

My contribution to the Helsinki group in DHH16

In order to gain understanding of the success of urban renewal projects, our plan was to explore the reality of the change that took place in one such project, the impressions the people had of it, and to compare these two. The case we studied was Länsi-Pasila in the 1970s and 1980s. Our aim was to compare the subjective with the objective, the people's impressions of the change to the actual, quantified change.

Naturally we all contributed ideas to the conversation. I may have introduced the idea of restricting ourselves to a definite time period or city district, but it was still a product of the exchange of ideas. In retrospect, with the understanding of the kind of data available, it might have been a better idea first to look at the data more carefully before deciding on a topic. On one hand, this would have given us the possibility of using digital techniques in more systematic ways. On the other hand, from the perspective of science, it would have been questionable to let ourselves be led by the data instead of the research question. I still think that a good understanding of the extent and the qualities of the data available is necessary to formulate realistic research questions.

My more concrete efforts in the hackathon were directed at getting something out of Finna.fi. Through my work expertise at the National Library I already had an idea of what was possible and how it could be done. In short, what I did was to write Python scripts to make programmatic queries to `api.finna.fi` and save the responses locally. This gave us the ability to form a data set of freely usable images with reasonable metadata.

What we still needed to do was to find a way to build a useful data set, a set related to our chosen area. This was not simple, because as the metadata does not have structured location information, there was no way simply to tell the computer that we only wanted records related to Pasila. We decided that it would be enough to have a list of street names in the area, both current and historical. Using street names in addition to the name of the area gave us more confidence that we were doing everything we could to retrieve all the relevant records we possibly could.

After we had all the metadata, we tested different ways of using it. It would have been interesting to use the approximate location information we got from the street name to visualize urban changes on a map, but this proved to be too difficult or time-consuming for the scope of the hackathon.

On the second morning we decided to focus on the keywords in the metadata to get a rough idea of whether the record represented the old or the new Pasila. By dividing the keywords used in the records in two categories we were able to calculate a value on the old-new axis for each

photo. This gave us a starting point to get the computer to understand one subjective view of the change, the documentary photographer's vision of it. We ran out of time so we were unable to explore the idea further, but the principle could be used in different ways. With more specific and carefully designed axes, the results could be quite interesting.

There are obvious caveats to using museum photographs in this way. First of all metadata has another subjective perspective in addition to that in the data itself: that of the person writing the metadata. Also, the photographer's vision is bound by the requirements of the job, and it is not simple to separate these two elements. In my mind, the most immediate scientifically fruitful approach would be to focus on exactly this: what patterns can we find in the work of museum photographers and how are they related to what is actually happening in the city? After we have an understanding of this, we know more precisely what kind of filter these photographers give us to the world.

The most important thing I gained from the course was a confidence in applying programmatic techniques to humanities materials. I had felt that the effort required to master the techniques was not justified by the advantages they provided. But the hackathon gave me confidence that the effort is not that significant and that the advantages are potentially immense. Even if my own field, philosophy, requires going far beyond what even the most sophisticated programmatic techniques can currently offer, I am sure I can benefit from what such techniques can extract from the texts I study. In particular, I am looking to learn more about named entity recognition and GIS.

Materials

- The [code on GitHub](#). The code is not pretty because it has not been designed as an application but a set of scripts that give a certain result. `FinnaRecords.py` gets records from `api.finna.fi` and `street-record.py` uses it to do queries based on a list of street names and store them as JSON.
- The [presentation](#).
- The [poster](#).