

An interactive & visual generic timeline encompassing (nearly) every Wikipedia article

Alexander Berntsen Olle Fredriksson
alexander@plaimi.net olle@plaimi.net

11th April 2015

Abstract

We can generate pleasant interactive timelines for most of the articles on Wikipedia. This is afforded through scraping DBpedia for semantic article data, putting it in tempuhs, and presenting it with mytimelines.org. The noble goal is an engrossing generic timeline of all of Wikipedia. As an added bonus, said timeline would be an excellent preview of the capabilities of mytimelines.org.

CONTENTS

1	Introduction	1
2	The problem	2
3	Our idea	2
4	Related work	3
5	Conclusions and further work	3

LIST OF FIGURES

1	A timeline without filtering results in an information overload.	2
2	A filtered timeline.	2
3	Using filter operations to get both German physicists and Swiss Chemists from the 1900s.	3
4	Adding a new filter.	3
5	A focussed timespan.	3

1. INTRODUCTION

DBpedia is an example of linked data, where structured data has been extracted from Wikipedia. This data may be semantically queried, and as such contains Wikipedia entries with metadata[1].

DBpedia is designed for semantic querying of data, not presenting data. If permitted to beg the question: data becomes more useful if used. So let's use it.

Some of the entries have time-related metadata. This data might be more absorbing if presented as an interactive visualisation.

We may scrape all of DBpedia, and record all of the entries with time metadata in tempuhs. This data may in turn be visualised using mytimelines.org.

2. THE PROBLEM

In fact, DBpedia is already scraped and put into tempuhs, with metadata describing and cataloguing the entries. The cataloguing is important for being able to filter the immense amounts of data.

Given a perfect universe with no computational bottlenecks, tempuhs is already fit for fight. Notwithstanding, due to the quantity of entries contained within DBpedia, a pragmatic timeline will require some improvements in tempuhs in order to serve it quickly enough. It needs to be partially served. It should furthermore in all likelihood be cached in a clever way. For this it might be fruitful to introduce the concept of read-only timespans to tempuhs.

As for mytimelines.org, it really is not up to par to handle this amount of data, or anything anywhere near this right now. timeline.js simply won't cut it. It is optimised for 20-30 entries[2] – DBpedia will be several millions of entries. mytimelines.org in its current state will not be able to present the required amount of data. Perhaps it will not even be able to **receive** the required amount of data.

The envisioned mytimelines.org user experience as presented in January[3] would however be more than sufficient. It is imperative that mytimelines.org is able to handle larger datasets.

3. OUR IDEA

A more mature mytimelines.org will be able to realise our idea. Our idea is elegantly simple and simply elegant: make a stupendous timeline of nearly all of Wikipedia.

The timeline will consist of merely “nearly” all articles, because not every article has time metadata. The ones that do may however be scraped for such metadata via DBpedia. We have already done that.

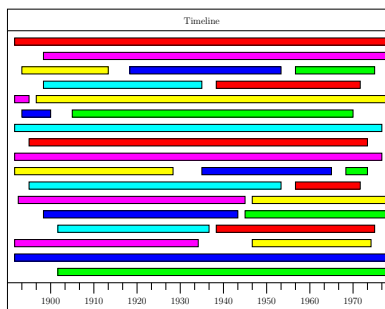


Figure 1: A timeline without filtering results in an information overload.

The reason we want one colossal timeline is so that the users themselves can judge what information is relevant through filtering the timespans contained within the timeline.

Figure 1 demonstrates how a timeline of all of DBpedia might look like. The programmer art is admittedly a slight degradation of the user interface presented in January. This figure demonstrates the problem of having Wikipedia as solely a visual timeline – too much data! It’s positively bewildering. Not cool.

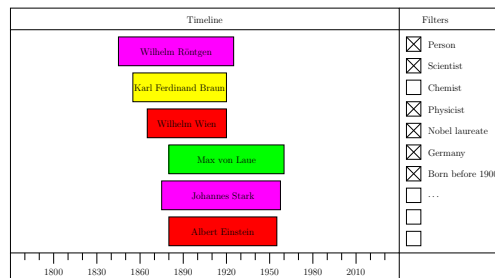


Figure 2: A filtered timeline.

Using all of DBpedia without the option to filter on categories would be cumbersome at best. While filtering was not present functionality in the January presentation, we have taken the liberty of augmenting the design in Figure 2. This allows fine-grained control of what data is presently presented. Now we have an interactive visual timeline – very cool.

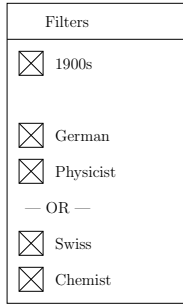


Figure 3: Using filter operations to get both German physicists and Swiss Chemists from the 1900s.

Additionally, it might be interesting to see articles that are related but not necessary in the same categories. Consequently there should be a way to make sets of filters, and express conjunctions and disjunctions of these. We have illustrated this in Figure 3, but concede that a professional user interface designer might be capable of expressing this more pleasantly.

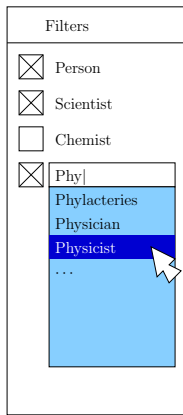


Figure 4: Adding a new filter.

The massive amount of data comes with a massive amount of categories for cataloguing purposes. A filter of filters is thus necessitated, as illustrated in Figure 4.

¹https://en.wikipedia.org/wiki/List_of_timelines

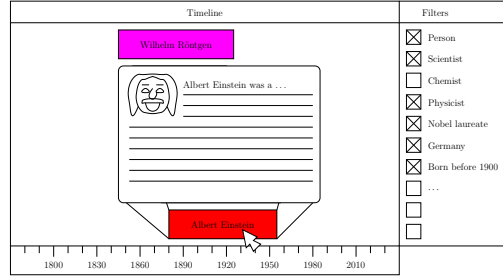


Figure 5: A focussed timespan.

Finally, Figure 5 shows an expanded timespan, wherein the preamble of the relevant Wikipedia article is displayed along with the relevant article image.

4. RELATED WORK

The Wikipedia project has several textual and tabular timelines¹. These are good examples of timelines we may visualise using tempuhs and mytimelines.org.

Our versions of those timelines are possibly superior for perusing in that they are visual and interactive – but doubly interesting inasmuch as they are all just subsets of a gigantic timeline which may be dynamically updated per filters.

5. CONCLUSIONS AND FURTHER WORK

An interactive visualisation of Wikipedia would be riveting from an information and infographics point of view. It would also serve as a neat preview of the capabilities of mytimelines.org; a motivating example for personal users to sign up to the service, and business users to perchance pay for similar timelines.

REFERENCES

[1] Christian Bizer, Jens Lehmann, Georgi Kobilarov, Sören Auer, Christian Becker, Richard Cyganiak, and Sebastian Hellmann. Dbpedia-a crystallization point for the web of data. *Web Semantics: science, services*

- and agents on the world wide web*, 7(3):154–165, 2009.
- [2] Northwestern University. How many entries will work best in TimelineJS? [http://](http://timeline.knightlab.com/#show-faq-04)
- timeline.knightlab.com/#show-faq-04, 2013.
- [3] Lene Kristin Thuseth. mytimelines presentasjon 16.01.2015, 2015.