

Querying and Visually Analysing Incomplete Spatio-Temporal Data: The Case of the Marsh's Library

Liliana Mamani Sanchez
School of Computer Science
University College Dublin
Belfield, Dublin 4, Ireland
liliana.mamanisanchez@ucd.ie

Michela Bertolotto
School of Computer Science
University College Dublin
Belfield, Dublin 4, Ireland
michela.bertolotto@ucd.ie

Marc Caball
School of History
University College Dublin
Belfield, Dublin 4, Ireland
marc.caball@ucd.ie

Jeffrey Cox
School of History
University College Dublin
Belfield, Dublin 4, Ireland
jeffrey.cox@ucd.ie

Abstract

We describe an open source prototype of a historical application based on records from Marsh's Library in Dublin over the period between 1826 and 1926. This system includes an empirical data base of readers and readership, and a user-friendly visualization interface which will aid the analysis of readership information by academic specialists. Both will help to provide a rich seam of evidence as to the broader evolution of Dublin's intellectual and literary history.

Keywords: Historical GIS; visualization

1 Introduction

We describe a historical geographic information system prototype which is being developed in the context of a digital humanities research project. Through an alignment of humanities research and GIS technology, it seeks to explore and reconstruct the role and scope of Marsh's Library as a knowledge node in Dublin's book and reading culture between 1826 and 1926, and as such, to transform academic and popular understanding of Dublin's cultural and literary history in the nineteenth and early twentieth.

Our aim is to build an open source historical GIS including an empirical data base of readers and readership, and a user-friendly visualization interface which will aid the analysis of readership information by academic specialists. Both will help to provide a rich seam of evidence as to the broader evolution of Dublin's intellectual and literary history. In addition, it will serve as means to make archival data and the results of an overarching analysis available to the general public.

Challenges related the digitisation of handwritten data are relevant to our research, particularly because of uncertainty, non-uniformity and incompleteness. The need for interactivity in information visualization was highlighted already over a decade ago (Badalamente and Greitzer 2005). However, not many systems for visualisation of spatio-temporal data provide a high level of interactivity that allows domain users to analyse their data. Regarding this, we are developing a system with an interactive map-based interface that allows

users (particularly domain experts) to investigate non-obvious details and relationships in the data.

Another objective of our research is to make the system easy to adapt to other archival data (e.g. data available in other libraries). Such a system will be of relevance to experts and professionals who curate and research data of a similar nature.

2 Data and methods

Marsh's Library houses around 25,000 rare books and 300 volumes of manuscripts, many dealing with the sixteenth and seventeenth centuries. It was the only public library in Dublin for the first century and a half of its existence.

The Library's archival records consist of a series of 'Visitors' Books' into which readers signed their names and addresses. For instance, the first people listed in the first Visitors' Book are Thomas Shaw of 19 Moore Street, John McCready of 43 Bride Street and Edward Burroughs of 13 Peter Street. Today, these addresses are in areas of socio-economic disadvantage. Therefore, these records provide a unique opportunity to chart the shifting socio-economic and cultural geography of Dublin. Entries in the visitors' books are not uniformly structured therefore this poses one important challenge of this dataset. For example, in some cases only the visitors' names are listed, while in other cases their address, profession as well as their referees' names is entered. Being handwritten, digitising entries presents the challenge of interpreting multiple handwriting styles. This cannot always be done accurately (if at all) and therefore, the

system needs to record the level of confidence with which data was entered. Furthermore, some historical addresses have now changed, so a mapping between new and old addresses is necessary.

In addressing these challenges, we have developed a prototype system comprising a spatial database to facilitate the development of spatial queries such as display of the number of readers by address, metropolitan district, and where known their reading material. We have deployed the following technologies: PostGIS for the database, OpenStreetMap (OSM) as the source of the most up-to-date spatial maps for the city of Dublin and Python and Javascript-based technologies for the prototype’s backend and GUI. The GUI visualization comprises spatial, temporal and thematic descriptions which are shown as an output of two kinds of querying tasks: a) one driven by the user’s specification of querying arguments, and b) another one driven by the concept of timeline.

Figure 1 shows the layout of the graphical interface of this prototype. This GUI consists of 6 components that will allow users to perform specialized searches.

These components support spatial search (component 2), temporal search (component 5), textual search by readers’ names (component 3), as well as searches that combine all criteria. The visualization component (component 1) maps readers’ addresses that match the query criteria. Interactivity is provided both at the level of textual output and visual output.

Different visualization options form part of component 4, including well-known interactive visualizations such as heatmaps, graphs, and charts. One methodology we are exploring is an extension of space-time cubes which will provide a more detailed representation of events (Bach et al, 2014) (Ding et al, 2016), for example to demonstrate popular

and non-popular areas of provenance of the Library’s readers and visitors.

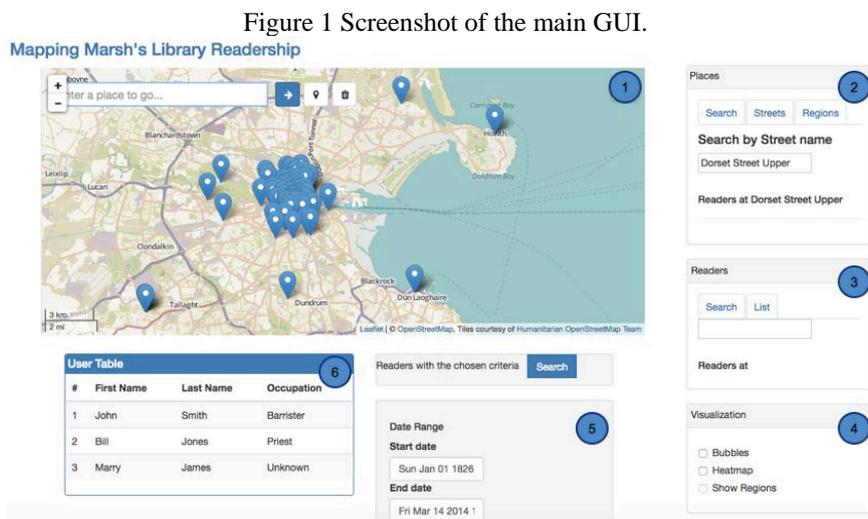
Timeline is a representation of a linear succession of events, which are represented by graphic objects along a graduated linear axis (Gautier et al, 2016). Thus, each library visit can be represented by a point. We are currently exploring the suitability of this representation depending on the frequency of visits.

3 Discussion and Future Work

Our future work will undertake the creation of advanced visualizations based on space-time cubes and timelines, and processing of complex queries. Beside challenges related to the effective representation and visualization of the spatio-temporal data, its inherent uncertainty increments the complexity of possible analyses. For instance, dates records have heterogeneous granularity: some dates were not recorded or not very detailed, this was probably due to lack of information recording standards at the time. Regarding the spatial aspect, uncertainty comes from a changing geography due to the city development. Changes in street names or disappearance of old streets affect the way addresses are represented in our database and subsequent queries. We are researching reliable methods that deal with this uncertainty as well as methods to keep track of geographic information changes as part of a historical knowledge repository.

4 Acknowledgments

This research was supported by an Irish Research Council grant to the project "Mapping readers and readership in Dublin, 1826-1926: a new cultural geography".



References

Badalamente, R.V., Greitzer, F.L., 2005. Top ten needs for intelligence analysis tool development, in: *Proceedings of the 2005 International Conference on Intelligence Analysis*.

Benjamin Bach, Emmanuel Pietriga, and Jean-Daniel Fekete. 2014. Visualizing dynamic networks with matrix cubes. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '14)*. ACM, New York, NY, USA, 877-886.

Ding, L., Krisp, J.M., Augsburg, U., Meng, L., Xiao, G., Keler, A., n.d. Visual exploration of multivariate movement events in space-time cube. In *proceedings of the 19th AGILE International Conference on Geographic Information Science-Geospatial Data in a Changing World*, 2016.

Gautier, J., Davoine, P.-A., Cuntty, C., Lyon II, I.R.G. Helical time representation to visualize return-periods of spatio-temporal events. In *proceedings of the 19th AGILE International Conference on Geographic Information Science-Geospatial Data in a Changing World*, 2016.