SPINOZA_VU: An NLP Pipeline for Cross Document TimeLines

Tommaso Caselli   **Antske Fokkens**   **Roser Morante**   **Piek Vossen**
Computational Lexicology & Terminology Lab (CLTL)
VU Amsterdam, De Boelelaan 1105
1081 HV Amsterdam Nederland
{t.caselli}{antske.fokkens}{r.morantevallejo}{p.t.j.m.vossen}@vu.nl

Abstract

This paper describes the system SPINOZA_VU developed for the SemEval 2015 Task 4: Cross Document TimeLines. The system integrates output from the News-Reader Natural Language Processing pipeline and is designed following an entity based model. The poor performance of the submitted runs are mainly a consequence of error propagation. Nevertheless, the error analysis has shown that the interpretation module behind the system performs correctly. An out of competition version of the system has fixed some errors and obtained competitive results. Therefore, we consider the system an important step towards a more complex task such as storyline extraction.

1 Introduction

This paper reports on a system (SPINOZA_VU) for timeline extraction developed at the CLTL Lab of the VU Amsterdam in the context of the SemEval 2015 Task 4: Cross Document TimeLines. In this task, a timeline is defined as a set of chronologically anchored and ordered events extracted from a corpus spanning over a (large) period of time with respect to a target entity.

Cross-document timeline extraction benefits from previous works and evaluation campaigns in Temporal Processing, such as the TempEval evaluation campaigns (Verhagen et al., 2007; Verhagen et al., 2010; UzZaman et al., 2013) and aims at promoting research in temporal processing by tackling the following issues: cross-document and cross-temporal

event detection and ordering; event coreference (in-document and cross-document); and entity-based temporal processing.

The SPINOZA_VU system is based on the News-Reader (NWR) NLP pipeline (Agerri et al., 2013; Beloki et al., 2014), which has been developed within the context of the NWR project[1] and provides multi-layer annotations over raw texts from tokenization up to temporal relations. The goal of the NWR project is to build structured event indexes from large volumes of news data addressing the same research issues as the task. Within this framework, we are developing a storyline module which aims at providing more structured representation of events and their relations. Timeline extraction from raw text qualifies as the first component of this new module. This is why we participated in Track A and Subtrack A of the task, timeline extraction from raw text. Participating in Track B would require a full re-engineering of the NWR pipeline and of our system.

The remainder of the paper is structured as follows: Section 2 provides an overview of the model implemented in the two versions of our system. Section 3 presents the results and error analysis, and Section 4 puts forward some conclusions.

2 From Model to System

Timeline extraction involves a number of independent though highly connected subtasks, the most relevant ones being: entity resolution, event detection, event-participant linking, coreference resolu-

---

[1] http://www.newsreader-project.eu

tion, factuality profiling, and temporal relation processing (ordering and anchoring).

We designed a system that addresses these subtasks, first at document level, and then, at cross-document level. We diverted from the general NWR approach and adopted an entity based model and representation rather than an event based one in order to fit the task. This means that we used entities as hub of information for timelines. Using an entity driven representation allows us to better model the following aspects:

- **Event co-participation**: the data collected with this method facilitates the analysis of the interactions between the participants involved in an event individually;

- **Event relations**: in an entity based representation, event mentions with more than one entity as their participants will be repeated in the final representation (both at in-document at cross-document levels); such a representation can be further used to explore and discover additional event relations[2];

- **Event coreference**: we assume that two event mentions (either in the same document or in different documents) are coreferential if they share the same participant set (i.e., entities) and occur at the same time and place (Chen et al., 2011; Cybulska and Vossen, 2013);

- **Temporal relations**: temporal relation processing can benefit from an entity driven approach as sequences of events sharing the same entities (i.e., co-participant events) can be assumed to stand in precedence relation (Chambers and Jurafsky, 2009; Chambers and Jurafsky, 2010).

2.1 The SPINOZA_VU System

The NWR pipeline which forms the basis of the SPINOZA_VU system consists of 15 modules carrying out various NLP tasks and outputs the results in NLP Annotation Format (Fokkens et al., 2014), a layered standoff representation format. Two versions of the system have been developed, namely:

---

[2]We are referring to a broader set of relations that we labeled as "bridging relations" among events which involve co-participation, primary and secondary causal relations, temporal relations, and entailment relations.

- **SPINOZA_VU_1** uses the output of a state of the art system, TIPSem (Llorens et al., 2010), for event detection and temporal relations;

- **SPINOZA_VU_2** is entirely based on data from the NWR pipeline including the temporal (TLINKs) and causal relation (CLINKs) layers.

The final output is based on a dedicated rule-based module, the TimeLine (TML) module. We will describe in the following paragraphs how each subtask has been tackled with respect to each version of the system.

Entity identification Entity identification relies on the entity detection and disambiguation layer (NERD) of the NWR pipeline. Each detected entity is associated with a URI (a unique identifier), either from DBpedia or a specifically created one based on the strings describing the entity. We extracted the entities by merging information from the NERD layer with that from the semantic role labelling (SRL) layer. We retained only those entity mentions which fulfil the argument positions of proto-agent (Arg0) or proto-patient (Arg1) in the SRL layer.

Event detection and classification The SPINOZA_VU_1 event module is based on TIPSem, which provides TimeML compliant data. We developed post processing rules to convert the TimeML event classes (OCCURRENCE, STATE, I_ACTION, I_STATE, ASPECTUAL, REPORTING and PERCEPTION) to specific FrameNet frames (e.g., *Communication*, *Being_in_operation*, *Body_movement*) and/or Event Situation Ontology (ESO) types (Segers et al., 2015) (e.g., *contextual*), which correspond to the event types specified in the task guidelines. For instance, not all mentions of TimeML I_STATE, I_ACTION, OCCURRENCE and STATE events can enter a timeline. The alignment with FrameNet and ESO is made by combining the data from the Word Sense Disambiguation (WSD) layer of the pipeline with Predicate Matrix (version 1.1) (Lacalle et al., 2014).

As for the SPINOZA_VU_2, we have used the NWR SRL layer to identify and retain the eligible events. In this case the access to the Predicate Matrix is not necessary as each predicate in the SRL layer is also associated with corresponding FrameNet frames and ESO types. Only the pred-

icates matching specific FrameNet frames and/or ESO types were retained as candidate events.

Factuality The factuality filter consists of a collection of rules in order to determine whether an event is within the scope of a factuality marker negating an event or indicating that it is uncertain, in which case the event is excluded from the set of eligible events. Factuality markers are different types of modality and negation cues (adverbs, adjectives, prepositions, modal auxiliaries, pronouns and determiners). For instance, if a verb has a dependency relation of type `AM-MOD` with a modal auxiliary is excluded from the candidate event in the timeline.

Coreference relations Two levels of coreference need to be addressed: in-document and cross-document. As for the former, both versions of the system rely on the coreference layer (COREF layer) of the pipeline. Concerning the cross-document level, two strategies have been implemented:

- Cross-document entity mentions are identified using the URI links associated with entity mentions; all entity mentions from different documents sharing the same URIs are associated with the same entity instance;
- Cross-document event coreference is obtained during a post-processing step of the timeline creation following the assumption that two event mentions denote the same event instance (i.e., they co-refer) if they share the same participants, time of occurrence and (possibly) location. Entity-based timelines are used as a basis to identify instances of cross-document event coreferential expressions.

Temporal Relations For the SPINOZA_VU_1 version, we used the Temporal Relations from TIPSem (TLINKs), including temporal expression detection and normalization. For the SPINOZA_VU_2 version, we used the TLINK and CLINK layers of the NWR pipeline. As for the CLINK layer, we converted all causal relations into temporal ones, with the value `BEFORE`. For both versions of the system we maximized temporal anchoring by recovering the beginning or end point of temporal expressions of type `DURATION` and resolving all TLINKs between a temporal expression and a target event except "`IS_INCLUDED`" relations into an anchoring relation.

TimeLine Extraction The TimeLine Extraction (TML) module[3] harmonizes and orders cross-document temporal relations (anchoring and ordering). It provides a method for selecting the initial (relevant) temporal relations (namely, all anchoring relations) and enhance an updating mechanism of information so that additional temporal relations (both anchoring and ordering relations) can be inferred. Timelines are first created at a document level and subsequently merged. The cross-document timeline model is event-based and aims at building a global timeline between all events and temporal expressions regardless of the target entities. This approach allows us to also make use of temporal information provided by events that are not part of the final timelines. Finally, the target entities for the timelines are extracted using two strategies: i) a perfect match between the target entities and the DBpedia URIs, and ii) the Levenshtein distance (Levenshtein, 1966) between the target entities and the URIs. For this latter strategy, an empirical threshold was set to maximize precision on the basis of the trial data.

3 Results and Error Analysis

In Table 1 we report the results of both versions of the system for Track A - Main. We also include the results of the best performing system and out of competition results of a new version of the system (OC_SPINOZA_VU), which obtained competitive results with respect the best system, WHUNLP_1.

System Version	Corpus 1	Corpus 2	Corpus 3	Overall
SPINOZA_VU_1	4.07	5.31	0.42	3.15
SPINOZA_VU_2	2.67	0.62	0.00	1.05
OC_SPINOZA_VU	7.50	6.64	6.59	7.12
Best system WHUNLP_1	8.31	6.01	6.86	7.28

Table 1: System Results (micro F1 score) for the SemEval 2015 Task 4 Task A - Main Track.

The OC_SPINOZA_VU system is based on SPINOZA_VU_2, and the main differences concern temporal relations identification at in-document and cross-document level, and entity extraction. In particular, we assume that: if a temporal expression

---

[3]https://github.com/antske/ BiographyNet/tree/master/ TimeLineExtraction

occurs in the same sentence of an event, the temporal expression is the event's temporal anchor; if no temporal expression occurs in the same sentence, we check if there are any temporal expressions in the two previous sentences or, if any, in the one following it. The event is then anchor to the closest temporal expression identified. Finally, if no temporal expression can be found in this sentence window, no temporal anchor is assigned to the event. As for event ordering, we have used the order of appearance of the event in the document to establish precedence relations. The final timeline is obtained by ordering cross-document event with a modified version of the TML module based on time anchors only. Entity extraction is extended by adding pure substring match.

Table 2 reports the results of the submitted systems and of the out of competition one. No other results are reported for Track A - Subtask A because only our system participated. The null results of the out of competition system are due to the modified version of the TML module.

System Version	Corpus 1	Corpus 2	Corpus 3	Overall
SPINOZA_VU_1	1.20	1.70	2.08	1.69
SPINOZA_VU_2	0.00	0.92	0.00	0.27
OC_SPINOZA_VU	0.00	0.00	0.00	0.00

Table 2: System Results (micro F1 score) for the SemEval 2015 Task 4 Task A - Subtrack.

Overall, the results of the submitted system are not satisfying. Out of 37 entity based timelines, the system produced results only for 31 of them. Three sources of errors occur in both versions of our system. Error analysis yields the following explanations:

Event detection We analyzed both entity-based event detection (all events associated with each target entity) and global event detection (all events regardless of the target entities). On entity-based event detection, SPINOZA_VU_1 scores an average F1 score on the 31 detected entities of 23.58 (38.7 precision and 17.35 for recall), whereas SPINOZA_VU_2 scores an average F1 of 20.46 (47.83 precision and 13.32 recall). As for global event detection, both versions of the system present a high recall and low precision pattern, although with substantial differences in terms of results. In particular, SPINOZA_VU_1 has an average recall of 44.96 and

an average precision of 25.5, while SPINOZA_VU_2 has an average recall of 77.03 and an average precision of 14.86;

Entity detection This layer is strictly connected to the event detection layer. The lower results are mainly due to the output of the COREF and SRL layers. Missing coreference chains (e.g. "the aircraft" not connected to a target entity like "Airbus A380")) and wrong spans of event arguments negatively impacts on the extraction of candidate events for the timeline;

Event ordering and anchoring The difference in performance between the submitted system and OC_SPINOZA_VU clearly indicates that there is room for improvement concerning the amount of temporal relations (anchoring and ordering ones) which are extracted. Furthermore, the difference in performance between the Main track and the Subtrack suggests that the main issues concern event ordering rather than their detection or anchoring.

4 Conclusions and Future Work

In this paper we presented the SPINOZA_VU system for timeline extraction system in the context of the SemEval 2015 Task 4: Cross Document TimeLines. The low ranking show not only that the task is very complex, but also that there is room for improving the system, as the results of the OC_SPINOZA_VU system show. The low performance is mainly a consequence of a combination of cascading errors and missing data from the different modules of the system, namely event detection, temporal relation extraction and entity detection. However, on the positive side, the theoretical model that has guided the development of the system can be further extended to address more complex tasks on top of the timeline extraction, such as storyline extraction.

References

Rodrigo Agerri, Itziar Aldabe, Zuhaitz Beloki, Egoitz Laparra, Maddalen Lopez de Lacalle, German Rigau, Aitor Soroa, Antske Fokkens, Ruben Izquierdo, Marieke van Erp, Piek Vossen, Christian Girardi, and Anne-Lyse Minard. 2013. Event detection, version 2. NewsReader Deliverable 4.2.2.

Zuhaitz Beloki, German Rigau, Aitor Soroa, Antske Fokkens, Piek Vossen, Marco Rospocher, Francesco Corcoglioniti, Roldano Cattoni, Thomas Ploeger, and Willem Robert van Hage. 2014. System design. NewsReader Deliverable 2.1.

Nathanael Chambers and Dan Jurafsky. 2009. Unsupervised Learning of Narrative Schemas and their Participants. In *Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP*, pages 602–610, Suntec, Singapore, August.

Nathanael Chambers and Dan Jurafsky. 2010. A Database of Narrative Schemas. In *Proceedings of the Seventh Conference on International Language Resources and Evaluation (LREC'10)*, Valletta, Malta, May.

Bin Chen, Jian Su, Sinno Jialin Pan, and Chew Lim Tan. 2011. A Unified Event Coreference Resolution by Integrating Multiple Resolvers. In *Proceedings of the 5th International Joint Conference on Natural Language Processing*, pages 102–110, Chiang Mai, Thailand.

Agata Cybulska and Piek Vossen. 2013. Semantic relations between events and their time, locations and participants for event coreference resolution. In *Proceedings of Recent Advances in Natural Language Processing (RANLP-2013)*, pages 156–163.

Antske Fokkens, Aitor Soroa, Zuhaitz Beloki, Niels Ockeloen, German Rigau, Willem Robert van Hage, and Piek Vossen. 2014. NAF and GAF: Linking linguistic annotations. In *Proceedings 10th Joint ISO-ACL SIGSEM Workshop on Interoperable Semantic Annotation*, pages 9–16.

Maddalen Lopez De Lacalle, Egoitz Laparra, and German Rigau. 2014. Predicate Matrix: extending SemLink through WordNet mappings. In *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*, Reykjavik, Iceland, May.

Vladimir I. Levenshtein. 1966. Binary codes capable of correcting deletions, insertions and reversals. In *Soviet physics doklady*, volume 10, pages 707–710.

Hector Llorens, Estela Saquete, and Borja Navarro. 2010. TIPSem (English and Spanish): Evaluating CRFs and Semantic Roles in TempEval-2. In *Proceedings of the 5th International Workshop on Semantic Evaluation*, pages 284–291.

Roxane Segers, Piek Vossen, Marco Rospocher, Luciano Serafini, Egoitz Laparra, and German Rigau. 2015. ESO: A Frame Based Ontology for Events and Implied Situations. In *Proceedings of Maplex2015*.

Naushad UzZaman, Hector Llorens, James Allen, Leon Derczynski, Marc Verhagen, and James Pustejovsky. 2013. SemEval-2013 Task 1: TempEval-3: Evaluating time expressions, events, and temporal relations. In *Proceedings of the 7th International Workshop on Seman- tic Evaluation*, pages 1–9.

Marc Verhagen, Robert Gaizauskas, Frank Schilder, Mark Hepple, Graham Katz, and James Pustejovsky. 2007. SemEval-2007 Task 15: TempEval Temporal Relation Identification. In *Proceedings of the 4th International Workshop on Semantic Evaluation*, pages 75–80, June.

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*, pages 57–62, Uppsala, Sweden.