Building a French Revolution Narrative from Wikidata

Inès Blin

Sony Computer Science Laboratories, Paris, France
Computer Science, Vrije Universiteit Amsterdam, The Netherlands
ines.blin@sony.com

Abstract

Historical reconstructions can help us better understand important events. This is essential for bettering our knowledge on specific topics and this enables us to build coherent sequences of information. Knowledge graphs like Wikidata contain information about generic knowledge that includes historical events, but the information is often entity-centric rather than event-centric. This makes the analysis and understanding of events related to a same topic harder, since it is less easily accessible. In this demonstration paper, we developed a system to build a timeline about the French Revolution in the form of a graph of a sequence of events, also called a narrative in this work. The system comes with an interface, where a user can first select the types of events to be retrieved, and is then walked through the steps of the system. The output of the system is a timeline of the French Revolution.

1 Introduction

We constantly create narratives to provide explanations or justifications for why and how somethings happens. This capability of creating such narratives is even suggested to be part of what makes us human [Boyd, 2010; Gottschall, 2012]. AI systems able to create narratives could therefore help in the realisation of human-centric systems. In this demonstration paper, we present a system that can build a timeline of the French Revolution by exploring Wikidata. The system takes the user's inputs to select nodes in Wikidata, walks the users through the different steps and outputs a timeline.

The system presented in this demonstration paper builds a narrative in the form of a graph. A narrative is defined here as a sequence of events. The main novelty of the paper is a system that walks the user through the different steps of the process and lets the user choose the events to include for the narrative. Earlier systems could let the user navigate through events and entities without a final timeline output [Boer *et al.*, 2017], or could build the timeline without interacting with the user, for example by retrieving most significant events for a biography description [Metilli *et al.*, 2019].

2 Related Work

A narrative is one way to understand or explain a series of related events. In this work, a narrative is technically defined as a sequence of events in the form of a graph. A narrative in this context is thus a type of knowledge graph. Therefore, our work is related to narratives and to how events are structurally represented, especially in event knowledge graphs structures. The output of our system is a timeline generated from a knowledge graph, hence our work is also related to timeline generation from knowledge bases. We describe the aforementioned research domains in this section.

Narratives are a means for better understanding how a situation unfolds over time. The events are the atomic concepts underlying these narratives. Towards formally defining such structures, Bartalesi and Meghini [2016] for example propose to define formal components of narratives using event calculus theory [Kowalski and Sergot, 1989].

From a resource perspective, the closest to our work would be event knowledge graphs that focus on describing events and links between them. Guan *et al.* [2021] survey event knowledge graphs from four different perspectives: history, ontology, instance and application. More recent research has focused on providing better access to semantic structured data about events, and has emphasised the need to have such event-centric information for applications such as timeline generation or history reconstruction. Available resources now include (i) EventKG [Gottschalk and Demidova, 2018] built from generic knowledge graphs, (ii) a knowledge graph built from textual news data [Rospocher *et al.*, 2016] (iii) a knowledge graph built from a Sherlock Holmes novel [Kawamura *et al.*, 2019] (iv) semantic data for cultural heritage [Dorobăț and Posea, 2019] and (v) GDELT [Leetaru and Schrodt, 2013] and ICEWS [Boschee *et al.*, 2015].

From an application perspective, the closest to our work would be systems that use knowledge bases for timeline or hypothesis generation. The aim of such systems is to help a user understand better one topic. For quality, Althoff *et al.* [2015] define three criteria for a timeline quality: relevance, temporally diverse and content diverse. For applications, Metilli *et al.* [2019] use Wikidata to build biographies, e.g. Dante's one. Kroll *et al.* [2020] furthermore build narratives as hypotheses on top of different knowledge bases, with an application in the biomedical domain.

3 Demonstrator Features

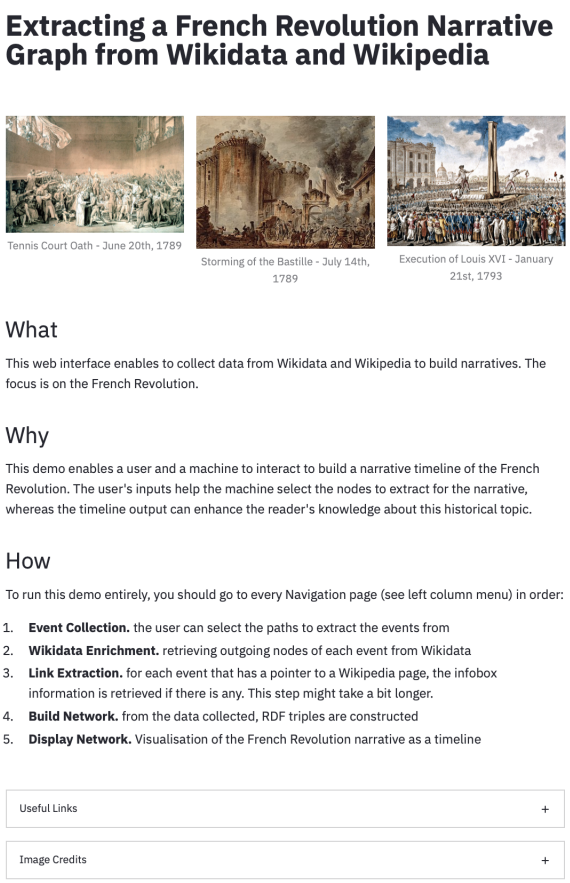

Figure 1: Home page of the application.

In the demonstration system, the user is walked through the different steps to build a narrative graph: 1) Collect events from Wikidata and select paths to extract nodes from 2) Extract features from Wikidata 3) Extract additional information from Wikipedia info boxes 4) Convert the data into RDF triples compatible with the Simple Event Model 5) Display the narrative graph as a timeline, with a description of each event as well as event type and participants. The demonstrator currently outputs a timeline for the French Revolution only, but could be extended to other types of revolutions for instance. The code is publicly available[1].

The backend is written in Python 3.9.4, the interface was developed with Streamlit[2]. Wikidata was accessed through the Wikidata SPARQL endpoint. Figure 1 shows the home page of the demonstrator that explains the purpose of the application and the main steps to extract the timeline.

3.1 Event Collection from Wikidata

The first step to build the timeline is to collect events from Wikidata. In the demonstrator, the user can select a collection of paths to extract events from, depicted in Figure 2. The user can then press the button "Collect events" to query the Wikidata knowledge graph.

Events and links to Wikidata and Wikipedia are stored in the backend, and the user is provided with some additional information, such as the number of events retrieved.

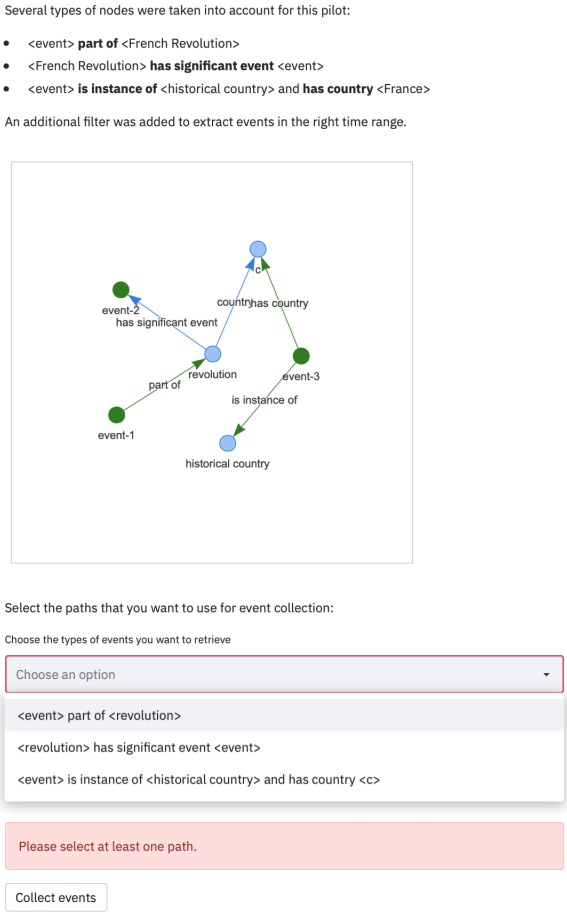

Figure 2: Collecting events from Wikidata

3.2 Extracting narrative features

Once the events have been extracted from Wikidata depending on the user's input paths, the next step is to extract narrative features for the events. A narrative feature for an event contains information about either a participant, a timestamp, a location, a type of event, or a temporal or causal link to another event. In Section 2. and 3. displayed in Figure 1, the user can press a button to extract these features.

In the demonstrator, the user can see the features that were extracted, as well as some additional statistics on the distribution of the types of features. We found that features extracted from Wikidata contained mostly information about time and location, whereas features extracted from Wikipedia contained more information about participants and causal links.

Features in Wikipedia also contain hyper links to other Wikipedia pages, which can therefore be linked back to Wiki-

[1] https://github.com/SonyCSLParis/building-fr-narrative-from-wikidata
[2] https://streamlit.io

data. We use and retrieve these corresponding Wikidata pages to have consistent URIs for the final graph.

3.3 Building the network

The next step is to build a narrative graph of the French Revolution events and their description, in the form of RDF triples. The rules to convert the extracted features for each event to triples were designed manually. The ontology used is the Simple Event Model [Van Hage *et al.*, 2011]. The four main classes of this model are: `sem:Event` (what), `sem:Actor` (who), `sem:Place` (where), and `sem:Time` (when). Further constraints classes `sem:Role`, `sem:Temporary` and `sem:View` can add information on the role of an actor, a temporal constraint or on a specific viewpoint respectively.

3.4 Timeline output

The final output in this demonstrator is an interactive timeline. This timeline contains ordered events extracted from Wikidata, as well as three additional components: (i) a brief description taken from Wikipedia (ii) event type information (iii) participant information. The events with no timestamps were therefore discarded from the timeline.

Figure 3 shows one event description for the 10 August. The user can read a summary of what happened during this event, as well as some additional information retrieved from Wikidata and Wikipedia. The user therefore understands that 10 August was a riot implying different actors like Louis XVI of France as a commanding officer, or the French First Republic as a combatant. Figure 4 lastly displays the interactive timeline overview. The user can manually slide events over time and click on each of them to better understand what happened. On the bottom of the figure, one can see that it is also possible to see which events happened during a coarser period of time, like for example the Kingdom of France.

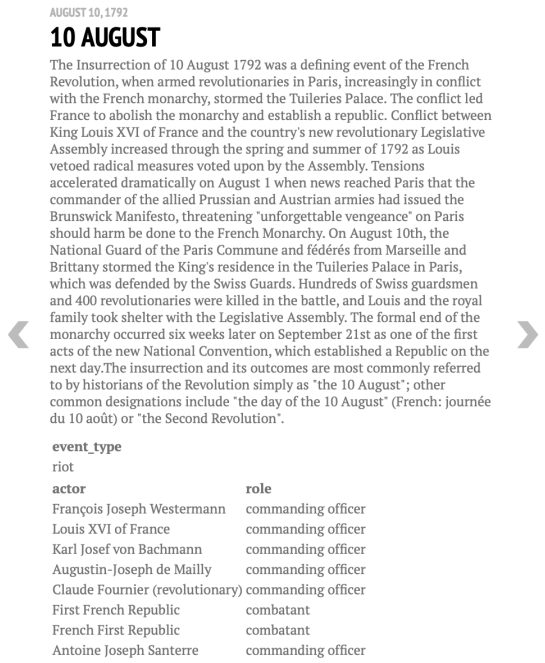

Figure 3: Event description example for 10 August

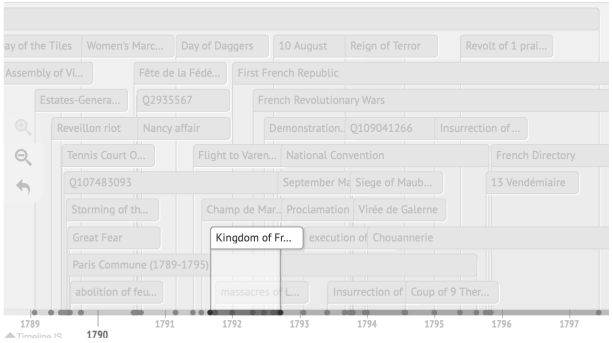

Figure 4: Output timeline overview

4 Conclusion

In this demonstration paper, we presented a system that is able to retrieve events and features of the French Revolution to construct a timeline of ordered events and descriptions. During this process, the user can choose how events can be selected and is walked through the different steps of the process, that makes the system transparent. The user lastly has access to an interactive timeline to better understand the series of events during the French Revolution.

The system described in this paper furthermore helped to identify some first challenges for building narratives from knowledge graphs. These challenges are linked to the main steps described in Section 3: 1) collect data 2) extract features 3) build the network. The main challenge for collecting data is to assess how to best choose the set of events that optimally describe the coarser event. Should such a set contain all events that happened, or only the most relevant, or only the ones that are interesting for one user? When extracting features, there is the challenge of completing missing information, like participants or locations. What additional resources or algorithms are best to complete the narrative graph? Lastly, the challenge when building the graph network is to find or define the ontology that is best suited for the application, and to populate this ontology.

Takeaways from this paper are also about possible future work directions. One first direction could be to study the temporality of the narrative, to better understand how events unfold. In the demonstrator, the timeline output provides a sequence of events ordered in time, but it does not provide explanations for how one event might have triggered changes that in turn caused another event. One second challenge and direction is about the evaluation of the narrative. For instance, one assumption here is that Wikidata and Wikipedia contain objective non-contradictory facts, but one interesting track would be to compare different input resources, or to add confidence scores to each triple. Another direction would be to see how well the system would scale to other events such as other revolutions. Lastly, one direction could be linked to pattern identification and event forecasting, to generalise from instantiated examples.

Acknowledgments

The work reported in this paper was funded by the European MUHAI project from the Horizon 2020 research and innovation programme under grant number 951846 and the Sony Computer Science Laboratories Paris.

We thank Ilaria Tiddi, Annette ten Teije, Frank van Harmelen and Remi van Trijp for their time and support.

References

[Althoff *et al.*, 2015] Tim Althoff, Xin Luna Dong, Kevin Murphy, Safa Alai, Van Dang, and Wei Zhang. Timemachine: Timeline generation for knowledge-base entities. In *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 19–28, 2015.

[Bartalesi and Meghini, 2016] Valentina Bartalesi and Carlo Meghini. Formal components of narratives. In *Italian Research Conference on Digital Libraries*, pages 112–121. Springer, 2016.

[Boer *et al.*, 2017] Victor de Boer, Liliana Melgar, Oana Inel, Carlos Martinez Ortiz, Lora Aroyo, and Johan Oomen. Enriching media collections for event-based exploration. In *Research Conference on Metadata and Semantics Research*, pages 189–201. Springer, 2017.

[Boschee *et al.*, 2015] Elizabeth Boschee, Jennifer Lautenschlager, Sean O'Brien, Steve Shellman, James Starz, and Michael Ward. ICEWS Coded Event Data, 2015.

[Boyd, 2010] Brian Boyd. *On the origin of stories: Evolution, cognition, and fiction*. Harvard University Press, 2010.

[Dorobăț and Posea, 2019] Ilie Cristian Dorobăț and Vlad Posea. Enriching the cultural heritage metadata using historical events: a graph-based representation. In *International Conference on Theory and Practice of Digital Libraries*, pages 344–347. Springer, 2019.

[Gottschalk and Demidova, 2018] Simon Gottschalk and Elena Demidova. Eventkg: A multilingual event-centric temporal knowledge graph. In *European Semantic Web Conference*, pages 272–287. Springer, 2018.

[Gottschall, 2012] Jonathan Gottschall. *The storytelling animal: How stories make us human*. Houghton Mifflin Harcourt, 2012.

[Guan *et al.*, 2021] Saiping Guan, Xueqi Cheng, Long Bai, Fujun Zhang, Zixuan Li, Yutao Zeng, Xiaolong Jin, and Jiafeng Guo. What is event knowledge graph: A survey. *arXiv preprint arXiv:2112.15280*, 2021.

[Kawamura *et al.*, 2019] Takahiro Kawamura, Shusaku Egami, Koutarou Tamura, Yasunori Hokazono, Takanori Ugai, Yusuke Koyanagi, Fumihito Nishino, Seiji Okajima, Katsuhiko Murakami, Kunihiko Takamatsu, et al. Report on the first knowledge graph reasoning challenge 2018. In *Joint International Semantic Technology Conference*, pages 18–34. Springer, 2019.

[Kowalski and Sergot, 1989] Robert Kowalski and Marek Sergot. A logic-based calculus of events. In *Foundations of knowledge base management*, pages 23–55. Springer, 1989.

[Kroll *et al.*, 2020] Hermann Kroll, Denis Nagel, and Wolf-Tilo Balke. Modeling narrative structures in logical overlays on top of knowledge repositories. In *International Conference on Conceptual Modeling*, pages 250–260. Springer, 2020.

[Leetaru and Schrodt, 2013] Kalev Leetaru and Philip A Schrodt. Gdelt: Global data on events, location, and tone, 1979–2012. In *ISA annual convention*, volume 2, pages 1–49. Citeseer, 2013.

[Metilli *et al.*, 2019] Daniele Metilli, Valentina Bartalesi, and Carlo Meghini. A wikidata-based tool for building and visualising narratives. *International Journal on Digital Libraries*, 20(4):417–432, 2019.

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

[Van Hage *et al.*, 2011] Willem Robert Van Hage, Véronique Malaisé, Roxane Segers, Laura Hollink, and Guus Schreiber. Design and use of the simple event model (sem). *Journal of Web Semantics*, 9(2):128–136, 2011.