Contemporary Newspaper Archives

Joshua Pacheco, Ludwig Frank, Bela Kurek

Problem Statement / Research Question

Can we use archives to contextualise newspaper articles for a clear and reliable understanding of events?

In an era of 'post-truth', it is increasingly difficult to distinguish between fact and fiction. It is therefore important that newspaper articles are read with a more holistic understanding of correlations and connections between events and actors. This understanding could be enhanced by situating newspaper articles within a timeline of events, generated by previously published articles in the archive. Alongside, a network of cross-references to parallel events in time could be introduced.

Articles of most current major newspapers are available in text format reaching back to at least the 1990s. However, they are only accessible by means of a specific search request. Despite recommending »related articles« to readers on the homepage, they are not referenced in the archive.

Concept

The desired result would be a functioning web application that generates timelines with cross references of events and articles within one newspaper's archive. An article should become more comprehensible by referencing the events leading up to it and gain its readers trust by showing consistency of coverage and events. Besides the technical execution of this concept, the main focus is on the design questions which arise concerning the visualisation of correlations and connections.

Enabling Technologies

We opt to create a web based application to enable a broad audience usage. The list of articles leading up to the selected article will, in the best case, be generated by Lucenes MoreLikeThis method. This would allow us to be independent on the language that is used. To start the prototyping phase es quick as possible with real data we will use ElasticSearch, which uses Lucene under the hood, as a SaaS, not worrying about back-end architecture.

On the front-end we will use React for UI components and state management (Redux). Other libraries that might be of interest are D3 (data visualization) and Slate (text editing). We will need to tweak them a lot to fit our use-case, but these libraries will serve as a good foundation.

Information Retrieval Software Libraries

Lucene

Seems to be the perfect candidate for our use-case. The MoreLikeThis method allows to generate a list of similar documents. It is used by wikipedia, among others and allows a big variety of data inputs (HTML, PDF, Text &). The following two libraries are based on this library.

Solr

Solr is a library based on Lucene that extends the project and allows us to use its versatile HTTP and JSON APIs.

ElasticSearch

Is a library based on Lucene that extends the project and offers a JavaScript interface. It also has an easy to set-up SaaS. This makes it the perfect candidate since it allows us to go straight into developing and testing the frontend without worrying too much about backend technologies and performance.

Frontend Libraries

React

React allows us to develop the project collaborative, since it is by far the most popular JavaScript front-end technology and easy to grasp for developers/designers familiar with JavaScript.
There are also a lot of resources online in form of libraries and documentation dealing with React.

D3

Slate

Should we, at some point need the ability for text editing, this library fixes the browsers 'contentEditable' and uses React components for its rendering.

Popmotion

Animation Library for state transitions.

Personae

The concepts target group are readers of digital newspapers, who are interested in a deeper contextual understanding of the presented information and in discovering unknown correlations between events in time.

Field overview

Timeline's claim [»Timeline puts our world in context, deepening the way we understand the news«] focuses on a similar concept, but the content is edited/curated, it is concerning single historic events and their apps timeline is currently not available.

Google's News Timeline was launched in 2009 but has been abandoned since.

Facebook is currently introducing a button on newspaper articles which provides relevant context [»such as information from the publisher's Wikipedia entry, trending articles or related articles about the topic, and information about how the article is being shared«].

Here are some examples of edited/curated timelines: New York Times, Washington Post, The Guardian, Wikipedia.

Timeline

Deadline	Task	Assigned to
March 7	first sprint-meeting in Berlin	@Joshua @Ludwig @Bela
	create personas and use-cases	@all
	field overview	@Joshua
	enabling technologies	@Ludwig
	prepare design sprint	@Joshua
March 14	4 days Design Sprint	@all
	establish contact to newspaper	@Bela
	present sprint findings	@all
March 21	wrap up personas and use cases	@all
	create journey map	@Joshua @Bela
	refine wireframes produced at sprint	@all
	sort out newspapers/stakeholders needs	@Bela
March 28	create first tangible prototype	@Ludwig
	prepare user test	@Joshua
April 4	user testing	@Joshua @Bela
	review user test	@all
April 11	adjust UX concept	@Joshua @Bela
April 18	*tbd*	
April 23	build first version of product	@Ludwig
	prepare presentation	@Joshua @Bela
May 2	finalise all processes	@all
	presentation at MIT	