HeidelToul: A Baseline Approach for Cross-document Event Ordering

Bilel Moulahi, Jannik Strotgen, Michael Gertz, Lynda Tamine

HAL Id: hal-01343033
https://hal.archives-ouvertes.fr/hal-01343033

HeidelToul: A Baseline Approach for Cross-document Event Ordering

Bilel Moulahi[1,3]* and **Jannik Strötgen**[2] and **Michael Gertz**[2] and **Lynda Tamine**[1]

[1] IRIT, University of Toulouse Paul Sabatier, France
[2] Institute of Computer Science, Heidelberg University, Germany
[3] Faculty of Science of Tunis, University of Tunis El-Manar, Tunisia
{bilel.moulahi,lynda.lechani}@irit.fr
{stroetgen,gertz}@informatik.uni-heidelberg.de

Abstract

In this paper, we give an overview of our participation in the timeline generation task of SemEval-2015 (task 4, TimeLine: Cross-Document Event Ordering). The main goals of this new track are, given a collection of news articles and a so-called target entity, to determine events that are relevant for the entity, to resolve event coreferences, and to order the events chronologically. We addressed the sub-tasks, in which event mentions were provided, i.e., no additional event extraction was required. For this, we developed an ad-hoc approach based on a temporal tagger and a coreference resolution tool for entities. After determining relevant sentences, relevant events are extracted and anchored on a timeline. The evaluation conducted on three collections of news articles shows that our approach – despite its simplicity – achieves reasonable results and opens several promising issues for future work.

1 Introduction

Due to the tremendous amount of documents being constantly published on the Internet, there is a need for more enhanced search facilities to retrieve relevant information. Consider, for example, a user looking for information about the "Golden Globe Awards". It might be possible that the user's information need is about the recent "72nd edition". However, it is also reasonable to assume that the user

---

*The work was done during an internship at Heidelberg University.

would appreciate relevant information about previous editions. Thus, presenting search results for time- and event-sensitive information needs in the form of a complete and updatable timeline would be a promising approach. While this issue is tackled by some applications, early techniques required manual effort (Shahar and Musen, 1992) and recent approaches rely on heavily structured information such as Google's entity-related search results, which are based on Google's knowledge graph (Singhal, 2012). However, instead of listing only structured knowledge on a timeline, e.g., winners of the 71st Golden Globes in our example, search results would become much more valuable when adding temporally anchored event information extracted from text documents (e.g., recent updates about the event).

In the SemEval task 4,[1] the goal is to detect all events in a document collection that are relevant for a target entity, and to anchor these events on a timeline. Thus, events are to be sorted chronologically, and, if possible, specific dates are to be assigned to the events. As in previous SemEval tasks addressing temporal relation extraction, namely in the Temp-Eval series (see, e.g., Verhagen et al., 2010), the TimeML event definition is used. However, a special focus is now put on the cross-document aspect, i.e., on cross-document event coreference resolution and cross-document temporal relation extraction. While the document collection contains news articles, target entities can be persons, organizations, products, or financial entities.

The organizers offered the task in two tracks. While the final goals of timeline construction are

---

[1] http://alt.qcri.org/semeval2015/task4/

identical in both tracks, systems addressing track A had to extract event mentions, while event annotations were provided to participants of track B. Furthermore, both tracks were evaluated with and without assigning explicit temporal information to the events. Since we participated in track B, the main challenges for our approach were to

- filter events relevant for the target entities,
- assign date information to relevant events,
- determine cross-document event coreferences,
- and to construct a timeline for each entity.

In the following section, we describe our approach and give an example for cross-document event ordering. In Section 3, we present and analyze the official evaluation results. Finally, we discuss open issues for future research in the context of cross-document timeline construction.

2 Cross-document Event Ordering

Given a set of documents and a set of target entities, the task is to build an event timeline for each entity. Documents are provided with annotated sentences, which may contain several event annotations.

2.1 System Architecture

We implemented an ad-hoc approach for both the retrieval and the anchoring of relevant events. Figure 1 illustrates the general architecture of our approach. Our system is based on five main components:

- **Matching**: In this step, we identify sentences in the document collection, in which parts of the target entity name occur. Furthermore, we use the cosine similarity matching function with a threshold to not select sentences that contain too few parts of the entity name. The result of this step is a list of sentences with event candidates for the timeline of the target entity.

- **Coreference resolution**: To avoid extracting event candidates only from sentences in which parts of the entity name occur explicitly, we apply entity coreference resolution using the Stanford CoreNLP tool (Lee et al., 2013; Manning et al., 2014). Thus, sentences, in which other terms, e.g., pronouns, are used to refer to the target entity, can be added to the list of sentences with event candidates.
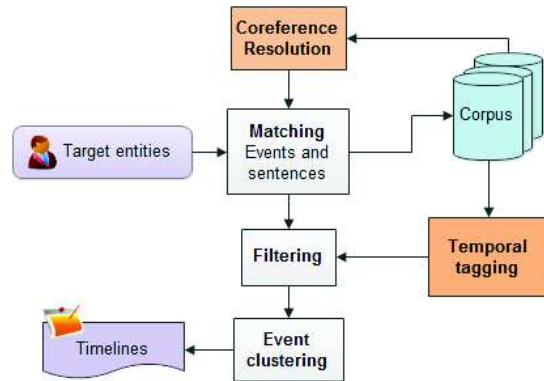

Figure 1: General architecture of our approach.

- **Temporal tagging**: To extract and normalize temporal expressions, HeidelTime (Strötgen and Gertz, 2013) is applied. If a temporal expression cooccurs with an event candidate in a sentence, the event is anchored at the respective point in time. If no expressions are detected in a sentence, we use the document creation time as anchor date for respective events.

- **Filtering**: Since the first steps result in many event candidates, we aim to filter out non relevant events to improve the precision of our approach. Using a threshold for the token distance between the event and the closest term referring to the target entity, we prune events for which it is unlikly that the entity is involved.

- **Cross-document event clustering**: Finally, all events anchored at the same point in time with identical covered text are clustered.

We apply several filtering techniques in order to prune non relevant sentences and events. These thresholds were tuned using the trial data provided by the organizers. Since the performance of our system depends on these parameters, we submitted two runs with different configurations:

- *HeidelToul_NonTolMatchPrune*: The first run uses a non-tolerant pruning setting with low values for the thresholds and distances.

- *HeidelToul_TolMatchPrune*: The second run performs a more tolerant pruning of events and sentences using quite high thresholds.

	anchor date	event	rel
1	1994	8983-13-flight	1
2	2000-02	4764-6-received	0
2	2000-02	4764-6-announced	1
3	2004-11-24	1173-6-negotiations	0
...	...	...	...
9	2008	8983-14-development	0
9	2008	8983-14-scheduled	0

Table 1: Timeline excerpt returned for *Boeing 777*. Events are either relevant (1) or not (0).

Evaluation results are discussed in Section 3.

2.2 Timeline Construction Example

Table 1 shows some events of the timeline constructed by our system for the entity *Boeing 777*. The listed events are extracted from the document parts depicted in Figure 2. Events mentioned in the timeline are surrounded by boxes, and (parts of) the entity mentions are underlined. In the following, we explain the timeline and the reasons for incorrectly returned and anchored events.

The first column of the timeline refers to the rank of the events, the second contains the dates in which events are anchored, and the third corresponds to the events that are detected as relevant for the target entity. Each event of the timeline is represented by the document *id* and sentence *id* from which it was extracted, and the covered text of the event mention. For instance, our system correctly determines the event *flight* as chronologically first relevant event (rank 1) occurring in 1994. It was extracted from sentence *13* of document *8983* (c.f. Figure 2c).

If two events are simultaneous, they can be associated with the same rank, as the second and third event. If a systems fails to extract the anchor dates of relevant events, these should be returned at rank 0 and are ignored in the evaluation.

Using the excerpts in Figure 2, we explain why events in Table 1 have been selected as relevant for *Boeing 777*. All sentences contain only substrings of the target entity name, i.e., the full entity name never occurs. For instance, sentence 13 of document 8983 contains the string *777* while sentence 14 contains the string *Boeing*. As explained above, we used substring matching with a threshold and coreference resolution to increase the number of poten-

```
<s id="6">The original
    <EVENT eid="e134">negotiations</EVENT>
    with Boeing were over a no-bid
    contract .
</s>
```

(a) Doc. #1173: Internal emails expose Boeing ...

```
<s id="6">So far , Boeing has
    <EVENT eid="e131">received</EVENT>
    five orders from two customers for
    the 777-200LR since it was
    <EVENT eid="e88">announced</EVENT> in
    February 2000 .
</s>
```

(b) Doc. #4764: Boeing unveils long-range 777.

```
<s id="13">The first
    <EVENT eid="e126">flight</EVENT> of
    the 777 was in 1994 .
</s>
<s id="14">The Boeing 787 , or Dreamliner
, is a mid-sized passenger airliner
currently under
    <EVENT eid="e85">development</EVENT>
    by Boeing Commercial Airplanes and
    <EVENT eid="e132">scheduled</EVENT>
    to enter service in 2008 .
</s>
```

(c) Doc. #8983: Boeing secures $11bn of aircraft deals.

Figure 2: Three document excerpts with sentences containing events returned for entity *Boeing 777*.

tially relevant events. While for many target entities, it is important to not require a full entity name (e.g., for persons), the term *Boeing* in the three document excerpts never refers to *Boeing 777*, resulting in some non-relevant events in our timeline. Note, however, that relying on strict entity matching, no event could be extracted from the sentences shown in Figure 2, and that some events considered as not relevant in the gold standard are at least debatable, e.g., *received*: Although our anchor date is incorrect (it should be the document creation time of the article due to *so far*), the event is relevant for the target entity since *Boeing 777* is the subject of the orders.

3 Experimental Results and Discussion

The evaluation data consists of 3 sets of 30 documents from Wikinews annotated with event men-

track	run	MICRO-FSCORE				details (overall)	
		corpus 1	corpus 2	corpus 3	overall	precision	recall
TrackB	GPLSIUA_1	22.35	19.28	33.59	25.36	21.73	30.46
TrackB	GPLSIUA_2	20.47	16.17	29.90	22.66	20.08	26.00
TrackB	HeidelToul_NTMP[2]	19.62	7.25	20.37	17.03	20.11	14.76
TrackB	HeidelToul_TMP[3]	16.50	10.94	25.89	18.34	13.58	28.22
SubTrackB	GPLSIUA_1	18.35	20.48	32.08	23.15	18.90	29.85
SubTrackB	GPLSIUA_2	15.93	14.44	27.48	19.18	16.19	23.52
SubTrackB	HeidelToul_NTMP	12.23	14.78	16.11	14.42	19.58	11.42
SubTrackB	HeidelToul_TMP	13.24	15.88	21.99	16.67	12.18	26.41

Table 2: Official results of participating groups in SemEval 2015 task 4. [2]NonTolMatchPrune: non tolerant matching and pruning setting; [3]TolMatchPrune: tolerant matching and pruning setting (cf. Section 2.1).

tions and a total of 38 target entities. Our system ranked second among only two participating groups.

While there have been a total of four teams participating in the task, only two participated in (sub)track B. Participants of (sub)track A additionally performed event extraction so that a comparison between results of all four participants is not possible. Thus, in Table 2, we only present the results of the two teams that addressed (sub)track B.

Table 2 (left) reports the results by means of Micro-FSCORE obtained by our runs and that of the other participating group. As shown, our system is outperformed by the system "GPLSIUA" for both settings. The performance difference is most significant for corpus 2, especially within TrackB. However, we notice that our tolerant setting gives better overall results than the non tolerant one. These improvements are less significant for corpora 1 and 2 than for corpus 3.

To get a deep understanding of the results, we report in Table 2 (right) the overall precision and recall values for our system configurations and that of the other participating group. Our non tolerant setting is slightly outperformed by the run "GPLSIUA_1" in terms of precision for trackB. However, it relatively enhances the other runs within the SubTrackB. This can be explained by the important number of relevant retrieved events due to the high values of distances and thresholds used to prune the events. In contrast, in terms of recall, our tolerant setting performs better than the non tolerant one in both subtracks. Actually, this is not surprising given that the filtering techniques are not strict.

Interestingly, an in-depth analysis of the nature of

the target entities and the types of temporal expressions in the documents for which our system fails to provide good timeline, may help to improve the overall performance of our system in the future. For instance, for the target entities "Boeing 777" and "Airbus A380" in corpus 1, we obtained the lowest values in terms of MicroFSCORE among all target entities. Clearly, this is due to the partial matching technique we used, which results in the extraction of many events related to other entities (e.g., "Boeing 787" instead of "Boeing 777"; cf. Table 1 and Figure 2). Moreover, all events that do not cooccur with a temporal expression in the same sentence are anchored at the document creation time by our system. This hurts the performance of our system in particular for TrackB, because many of those events are placed at rank 0 in the gold standard.

4 Conclusions

In this paper, we presented an overview of our participation in the timeline generation task of SemEval-2015. We proposed a baseline approach for the extraction and anchoring of events. Our system is evaluated using three corpora of news articles and shows reasonable results.

Interesting future work to improve our approach could include a fine tuning of the matching function as well as the filtering parameters used to prune non relevant events. In addition, more sophisticated entity disambiguation could further improve the performance of our system.

Acknowledgments

References

Heeyoung Lee, Angel Chang, Yves Peirsman, Nathanael Chambers, Mihai Surdeanu, and Dan Jurafsky. 2013. Deterministic Coreference Resolution Based on Entity-centric, Precision-ranked Rules. *Computational Linguistics*, 39(4):885–916, December.

Christopher D. Manning, Mihai Surdeanu, John Bauer, Jenny Finkel, Steven J. Bethard, and David McClosky. 2014. The Stanford CoreNLP Natural Language Processing Toolkit. In *Proceedings of 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, pages 55–60.

Yuval Shahar and Mark A. Musen. 1992. A Temporal-Abstraction System for Patient Monitoring. In *Proceedings of the Annual Symposium on Computer Application in Medical Care*, pages 121–127.

Amit Singhal. 2012. Introducing the Knowledge Graph: Things, not Strings". *Official Google Blog*, May.

Jannik Strötgen and Michael Gertz. 2013. Multilingual and Cross-domain Temporal Tagging. *Language Resources and Evaluation*, 47(2):269–298.

Marc Verhagen, Roser Saurí, Tommaso Caselli, and James Pustejovsky. 2010. SemEval-2010 Task 13: TempEval-2. In *Proceedings of the 5th International Workshop on Semantic Evaluation (SemEval '10)*, pages 57–62.