The 15th Annual Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies

LIMSI-COT at SemEval-2016 Task 12: Temporal relation identification using a pipeline of classifiers

Julien Tourille[1,2], Olivier Ferret[3], Aurélie Névéol[1], Xavier Tannier[1,2]

[1]LIMSI, CNRS, Université Paris-Saclay, F-91405, Orsay, [2]Université Paris-Sud, [3]CEA, LIST, F-91191, Gif-sur-Yvette

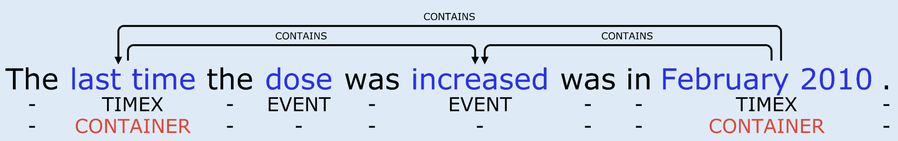
surname.name@limsi.fr ; surname.name@cea.fr

Container Relation Subtask (CR)

Task Objective: identify narrative container relations.

Container Classifier

Objective: classification of entities according to whether or not they are the source of one or more CONTAINS relations

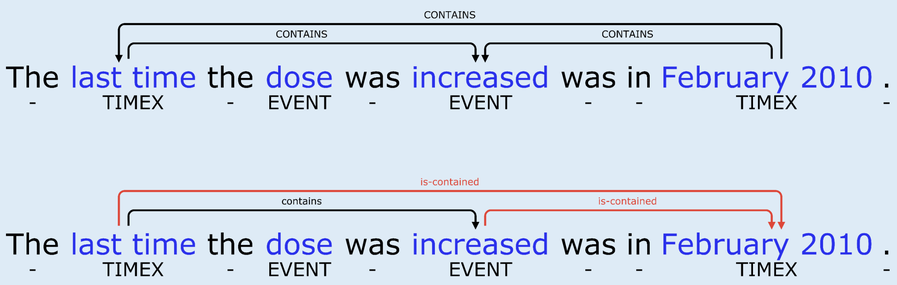


Intra-Sentence Relation Classifier

Objective: classification of entity pairs within sentences

Method:

- Transformation of a 2-category problem (contains, no-relation) into a 3-category problem (contains, no-relation, is-contained) to reduce the number of pairs.

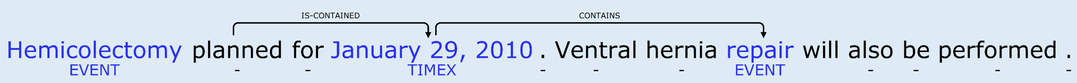


Inter-Sentence Relation Classifier

Objective: classification of entity pairs across sentences

Method:

- 3-category problem (contains, is-contained, no-relation)
- 3-sentence window



\+

List detection module

Objective: automatic recognition of laboratory results

Method: regular expressions

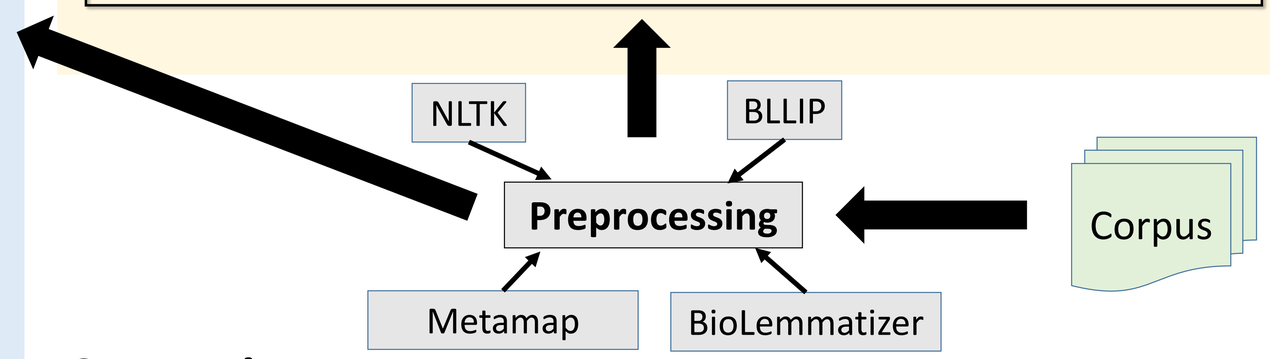
Document Creation Time Relation Subtask (DR)

Task objective: identify the relation between an event and the document creation time.

DocTime Relation Classifier

Objective: EVENT classification according to their relation to the Document Creation Time

Classes: *before*, *before-overlap*, *overlap*, *after*



Strategies

- RUN 1: Plain lexical features: surface forms
- RUN 2: Word Embeddings: vectors calculated on the MIMIC II corpus using word2vec

Machine Learning Algorithms

Run	Classifier	Algorithm	% of feat. space
1	CONTAINER	SVM (RBF)	60
	INTRA	SVM (RBF)	60
	INTER	SVM (RBF)	100
	DCT	SVM (Linear)	100
2	CONTAINER	SVM (Linear)	100
	INTRA	SVM (Linear)	100
	INTER	SVM (Linear)	100
	DCT	Random Forests	100

Machine learning algorithms used for the final submission

Features

Feature	DocTime Classifier	Container Classifier	Intra-sent. Classifier	Inter-sent. Classifier
Surface form	✓	✓	✓	✓
Gold standard attributes	✓	✓	✓	✓
Lemma	✓	✓	✓	
POS and CPOS tags	✓	✓	✓	
Semantic types and semantic groups	✓	✓	✓	✓
Entity type	✓	✓	✓	✓
Token count between the two entities			✓	✓
Entity count between the two entities			✓	✓
Syntactic paths between the two entities			✓	
Container model prediction			✓	
Intra-sentence model prediction				✓
Sentence context				
Gold standard entities – Lemma, surfaces form, POS and CPOS tags, semantic types and semantic groups	✓	✓		
Gold standard entities in-between – type, attributes, semantic types and semantic groups, container model prediction or intra-sentence model prediction, count			✓	✓
Tokens – Lemmas, POS and CPOS tags	✓	✓		
Gold standard entities – count before and after	✓	✓		
Section context				
Gold standard entities – Lemmas, surface forms, POS and CPOS tags, semantic types and semantic groups	✓			
Relative position of the sentence(s)	✓			✓
Tokens – count before and after, lemmas, POS and CPOS tags	✓			
Document context				
Gold standard entities – count before and after, semantic types and semantic groups, type, attributes	✓			

Results

Run	Ref	Pred	Corr	P	R	F1
1	18,990	18,989	14,603	0.769	0.769	0.769
2	18,990	18,989	15,317	0.807	0.807	0.807

DR subtask performance

Run	Ref	Pred	Corr	P	R	F1
1	5,894	3,755	2,642	0.704	0.436	0.538
2	5,894	2,544	1,911	0.751	0.320	0.449

CR subtask performance