Chapter 5

Visualising Interdisciplinary Research: Algorithmic Treatment of Museum Case-study Information Sets

David Gauthier, Jakob Bak and Jamie Allen

Introduction

The practice and theorising of museum spaces is a rich subject for both analytic and synthesising interdisciplinary research. Interests in this area involve strands of contemporary practice-based research in cultural and social studies, and their application in media, design and other creative digital practices, including burgeoning pursuits in the digital humanities. Study of the museum space necessitates an interdisciplinary approach, with the inevitable problems of developing common languages, values (in terms of emphasis, publication outputs and communications) and knowledge resources for both collaborating researchers and external audiences of research outputs.

One way of dealing with the provision of multiple perspectives on large combined quantitative, qualitative and descriptive data sets is through computational visualisation. This interaction design research area has been one subject of the Copenhagen Institute of Interaction Design's (CIID) work within the pan-European collaborative research project 'European Museums in an Age of Migrations' (MeLa). Based on case-study research data of technology applications within museum spaces, collected during early stages of the MeLa project, the work presented in this chapter is inspired by what has been called 'distant reading' (Moretti 2000). Rather than operating on single cases, information is extracted from a set of case study descriptions and metadata to produce a diagrammatic interpretation of an entire, often large (>100 cases) data set. Using sorting and semantic algorithms to automate the reading of data, it is possible to 'focus on units that are much smaller or much larger than the text: devices, themes, tropes or genres and systems' (Moretti 2000: 57).

Metadata Visualisation

As a first visualisation, the authors looked at the geographical distribution of the institutions gathered in their survey. In total, 123 cases concerning uses of

Figure 5.1   World map of cases and fields of activity

Source: Copenhagen Institute of Interaction Design.

technology in or by museums institutions were collected and catalogued according to 41 parameters and questions (multiple choice, scales and free-text). Aligned with MeLa project's themes, the survey's focus was directed towards methods of representing and communicating cultural tropes, dialogues and narratives that had been mediated through the use of technology in different exhibition contexts. Most submitted cases originated from Europe and North America, where cities such as New York, London and Amsterdam act as poles comprising the most catalogued cases. Each institution was identified with a particular field of activity and rendered as a coloured dot. These institutions shown included the Museum of Copenhagen, the National History Museum of London and the Netherlands Architecture Institute.

The visualisations and algorithms used are devised to manipulate the case-study data, based on the input fields required for each case study. Many of these input fields, descriptors and prompts were intended to produce information most suitable for a distant-reading approach, for example, to enable cases to be indexed with a finite set of categories and/or mutually exclusive numerical inputs. Using this researcher-input metadata, analyses can more easily correlate cases and visualise relationships. However, metadata introduces various descriptive biases. Some level of taxonomies must necessarily be defined *a priori* in order to allow for the practical collection of data, for anticipated cases. Although this pre-defined taxonomy produces a common language for all researchers, each of these have their own bias – that is, how and what to categorise using the taxonomy. During our investigation, simple initial visualisations helped the iteration and revision of initially researcher-chosen taxonomies. These bias considerations led to increased use of textual input by the MeLa group of researchers (in line with more typical museum studies and other humanities-based textual analyses used to describe and analyse museum case studies) and to the use of semantic analyses as a basis for further case-study visualisations.

Latent Semantic Visualisation

In latent semantic visualisation, an approach is taken in which text is dealt with as an object-of-analysis, from which algorithms extract overall topics and themes, automatically from 'bags of words', using various connectivity, frequency and logical assumptions about language. Input fields in the survey of case studies include free-text inputs in which researchers, rather than categorising cases according to a predefined and limited typology (metadata), are asked to reflect and describe in words their view and analysis of cases.

Two specific topic modelling algorithms, Latent Dirichlet Allocation (LDA) (Blei, Ng and Jordan 2003) and Correlated Topic Models (CTM) (Blei and Lafferty 2007), were employed to extract topics from the free-text portions of the case-study questionnaire results. A single latent topic consists of a single list of words, which are (as a group) algorithmically and statistically selected as representative of
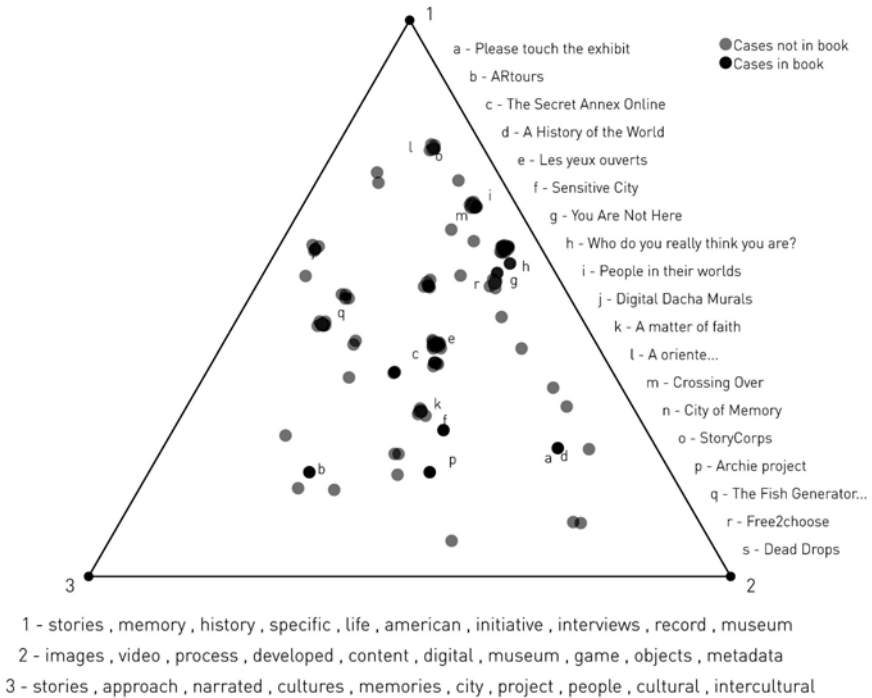
Figure 5.2　Topical triangle

Source: Copenhagen Institute of Interaction Design.

the content of the texts. Hence the meaning of a given latent topic is derived from its constituting list and interpreted by researchers after the algorithm runs. Both of the above algorithms yield a multitude of these latent topics (a triplet selection in the case of Figure 5.2) against which all documents in the case-study data set are measured automatically and given proportion as a scale from 0 to 10. We used this scale information in Figure 5.2 to place the group of case studies inside a triangle where each of the vertices represents an extracted latent topic. With this approach, correlation is assumed between the words of a text and its latent (or unexpressed) semantic themes. This method is interesting as it differs from other kinds of metadata (for example, human-generated 'tagging' of cases with keywords), as it is more suggestive than deterministic and hierarchical; topic modelling offers better dynamism and flexibility in abstracting textual case-study data.

Figure 5.2 offers a view of the case studies, spatially dispersed according to three extracted latent topics listed beneath the triangle. From the extracted list of words we can interpret the meaning of latent topics such as: (1) stories as specific historical and lived memories, (2) media and content oriented and (3) approaches in presenting stories and memories. From this diagram, we can read

that most of our cases have strong ties with topic (1), while oscillating between topic (2) and topic (3).

Conclusions

Conducting highly interdisciplinary research, such as that undertaken in the MeLa project, brings into perspective the contours and boundaries of one's own field of expertise, and what types of artefacts need to be created to help bridge these boundaries. Multiple discussions on how to approach the task of designing surveys, their generality, differences, relevance and appearance contributed to defining common objects and languages relevant to each research field under the MeLa core themes of migration, cultural identities and dialogue. In turn, the graphics and dynamic visualisations created should be regarded as computationally -derived boundary objects for interdisciplinary research of the kind being undertaken here, that is:

> … objects that both inhabit several communities of practice and satisfy the informational requirements of each of them. Boundary objects are thus both plastic enough to adapt to local needs and constraints of the several parties employing them, yet robust enough to maintain a common identity across sites. They are weakly structured in common use and become strongly structured in individual-site use. (Star and Griesemer 1989: 297)

Digital humanities techniques occupy a potentially important position in interdisciplinary and multidisciplinary contexts in that they are capable of elaborating and researching such objects (Manovich 2008). They do not claim to represent universal truths about the information at hand, but rather evoke and resonate boundaries of meaning on the intersecting borders of various research fields and communities.

Acknowledgements

References

Blei, D.M. and Lafferty, J.D. 2007. 'A correlated topic model of Science'. *Annals of Applied Statistics*, 1, 17–35.

——, Ng, A.Y. and Jordan, M.I. 2003. 'Latent Dirichlet Allocation'. *Journal of Machine Learning Research*, 3, 993–1022.

Manovich, L. 2008. 'Cultural analytics: analysis and visualization of large cultural data sets' [Online: Software Studies Initiative]. Available at http://softwarestudies.com/cultural_analytics/cultural_analytics_2008.doc [accessed 11 February 2013].

Moretti, F. 2000. 'Conjectures on World Literature'. *New Left Review*,1, 54–68.

Star, S.L. and Griesemer, J.R. 1989. 'Institutional Ecology, "Translations" and Boundary Objects: Amateurs and Professionals in Berkeley's Museum of Vertebrate Zoology, 1907–39'. *Social Studies of Science*, 19(3), 387–420.