

How to Search and Contextualize Scenes inside Videos for Enriched Watching Experience: Case Stories of the Second World War Veterans

Eero Hyvönen^{1,2}, Esko Ikkala¹, Mikko Koho¹, Rafael Leal¹, Heikki Rantala¹,
and Minna Tamper^{1,2}

¹ Semantic Computing Research Group (SeCo), Aalto University, Finland
<https://seco.cs.aalto.fi>, firstname.lastname@aalto.fi

² HELDIG – Helsinki Centre for Digital Humanities, University of Helsinki, Finland
<https://heldig.fi>, firstname.lastname@helsinki.fi

Abstract. This demo paper demonstrates the idea of publishing and watching videos on the Semantic Web. An in-use application, WARMEMOIRSAMPO, is presented that enables scene segments in videos to be searched by their semantic content. While watching a video, additional contextual information is provided dynamically. The system is based on a SPARQL endpoint whose knowledge graph has been extracted automatically from timestamped natural language descriptions of the video contents.

1 Introduction and Related Work

More and more content on the Web is published as videos³. Often the videos are long and may contain heterogenous scene segments, such as news in a longer broadcast or scenes of different topics in a film or an interview. Traditional search for whole videos is then not enough but tools for searching and accessing particular scene segments inside the videos are needed.

This paper presents an in-use semantic portal for annotating and searching scenes inside videos. When viewing a video, the annotations can be used for contextualizing the scenes with additional linked data for enriched watching experience. As a case study, a collection of interviews of the Second World War (WW2) veterans is considered. Here timestamped textual descriptions of the videos were available written by the interviewer. Our challenge was to create a knowledge graph of the videos based on their textual descriptions, and on top of it a portal for searching, exploring, and viewing the videos.

Related Work The research area of video indexing is surveyed in [1]. Indexing can be done by analyzing the frames and/or audio of the recording to find, e.g., the spots where goals are made in a football match. Another option is to use the textual subtitles (dialogues, commentaries) of the video. In some cases, e.g., in historical film archives, manually curated textual descriptions or

³ E.g., <https://youtube.com/>, <https://tiktok.com/>, <https://vimeo.com/>, etc.

commentaries of the videos may be available for preserving cultural heritage—they can be used for annotations and indexing, too, as in our case study. Various methods and tools are available for extracting linked data from texts [6]. Providing contextual information and ads while watching videos has been suggested already in the 80’s in systems, such as Hypersoap⁴. Works on enriching video watching experience using linked data-based recommendations include, e.g., [7].

2 WarMemoirSampo System: Data Service and Portal

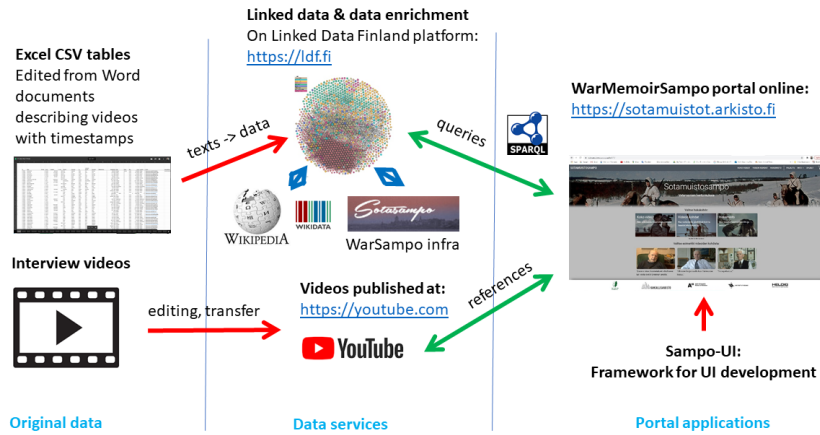


Fig. 1. Publication pipeline from source data to the semantic video viewing portal

Publishing Model The publication pipeline of the WARMEMOIRSAMPO system is illustrated in Fig. 1. The original data on the left are 1) a set of interview videos and 2) timestamped notes made by the interviewer about the contents of the videos, transformed into spreadsheets with timestamps and corresponding texts. The videos were published on YouTube and its IFrame player⁵ is used. The spreadsheets were transformed into a knowledge graph (KG) of 323 000 triples. Its data model contains the core class for an interview whose instances refer to sets of timestamped scene instances. The scenes are annotated with instances of named entities, based on their mentions in the texts describing the scenes, and topical keywords linked to an ontology. The entities were linked to resources in the WarSampo KG [4], an extensive data infrastructure for the

⁴ www.media.mit.edu/hypersoap/

⁵ https://developers.google.com/youtube/iframe_api_reference

Finnish WW2, and Wikidata/Wikipedia for further information. The semantic scene annotation and indexing process using natural language processing tools is presented in [5].

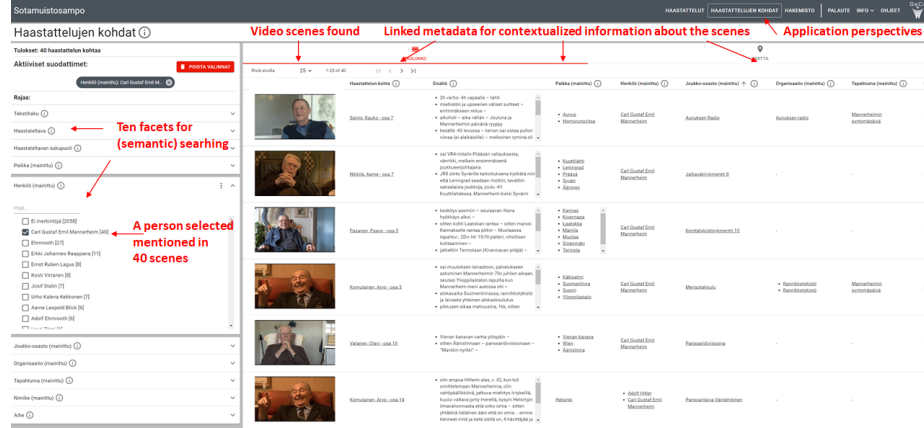


Fig. 2. Faceted search for scenes inside videos.

Semantic Portal Based on the Sampo model [2] and Sampo-UI framework [3], the landing page of WARMEMOIRSAMPO portal provides three application perspectives to the KG with faceted semantic search: 1) *Interviews perspective* is used for searching whole videos based on their nine key properties: Interviewee, Interview notes, Gender, and mentioned Place, Person, Military unit, Organization, Event, Other entity, and Topic. There is also a facet for traditional text-based search. 2) *Scenes perspective* is used for searching video scenes using the same facets. 3) *Directory perspective* contains all ca. 3000 entities mentioned in the texts with direct links to scenes where the entities were mentioned.

Fig. 2 depicts the Scenes perspective, where the user has selected “Carl Gustav Emil Mannerheim” on the facet Person: the 40 scenes mentioning this marshal are shown on the right with metadata links for further information. By clicking on a video, its is opened for dynamic viewing as depicted in Fig. 3. Links to additional information are provided on the fly. Selecting the tab Map on top shows the places mentioned on a map; Fig. 4 shows all 4566 of them. Clicking a marker on the map opens a popup with links to all related scenes. Finally, the tab Word cloud summarizes the topics of the video interview.

Technology The data model of the underlying knowledge graph contains classes for videos and their scenes annotated by entities extracted from their natural language descriptions, including people, places, military units, organizations, and events. In addition, metadata about the videos, such as the names of the interviewee and interviewer and the date of interview, were available. The facets ontologies for entities were created bottom-up based on the entities

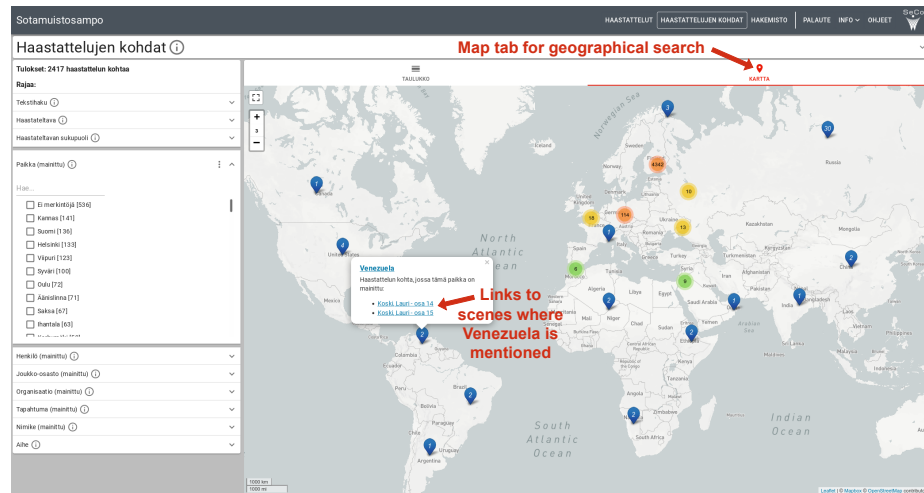


Fig. 4. Map view to access scenes that mention a place.

References

1. Hu, W., Xie, N., Li, L., Zeng, X., Maybank, S.: A survey on visual content-based video indexing and retrieval. *IEEE Transactions on Systems, Man, and Cybernetics* **41**(6), 797–819 (2011). <https://doi.org/10.1109/TSMCC.2011.2109710>
2. Hyvönen, E.: Digital humanities on the semantic web: Sampo model and portal series. *Semantic Web – Interoperability, Usability, Applicability* (2022), <http://www.semantic-web-journal.net/content/digital-humanities-semantic-web-sampo-model-and-portal-series-0>, accepted
3. Ikkala, E., Hyvönen, E., Rantala, H., Koho, M.: Sampo-UI: A full stack JavaScript framework for developing semantic portal user interfaces. *Semantic Web – Interoperability, Usability, Applicability* **13**(1), 69–84 (January 2022). <https://doi.org/10.3233/SW-210428>
4. Koho, M., Ikkala, E., Leskinen, P., Tamper, M., Tuominen, J., Hyvönen, E.: WarSampo knowledge graph: Finland in the Second World War as linked open data. *Semantic Web – Interoperability, Usability, Applicability* **12**(2), 265–278 (January 2021). <https://doi.org/10.3233/SW-200392>
5. Leal, R., Rantala, H., Koho, M., Ikkala, E., Merenmies, M., Hyvönen, E.: WarMemoirSampo: A semantic portal for war veteran interview videos. In: *6th Digital Humanities in Nordic and Baltic Countries Conference* (2022), submitted paper
6. Martinez-Rodriguez, J.L., Hogan, A., Lopez-Arevalo, I.: Information extraction meets the semantic web: A survey. *Semantic Web – Interoperability, Usability, Applicability* **11**(2), 255–335 (2020). <https://doi.org/10.3233/SW-180333>
7. Nixon, L., Bauer, M., Bara, C.: Connected media experiences: Web based interactive video using linked data. In: *Proceedings of the 22nd International Conference on World Wide Web*. p. 309–312. ACM, New York (2013). <https://doi.org/10.1145/2487788.2487931>