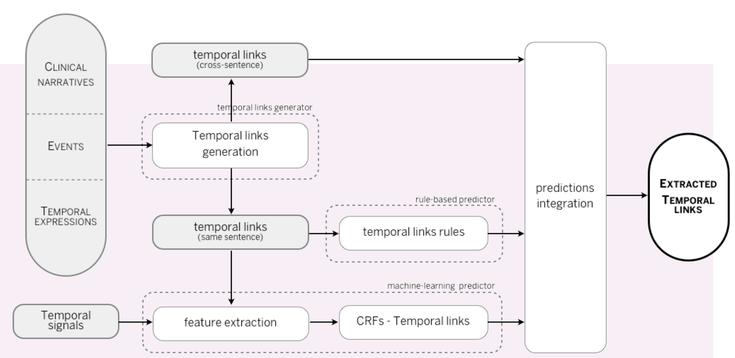
  The temporal links are **generated over every pair** of:

  a) events in the same sentence;
  b) events and temporal expressions in the same sentence;
  c) events sharing a common MetaMap[1] concept;
  d) events referring to the same temporal expression;
  e) events connected with a cTAKES[4] co-reference relation;
  f) events and appropriate section date (admission or discharge).

- **categorise** them in one of After, Before and Overlap.

The rule-based component uses **lexical collocations** and **temporal signals** to predict the link type whereas the ML one uses **CRFs**[3].

- **Run 1** uses recall-optimised rules and the machine learning predictor
- **Run 2** uses precision-optimised rules and machine learning predictor (step a, b, c and f)
- **Run 3** maximises the F-measure using recall-optimised rules and machine learning predictor.



# Run	Submitted *			Corrected ;)		
	P	R	F	P	R	F
Run 1	77.06*	**25.02***	**37.98***	37.84	**55.42**	44.57
Run 2	**80.57***	24.07*	37.07*	**51.45**	48.57	**49.97**
Run 3	80.54*	24.06*	37.05*	38.06	54.78	44.91

test set - micro-averaged results; * partial results

TASK 4: END-TO-END

The end-to-end task aimed to test the entire pipeline by using the events and temporal expressions predicted, instead of the gold standard.

Task 1 Temporal expressions → **Task 2** Events → **Task 3** Temporal relations

The architecture of the system is the same as Task 3.

- **Run 1** maximises the recall by using temporal expressions and events predicting with the recall-optimised runs (second runs respectively) along with temporal relations generated from steps a, b, c, d and f in Task 3
- **Run 2** maximised the F-measure by using the first runs of tasks 1 and 2 along with the same previous temporal relations settings
- **Run 3** aims at maximise the precision.

Temporal expressions

P	R	P&R	F	Type	Value	Modifier
85.93	86.21	86.07	86.07	80.71	67.42	79.01

Events

P	R	P&R	F	Type	Polarity	Modality
88.15	84.17	86.11	86.11	81.15	81.15	80.43

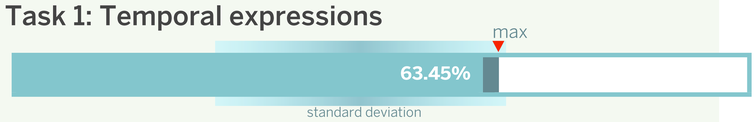
Temporal links

P	R	P&R	F
32.81	36.32	34.10	34.48

test set - micro-averaged results

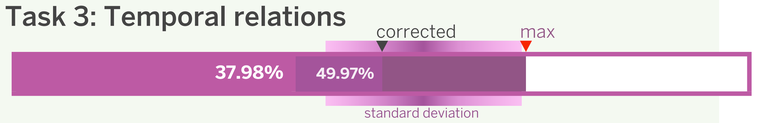
SUMMARY

Task 1: Temporal expressions

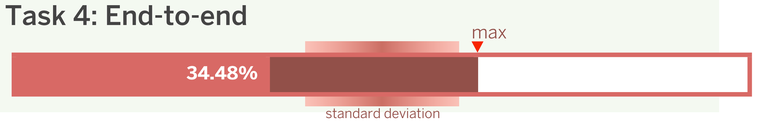


Task 2: Events

Task 3: Temporal relations



Task 4: End-to-end



[1] Aronson, A. R. Effective mapping of biomedical text to the UMLS metathesaurus: the MetaMap program. Proceedings / AMIA ... Annual Symposium. AMIA Symposium (2001), 17–21.
[2] Cunningham, H., Maynard, D., Bontcheva, K., and Tablan, V. Gate: A framework and graphical development environment for robust NLP tools and applications. In Proceedings of the 40th Anniversary Meeting of the Association for Computational Linguistics (ACL'02) (2002).
[3] Lafferty, J. D., McCallum, A., and Pereira, F. C. N. Conditional random fields: Probabilistic models for segmenting and labelling sequence data. In ICML (2001), pp. 282–289.
[4] Savova, G. K., Masanz, J. J., Ogren, P. V., Zheng, J., Sohn, S., Schuler, K. K., and Chute, C. G. Mayo clinical text analysis and knowledge extraction system (cTAKES): architecture, component evaluation and applications. JAMIA 17, 5 (2010), 507–513.
[5] Uzuner, O., South, B. R., Shen, S., and DuVall, S. L. 2010 i2b2/va challenge on concepts, assertions, and relations in clinical text. JAMIA 18, 5 (2011), 552–556.

EPSRC Engineering and Physical Sciences Research Council
CDT CS Centre for Doctoral Training in Computer Science
Linked2Safety
NHS National Institute for Health Research