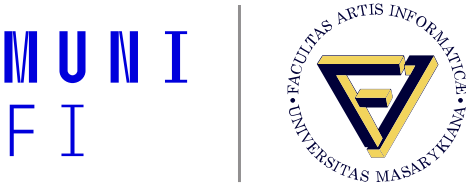




# Multilingual Recognition of Temporal Expressions

Michal Starý, Zuzana Nevěřilová, and Jakub Valčík

**Michal Starý**

**xstary1@fi.muni.cz**

Konica Minolta Laboratory Europe &

NLPC, Faculty of Informatics, Masaryk University

December 8, 2020

# Outline

- Introduction
- Related work
- Our Goal
- Our approach
- Results
- Analysis
- Conclusion
- Future work

# Introduction

## Temporal information

Overview:

- Temporal information plays a significant role in several NLP tasks such as information retrieval, summarization, question answering, or event extraction.
- The ultimate goal of temporal processing is to extract *events* from unstructured text, i.e., *what* happens, *when* and *how* it relates to some other events.

Tasks:

1. temporal expression recognition (TER)
2. temporal expression normalization
3. event detection
4. temporal relation extraction

# Introduction

## Example

Send: 02-11-2019

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

Figure: Plain text email

# Introduction

## Example

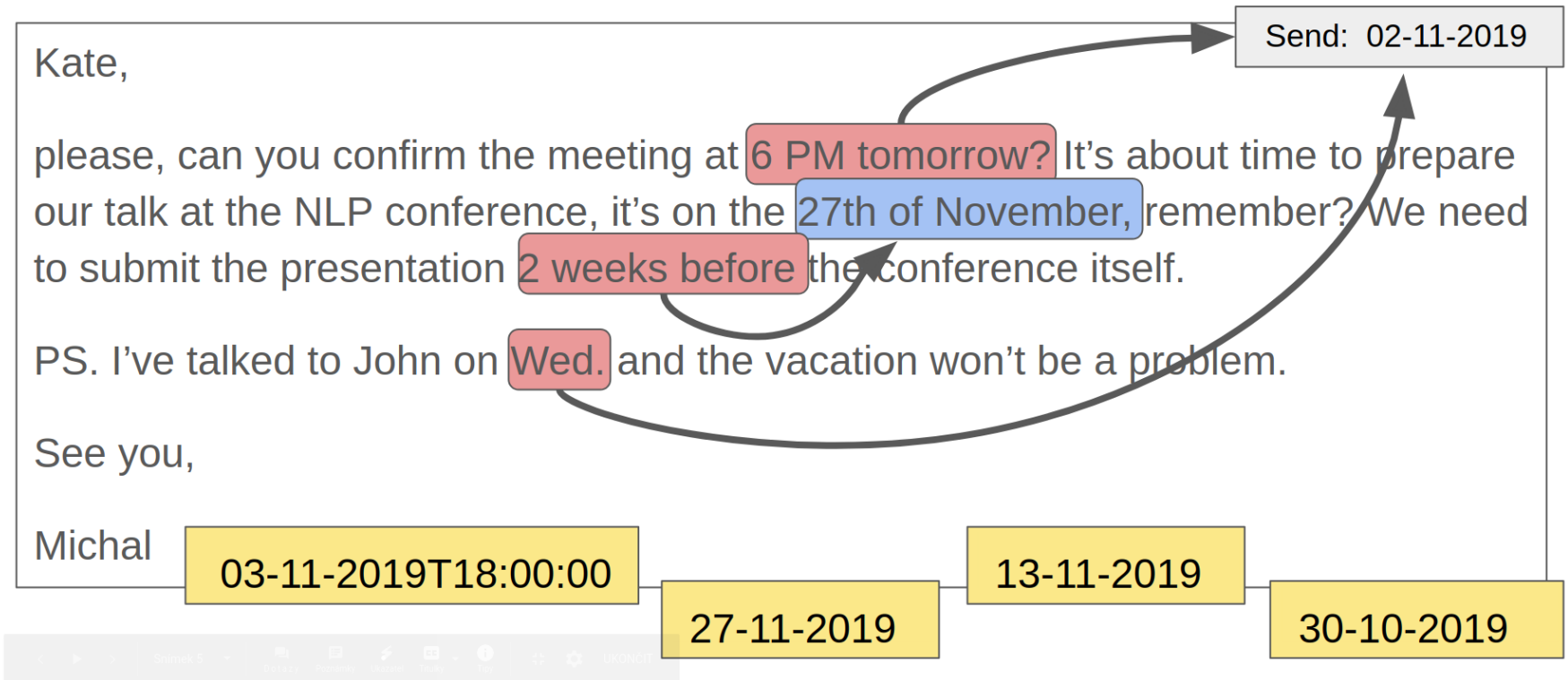


Figure: Temporal expressions recognized and normalized

# Introduction

## Example

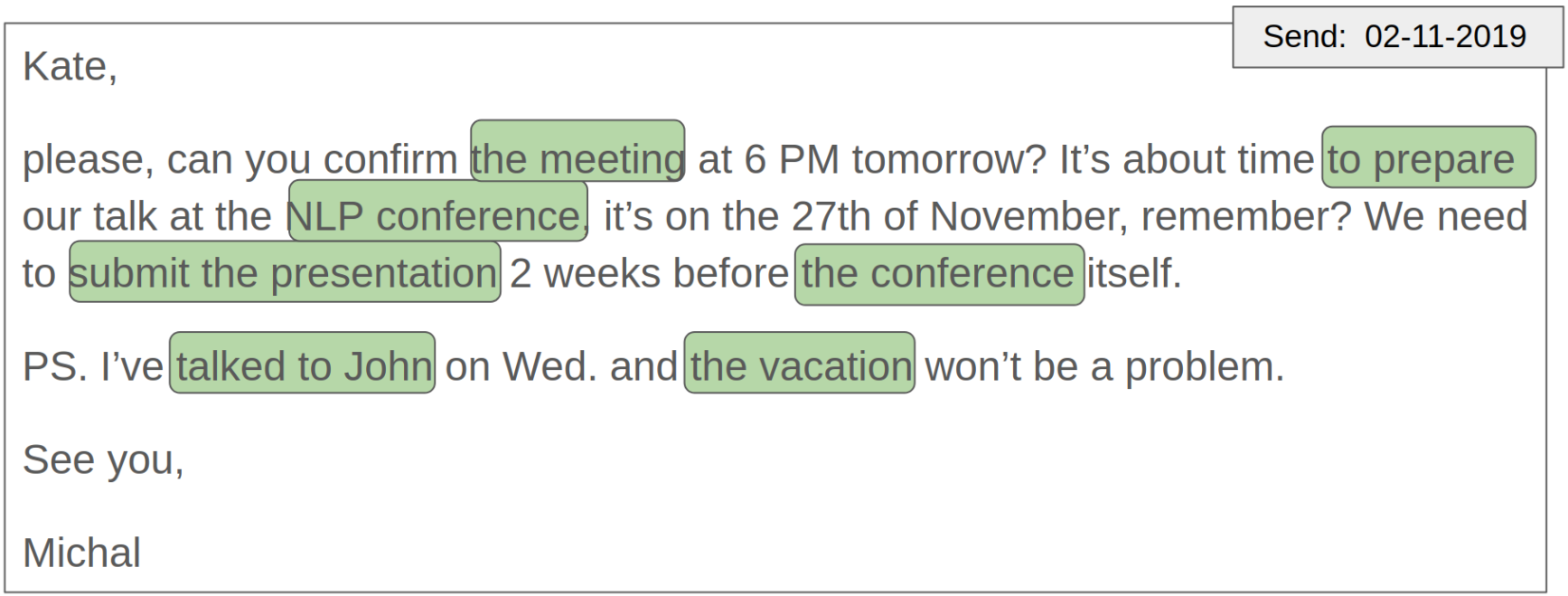


Figure: Events candidates detected

# Introduction

## Temporal expression recognition (TER)

Recognize different kinds of temporal expressions:

- well-structured (*November, 12*)
- numeric (*10/12/2005*)
- compound (*the 2012 through 2016 tax years*)
- domain specific (*3rd Quarter*)
- expressions with less common modifiers (*prior year*)

# Related work

## Tools, models for TER

- Rule based
  - HeidelTime
  - SynTime
  - SUTime
- ML based
  - TOMN
  - PTIME
- (multilingual)BERT based
  - BERT TER by Chen et al. [1]
  - multilingual BERT TER by Lange et al. [2]

# Related work

## Temporal datasets, Markup languages

Markup languages

- TimeML

Datasets are mostly EN, few in other languages.

- TBAQ (TimeBank and AQUAINT) [6] - EN
- TE3-platinium [6] - EN
- KRAUTS [5] - DE
- PolEval2019 [3] - PL

# Our goal

Multilinguality

- recognize temporal entities in more than one hundred languages

Cross domain support

- cover not only generic expressions, but also the temporal expressions specific for different domains

Competitiveness

- comparable performance to state-of-the-art monolingual methods in monoligual setting

# Our approach

- Use existing tools and models
- Exploit that TER $\subset$ NER
- Get the best from the rule based tools
- Get the best from the data driven tools
- Focus on Recall (customizable second stage classifier)

# Our approach

## Base models

- multilingual BERT NER by DeepPavlov [1]
  - based on multilingual BERT model
  - NER model trained on OntoNotes dataset
  - trained on thousands temporal expressions from 6 domains
  - impressive results in zero-shot learning
- HeidelTime [4]
  - rule based system
  - native support for 13 major languages
  - automatic multilingual support (automatically translated rules)
  - mostly for the news domain

[1] http://docs.deeppavlov.ai/en/master/features/models/ner.html

# Our approach

## Combined model

1. Recognize temporal expressions using HeidelTime with appropriate rule set
2. Recognize named entities using multilingual BERT NER
3. Select temporal expressions from recognized named entities
4. Combine temporal expressions from HeidelTime and BERT NER
   - Take union of non-overlapping expressions from both methods
   - Greedily merge overlapping expressions from both methods into longer expressions
5. Return final expressions

# Our approach

## Combined model

### Example

Bubbly Day

---

HeidelTime: Pour a glass of sparkling sunshine to celebrate National Bubbly **Day** every first **Saturday** in **June**!

$H = \{(Day, 65, 67), (Saturday, 81, 88), (June, 93, 96)\}$

BERT NER: Pour a glass of sparkling sunshine to celebrate National Bubbly Day **every first Saturday in June**!

$B = \{(every\ first\ Saturday\ in\ June, 69, 96)\}$

$B_{non-overlap} = \emptyset$
$H_{non-overlap} = \{(Day, 65, 67)\}$
$C_1 = \{(Day, 65, 67)\}$
$C_2 = \{(every\ first\ Saturday\ in\ June, 69, 96)\}$

Combined: Pour a glass of sparkling sunshine to celebrate National Bubbly **Day every first Saturday in June**!

$C = \{(Day, 65, 67), (every\ first\ Saturday\ in\ June, 69, 96)\}$

---

# Results

## Mentioned models comparison in monolingual setting

Mostly EN, few in other languages.

| Language | Dataset | Method | Strict | | | Relaxed | | |
|---|---|---|---|---|---|---|---|---|
| | | | P | R | F1 | P | R | F1 |
| English | TBAQ-gold | HeidelTime | **83.64** | **83.32** | **83.48** | **91.68** | 91.33 | **91.5** |
| English | TBAQ-gold | BERT NER | 72.52 | 66.9 | 69.6 | 90.07 | 83.1 | 86.44 |
| English | TBAQ-gold | Combined | 69.74 | 75.91 | 72.69 | 86.79 | **94.46** | 90.46 |
| English | TE3-platinium | HeidelTime | **83.85** | **78.99** | **81.34** | **93.08** | 87.68 | **90.3** |
| English | TE3-platinium | BERT NER | 76.07 | 64.49 | 69.8 | 92.31 | 78.26 | 84.71 |
| English | TE3-platinium | Combined | 73.19 | 73.19 | 73.19 | 89.13 | **89.13** | 89.13 |
| German | KRAUTS | HeidelTime | **80.29** | **65.05** | **71.87** | **90.15** | 73.03 | 80.69 |
| German | KRAUTS | BERT NER | 53.92 | 35.96 | 43.15 | 64.94 | 54.13 | 64.94 |
| German | KRAUTS | Combined | 70.53 | 64.13 | 67.18 | 84.76 | **77.06** | **80.73** |

Table: Evaluation in monolingual setting

# Results

## Multilingual setting

Mostly EN, few in other languages.

| Language | Dataset | Method | Strict | | | Relaxed | | |
|---|---|---|---|---|---|---|---|---|
| | | | P | R | F1 | P | R | F1 |
| German | KRAUTS | HeidelTime-A | **59.47** | 24.5 | 34.7 | **91.31** | 37.61 | 53.28 |
| German | KRAUTS | BERT NER | 53.92 | 35.96 | 43.15 | 64.94 | 54.13 | 64.94 |
| German | KRAUTS | Combined | 57.56 | **45.41** | 50.77 | 82.33 | **64.95** | 72.62 |
| German | KRAUTS | BERT aligned[2] | ? | ? | **66.53** | ? | ? | **77.82** |
| Polish | PolEval2019 | HeidelTime (EN) | 39.59 | 19.88 | 26.47 | 88.5 | 44.44 | 59.17 |
| Polish | PolEval2019 | HeidelTime (Auto) | 61.13 | 12.14 | 20.26 | **91.34** | 18.14 | 30.27 |
| Polish | PolEval2019 | BERT NER | **62.63** | 41.19 | 49.7 | 91.2 | 59.98 | 72.37 |
| Polish | PolEval2019 | Combined (HT-EN) | 57.3 | 42.63 | 48.89 | 84.7 | 63.02 | 72.27 |
| Polish | PolEval2019 | Combined (HT-Auto) | **62.63** | **46.7** | **53.37** | 89.89 | **67.42** | **77.05** |

Table: Evaluation in multilingual setting. N.B. [2] only publishes F1 scores.

# Analysis on Different Document Types

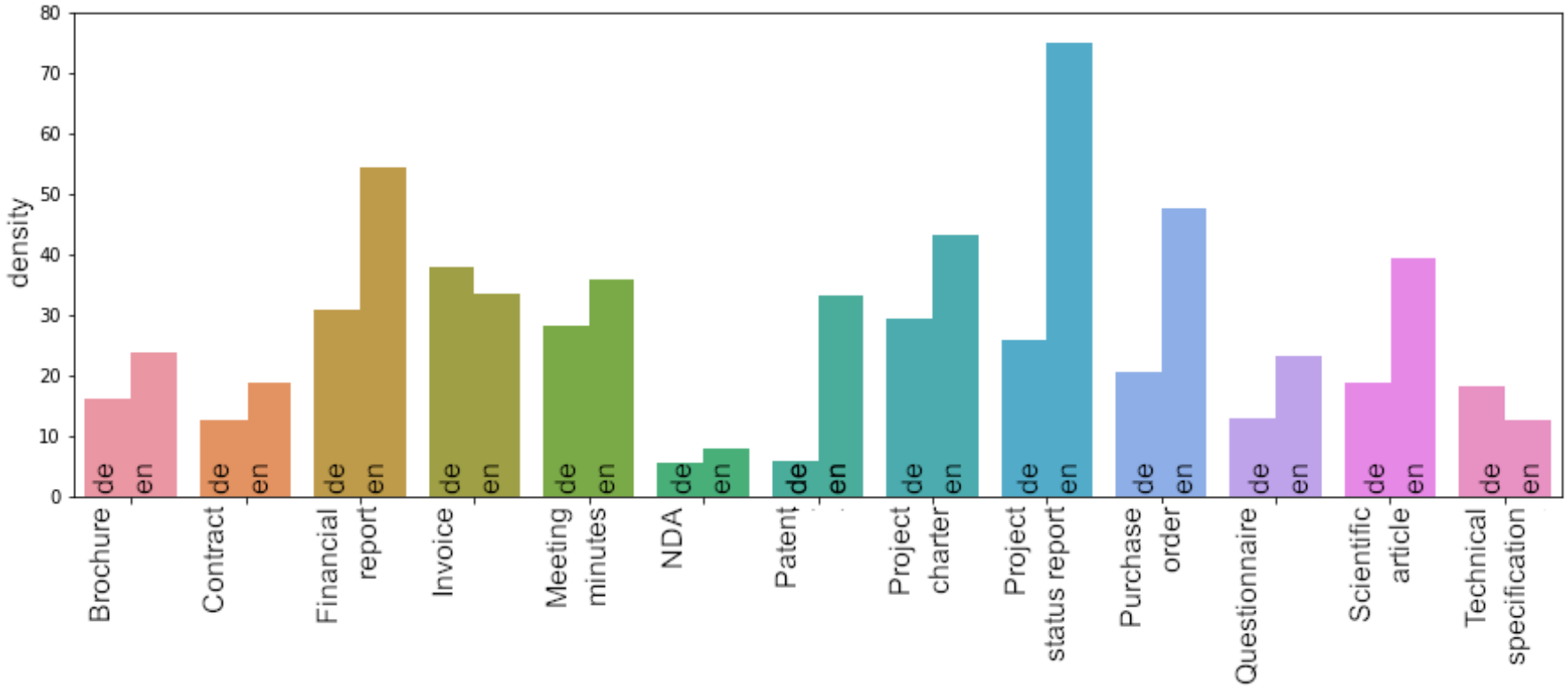


Figure: Temporal expressions by document type

# Conclusion

- Inspected and summarized the contemporary situation in multilingual TER
- Proposed a new method to improve the overall performance in multilingual and cross-domain setting
- Shown that the combination of rule based and BERT based models is a valid approach to improve the overall performance, mainly for languages different from English
- Analysed the temporal characteristics of several document types

# Future work

- Fine-tune BERT NER on temporal datasets
- Evaluate on more datasets
- Semantic classification schema

MUNI
FACULTY
OF INFORMATICS