

Locating the CTA

How can the Centraal Tekeningen Archief (CTA) best be made public, so that researchers, project developers and every other interested person can access the data?

The problem

The Centraal Tekeningen Archief (CTA) was recently digitized in order to make the archives public. Before making it public, a logical system needs to be implemented in order to search for certain files. The idea was to add the location of the subject of each file to the actual file, to make requests based on location possible. However, the locations are embedded in text strings with no logical order or whatsoever. The first challenge of this project was to extract the locations from these text strings. After this, a system needs to be made with the locations to present the digitized CTA archive to the public.



Authors

Hedwig Overwater, 2579976
Tomas van Daalen, 2598668

Institution

City Archives Amsterdam

Supervisors

Nelleke van Zeeland
Marc Holtman
Ivo Zandhuis

Course

Digital Humanities in Practice



Analyzing and ordering the data

Before actually some analysis of the data needs to be executed. In the analysis we looked if there were any regularities in the data. This might be useful before actually programming to gain some knowledge about the data. After this, we put the dataset in a SQL-operated system called phpMyAdmin

Extracting the data

We set up a script in the php language that was able to extract the streetnames from the database. It was able to do this by comparing the streetnames in another database called Adamlink to the text strings. The end result was a database with the file number, the found street and a URI of that specific street. This new database was implemented in phpMyAdmin to allow SQL queries.

Checking the new data

To check if the data was correct, we did a systematic sample test. This test checked if the found street was indeed mentioned in the description. In our first sample test we got that about 90% of our results was correct. There were some mistakes in streets consisting of two words, like Nieuwe Herengracht. In that case the script assigned both Herengracht and nieuwe Herengracht. A adjustment in the script caused the results increase to 97% of the matches being correct.

Making linked data

In order to actually make a useful interface, the data needs to be linked to other data about Amsterdam. This has been done by making the database into a turtle (.ttl) file and implementing in the SPARQL endpoint of the Amsterdam City Archives.

Visualizing the data

With the data in the SPARQL endpoint, we got multiple possibilities. We chose to make a visualization with a map. This was possible with the data of Adamlink, which contains streetnames and their coordinates. The visualization is a map of Amsterdam with clickable highlighted streets. When the user clicks on a street, a list of all the files of that street appears.

Conclusion

We found that one of the most effective ways of publishing the CTA archive is by making use of the streetnames mentioned in the descriptions. Since most of the files are concerning a street, search operations based on the streetnames should be the most effective. We chose to do a visualization with the city map of Amsterdam, because in our opinion this gives a good overview of the files separated over the city. However, this is not the only possibility. For example, a search engine with filters could be just as effective.

References

Adamlink, <https://adamlink.nl/>

All 9,866,539 buildings in The Netherlands project.
<http://code.waag.org/buildings/#52.3728,4.8557,13>

Block by Block, Brooklyn's Past and Present project.
<https://www.bklynr.com/block-by-block-brooklyn-s-past-and-present/>

CTA archief, Gemeente Archief Amsterdam,
<https://archief.amsterdam/inventarissen/overzicht/10057.nl.html>

Holtman, Marc (2017) Digitalisering Geïndustrialiseerd: Een Nieuwe Kijk op Uitvoering en Kwaliteitsbeheersing van Massa Digitaliseringen-projecten. Gemeente Amsterdam Stadsarchief, Amsterdam.

Ward, Matthew, Georges G. Grinstein, and Daniel Keim, Interactive Data Visualization: Foundations, Techniques, and Applications, Peters, 2010.