Visualization of geo-referencialised data of Dingler's »Polytechnisches Journal«

Patent entries on the TimeMap

Marius Hug – Humboldt-Universität zu Berlin

Martina Gödel – p. p. Editura GmbH CO. KG, Berlin

April, 2011

1 Abstract

This poster is to introduce a specific data presentation of a digitization project located at the Institut für Kulturwissenschaft, a department of the Humboldt-Universität zu Berlin. In this project we digitize Dingler's »Polytechnisches Journal«, including thousands of patent application. In order to beeing able to visualize these patents as geo-referencialised data on a TimeMap we extensively encode names, dates and places.

2 What is Dingler's »Polytechnisches Journal«?

In 1820 the German chemist and industrialist J. G. Dingler released the »Polytechnisches Journal«. This journal was to include his personal but representative selection of a broad variety of articles. Originally most of these articles had been published in magazines all over Europe.

The »Polytechnisches Journal« was published over a period of 111 years. It is unique and highly relevant for very different research fields which focus on the cultural history as it emerged from Europe's technical transformations. Therefore it is significant not only for people engaged in the history of science, but for anyone interested in the cultural heritage of Europe.

3 Our digitization project

This project, funded by the DFG (German Research Foundation) and located at the Institut für Kulturwissenschaft, a department of the Humboldt-Universität zu Berlin, sets out to digitize »Dingler's Polytechnisches Journal«. Aside from the digitization of the journal's images, which already has been done by our cooperation partner SLUB-Dresden, we encode the OCRed text according to the Text Encoding Initiative Guidelines TEI-P5. Our online edition of Dingler's journal is freely available on the web via `www.polytechnischesjournal.de/`.

4 Patents – an interesting data model

The »Polytechnisches Journal« introduces a broad variety of inventions of the 19th century, including: locomotives, steam boats, photo cameras, electric light bulbs, telegraphs and telephones as well as developments concerning electro-magnetism, steel production, gutta-percha or dynamite, just to mention a few. A main source for all these are patent applications. The »Polytechnisches Journal« includes both articles, based on the relevant patent application, and voluminous patent lists, giving a summary of applications for a specific time.

Since patent applications are a very interesting source in a cultural-historical research field, our project intends to allow very different research approaches and therefore a deeply granulated tagging is absolutely essential.

A typical list entry, after standard text encoding by the service provider Editura GmbH, will look like this and only serves as the basis for our more detailed encoding:

```xml
<div type="patent" xml:id="p058067_17">
    <p>
        Dem <hi rendition="#wide">Samuel Colt</hi>,
        Gentleman am Ludgate Hill, in der City von London:
        auf gewisse Verbesserungen an Feuergewehren.
        <hi rendition="#roman">Dd</hi>. 22. Okt. 1835.
    </p>
</div>
```

At this time our patentlists cover more than 24.000 entries. So we are dealing with a great collection of data-sets, which are both compact and expansive in the sense, that each single entry is short, but still contains persons as patent applicants, titles, places and dates.

Therefore apart from the patentlists' importance for research activities, they are indeed a great source for extensive TEI-tagging and a challenge for visualization.

After our semi-automatical encoding our XML-file includes following elements:

```
- <persName role="patent_app" ref="&persons;#pers05725">Samuel Colt</persName>
- <title>Verbesserungen an Feuergewehren.</title>
- <placeName ref="&places;#pl000004">London</placeName>:
- <date when="1835-10-22" type="dd">22. Okt. 1835</date>.
```

5 Visualizing I: Simile timeline

Our data is stored in tables in a postgres database. After selecting all the relevant fields we can write our data to a XML-file which will feed the Simile Timeline, developed by the MIT (http://www.simile-widgets.org/timeline/). The syntax of the timeline.xml is simple and very restricted. Each event askes for only two attributes:

attribute	value (taken from dingler-DB)
start	date
title	persName: title

After filling in our data a typical entry corresponds to following pattern:

```
<event start="1835-10-22" title="Samuel Colt: Verbesserungen an Feuergewehren." />
```

Visualizing our data on a Simile Timeline will then result in:

3
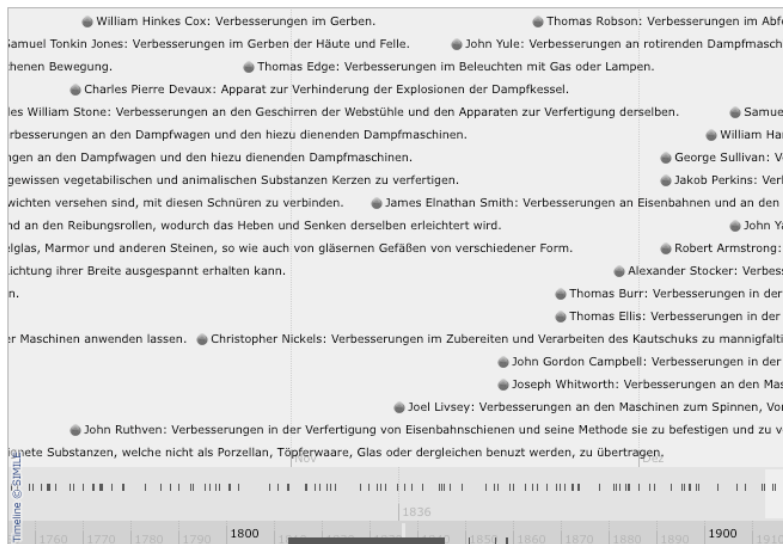
Next step is matching these patent events to a patent database. (In case of US patents, even original patent applications are available for download.) This gives us not only the original title (mostly a short form) and the patents' number, which is important for researching the original files, but each patent application is matched to different categories such as agriculture, weapons etc.

These categories will be adopted to our data-set and therefore our users can access the patent collection more effectively. We can provide access by

1. a kind of browsing experience via categories divided by colors

2. a filtering function, highlighting different search results for comparing data

3. a search field, which will result in showing only search hits

Finally the presentation of our data on a timeline will look the following way

Figure 1: http://dingler.culture.hu-berlin.de/timeline

The underlying data model is this:

```
<event start="date"
       title="persName: title"
       link="xml:id"
       icon="category_color">
       Original title (country code plus patent number).
       Link to download (if available)
</event>
```

4

As you can see we added two more attributes: the link-attribute will allow direct access to our usual online presentation and the icon-attribute which is to store the category-data.

Visualizing our patent data on a timeline obviously ought not to be a one way solution. Just think of following scenario: Making use of the projects' standard search field leads you straight into a patent-list where there is a certain patent entry which matches your specifications. Now you have the opportunity to visualize your search hit on the timeline and then load additional data filtered by categories. This will give you the chance of finding similar inventions made at the same time as well as very different events, which might be interesting as well.

6 Visualizing II – Dingler Online on a TimeMap

As you can see on the following code snippet, our patent data model allows even more:

```xml
<div type="patent" xml:id="p058067_17">
    <p>
        Dem <persName role="patent_app" ref="&persons;#pers05725">Samuel Colt</persName>,
        Gentleman am Ludgate Hill, in der City von
        <placeName ref="&places;#pl000004">London</placeName>:
        auf gewisse Verbesserungen an Feuergewehren. <hi rendition="#roman">Dd</hi>.
        <date when="1835-10-22" type="dd">22. Okt. 1835</date>.
    </p>
</div>
```

Additional to *names* and *dates* every patent entry includes a *placeName*. After tagging these places and mapping them to there geographical data via a perl script — based on two modules (Geo::KML and XML::LibXML) — we get following result, stored in an external file places.xml:

```xml
<place xml:id="pl000004">
    <placeName>London</placeName>
    <location>
        <geo>51.5 -0.116667</geo>
    </location>
</place>
```

All these information finally go together in a KML-file which serves as data-file for visualizing the patent-entry on a Timemap, which is to combine our Simile timeline presentation with a google map.

Visualizing the data on a timemap will result in following presentation:
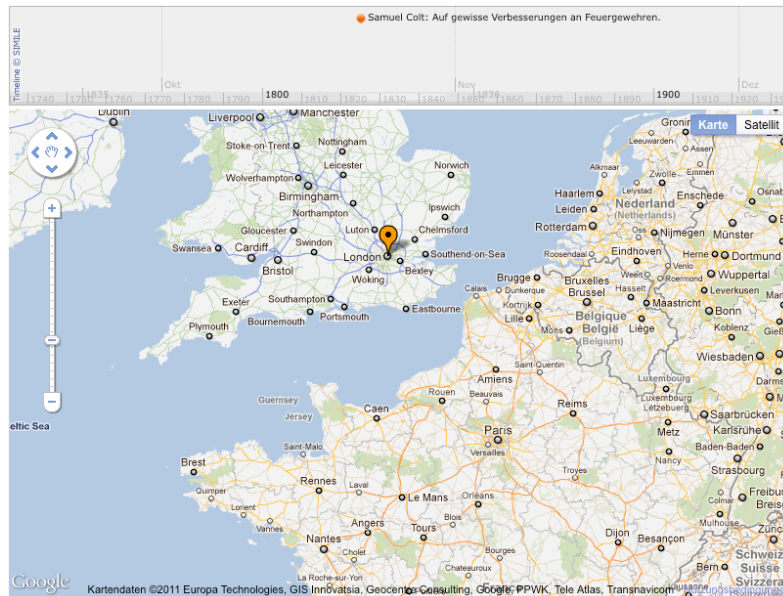
The syntax of the KML-file is as easy as this:

```
<PlaceMark>
    <name>Samuel Colt: Auf gewisse Verbesserungen an Feuergewehren.</name>
    <TimeStamp>
        <when>1835-10-22</when>
    </TimeStamp>
    <Point>
        <coordinates>51.5 -0.116667</coordinates>
    </Point>
    <description>
        <title_orig>Fire Arms</title_orig>
        <patent_no>GB 6009</patent_no>
        <category>MIL</category>
    </description>
</PlaceMark>
```

Having all patents visualized on a timemap will be a uniqe possibility getting to know the technical horizon of our European culture. When and where was the steam engine mentioned first. How did this invention make its way to become one of the most important technical innovation all over Europe? And which inventions have been made at the same time nearby or further away.