Multilingual Recognition of Temporal Expressions

Michal Starý[1,2], Zuzana Nevěřilová[1,2], and Jakub Valčík[2]

[1] Natural Language Processing Centre, Faculty of Informatics, Masaryk University
Botanická 68a, Brno, Czech republic
[2] Konica Minolta Laboratory Europe, Holandská 7, Brno, Czech republic

Abstract. The paper presents a multilingual approach to temporal expression recognition (TER) using existing tools and their combination. We observe that the rules based methods perform well on documents using well-formed temporal expressions in a narrower domain (e.g., news), while data driven methods are more stable within less standard language and texts across domains.
With combination of the two approaches, we achieved F1 of 0.73 and 0.9 for strict and relaxed evaluations respectively on one English dataset. Although these results do not achieve the state-of-the-art on English, the same method outperformed the state-of-the-art results in a multilingual setting not only in recall but also in F1. We see this as a strong indication that combining rule based systems with data driven models such as BERT is a valid approach to improve the overall performance in TER, especially for languages other than English.
Further observations indicate that in the domain of office documents, the combined method is able to recognize general temporal expressions as well as domain specific ones (e.g., those used in financial documents).

Keywords: temporal expression, multilingual, date recognition

1 Introduction

Temporal information plays a significant role in 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.

We divide the ultimate goal into four tasks: 1. temporal expression recognition (TER), 2. temporal expression normalization, 3. event detection, and 4. temporal relation extraction. Attached together these tasks form the temporal processing pipeline that is able to process text to events.

In this work, we focus on *temporal expression recognition* in multilingual cross-domain setting. By *multilingual*, we consider a system able to recognize temporal entities in more than one hundred languages, *cross-domain* means to cover not only generic expressions, but also the temporal expressions specific for specific domains. Even though the practical impact of such universal system

A. Horák, P. Rychlý, A. Rambousek (eds.): Proceedings of Recent Advances in Slavonic Natural Language Processing, RASLAN 2020, pp. 67–78, 2020.

is tremendous, the both cross-domain and multilingual setting has received a little attention by a majority of previous work. This is probably caused by the lack of resources with annotated temporal expressions in other languages than English.

We overcome the missing annotated resources by combining two conceptually different approaches. First is the use of a rule based system with multilingual support, HeidelTime. Its multilinguality was achieved by selection and automatic translation of the rules. Second approach is use of BERT NER, a named entity recognition (NER) tool based on multilingual BERT model. Temporal expression recognition is a subset of NER.

We present a combined approach that outperforms the two approaches in recall. We compare the results across different languages and documents types. We point out document types where the combined approach achieves the best results.

2 Related work

Temporal processing was defined in 2002 by specifying the TimeML standard [10]. Further on, temporal processing has been a topic of seven SemEval[3] challenges (2007, 2010, 2013, 2015, 2018, and partly 2016, 2017). As a result, several temporal datasets and temporal processing tools have been developed over last decade.

Later, it was shown that TimeML standard is not sufficient to cover all nuances of temporal information. New annotation standard SCATE was developed [1]. Although being more accurate, it is more complicated. Only two datasets and one tool supporting SCATE annotation schema have been developed so far.

Recent research mostly focuses on (contextualized) word representations and other learning techniques in temporal processing to improve either the performance, universality, or both.

2.1 TimeML and the `TIMEX3` tag

TimeML is a markup language for temporal events in documents. It addresses four problems regarding event markup, including time stamping (that anchors an event to a time), ordering events mutually, reasoning with contextually underspecified temporal expressions, and reasoning about the length of events and their outcomes[4].

In this work, we mostly focus on the time stamping, which is realized through a `TIMEX3` tag. The `TIMEX3` tag[5] is primarily used to mark up explicit

[3] https://www.aclweb.org/anthology/venues/semeval/
[4] http://www.timeml.org
[5] http://www.timeml.org/publications/timeMLdocs/timeml_1.2.1.html#timex3

temporal expressions, such as times, dates, durations, etc. The temporal expressions have values of the following `TIMEX3` types: *DATE*, *TIME*, *DURATION*, *SET* (following the TIDES 2002 guidelines [6]).

2.2 Temporal datasets

Several temporal datasets have been developed and made publicly available. They mostly follow the TimeML annotation schema. Concerning the languages covered, English is by far the most covered with datasets from multiple domains. Followed by major Romance languages, German and a few other languages with minor resources. These resources usually contain hundreds to lower thousands of `TIMEX3` tags. Not only the number of temporal expressions but also the diversity of them is an important factor of the dataset relevance. This diversity is much higher in cross-domain datasets compared to the single domain ones.

Unfortunately, many of these resources have become unavailable due to lost of support by data providing sites. A presentation of the available temporal datasets used in this work follows.

TBAQ – TimeBank and AQUA [14] is the dataset for temporal processing. It has been developed during TempEval competitions and contains about 200 fully tagged English news documents with 1,243 `TIMEX3` tags in total.

TE-3 Platinium [14] is a small dataset containing of only 20 English news documents, that were originally used for evaluating tools developed with access to the TBAQ dataset. It contains 138 `TIMEX3` tags.

KRAUTS [13] is a German dataset containing news documents from three different sources, with a high diversity. It contains 1,140 `TIMEX3` tags.

PolEval2019 [8] is a recently created Polish dataset that contains more than 1,500 documents from various domains with 6,116 `TIMEX3` tags. It is built on top of the Polish KPWr corpus, by taking the named entities annotations and porting those labeled as DATE into `TIMEX3` in the TimeML schema.

2.3 Existing Tools

The existing tools for temporal processing are rule based, data driven or hybrid. They also vary in the extent of temporal processing they cover – from focus on TER only to the whole event processing pipeline.

From the rule based category, **HeidelTime** [11] has a growing diversity of supported languages and domains and also ongoing support by creators.

[6] `http://www.timeml.org/timeMLdocs/TimeML.xsd`

Somewhat similar **SUTime** [3], which is now inferior to HeidelTime, is still used due to its integration in the StanfordCoreNLP package.

Another rule based tool is **SynTime** [17] that uses a set of heuristics over token types to reach state-of-the-art performance on English TimeBank dataset. Moving from rule engineered to data driven tools, **TOMN** [16] is using a conditional random fields that operate on well designed features extracted by predefined pattern matching. A different learning strategy – learning of patterns – have been created by authors of **PTime** [6]. Such a pattern learning was shown to be especially useful in the colloquial domain by setting current state-of-the-art results on English Tweets dataset.

Recent research work used BERT models [5] for TER. On English only, **Chen et al.** [4] have reached comparable results to the existing tools. In the multilingual setting, the fine-tuned BERT model with adversarial alignment by **Lange et al.** [7] has surpassed the previous state-of-the-art multilingual tool HeidelTime with automatic rules by a large margin. Unfortunately, neither of these models has been made publicly available.

3 Multilingual Temporal Expression Recognition

Recent research suggests that monolingual (English) BERT based models perform on English TER comparably to existing tools [4]. Moreover, it has been shown that fine-tuning multilingual BERT model on TER can outperform current state-of-the-art multilingual tools – HeidelTime with automatic rules [12] by a margin. On the languages supported by HeidelTime, the multilingual BERT model was not able to catch up [7]. Also, one has to keep in mind that this multilingual model has been fine-tuned and tested on the same domain (news) and on the same families of languages (Germanic, Romance), therefore the good performance on these languages and domains do not imply good cross-domain and real cross-lingual performance.

We believe that both approaches – rule based and based on BERT have their own advantages in a multilingual setting. The rule based methods work well with well-structured (e.g., *November, 12*) or numeric (e.g., *10/12/2005*) temporal expressions that are easily expressible in terms of rules, either handcrafted or translated. On the other hand, more compound expressions (e.g., *the 2012 through 2016 tax years*, *the end of each year*), expressions with less common modifiers (e.g., *prior year*) or domain specific expressions (e.g., *3rd Quarter*) are handled much better by the multilingual BERT based model.

To exploit them both, we have combined outputs from HeidelTime with outputs from the multilingual BERT based model.

3.1 Rule based HeidelTime [11]

When working in a multilingual setting, the best rule based tool by far is HeidelTime. Not only that it can be seen as a gold standard in the area of TER and normalization, it also has a native support for 13 languages. Event though there are systems which achieve better scores on English [17,6,16,4], the diversity of supported languages and availability makes HeidelTime still the practical winner.

Moreover, it has been automatically ported by translating the rules to more than 200 languages, and therefore become multilingual in a sense [12]. Nevertheless, the performance of translated rules ranges from bad to extremely bad and is deemed to be useful mostly as a baseline.

Further on, we work with both of these types of rules – the handcrafted as well as the automatically translated.

3.2 Multilingual BERT NER by DeepPavlov

As a representative of BERT based models, we have used publicly available multilingual BERT [5] based NER model by DeepPavlov[7] that was fine-tuned on the OntoNotes dataset [15]. OntoNotes annotations specify 19 tags, from which one is a *DATE* tag with almost 11,000 occurrences in texts from six different domains. Even though the fine-tuning was done solely on English data, the multilinguality stamped into the BERT model during the pre-training [9] allows the model to recognize named entities in 104 languages.

The vast majority of former research was relying on specialized temporal corpora (e.g., the TBAQ corpus). TBAQ contains less then 1,500 annotated temporal expressions from just one domain. Even though it has been shown by Chen et al. that BERT based recognizers already perform well in a low-resource conditions [4], the cross-domain differences are still hard to cover. Higher number of temporal expressions from multiple domains seen during the training shall lead to higher recall, compared to methods that used just TBAQ dataset.

In addition, there are practical advantages of using a generic NER model instead of specific temporal expression tool or custom trained model. The out-of-box functionality (no custom training necessary) and one step extraction of both named entities and temporal expressions are valuable qualities in real world use-cases.

Needless to say, these benefits are not for free, the OntoNotes annotations do not follow the TimeML standard exactly, which means that additional post-processing tasks arise to correctly deal with the type of temporal expressions and with composition and decomposition of recognized expressions.

3.3 The combined approach

To exploit the best of both these approaches we recognize the text by both systems and combine the outputs. We have used a simple algorithm, which prefers recall over precision.

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

Definition 1. *Let $pTE = (t, s, e)$ be a* positional temporal expression *where t stands for the text value, s for the position of the first character of t in the text and e for the last position of t in the text.*

Definition 2. *Let H be a set of positional temporal expressions recognized by HeidelTime and B those recognized by BERT NER.*

To get the final C set of combined pTEs we deal distinctly with overlapping and non-overlapping pTEs.

- The combined set C_1 is composed by all pTEs from H that do not overlap with any pTE from B and all pTEs from B that do not overlap with any pTE from H. Simply put, $C_1 = B_{non-overlap} \cup H_{non-overlap}$.
- The combined set C_2 is composed by merging overlapping pTEs together. Merging is done by taking particular overlapping pTEs and creating one longer pTE, that covers all formerly overlapping pTEs.
- Finally, we take the union of both sets to get the final C set, $C = C_1 \cup C_2$

An example result can be seen in Example 1.

Example 1. 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**!

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

---

4 Evaluation

We divided the evaluation into two categories. For languages supported natively by HeidelTime, we have used the TBAQ-gold and TE3-platinium English and KRAUTS German datasets. For languages not supported natively by HeidelTime, we have used PolEval2019 Polish dataset and to compare with previous work, we have also included German KRAUTS dataset again, but with multilingual setting. We have used evaluation script developed for TempEval-3 [14].

We report the precision, recall and F1 score in both strict and relaxed matching. Strict matching means that both start and end of the entity have to match, whereas relaxed matching requires only partial overlap of predicted entity with the reference one.

For our work, the most important part is the relaxed matching F1 score with additional focus on high recall. These requirements are based on the planned usage of a temporal entity classifier, currently under development. Higher recall leads to recognition of uncommon entities and we can subsequently decide how to deal with them with respect to our goal.

4.1 Monolingual setting

We are using English and German temporal datasets both from the news domain. Note that performance may be different on other domains, since HeidelTime is tailored for the news domain.

Table 1 shows that the BERT NER model performs worse than HeidelTime in both strict and relaxed matching. Also, the performance of the combined model is inferior in strict matching. Still the performance of the combined model in the relaxed matching is comparable to HeidelTime. The combined model achieves the best recall in the relaxed matching.

Table 1. Evaluation in monolingual setting

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**

4.2 Multilingual setting

We evaluate the approaches in the multilingual setting, since it is an important area for our purposes. We present the result in Table 2.

For the German KRAUTS dataset, we compared the combined method with adversarialy aligned BERT model [7] that was trained on English, Spanish, and Portugese TimeBanks. Even though our combined method outperformed the individual base models, it was inferior to [7].

For the Polish PolEval2019 dataset, we experimented with different HeidelTime settings. Surprisingly, the English HeidelTime outperforms the HeidelTime with automatic German rules by a margin. We believe this is due to a low coverage of the translated rules and a similarity between these languages. Still, the BERT NER outperformed both HeidelTimes. When combined, the HeidelTime with automatic rules turned out to be much better base model. Overall, the combined method reaches the best scores compared to all bases in almost every metric.

We can clearly observe that combining HeidelTime with BERT NER significantly improves the performance compared to the individual models. Still, adversarialy aligned BERT model, fine-tuned on specific temporal datasets from the news domain outperforms our combined method on German news dataset.

Table 2. Evaluation in multilingual setting. N.B. [7] only publishes F1 scores.

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[7]	?	?	**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**

5 Multilingual TER by document type

We experimented with a collection of English, German and French office documents categorized by their type. For each language, we collected about 1000 documents, divided them into 10 categories with about 100 documents each. These categories are the same throughout the languages.

A preliminary observation has shown that temporal entities in office documents may differ significantly from those contained in other genres such as news. Apart from absolute dates such as *2020/10/31*, office documents contain expressions such as *Q3* meaning *3rd quarter*, or *FY2020* meaning *fiscal year 2020*.

The experiment has two goals:

1. determine what document types contain the most temporal entities
2. which of the three methods is most suitable for TER in these types of documents (again recall is preferred over precision)

We define the *density* of temporal expressions as the number of such entities per 10,000 characters of text. This method is more realistic than the absolute number since it normalizes absolute numbers per document length. Figure 1 shows the density per document type among English and German documents respectively. It can be seen that for English documents, the category with most temporal entities is project status report and financial report, while for German documents, the highest number of temporal entities is found within invoices and financial reports. It is not surprising the highest number of temporal entities is in financial documents, however, since we have used the combined methods, we need to check that the entities specific to the financial domain were recognized.

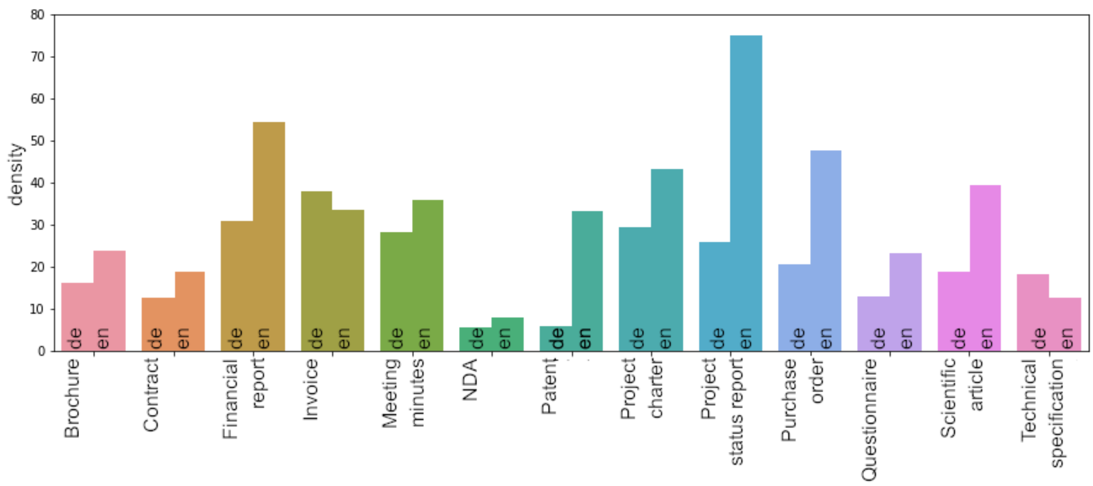

Fig. 1. Number of temporal expressions in English and German documents per document type

We observed that BERT NER is sometimes not able to detect partial dates (e.g., year) when present in an unusual context or, more precisely, in a context typical for office documents. On the other hand, HeidelTime is not able to capture entities from the financial domain since is has not rules for them. The bar charts in Figure 1 shows the number of entities recognized by the combined method.

Examples 2 and 3 show domain-specific temporal entities. The recognized entities are in bold. In English, the combined method is able to recognize such entities. For German, some entities are missed (e.g., *accounting year*) or matched only in some cases (e.g., *financial year*).

Example 2. Temporal entities specific to financial documents

The Company expensed $0.2 million for employee participation in this plan in **fiscal year 2014**.

Lagebericht der AG und Konzernlagebericht für das Geschäftsjahr **2004**.
(Management report of the AG and group management report for the financial year **2004**)

Der strategischen "Wegweiser" wollen wir uns im **Geschäftsjahr 2018** stabil und zukunfts
(We want to be the strategic stable and future "guide" in the **financial year 2018**)

Example 3. Temporal entities specific to legal documents.

For Model numbers 1200, 1300, and 1600 Combo Machines, the warranty is a one **(1) year** parts only, no labor warranty.

Entnehmen Sie dieser Seite alle Informationen zu den **monatlichen** Abschläge im neuen Abrechnungsjahr.
(On this page you will find all information about the **monthly** payments in the new accounting year.)

---

6 Conclusion and future work

This paper presents our research of temporal entities recognition in general and within the area of office documents. Since the follow-up task is to extract events related to temporal entities, the algorithm prefers recall over precision.

The main contribution of this work is the combination of rule based and data driven methods. For non-English texts, the combined approach performs the best.

Although we are interested only in some types of office documents (legal and financial), we could not evaluate the performance of the combined methods in this domain because of lack of annotated data.

In further work, we plan to take the following steps.

6.1 Fine-tune BERT NER on temporal datasets

Fine-tuning of BERT model using adversarial alignment as shown by Lange et al. [7] can further improve the performance of our combined system. Still, this fine-tuning has to be done on specialized temporal datasets such as TBAQ. Even though there are ways not to lose the one-step recognition of Named Entities and Temporal Expressions, it remains an open question how will this fine-tuning affect the cross-domain performance.

6.2 Evaluate on more datasets

We believe the differences between rule based and BERT based approach is much more obvious is some domains. Very formal or structured documents shall be well very suitable for rule based HeidelTime. Whereas in domains such as voice assistant or colloquial speech, the BERT based model is deemed to be much better. We aim to quantify these impressions.

6.3 Semantic classification schema

Moving forward from recognition, we are also trying to better understand the temporal expressions. It was shown the TimeML standard is not capturing all important aspects of temporal information [1], so getting the temporal expressions in the TimeML schema is not sufficient. New tagging schemas has been already proposed [1,2], but due to our focus on multilinguality, the tools we are using are following the TimeML or even general NER tagging specifications.

To overcome this issue we are currently working on semantic classification schema, that further granulates the TimeML output of existing systems and, thereby, allows much finer control over detected entities. We see this control as a critical component for improving the performance of a complete temporal processing pipeline.

Acknowledgments. This work has been partly supported by the Ministry of Education of CR within the LINDAT-CLARIAH-CZ project LM2018101.

References

1. Bethard, S., Parker, J.: A semantically compositional annotation scheme for time normalization. In: Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16). pp. 3779–3786. European Language Resources Association (ELRA), Portorož, Slovenia (May 2016), https://www.aclweb.org/anthology/L16-1599
2. Caselli, T., Bartalesi Lenzi, V., Sprugnoli, R., Pianta, E., Prodanof, I.: Annotating events, temporal expressions and relations in Italian: the it-timeml experience for the ita-TimeBank. In: Proceedings of the 5th Linguistic Annotation Workshop. pp. 143–151. Association for Computational Linguistics, Portland, Oregon, USA (Jun 2011), https://www.aclweb.org/anthology/W11-0418
3. Chang, A.X., Manning, C.: SUTime: A library for recognizing and normalizing time expressions. In: Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC'12). pp. 3735–3740. European Language Resources Association (ELRA), Istanbul, Turkey (May 2012), http://www.lrec-conf.org/proceedings/lrec2012/pdf/284_Paper.pdf
4. Chen, S., Wang, G., Karlsson, B.: Exploring Word Representations on Time Expression Recognition. Tech. rep., Microsoft Research Asia (2019)
5. Devlin, J., Chang, M.W., Lee, K., Toutanova, K.: BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. arXiv:1810.04805 [cs] (May 2019), http://arxiv.org/abs/1810.04805, arXiv: 1810.04805
6. Ding, W., Gao, G., Shi, L., Qu, Y.: A Pattern-Based Approach to Recognizing Time Expressions. Proceedings of the AAAI Conference on Artificial Intelligence **33**, 6335–6342 (Jul 2019). https://doi.org/10.1609/aaai.v33i01.33016335, http://www.aaai.org/ojs/index.php/AAAI/article/view/4595
7. Lange, L., Iurshina, A., Adel, H., Strötgen, J.: Adversarial Alignment of Multilingual Models for Extracting Temporal Expressions from Text. arXiv:2005.09392 [cs] (May 2020), http://arxiv.org/abs/2005.09392, arXiv: 2005.09392

8. Ogrodniczuk, M., Kobyliński, L. (eds.): Proceedings of the PolEval 2019 Workshop. Institute of Computer Science. Polish Academy of Sciences, Warszawa (2019), oCLC: 1107599814
9. Pires, T., Schlinger, E., Garrette, D.: How Multilingual is Multilingual BERT? In: Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. pp. 4996–5001. Association for Computational Linguistics, Florence, Italy (2019). https://doi.org/10.18653/v1/P19-1493, https://www.aclweb.org/anthology/P19-1493
10. Pustejovsky, J., Castaño, J.M., Ingria, R., Saurí, R., Gaizauskas, R.J., Setzer, A., Katz, G., Radev, D.R.: Timeml: Robust specification of event and temporal expressions in text. In: Maybury, M.T. (ed.) New Directions in Question Answering, Papers from 2003 AAAI Spring Symposium, Stanford University, Stanford, CA, USA. pp. 28–34. AAAI Press (2003)
11. Strötgen, J., Gertz, M.: Multilingual and cross-domain temporal tagging. Language Resources and Evaluation **47**(2), 269–298 (2013), https://www.jstor.org/stable/42636370, publisher: Springer
12. Strötgen, J., Gertz, M.: A Baseline Temporal Tagger for all Languages. In: Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing. pp. 541–547. Association for Computational Linguistics, Lisbon, Portugal (Sep 2015). https://doi.org/10.18653/v1/D15-1063, https://www.aclweb.org/anthology/D15-1063
13. Strötgen, J., Minard, A.L., Lange, L., Speranza, M., Magnini, B.: KRAUTS: A German Temporally Annotated News Corpus. In: Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018). European Language Resources Association (ELRA), Miyazaki, Japan (May 2018), https://www.aclweb.org/anthology/L18-1085
14. UzZaman, N., Llorens, H., Derczynski, L., Allen, J., Verhagen, M., Pustejovsky, J.: 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. Association for Computational Linguistics, Atlanta, Georgia, USA (Jun 2013), https://www.aclweb.org/anthology/S13-2001
15. Weischedel, R., Palmer, M., Marcus, M., Hovy, E., Pradhan, S., Ramshaw, L., Xue, N., Taylor, A., Kaufman, J., Franchini, M., El-Bachouti, M., Belvin, R., Houston, A.: OntoNotes Release 5.0 LDC2013T19. Linguistic Data Consortium (October 2013), https://catalog.ldc.upenn.edu/LDC2013T19
16. Zhong, X., Cambria, E., Hussain, A.: Extracting Time Expressions and Named Entities with Constituent-Based Tagging Schemes. Cognitive Computation **12**(4), 844–862 (Jul 2020). https://doi.org/10.1007/s12559-020-09714-8, http://link.springer.com/10.1007/s12559-020-09714-8
17. Zhong, X., Sun, A., Cambria, E.: Time Expression Analysis and Recognition Using Syntactic Token Types and General Heuristic Rules. In: Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). pp. 420–429. Association for Computational Linguistics, Vancouver, Canada (Jul 2017). https://doi.org/10.18653/v1/P17-1039, https://www.aclweb.org/anthology/P17-1039