Historical Knowledge Modelling and Analysis through Ontologies and Timeline Extraction Operators: Application to Computing Heritage

Christophe Ponsard[1], Aurélien Masson[2] and Ward Desmet[3]
[1]*CETIC Research Centre, Charleroi, Belgium*
[2]*ESEO Engineering School, Angers, France*
[3]*NAM-IP Computer Museum, Namur, Belgium*

Keywords: Domain Specific Modelling, Ontology, Historical Knowledge, Visualization, Timeline, Case Study.

Abstract: Cultural heritage as human science has a long tradition of text-based reporting and analysis. This domain has a very rich semantic structure, especially to relate many different types of entities anchored in some time period with more or less strong temporal inter-dependencies. Various modelling approaches, largely based on ontologies, have been proposed to capture and structure this kind of knowledge. In this paper, we are interested in easing the analysis capabilities on historical knowledge using the timeline as central concept that can be extracted and manipulated in various ways through specific operators sharing some similarities with multi-dimensional analysis in business intelligence. We propose zooming on a specific aggregates, pivoting from a timeline to another one or drilling-across to compare different timelines. Our work is illustrated on a concrete implementation targeting the Computing Heritage of the micro-computer period, including machines, operating systems, companies, people and applications. The information is extracted from a museum information system combined with DBPedia. We also developed a specific visualisation tool under the form of a mobile application which can also be used as museum guide.

1 INTRODUCTION

As defined by the SAGE dictionary (Thorpe and Holt, 2007): "Historical analysis is a method of the examination of evidence in coming to an understanding of the past. It is particularly applied to evidence contained in documents, although it can be applied to all artefacts. The historian is, first, seeking to gain some certainty as to the facts of the past. Establishing the facts also gives the researcher a chronology. The second task is to seek to establish cause and effect between those facts in order to understand why things happened. It is important to remember that while the past is the immensity of everything that has happened, history is what we know of the past". The way the past is presented is usually based on a narrative because "Human beings are story tellers who exist ontologically in a universe of narrative making" (Ankersmit, 2005).

The above references highlight fundamental aspects of historical analysis and reporting:

- history is what we know about the past,
- its analysis is based on the identification of the chronology and casual relations across facts, and
- its reporting is done through some form of narrative referring to an ontology.

The development of computer technologies and networks has led to the emergence of digital history which goes far beyond the pure digital formatting of content. It also enables new forms of historical analysis and visualisations (Burton, 2005), especially "visual historiography" which integrates non-textual materials into historical representations and hypertexts with social, spatial, and chronological perspectives (Roegiers and Truyen, 2006).

This paper focuses on a such visualisations through a common form of visual representation for time-based narrative: the timeline. Beyond the mere chronological representation of events in chronological order, the way it is build and displayed can exploit a rich set of relationships such as mereology (is part of), causality (is cause/consequence) and thematic aspects (tags). This requires conceptualisation and the explicit modelling of event structure and relationships through an adequate meta-model or ontology. Providing flexible and dynamic timeline visualisation means is useful both for:

- understanding history by engaging with the model

Ponsard, C., Masson, A. and Desmet, W.
Historical Knowledge Modelling and Analysis through Ontologies and Timeline Extraction Operators: Application to Computing Heritage.
DOI: 10.5220/0010900400003119
In *Proceedings of the 10th International Conference on Model-Driven Engineering and Software Development (MODELSWARD 2022)*, pages 302-309
ISBN: 978-989-758-550-0; ISSN: 2184-4348

to produce explanations for some events by tracing its connections with other events and the historical context (Walsh, 1951), and
- providing specific views and explanations to the general public in the context of an exhibition to stress the importance/influence of specific events or of a more global context, e.g. through parallel related timelines.

The use case which triggered this work is an exhibition about the micro-computer revolution from the 1970's to the 1990's which is technologically anchored in the development of microprocessors and graphical user interfaces, and socially in the hacker counter-culture. This work involved many timeline visualisation both in the preparation phase and as visualisation support for the exhibition. In order to go beyond static paper posters, we also designed an application providing a dynamic form of navigation through timelines. This work has led us to refine and formalise our approach with the aim to define a more general framework and tooling. The result could be applied in a wider context: in a computer museum but also in other kinds of museums or organisations involved in historical studies. We report here about our current progress and the following contributions:
- the design of the underlying ontology, based on a comparative survey,
- the extraction of interesting timelines from the resulting knowledge base, and
- the illustration of specific visualisations though our application.

This paper structure follows the above contributions. Our micro-computer case study is used as running example. Sections 2, 3 and 4 will respectively cover ontology design, timeline extraction and visualisation. Section 5 will discuss our results. Finally, Section 6 will conclude and give some perspectives.

2 ONTOLOGIES FOR HISTORICAL KNOWLEDGE

Many ontologies and meta-models have been published and provide solid grounds to build a knowledge base. This section reviews the main representatives and compares them against key concepts for digital history processing.

2.1 Simple Event Model (SEM)

Events have become central elements in the representation of data from domains such as history, cultural heritage, multimedia and geography.

The Simple Event Model (SEM) is an ontology created to model events in various domains such as history, cultural heritage, multimedia and geography, without making any assumptions about the domain-specific vocabularies used (W3C, 2009)(van Hage et al., 2011). It is designed with a minimum of semantic commitment to guarantee maximal interoperability. SEM is structured around four core concepts depicted in Figure 1:
- *Event* is the central concept. It has a loose definition (e.g. possibly fictional) and can originate from different sources. To cope with this, events may have bounded roles, related facts may have time bounded validity and an authoritative source may be associated to views.
- *Actor:* who or what participated in an *Event*.
- *Place:* where the an *Event* took place.
- *Time:* when the *Event* took place.

In this case study statement "The UK company Sinclair Radionics Ltd was founded by Clive Sinclair in 1961": the *Event* is the foundation, the *Actor* is Clive Sinclair, the *Time* is 1961 and Place is UK.

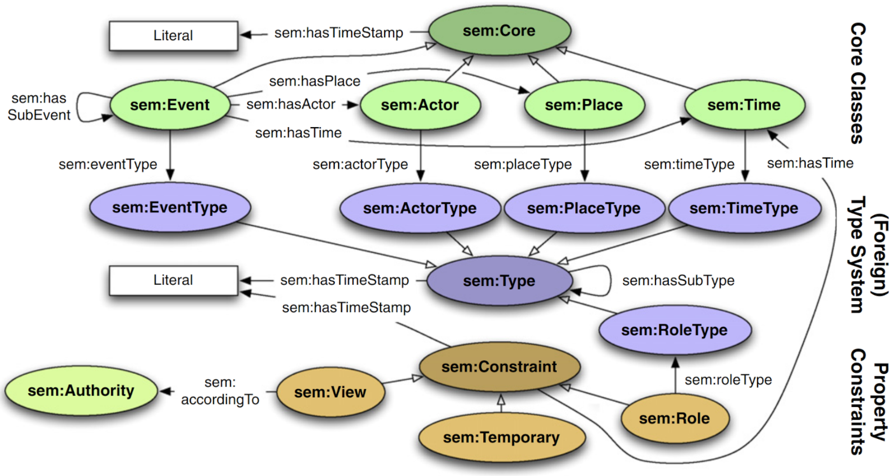

Figure 1: Simple Event Model.

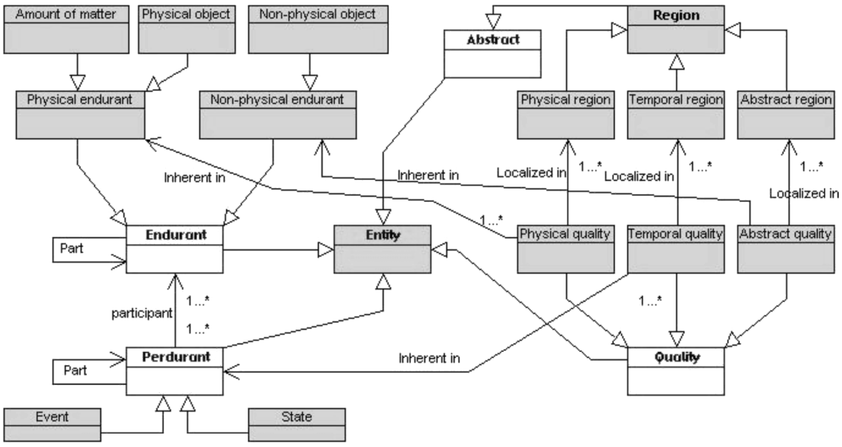

Figure 2: DOLCE Foundational Ontology.

2.2 DOLCE and Extensions

DOLCE (Descriptive Ontology for Linguistic and Cognitive Engineering) is a foundational ontology (FO) developed originally in the EU WonderWeb project (Masolo et al., 2003). FOs are domain-independent axiomatic theories that provide solid grounds to capture and reason about knowledge. Figure 2 depicts the meta-model using an UML class diagram. It is structured using the dual concepts of *Endurant* (including *Objects* or *Substances*) and *Perdurant* (including *Events*, *States*, or *Processes*) which are linked by the relation of participation:

- *Endurants:* localized in space, they get their temporal location from *Perdurants* they participate in.
- *Perdurants:* localized in time, they get their spatial location from *Endurants* participating in them.
- Additionally, specific qualities can characterise either *Endurants* (as *Physical* or *Abstract Qualities*) or *Events* (as *Temporal Qualities*). Each kind of *Quality* is associated to a *Quality Space* representing its range of possible values.

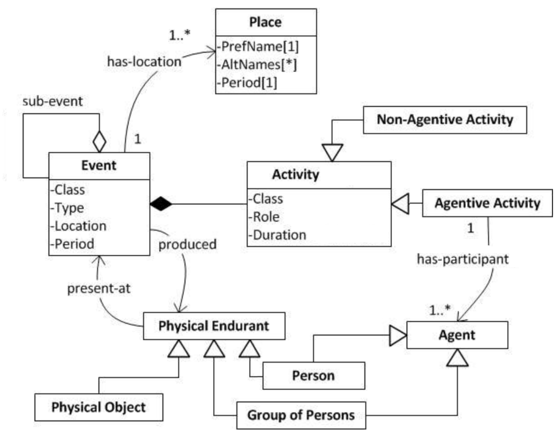

Figure 3: Spatial History Ontology.

Examples of *Endurants* in our exhibition are micro-computers (e.g. IBM-PC, ZX81) and people like computer scientists or entrepreneurs (e.g. Bill Gates, Steve Jobs). However, the later have can initiate events while the former are pure physical objects. This finer-grained level is supported by extended ontologies such as Spatial History Ontology (SHO) (Grossner, 2010) shown in Figure 3.

2.3 Constructed Past Theory (CPT)

Constructed Past Theory is an epistemological theory about how we come to know things that happened or existed in the past (Thibodeau, 2019). It is more elaborated in making the distinction between constructed and target past. Is it specified as UML class diagram partly depicted in Figure 4.

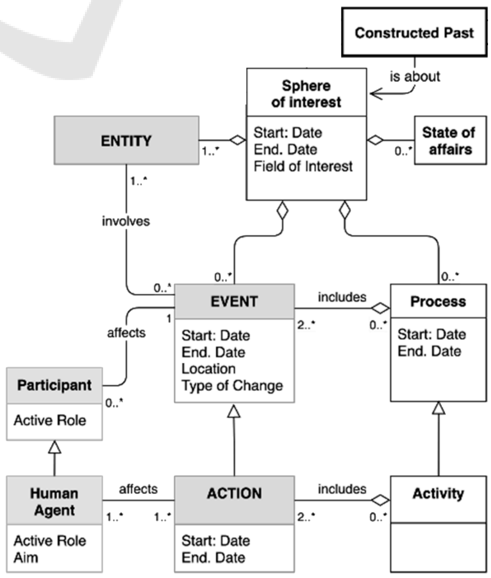

Figure 4: Constructed Past Theory (UML meta-model).

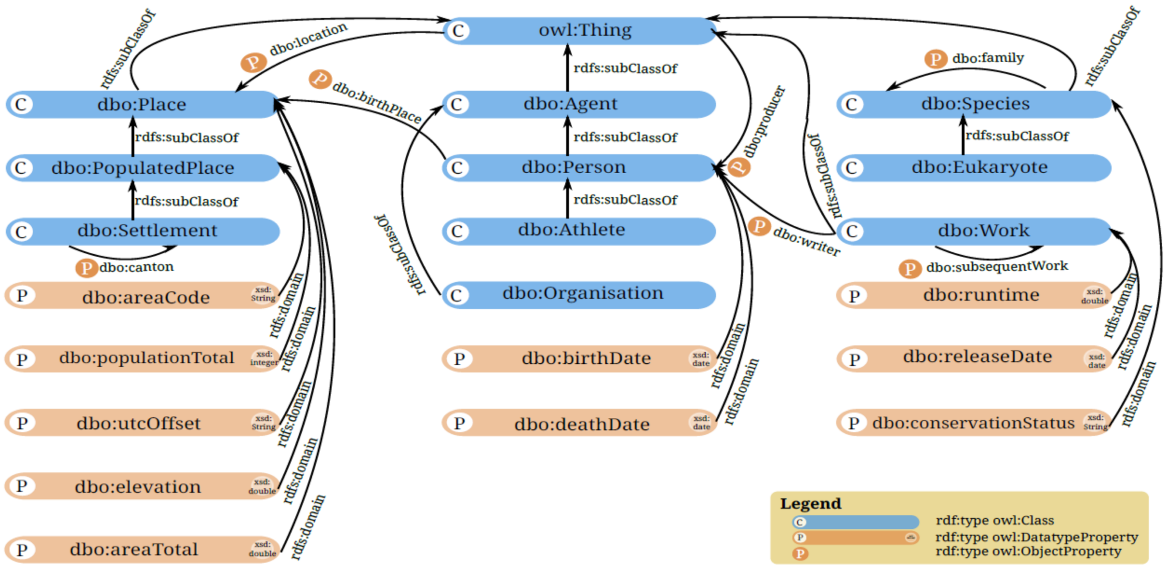

Figure 5: Main Structure of DBpedia.

It covers the SEM concepts of *Event* and *Entity*, including active. However, space and time are captured as attributes and not as first class concepts as in SEM and DOLCE. More complex events can be captured through processes which is also present in DOLCE (aggregated *Perdurant*) and SEM (subevents).

2.4 DBPedia Ontology

The previous meta-models are interesting but in order to be practical, data must be available. A rich form of data is provided by wikipedia. Structured information (e.g. infoboxes) is collected by DBPedia (Bizer et al., 2009)(Cvjetkovic et al., 2013)(Lehmann et al., 2015). It is organised into a dedicated semantic web ontology depicted in Figure 5. It shows the presence of the key concepts discussed before: general *Thing* and a more specialised form of *Agent*. The notion of *Space* is also explicit.

However, the notion of *Event* is not a first class concept: time aspects are only attached to specific concepts as attributes, e.g. birth and death date of a *Person*, release date or a *Work*. Consequently, building timelines requires to perform complex queries to gather the relevant information possibly across different concepts. DBPedia can be queried using SparQL (SPARQL Protocol and RDF Query Language), a standardised semantic query language for databases able to retrieve and manipulate data stored in Resource Description Framework (RDF) format (OMG, 2008).

Listing 1: SparQL query over computer companies.

```
SELECT DISTINCT ?date ?name
WHERE {
  ?company foaf:name ?name.
  ?company dbo:industry dbr:Computer_hardware.
  ?company dbo:foundedBy ?founder.
  ?company dbo:foundingYear ?date.
  FILTER(?date <= "19900101"^^xsd:date)
}
ORDER BY ASC(?date)
LIMIT 100
```

Listing 1 shows how to retrieve companies active in computer hardware founded before 1990. The formulation of the query requires careful design as data can be quite heterogeneous. This query yields the result partially depicted in Listing 2 which represents a timeline. Note that extra attributes retrieved by the related query are not shown but can be used e.g. founder and other company attributes. Note the interesting duplicate: HP Inc. and Hewlett-Packard. It can be explained by the split of this company in 2015. Such evolution is not yet captured by our current framework.

Listing 2: SparQL query over computer companies.

```
date name
1911 International Business Machines Corporation
1939 HP Inc.
1939 Hewlett-Packard Company
1958 Commodore International Corporation
1959 Kontron S\&T AG
1960 Augmentation Research
...
```

3 TIMELINE PROCESSING

This section details how our timelines are represented and manipulated. First, we present our meta-model design which is strongly inspired from the ontologies reviewed in Section 2. Second, we identify the main timeline extraction and navigation operations in order to provide interesting ways to walk through a domain using the time dimension as main representation but also by exploiting various filters and interconnections present in our meta-model.

3.1 Meta-model Design

The survey in the previous section illustrated that the identified ontologies share key concepts of event, active/passive objects and links with space and time information but with some differences on how to generalise or specialise them. We did not select a specific ontology literally but produced a customised meta-model strongly aligned with them, only keeping common and relevant concepts fitting our need to reason on timelines. Figure 6 depicts its current version.

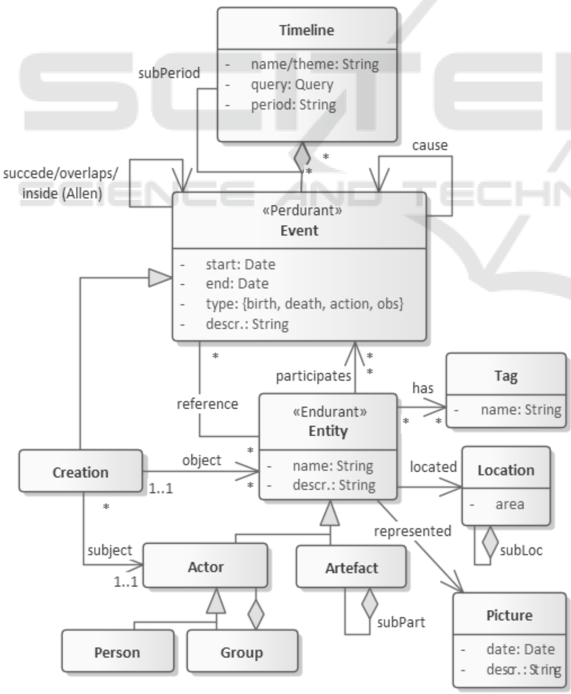

Figure 6: Current Timeline-oriented Meta-model.

Our approach was to combine the strong theoretical background based on DOLCE/SHO while keeping it as simple as possible like SEM while enabling to rely on information extracted from DBPedia. Our goal is to build a high quality knowledge base by combining first class data sources like museum inventory and implementing a cleaning and validation process by relying on open data. We explicitly modelled the *Timeline* as first class concept because it is central to our time-oriented approach while *Space* is captured using a *Location* concept. We refer to *Perdurant* (*Events*) and *Endurant* (either passive: *Objects* or active: *Actors*) using stereotypes. We support aggregation for all concepts with possible sub-timelines (e.g. specific phases of micro-computer development), sub-locations (e.g. a specific country, town), groups for actors (to capture the notion of company, department, e.g. XEROX PARC) and objects sub-parts (e.g. computer monitor, central unit, keyboard). An *Object* can be of physical but also logical nature, e.g. software in our case.

3.2 Timeline Extraction and Navigation

Assuming we have achieved data collection and validation according to our meta-model, we can consider the user experience going through the knowledge base using timelines. It is interesting to explore what kind of timeline can be extracted and how they can be connected together. We could identify different meaningful timelines that collect related events by focusing on:

- *actor(s)* at different granularity levels: it can be the life of a person, a group or a company, possibly with a focus on common event characteristics,
- *object(s)*, also at different granularity levels: it could relate to the precise history of a specific object (e.g. the design of the LISA computer) but also of a family of objects according to specific criteria, e.g. micro-computer of a specific period, manufacturer, using a specific CPU,...
- *temporal, spatial or thematic contexts*, respectively through specific *Event* (dates), *Location* or *Tag* characteristics. Different granularity levels can also be considered, e.g. to reflect the computer history related to micro-computer in France from 1970 to 1985. This can come as additional filter for the previous types of timelines.

Moreover timelines can be enriched by following specific dependencies such as causality links or expressed by explicit references to secondary entities. However, this can result in losing focus so we favoured a more dynamic mechanism through the use of different navigation operators allowing the user to easily “switch”between timelines according to specific aspects he would like to investigate. This shares some similarities with OLAP operators in multidimensional analysis, although we only focus on time here (Salley and Codd, 1998).

Figure 7 reflects this idea by showing 6 timelines with different scopes (Person, Company, Operating System,...) with intersecting points through specific events which can be used as pivot to change perspective. Some possible operators are:

- *event pivoting* between related entities or features: e.g. from Amiga 500 computer to Commodore Company or the 68k CPU or GUI timeline.
- *time zoom in/out* based on a defined period, e.g. the micro-computer history can be divided in early, golden age and standardisation periods.
- *actor zoom in/out*, from person level to company.
- object zoom in/out, e.g. down to version/variant level and up to product family level.
- *relations inclusion*, possibly iterative and with closure, e.g. to follow casual relation to look for causes/consequences related to some events.
- *combining multiple timelines together* either merged or keeping them separated with an adequate visualisation (temporal alignment, shared events, specific relations...)

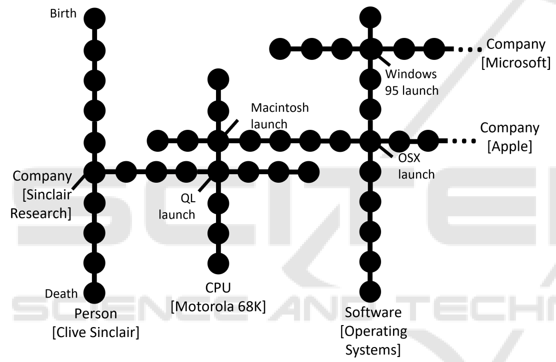

Figure 7: Navigation through Pivoting.

4 IMPLEMENTATION

Our current implementation is a monolithic prototype mobile application composed of two main components: a database and a visualisation module. The database implements our meta-model using a lightweight SQLite relational database (Hipp, 2000) fed by local computer inventory and DBPedia. The visualisation relies on React Native (Facebook, 2015) which provides a nice interface for a museum guide. It can later evolve towards a client/server application and is already available in Open Source (NAM-IP, 2021). Figure 8 shows a typical usage scenario with a pivoting between two timelines, from left to right:

1. a timeline showing some computers of the second phase (1977-1990),
2. a navigation to the Amiga 500 computer launched in 1987 displaying features such as the manufacturer, location, Operating System (OS), all of which are possible entry points to other timelines,
3. as the user decided to pivot to OS, the OS timeline is shown with a focus on the Amiga OS.

The user is also able to configure its timeline by selecting a specific period or the whole history, and by enabling or disabling different thematic features such as machines (MICRO), operating systems (OS), processors (CPU), user interface (IHM) or applications (APP). At this point, all events are merged and only the start date is considered so possible overlaps are not reflected. In order to have a better view, especially on smaller displays, a compact mode not showing picture nor first lines of description is available.

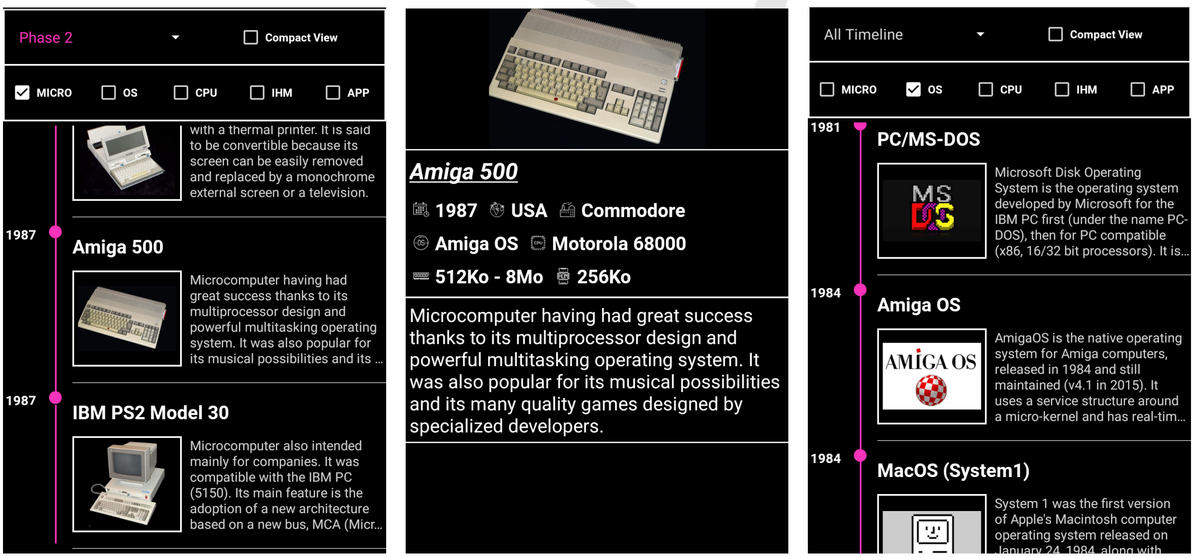

Figure 8: Navigation Scenario through our Mobile Guide.

5 DISCUSSION OVER RELATED WORK

The use of timelines in visual historigraphy has been discussed in the introduction. From a research perspective, the importance of tools providing visual contextualization of events is well-known (Shneiderman, 1996). We focus here on software development sharing similarities with our approach.

ChronoZoom is an open source data visualization tool that supports zooming through time to explore timelines across the whole history of the universe down to present history of humanity (Walter et al., 2013). It is fully web-based and has an efficient Cloud-based hosting providing efficient "infinite" zooming capabilities. It allows the user to create or customize timelines and to plot various time series data next to each other for comparison. Our approach is less ambitious as we only focus on specific thematic typically the domain of a museum. Although we focused on a specific exhibition, our approach is general and inspired by major ontological frameworks but also practical semantic web technologies, which opens the way to grab information from Wikipedia contributions. As ChronoZoom, we target both education and research but our current front-end is more focused on education. At the technological level, our solution can grab data from a locally embedded database but possibly also from a web API although not optimised for large scale use.

Different mobile applications have been developed more specifically for supporting computer history in general or in the context of a museum collection. They rely on the notion of timeline more or less explicitly. The Nexon Computer Museum, the first Korean Computer Museum with approximately 7,000 artifacts has a complete application to browse through the 400 artefacts on display and relying on a chronological ordering (Nexon, 2013). It gives details but few navigation operations. A computer history timeline developed by a company specialised in guides is also available for Android (Every Time Apps Studio, 2018). It is close to a poster timeline with however a practical slider for quick access. Different websites also publish computer timeline, including the famous Computer History Museum (CHM, 2021). It recaps the main events with a zoom by decades and years. It also proposes classification in different themes as we do, and additionally a search engine.

Gathering and qualifying data is a difficult task and can benefit from the semantic web. Our current approach relies on DBPedia (Bizer et al., 2009) but other knowledge graphs could be used such as the one provided by Google (Google, 2012), although it is itself largely based on Wikipedia. A difficulty is to achieve a mapping towards on our meta-model given the diversity of fields which are not uniformly used. Moreover the content is largely pure text, which means many interesting information cannot be automatically extracted with semantic search engines. A solution to this is the use of a semantic extension enabling to make semantic content explicit and thus make it visible to DBPedia (Vrandecic and Krotzsch, 2005). Another/complementary approach is to rely on natural language processing to extract such knowledge provided some form of reliability can be enforced. In our approach, we introduced a basic form of semantic processing by detecting synonyms in our text description, e.g. "First Macintosh" or "Macintosh 128K" or "Apple Macintosh" which can be extracted from DBPedia and manually enriched as required. Specialised work on the formulation and recognition of temporal patterns of events in English could also be considered (Saygi et al., 2018).

6 CONCLUSION & NEXT STEPS

In this paper, we presented our current progress in developing a framework for modelling and analysing historical knowledge by relying on timeline extraction, navigation and visualisation techniques. Although anchored in a specific case study of computer heritage, we took care of setting solid foundations based on a survey of relevant ontologies. We could successfully implement a prototype application based on a designed meta-model populated by a validated dataset mixing in-house and DBPedia information. The application, currently under validation in our museum, is quite appreciated.

Our framework is already available in Open Source on Github (NAM-IP, 2021). In the future, we plan to extend it in different directions: introduce more pivot points in the user interface, support parallel visualisation of timelines, provide location filtering and new categories for more cultural or political context and references. We also plan to elaborate a web-service API with a robust and scalable graph database as backend. At longer term, a collaborative web client with semantic editing capabilities will be considered.

ACKNOWLEDGEMENTS

We thanks the volunteers and anonymous visitors of the NAM-IP computer museum for their feedback on our current prototype.

REFERENCES

Ankersmit, F. (2005). *Sublime Historical Experience*. Cultural memory in the present. Stanford University Press.

Bizer, C. et al. (2009). DBpedia - A crystallization point for the Web of Data. *Journal of Web Semantics*, 7(3):154–165. The Web of Data.

Burton, O. V. (2005). American Digital History. *Social Science Computer Review*, 23(2):206–220.

CHM (2021). Timeline of computer history. https://www.computerhistory.org/timeline.

Cvjetkovic, V., Djokic Petrovic, M., and Arsic, B. (2013). Wikipedia Browsing With DBpedia. In *The 2nd Electronic International Interdisciplinary Conference*.

Every Time Apps Studio (2018). History of Computer. Google Play Store.

Facebook (2015). React native. https://reactnative.dev.

Google (2012). Knowledge Graph. https://developers.google.com/knowledge-graph.

Grossner, K. (2010). *Representing Historical Knowledge in Geographic Information Systems*. PhD thesis.

Hipp, D. R. (2000). Sqlite. https://www.sqlite.org.

Lehmann, J. et al. (2015). Dbpedia - a large-scale, multilingual knowledge base extracted from wikipedia. *Semantic Web*, 6(2):167–195.

Masolo, C., Borgo, S., Gangemi, A., Guarino, N., and Oltramari, A. (2003). Ontology Library D18. WonderWeb project.

NAM-IP (2021). Mobile museum guide. https://github.com/NAMIP-Computer-Museum/guideApp.

Nexon (2013). Mobile application of the computer museum. https://www.nexoncomputermuseum.org/english.

OMG (2008). SPARQL Query Language for RDF. https://www.w3.org/TR/rdf-sparql-query.

Roegiers, S. and Truyen, F. (2006). History is 3D: presenting a framework for meaningful historical representations in new media. *New heritage: beyond verisimilitude*, pages 49–57.

Salley, C. and Codd, E. F. (1998). Providing OLAP to User-Analysts: An IT Mandate.

Saygi, G., Blaise, J.-Y., and Dudek, I. (2018). Temporal hints in the cultural heritage discourse: what can an ontology of time as it is worded reveal? *Revue des Nouvelles Technologies de l'Information*.

Shneiderman, B. (1996). The eyes have it: A task by data type taxonomy for information visualizations. IEEE Computer Society.

Thibodeau, K. (2019). The construction of the past: Towards a theory for knowing the past. *Information*, 10.

Thorpe, R. and Holt, R. (2007). *The SAGE Dictionary of Qualitative Management Research*. SAGE Publications.

van Hage, W. R., Malaisé, V., Segers, R., Hollink, L., and Schreiber, G. (2011). Design and use of the Simple Event Model (SEM). *Journal of Web Semantics*, 9(2):128–136.

Vrandecic, D. and Krotzsch, M. (2005). Semantic wiki. https://www.semantic-mediawiki.org.

W3C (2009). "simple event model". https://semanticweb.cs.vu.nl/2009/11/sem/.

Walsh, W. H. (1951). An Introduction to Philosophy of History.

Walter, R. L., Berezin, S., and Teredesai, A. (2013). Chronozoom: Travel through time for education, exploration, and information technology research. Association for Computing Machinery.