

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

Julien Tourille<sup>1,2</sup>, Olivier Ferret<sup>3</sup>, Aurélie Névéol<sup>1</sup>, Xavier Tannier<sup>1,2</sup>

<sup>1</sup> LIMSI, CNRS, Université Paris-Saclay, F-91405, Orsay

<sup>3</sup> CEA, LIST, F-91191, Gif-sur-Yvette

<sup>2</sup> Université Paris-Sud



# Outline

---

1. Introduction

2. Document Creation Time Relation Subtask

3. Container Relation Subtask

4. Results

5. Conclusion and Perspectives

# Task Description

---

## THYME Corpus

- Clinical notes and Pathological Notes from the Mayo Clinic
- Manually annotated with **events**, **temporal expressions** and **narrative container relations**

## Six Subtasks

1. **TS**: identifying the **spans of time expressions**
2. **ES**: Identifying the **spans of event expressions**
3. **TA**: identifying the **attributes of time expressions**
4. **EA**: identifying the **attributes of event expressions**
5. **DR**: identifying the **relation between an event and the document creation time**
6. **CR**: identifying **narrative container relations**

# Task Description

---

## THYME Corpus

- Clinical notes and Pathological Notes from the Mayo Clinic
- Manually annotated with **events**, **temporal expressions** and **narrative container relations**

## Six Subtasks

1. **TS**: identifying the **spans of time expressions**
2. **ES**: Identifying the **spans of event expressions**
3. **TA**: identifying the **attributes of time expressions**
4. **EA**: identifying the **attributes of event expressions**
5. **DR**: identifying the **relation between an event and the document creation time**
6. **CR**: identifying **narrative container relations**

# Temporal relation subtasks (1/2)

---

## Document Creation Time Relation Subtask (DR)

→ **Objective:** identify the relation between an event and the document creation time

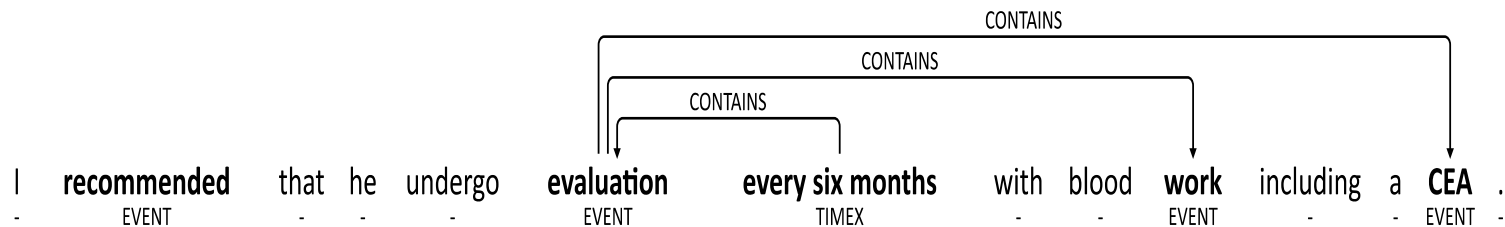
→ **Classes:** {before, before-overlap, overlap, after}

She	<b>feels</b>	slightly	<b>weak</b>	but	has	<b>resumed</b>	most	of	her	normal	<b>activities</b>	.
-	EVENT	-	EVENT	-	-	EVENT	-	-	-	-	EVENT	-
-	OVERLAP	-	OVERLAP	-	-	BEFORE	-	-	-	-	BEFORE-OVERLAP	-

# Temporal relation subtasks (2/2)

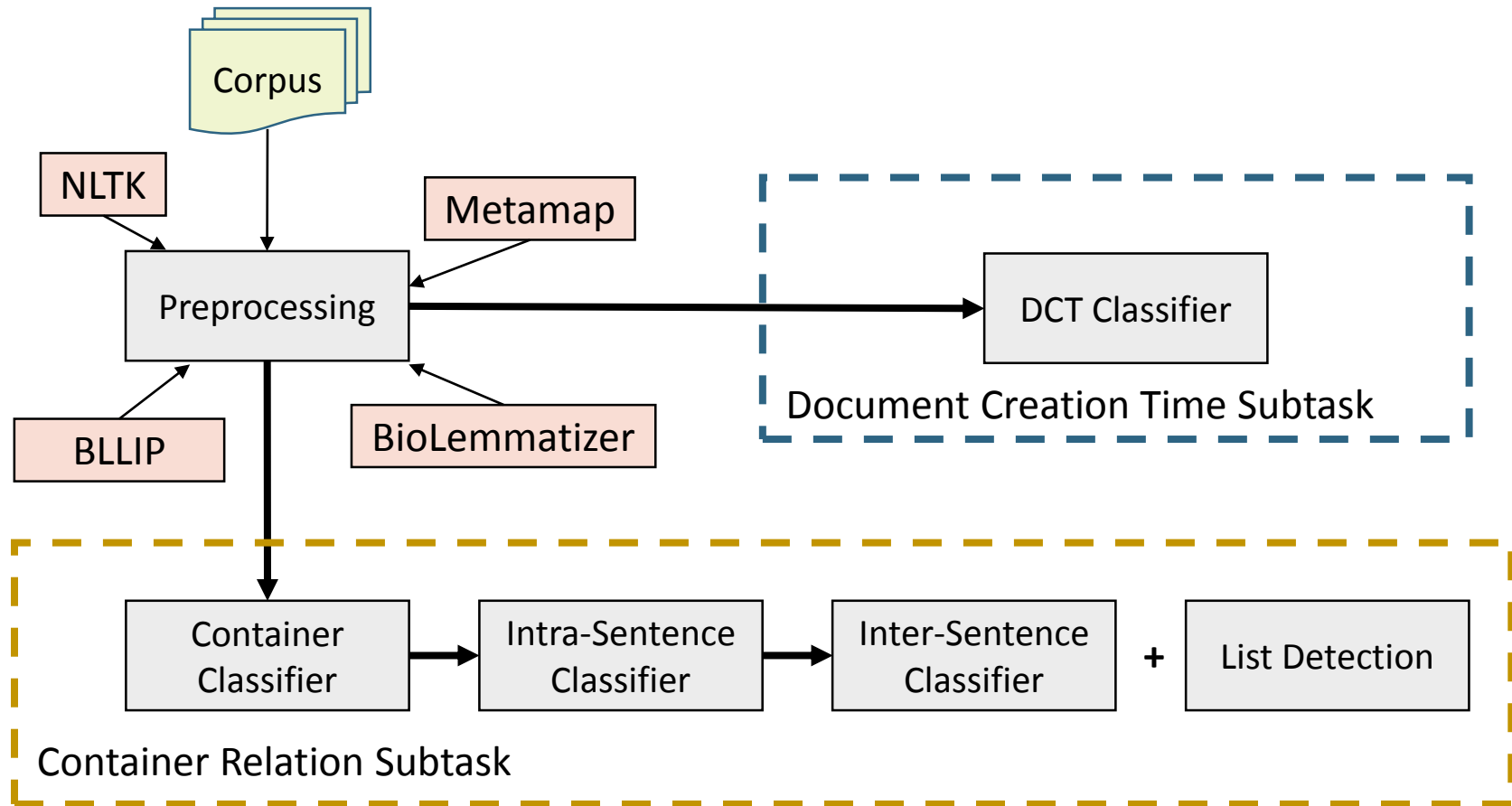
## Container Relation Subtask (CR)

→ Objective: identify narrative container relations

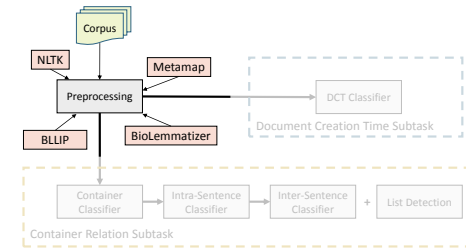


Every six months CONTAINS evaluation CONTAINS (blood work AND CEA)

# System Overview



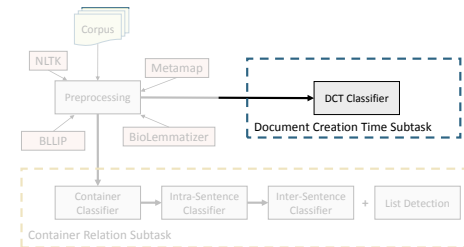
# Preprocessing



1. **Sentence segmentation:** NLTK – Punkt sentence Tokenizer (Loper and Bird, 2002)
2. **Parsing:** BLLIP Reranking Parser (Charniak and Johnson, 2005) + Pre-trained biomedical parsing model (McClosky, 2010)  
→ POS and CPOS tags + syntactic dependencies
3. **Lemmatization:** BioLemmatizer (Liu et al., 2012)
4. **Medical entity recognition:** Metamap (Aronson and Lang, 2010)  
→ Semantic types and semantic groups



# DR Subtask Overview



**Method:** supervised classification problem

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

**Features:**

## 1. Entity:

- surface form, gold-standard attributes, lemma(s), POS and CPOS tags, semantic types and semantic groups

## 2. Sentence context:

- gold-standard entities: lemma, surface form, POS and CPOS tags, semantic types and semantic groups, count before and after
- tokens: lemma, POS and CPOS tags

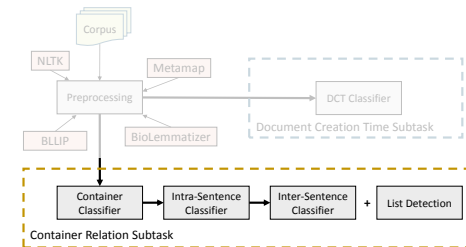
## 3. Section context:

- gold-standard entities: lemma, surface form, ...
- relative position of the sentence
- tokens: count before and after, lemmas, POS and CPOS tags

## 4. Document context:

- gold standard entities: count before and after, semantic types and semantic groups, type, attributes

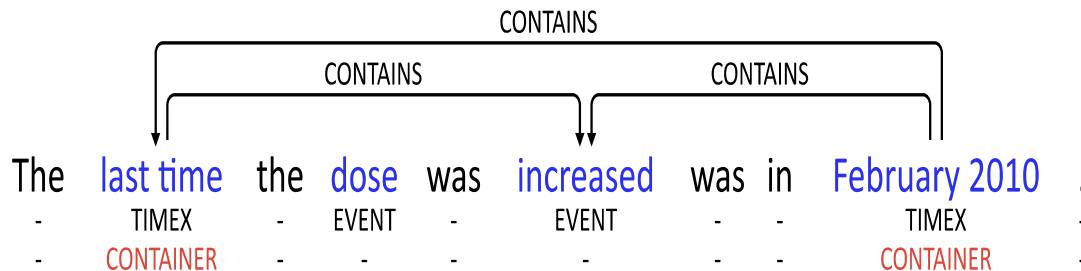
# Container Classifier



**Intuition:** some entities are more likely to be containers e.g. TIMEX

## Container Classifier

Classify each EVENT/TIMEX according to whether or not they are likely to be a container (contains other EVENT/TIMEX)



Used as **feature** for the **intra-sentence classifier**

# Container Relations

## Quantitative analysis:

Total number of CONTAINS relations: 17,474

→ **13,304** intra-sentence relations ( $\approx 76\%$ )

→ **4,170** inter-sentence relations ( $\approx 24\%$ )

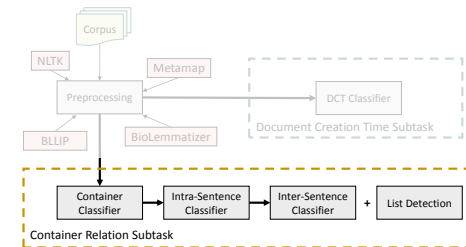


## Task decomposition

- Intra-sentence classifier:** allow the use of fine-grained features at the sentence level provided by sentence analysis tools such as syntactic analyzers
- Inter-sentence classifier**

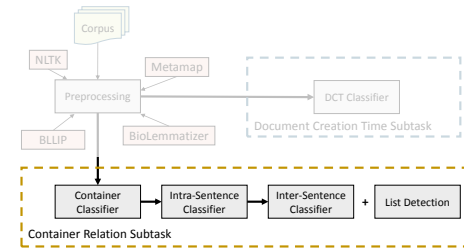
**Problem:** inter-sentence level event combination is huge

→ **Inter-sentence dataset is unbalanced**



# Inter-sentence relations

## Container relation by window size

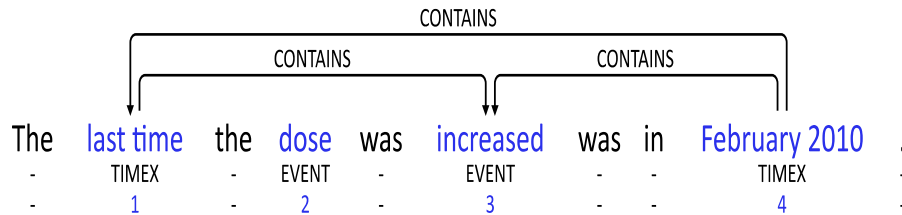


Window	Number of relations	Total
1	13,304	13,304 (76.30%)
2	1,463	14,767 (84.69%)
3	752	15,519 (89.00%)
4	497	16,016 (91.85%)
5	364	16,380 (93.94%)
6	151	16,531 (94.80%)

- Intra-sentence candidate pairs: 222,698
- Inter-sentence candidate pairs: 622,568
- **Inter-sentence dataset** remains **strongly unbalanced**

# Complexity Reduction

## All permutations

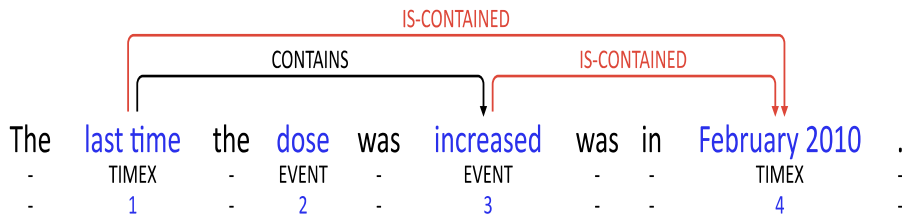


**Classes:** *contains, no-relation*

**Pairs candidates:** 12

**Pairs:** 1-2, 2-1, 1-3, 3-1, 1-4, 4-1, 2-3, 3-2, 2-4, 4-2, 3-4, 4-3

## All combinations from left to right



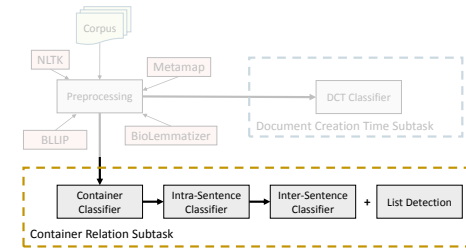
**Classes:** *contains, no-relation, is-contained*

**Pairs candidates:** 6

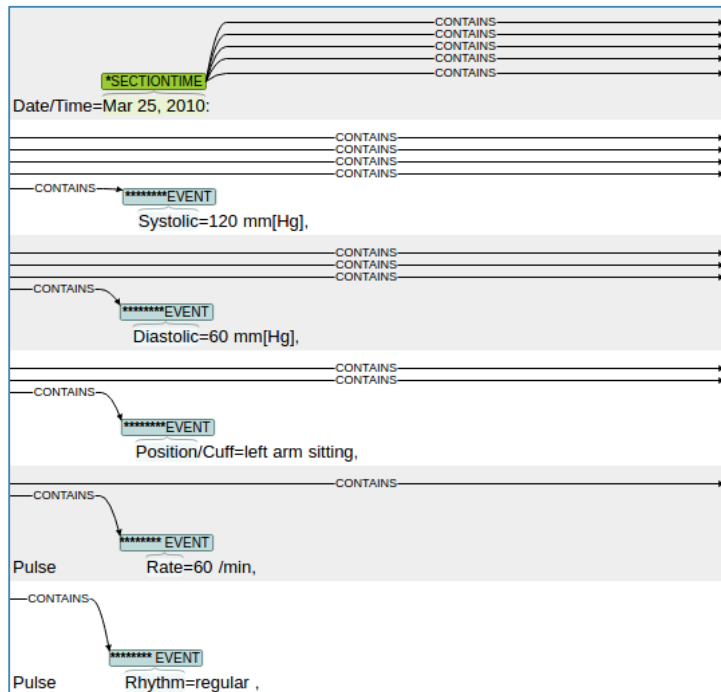
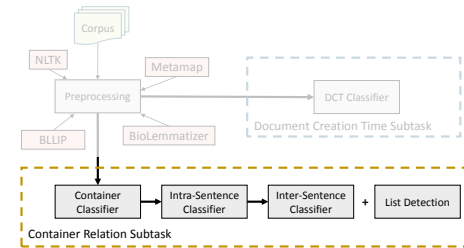
**Pairs:** 1-2, 1-3, 1-4, 2-3, 2-4, 3-4

Intra-sentence candidate pairs: from 222,698 to **111,349**

Inter-sentence candidate pairs: from 622,568 to **311,284**



# List Detection



**Objective:** increase recall at inter-sentence level

**Method:** regular expressions to detect structured parts of texts related to laboratory results

# CR Subtask overview

## Three Classifiers

1. Container
2. Intra-sentence relations
3. Inter-sentence relations

+ One list detection module based on regular expressions

## Features:

### 1. Entity:

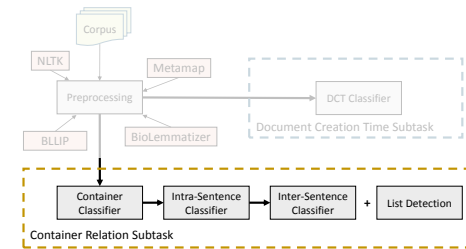
- surface form, gold-standard attributes, lemma(s), POS and CPOS tags, semantic types and semantic groups, token count between the two entities, entity count between the two entities, syntactic paths between the two entities, model predictions

### 2. Sentence context:

- gold-standard entities: lemma, surface form, POS and CPOS tags, semantic types and semantic groups, count before and after
- tokens: lemma, POS and CPOS tags

### 3. Section context:

- relative position of the sentence



# Parameters

---

## Strategies

- **Run 1:** plain lexical features
- **Run 2:** word embeddings computed on the MIMIC II corpus (Saeed et al., 2011)

## Machine learning algorithms

Run	Classifier	Algorithm	% of feature space
Plain lexical features	CONTAINER	SVM (RBF)	60
	INTRA	SVM (RBF)	60
	INTER	SVM (RBF)	100
	DCT	SVM (Linear)	100
Word embeddings	CONTAINER	SVM (Linear)	100
	INTRA	SVM (Linear)	100
	INTER	SVM (Linear)	100
	DCT	Random Forests	100



# DR Subtask - Performance

---

	P	R	F1
Plain lexical feature	-	0.769	-
Word embeddings	-	<b><u>0.807</u></b>	-
Max	-	0.843	-
Median	-	0.724	-
Baseline	-	0.675	-

# Plain Lexical Features - Performance

	Pred	Corr	P	R	F1	
Intra Classifier	3229	2468	0.764	0.409	0.533	
+ Inter Classifier	3651	2619	0.717	0.432	0.539	↗
+ List Detection	3755	2642	0.704	0.436	0.538	↘
Max	-	-	0.823	0.564	0.573	
Median	-	-	0.589	0.345	0.449	
Baseline	-	-	0.459	0.154	0.231	

Container classifier accuracy on dev corpus = 0.917

# Word Embeddings - Performance

	Pred	Corr	P	R	F1	
Intra Classifier	2296	1845	0.804	0.310	0.447	
+ Inter Classifier	2440	1888	0.774	0.317	0.449	↗
+ List Detection	2544	1911	0.751	0.320	0.449	=
Max	-	-	0.823	0.564	0.573	
Median	-	-	0.589	0.345	0.449	
Baseline	-	-	0.459	0.154	0.231	

Container classifier accuracy on dev corpus = 0.924

# Conclusion & Perspectives

---

- **Efficient model** based on **simple modules**
  - Document Creation Time Relation subtask: multiclass classifier
  - Container Relation subtask: pipeline of classifiers
- Complexity can be handled by **problem transformation** and **recall/complexity trade-off**
  - 2-class problem → 3-class problem
  - Limited window size for inter-sentence relations
- **Word embedding does not improved systematically performance**  
→ Further investigation is needed
- **The model does fit on other languages**: similar results on **French**

# Thank you !

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

Julien Tourille<sup>1,2</sup>, Olivier Ferret<sup>3</sup>, Aurélie Névéol<sup>1</sup>, Xavier Tannier<sup>1,2</sup>

<sup>1</sup> LIMSI, CNRS, Université Paris-Saclay, F-91191, Orsay

<sup>3</sup> CEA, LIST, F-91191, Gif-sur-Yvette

<sup>2</sup> Université Paris-Sud

