Patent policies in Dingler's »Polytechnisches Journal«
Exemplary tagging of names, dates and places

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 highly divergent articles. Most of these articles originated from 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.

Why patents?

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 electromagnetism, 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 lists of patents, 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.

The tagging workflow

Our TEI-encoding (it corresponds to encoding level 5, according to the »Guidelines for Best Encoding Practices«) provides direct access to patent applications. All patent lists are distinguished between their geographical origins, and we are to encode not only the *persons* as patent applicants, but also *dates* and *places* whenever available.

The consistency in which the original pattern was set up in type is a great help for our script-based workflow. Therefore we are enabled to make extended use of xpath queries and regular expressions.

A typical entry in a patent-list, after standard text encoding by the service provider Editura GmbH, will look like this:

```
<div type="patent" xml:id="p064078_7">
    <p>
        Dem <hi rendition="#wide">William Hinkes Cox</hi>,
        Gerber in Bedminster bei Bristol:
        auf Verbesserungen im Gerben.
        <hi rendition="#roman">Dd</hi>. 14. Okt. 1836.
    </p>
</div>
```

Especially the tagging of dates is done pretty much automatically using a little perl-script which can even match historic month-names (Hornung means February, for instance). Following code-snippet is taken from this script:

```
foreach my $journal (@ARGV) {
    debug( "lese $journal\n" );
    open( my $fh, '<', $journal ) or die $!;
    my $content = do { local $/, <$fh> };
    $content =~ s/(<text[^>]*type="art_patents"[^>]*>\s*)(.+?)(?=<\/text>)/$1 . findartpatent
    $content =~ s/(<div[^>]*type="misc_patents"[^>]*>\s*)(.+?<\/div>)\s*(?=<\/div>)/$1 . fin
    $content =~ s/(?<![0-9]{4});(<\/date>)/$1;/g;
    $content =~ s/;(<\/date>)/$1;/g;
    debug( "$patents Patente, $hits Ersetzungen\n", '-'x30, "\n" );
    $total += $hits;
    $hits = 0;
    $patents = 0;
    my ($tempfh, $tempname) = tempfile( UNLINK => 1 );
    print $tempfh $content;
    close $fh or die $!;
    move( $tempname, $journal );
}
debug( "$total Ersetzungen gesamt\n" );
```

Once the persons, dates and places of the patent entry are tagged, our TEI-file will look the following way:

```
<div type="patent" xml:id="p064078_7">
    <p>
        Dem <persName role="patent_app" ref="&persons;#pers05033">William Hinkes Cox</persName>,
        Gerber in Bedminster bei <placeName ref="&places;#p1000028">Bristol</placeName>:
        auf Verbesserungen in Gerben. <hi rendition="#roman">Dd</hi>.
        <date when="1836-10-14" type="dd">14. Okt. 1836</date>.
    </p>
</div>
```

The presentation strategy

Our encoding provides an advanced search restricted to patent-lists, furthermore it is possible to search specifically for patents with a certain origin as far as space and time are concerned. Our users can search for articles explicitly referring to patent applications and for persons restricted to the role of patent applicants.

A very promising visualization is based on an open-source solution from the Simile project at MIT: Timeline (http://www.simile-widgets.org/timeline/). Our data can be presented on a timeline very effectively. So having all UK patent applications of the years between 1810 and 1860 visualized on a timeline gives you the chance of finding correlations between beforehand different fields.
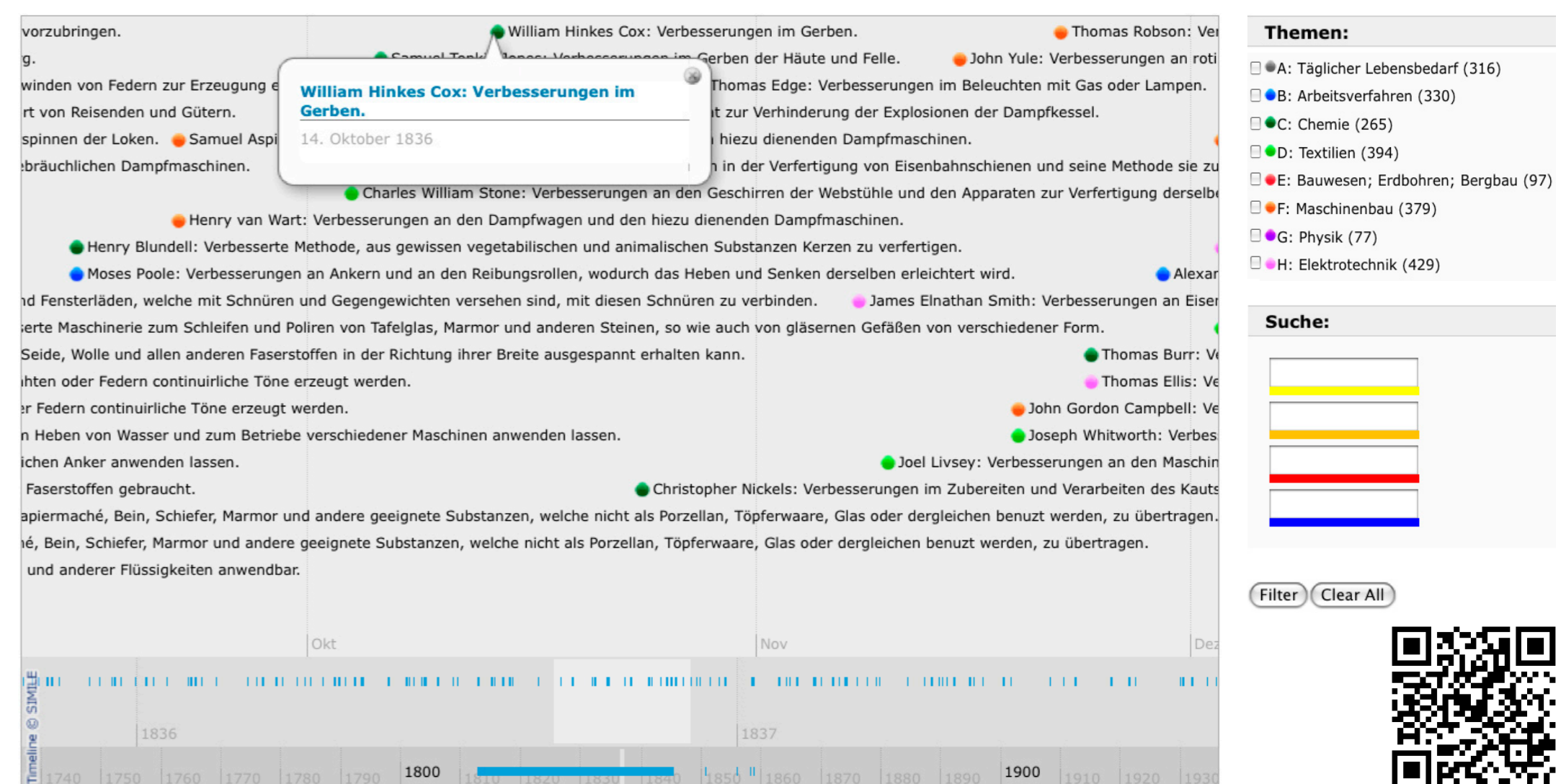
The syntax of the timeline is XML-based and very restricted. A typical entry corresponds to following pattern:

```
<event start="1836-10-14"
        title="William Hinkes Cox: Verbesserungen im Gerben"
        link="http://dinglr.de/ar064078">
</event>
```

The value of the start-attribute matches with our date-element in the patent-list and the persName-element can be matched to the first part of the @titles-value in our timeline.xml.

A great challenge are the patents' titles, whose encoding is not trivial since the editors of the journal translated and shortened them. Here again we can rely on a couple of regex-expressions which are to extract that piece of the patent-entry which is to match the entry's title.

The digitization project

Our project is funded by the DFG (German Research Foundation) and is located at the Institut für Kulturwissenschaft, a department of the Humboldt-Universität zu Berlin. Aside from the digitization of the journal's images, we encode the OCRed text according to the Text Encoding Initiative Guidelines TEI-P5.

Dingler-Online is an enriched digitization, which aims to be a user-friendly platform inspiring a broad use not restricted to historians of technology or, come to that, researchers, but is open to the interested public in general.

Concerning patents, at the moment our patent-lists 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, titles, places and dates. Therefore apart from the patent-lists' importance for research activities, they are indeed a great source for extensive TEI-tagging and a challenge for visualization.

Prospectively we are working on a way of matching each single patent entry to its corresponding section of the International Patent Classification (IPC). This will give users the chance of accessing our patent collection even more effectively.

Dingler-Online | www.polytechnischesjournal.de

A Project of the Department for the History and Theory of Culture at Humboldt-Universität zu Berlin

Deutsche Forschungsgemeinschaft

DFG · SLUB · Humboldt-Universität zu Berlin