

Location Extraction Tool

Rosa Merino Claros
University of Amsterdam
R.MerinoClaros@uva.nl

Alex Olieman
University of Amsterdam
olieman@uva.nl

Our aim is to build a location extraction tool for text files. The main idea of this engine is the following: to locate within a text file the words that point to a location - a street, a city, a region or a country - and create as output a map with all the locations found in the text. This kind of tool has many possible interesting applications in the digital humanities: for journalists and social science researchers exploring newspapers, for historians working with historical documents, for the analysis of biographies, etc. But the main advantage of such an engine consists of its potential as visualisation tool: at a glance the researcher gains an idea of the spatial interaction described in the text file.

Location extraction encompasses much more than just a lookup of every single word of a text in a gazetteer. The result of the place name lookup –geoparsing– then needs to be disambiguated. The main difficulty is to disambiguate place names by distinguishing places from persons and by selecting the most likely place out of a list of homographic place names.

In order to adapt this tool to the concrete needs of the CREATE project –which offers an interdisciplinary platform for analysing and visualising the making of creative cities like Amsterdam– we have added a semantic filtering feature. In this sense our goal is also the possibility to filter out all mentions of a given city that are related to its cultural activity by spotting names associated with the creative industry e.g. cinemas, painters, theatres, museums, etc., and highlighting the locations where such entities have operated. The final output is a map marking all named locations linked to the text fragment where the creative industry mentions appear.

References

Pouliquen B., Kimler M., Steinberger R., Ignat C., Oellinger T., Fluart F., Zaghouni W., Widiger A., Forslund A., Best C. (2006) 'Geocoding multilingual texts: Recognition, disambiguation and visualisation', *In Proceedings of LREC-2006*.