TimeFrame: Querying and Visualizing Event Semantic Frames in Time

Davide Lamorte[1], **Marco Rovera**[2], **Alfio Ferrara**[1], **Sara Tonelli**[2]
[1] Università Statale di Milano, Italy
[2] Fondazione Bruno Kessler, Trento, Italy
davide.lamorte@unimi.it, m.rovera@fbk.eu, alfio.ferrara@unimi.it, satonelli@fbk.eu

Abstract

In this work we introduce *TimeFrame*, an online platform to easily query and visualize events and participants extracted from document collections in Italian following a frame-based approach. The system allows users to select one or more events (frames) or event categories and to display their occurrences on a timeline. Different query types, from coarse to fine-grained, are available through the interface, enabling a time-bound analysis of large historical corpora. We present three use cases based on the full archive of news published in 1948 by the newspaper "Corriere della Sera". We show that different crucial events can be explored, providing interesting insights into the narratives around such events, the main participants and their points of view.

Keywords: Text Visualization, Event Extraction, Frame Parsing

1. Introduction

Event-based analysis of corpora has proven to be an effective way to distill information from large amounts of text, supporting tasks such as storytelling (Liu et al., 2020), text simplification (Barlacchi and Tonelli, 2013) and knowledge modelling (Rospocher et al., 2016; Vossen et al., 2016). However, making event information available to users, especially non-expert ones, represents a challenge due to the complex structure of event annotation paradigms, which typically include both event mentions and participants, often structured in some sort of taxonomy. Nevertheless, visualization is crucial to enable users to perform rapid, targeted and customized analysis of large event collections.

We address this limitation by presenting *TimeFrame*, an online tool that allows event-based browsing of large corpora in Italian based on the output of the EventNet-ITA's frame parser (Rovera, 2024). While existing text visualization tools often focus on term-based (Handler et al., 2022) or entity-based (Düring et al., 2021) queries, *TimeFrame* explores the opportunities events and their argument structure offer for event-based text mining and visualization, with particular reference to newspaper textual data. The tool takes in input documents analysed according to the FrameNet paradigm (Fillmore and Baker, 2001), i.e. annotated with semantic frames, each consisting of a trigger (so-called *lexical units – LU*) and a set of semantic roles (*frame elements – FE*). This annotation is further enriched with the temporal information related to the publication of each document, a crucial dimension to navigate historical archives and reconstruct stories.

Through *TimeFrame* it is therefore possible to perform event-based search of a document collection over time. Users can select the granularity of events they are looking for by choosing between the original (i.e. fine-grained) frames or opt for event categories, which have been manually created to group semantically similar events and provide different perspectives on the same event type.

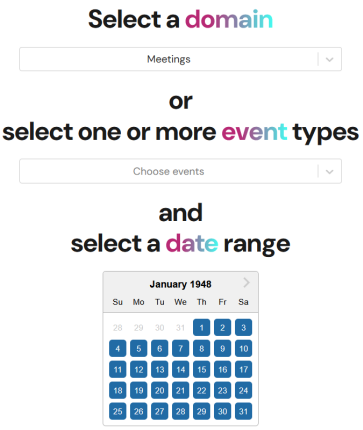
TimeFrame is structured on a back-end component, the frame parser, and on a front-end web application, providing the user interface and interaction. While the frame parser can be applied to any Italian text and the front-end can be used virtually for any frame-based dataset, in this paper we adopt the 1948 edition of the Italian newspaper *Corriere della Sera* as a case study.

2. Related Work

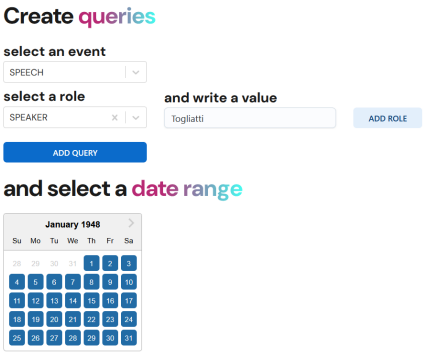
Searching and visualising textual corpora enriched with frame information is a challenging task, since this annotation framework foresees different information layers and a high number of frames and FEs. A notable example is the *Sociofillmore* tool (Minnema et al., 2022), which has been designed to analyse texts in different languages after performing frame-semantic annotation. Its main goal, however, is to study perspectives in written texts and give users the possibility to perform different linguistic analyses. *TimeFrame*, instead, is designed for a more serendipitous exploration of corpora and targets users interested also in the temporal dimension of events such as archivists and historians. Another available tool is *Smell Explorer* (Menini, 2024),[1] which annotates olfactory events inspired by frame semantics, but which is however only limited to one event type. Other search and

[1] https://smell-extractor.tools.eurecom.fr/

First Workshop on Reference, Framing, and Perspective, pages 13–17



(a) Interface for 'easy' search



(b) Interface for 'advanced' search

Figure 1: Different query options offered by *TimeFrame*.

visualization tools, albeit powerful, focus on specific aspects of investigation. *Clioquery* (Handler et al., 2022), for instance, is focused on term-based queries for historical investigation, while *Impresso* (Düring et al., 2021) is tailored to an entity-based search of archives from the past.

3. TimeFrame Platform

The *TimeFrame* platform is available at this link: `https://eventnetdemo.islab.di.unimi.it/`. Users can access the underlying database, where the frame annotated documents are saved, through different query types. We detail below the pre-processing step to create the corpus database, the query types to access it and the technical details of the platform implementation.

3.1. Data Pre-Processing

In order to use *TimeFrame*, it is necessary to pre-process the document collection of interest using a frame parsing tool. To analyse Italian data, we rely on EventNet-ITA (Rovera, 2024), a frame parser for Italian able to recognize and classify over 200 different types of event-denoting frames, along with their specific FEs. The model has been trained on a large, manually annotated corpus, sampled from the Italian Wikipedia edition. This annotated corpus is designed to cover fine-grained event frames over a bunch of different domains: movements, communications, war and conflicts, economics, geopolitics, arts, biographies, among others. The model takes in input one sentence at a time and parses it in a full-text fashion, extracting as many frames and FEs as there are in the sentence. On the Wikipedia corpus, the model achieves macro F1 = 0.90 for event frame classification and macro F1 = 0.73 on FE classification (at span level). After this pre-processing step, the analysed corpus is stored in a MongoDB database (see Section 3.3) and can be queried through the *TimeFrame* interface.

3.2. Query Types

After accessing the system, a user is displayed the possibility to navigate the processed corpus starting from event categories and frame(s). The 'easy' interface to perform different query types is displayed in Fig. 1 (a). An 'advanced' search is also available, which can be activated through the interface, that allows users to select also specific FEs and the corresponding fillers (i.e. strings) to be retrieved and displayed (see Fig. 1, b). Each search can be bound to a time period specified by the user. It is also possible to select whether a search should be performed only on document titles, to capture only the major events, or also on the text body.

Four types of queries with different granularities are made available in *TimeFrame*:

1) Query *by domain*: the user can select a domain, corresponding to a set of pre-defined event types in a thematic area, such as Economics, Geopolitics, Crime, Movements, Arts, Communication, Conflicts, among others. This query level provides an entry point for the user approaching the dataset, returning a general, domain-driven view on the data; these 13 categories are currently hardcoded and have been created by adapting the macro-categories that characterise the EventNet-ITA corpus. This query option is the first one displayed in Fig. 1 (a).

2) Query *by single or multiple event types* (or frames): in this case the user can pick one or more event types, in order to explore a more specific phenomenon; the set of event types chosen will define the particular view the user wants to elicit from the dataset. For example, a user could elicit information about declarations related to an election

(or an electoral campaign) by using a combination like ELECTION, ELECTORAL_CAMPAIGN, ANNOUNCEMENT, STATEMENT, COMMITMENT, SPEECH. This query option is the second one displayed in Fig. 1 (a), available in the 'easy' search.

3) Query *by single event, constrained by FE*: support for query composition is provided at this stage, where the user can constrain the query, based on a specific event, with one or more frame elements. The system will output all documents containing the target event where the selected frame element is realized. For example, we could ask the system to provide all documents where a SPEAKER is reported holding a SPEECH, or where a DESIGNER is reported having designed some OBJECT. This query option is the first one displayed in Fig. 1 (b) and is available in the 'advanced' search, like the following one.

4) Query *by single event, constrained by FE and term search*: this query model works as a further refinement of the previous one. In addition to role-based constraints, the user can provide a specific term that has to appear as filler of the given role. Continuing with the previous example, we could ask the system to return all OBJECTS designed by 'Pininfarina' (a well-known Italian designer) or all speeches held by 'Togliatti' (leader of the Communist Party in Italy in 1948).

After a query is performed, results are displayed on a timeline, where events are chronologically ordered. Each retrieved document which contains an event matching the query is displayed via its title, and a button allows to access the full text of the article. This visualisation is introduced by a temporal heatmap, providing a synthetic overview of the distribution of the target phenomena over the chosen timespan (see the example in Fig. 2).

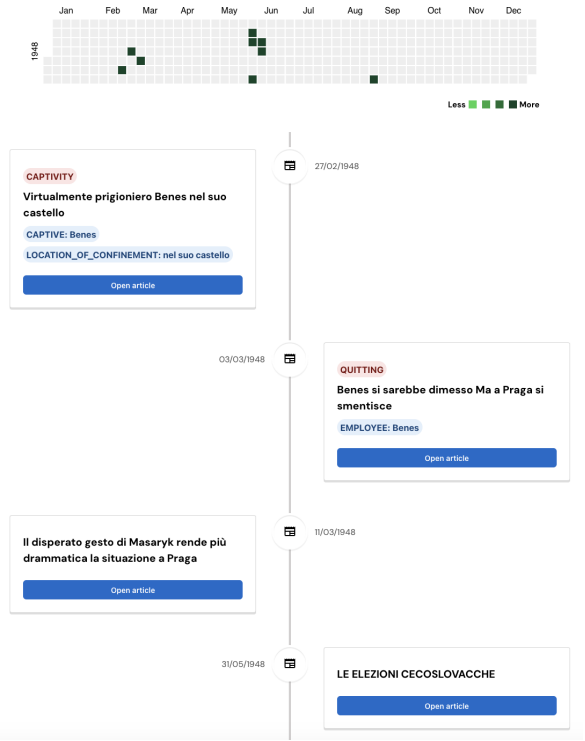


Figure 2: Example of query output after searching for the QUITTING of Edvard Beneš, the former President of Czechoslovakia.

3.3. Implementation Details

TimeFrame is developed as a web application that exploits MongoDB as a DBMS, Node JS and Express as the server back-end and React for the front-end. In order to efficiently execute the types of queries illustrated in Section 3.2, the events extracted from EventNet-ITA are stored in the MongoDB database by associating them with the metadata and text available for each analyzed document.

Document metadata and text are stored in the `date`, `id`, `title` and `body` fields, respectively, while the list of events extracted from the document is stored in the `events` array. For each event in particular, in addition to information on its location in the document (`location` field), the record provides the `label` associated with the event and the list of `roles` associated with it, each characterized by a corresponding `label` and `text`.

This data structure makes it possible to formulate the queries supported by *TimeFrame* as simple searches in MongoDB. In particular, the *TimeFrame* interface supports finding documents whose `events` array contains at least one event with a label corresponding to the type of event specified by the user (e.g., `'events': {'$elemMatch': {'label ': 'DEATH'}}`). Starting from this query it is then possible to dynamically add one or more clauses concerning the semantic roles associated with the event, by selecting the documents whose events contain `roles` with labels and text corresponding to the user input.

4. 1948 Archive of Corriere della Sera

1948 was a crucial year in world history and dense with events that would have far-reaching implications for future history. In Italy, the first republican constitution took effect in January, while in April the first democratic elections, after a fierce electoral campaign, resulted in the victory of the Christian Democratic party. In Europe, on the other hand, postwar political readjustments were in full swing. In March, the Brussels Treaty was signed, setting the stage for the future creation of international organizations like the European Union and NATO. In Eastern Europe, meanwhile, Soviet influence was growing and the region was shaken by coups and regime changes. In the United States, Harry Truman became president after defeating the Republi-

can candidate Thomas Dewey. In India nonviolent leader Gandhi was assassinated in January, while early signs of decolonization were showing up in Burma and Indonesia.

With *TimeFrame* it is possible to perform a time-based exploration of different ways in which events are represented in public discourse. We do so by taking into exam the collection of all news articles published in 1948 by the Italian newspaper *Corriere della Sera*. This collection is currently not publicly available and has been granted for demonstration purposes by the newspaper's archive. The corpus contains 10,418 documents (5,111,000 tokens) which have been processed with EventNet-ITA, producing a database of 146,787 event occurrences (i.e. frame mentions) and 198,370 related FEs.

In public discourse, and particularly in journalistic narrative, different linguistic devices can be used to refer to a particular event, for example by mentioning only a specific aspect, or the outcome of such event. For example, there may be differences between the case where the event is the focus of the discourse (e.g., in the title of an article) or when the event is only mentioned, as part of a larger discourse focusing on a different topic. In some cases this is more or less intentional, while in other cases it is simply the result of the linguistic variety with which language refers to circumstances in the world.

We show below how *TimeFrame* can enable users to explore such referential variety, by retrieving different mentions of the same event. In all three use cases we employ the query type 4 described in Section 3.2. In brackets we provide the number of matches for each query.

4.1. UC 1: Assassination of Gandhi

The first use case concerns the assassination of Mohāndās Karamchand Gāndhī (Gandhi), the 30th January in New Delhi. Events like this (assassinations), can alternatively be presented as agentive (KILLING) or non-agentive (DEATH). We therefore perform two queries:

1. Frame: KILLING | FE: VICTIM, term: *Gandhi* (3 matches)

2. Frame: DEATH | FE: PROTAGONIST, term: *Gandhi* (4 matches)

By comparing the results, we observe that the DEATH frame is preferably used to factually describe the killing and the following sub-events (for instance how the British government reacted), while KILLING is mainly used for comments and more emotive reactions (for instance, he is described as 'glorious and heroic victim of world peace').

4.2. UC 2: U.S. Presidential Elections

Elections are often referred to from multiple perspectives, notably from the point of view of the winner, or of the loser, or, still, of the candidate appointed to the position. US presidential elections in 1948, covered by our newspaper corpus, provide a suitable example of how *TimeFrame* is able to capture such subtle variations in perspective:

1. Frame: ELECTION | [FE: ROLE, term: *presidenziali*] [FE: PLACE, term: *americane*] (4 matches)

2. Frame: APPOINTING_ELECTION | FE: APPOINTEE, term: *Truman* (6 matches)

3. Frame: WIN_ELECTION | FE: WINNER, term: *Truman* (5 matches)

4. Frame: LOSE_ELECTION | FE: LOSER, term: *Dewey* (2 matches)

As expected, much more space is given to discussing the elections from the point of view of the winner, while only two articles mention Dewey (one of which was written months before the election and presented the event as hypothetical).

4.3. UC 3: Czechoslovak Coup d'état

In many cases, events narrated in the press, or even in historical discourse, do not refer to a single occurrence, but to a chain of facts. In such cases, while events are still unfolding and therefore are not clear, journalistic practice resorts to general utterances, like "the recent events in Paris" or "the current happenings in Asia". This is the case, for example, with the "Czechoslovak coup d'état", a sequence of events that led to a regime change in the country on February 25.

1. Frame: COUP | FE: PLACE, term: *Cecoslovacchia*, *Praga* (5 matches)

2. Frame: CHANGE_OF_LEADERSHIP | FE: PLACE, term: *Cecoslovacchia*, *Praga* (2 matches)

3. Frame: QUITTING | FE: EMPLOYEE, term: *Benes* (9 matches)

4. Frame: EVENT | FE: PLACE, term: *Cecoslovacchia* (9 matches)

The output of the third Query is reported in Figure 2, and visually displays how the situation evolves over time. If we check the timeline, we observe that the four frames appear in sequence in the archive.

5. Conclusions

In this work we introduced the *TimeFrame* platform, which allows users to query corpora that have been analysed with a frame semantic parser. The tool makes it possible to perform both coarse-grained queries, at the level of event categories, and fine-grained ones, up to role fillers. Based on three use cases from the 1948 archive of Corriere della Sera newspaper, we show that the platform can be easily used to track specific events over time and capture different sub-events and points of view. The fact that each query type is bound to the temporal dimension makes this platform particularly valuable to users interested in historical analysis such as archivists and history scholars.

6. Acknowledgements

We thank the Archive of Corriere della Sera for making their documents available for our analysis.

7. Bibliographical References

Gianni Barlacchi and Sara Tonelli. 2013. Ernesta: A sentence simplification tool for children's stories in italian. In *Computational Linguistics and Intelligent Text Processing: 14th International Conference, CICLing 2013, Samos, Greece, March 24-30, 2013, Proceedings, Part II 14*, pages 476–487. Springer.

Marten Düring, Roman Kalyakin, Estelle Bunout, and Daniele Guido. 2021. Impresso inspect and compare. visual comparison of semantically enriched historical newspaper articles. *Information*, 12(9):348.

Charles J. Fillmore and Collin F. Baker. 2001. Frame semantics for text understanding. In *In Proceedings of WordNet and Other Lexical Resources Workshop*.

Abram Handler, Narges Mahyar, and Brendan O'Connor. 2022. Clioquery: Interactive query-oriented text analytics for comprehensive investigation of historical news archives. *ACM Transactions on Interactive Intelligent Systems (TiiS)*, 12(3):1–49.

Bang Liu, Fred X Han, Di Niu, Linglong Kong, Kunfeng Lai, and Yu Xu. 2020. Story forest: Extracting events and telling stories from breaking news. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, 14(3):1–28.

Stefano Menini. 2024. Semantic frame extraction in multilingual olfactory events. In *The 2024 joint international conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING)*.

Gosse Minnema, Sara Gemelli, Chiara Zanchi, Tommaso Caselli, and Malvina Nissim. 2022. Sociofillmore: A tool for discovering perspectives. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, pages 240–250.

Teresa Paccosi, Stefano Menini, Elisa Leonardelli, Ilaria Barzon, and Sara Tonelli. 2023. Scent and sensibility: Perception shifts in the olfactory domain. In *Proceedings of the 4th Workshop on Computational Approaches to Historical Language Change*, pages 143–152.

Marco Rospocher, Marieke Van Erp, Piek Vossen, Antske Fokkens, Itziar Aldabe, German Rigau, Aitor Soroa, Thomas Ploeger, and Tessel Bogaard. 2016. Building event-centric knowledge graphs from news. *Journal of Web Semantics*, 37:132–151.

Marco Rovera. 2024. EventNet-ITA: Italian frame parsing for events. In *Proceedings of the 8th Joint SIGHUM Workshop on Computational Linguistics for Cultural Heritage, Social Sciences, Humanities and Literature (LaTeCH-CLfL 2024)*, pages 77–90, St. Julians, Malta. Association for Computational Linguistics.

Nicole Sultanum, Anastasia Bezerianos, and Fanny Chevalier. 2021. Text visualization and close reading for journalism with storifier. In *2021 IEEE Visualization Conference (VIS)*, pages 186–190. IEEE.

Piek Vossen, Rodrigo Agerri, Itziar Aldabe, Agata Cybulska, Marieke van Erp, Antske Fokkens, Egoitz Laparra, Anne-Lyse Minard, Alessio Palmero Aprosio, German Rigau, et al. 2016. Newsreader: Using knowledge resources in a cross-lingual reading machine to generate more knowledge from massive streams of news. *Knowledge-Based Systems*, 110:60–85.

Yunchao Wang, Guodao Sun, Zihao Zhu, Tong Li, Ling Chen, and Ronghua Liang. 2024. E2 storyline: Visualizing the relationship with triplet entities and event discovery. *ACM Transactions on Intelligent Systems and Technology*, 15(1):1–26.