EUROGRAPHICS Workshop on Graphics and Cultural Heritage (2014)
R. Klein and P. Santos (Editors)

Towards a query-driven heritage-data visualization tool

Lee F. Williams[1] and Jonathan C. Roberts[1]

[1]School of Computer Science, Bangor University, UK
[2]Gwynedd Archaeological Trust, UK

Abstract

Many heritage organisations hold databases that contain many millions of records. These are huge datasets that store historic records and often focus on sites and monuments; especially sites of archaeological and historic importance. This data is often placed online, but, especially for research purposes there is a need and opportunity to analyse, interact and visualize this data in new ways. In this paper we present an initial prototype to create an advanced query tool, that can represent and visualize Heritage Environment Record data. We discuss the potential of this tool over conventional methods and present future developments.

Categories and Subject Descriptors (according to ACM CCS): I.3.3 [Computer Graphics]: Picture/Image Generation—

1. Introduction

Heritage organisations store huge databases of structured and unstructured data. This information is gathered from excavations, terrestrial surveying and airborne surveying and monitoring process that have been achieved through human observation. The data gathering process has been taking place for many years, and there are many records that are stored in hard-copy form and are yet to be digitized. The digitization process is often ongoing, and mixed with newly gathered data. Therefore while the data itself may have been newly entered into the database (such as from an excavation on a hillfort site, excavated in 2014) the data itself could represent a period of time in the past.

Currently users do have at their disposal many tools. For instance, GIS systems, maps and statistical tools are used to help researchers investigate this data. In fact, the GIS tools provide a massive array of functions. But due to the diverse nature of the data, and the need to integrate many different types of data together, it can be difficult to use current tools to explore the data.

What is required are visual analytic tools that integrate this multi-modal disparate data to provide an expressive set of functions to help the users find correlations and understand the data in a new way.

Such a tool would enable users to analyse, interact and visualize this data better. Our goal is thus to provide tools to aid the user search, query and visualize the information. It may be that an expert user, with specific knowledge could do similar operations, and make different correlations and discoveries, but would take them several weeks of data processing using standard techniques; our aim is to provide something that could do such a job in a matter of minutes, thus saving time and money for the end users.

This paper presents work in progress to achieve these goals. Here we present our design ideas and introduce a prototype that is under development.

2. Related Work

Interaction is an important component of information visualization [YaKSJ07], it allows users to explore the data and discover new correlations or values within the data. In this work we utilize the principles of exploration, where users can view several different views and forms of the information. Through this multiple view approach the user can gain a better understanding of the information [Rob07].

However, many current systems, that are used by the heritage organizations, focus on a single view. Often a map visualization is used along with a query interface (e.g., using SQL) and plot the results on a map layer [MK01, DMK05]. While this is effective, it does not allow the user to explore the data, see multiple results and compare them side-by-side.

DOI: 10.2312/gch.20141324

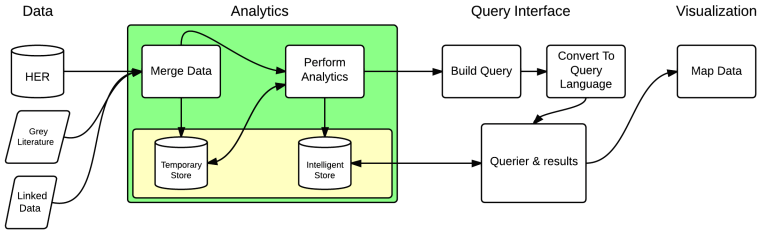

Figure 1: Schematic diagram showing the four main stages of the systems internal process, highlighted in green is the main heart of the system, notice the transition of the data from a generic DB into an intelligent data store.

3. Design

In our work, we take an ensemble approach. Where the user can integrate several constraints, over different data types. This visual analytic strategy enables the user to query and filter the data based on several different types of criteria.

We separate the design into four parts (see figure 1): (1) Data storage, that contains the Heritage Environment Record data (HER) and links to other datasets. (2) Analytic processes, (3) Query interface and finally (4) data visualization.

The raw, core data is stored in multiple different locations, the main bulk of the data comes in the form of a MySQL Database, we also have associated to the HER grey literature files that correspond to records within the HER which we apply OCR operations on them. Linked data comes in the form of third part API data which is linked to the HER data at runtime.

The analytics phase of the system is the heart of the system, this is where all the knowledge is applied to the data and we perform complex operations on the data to give us a deeper insight into the data.

Our query interface allows the user to build up a query in plain English then turns it into a query that is understandable to the data store.The interface is based on the five w's, Who, What, When, Where and Why. Currently only three of the five elements have been implemented. These elements include a 'When', 'What' and 'Where' search.

Currently our prototype provides a visual search environment which allows users to search for items that 'look like' something. Figure 2 shows how results matching the visual query are displayed on the map.

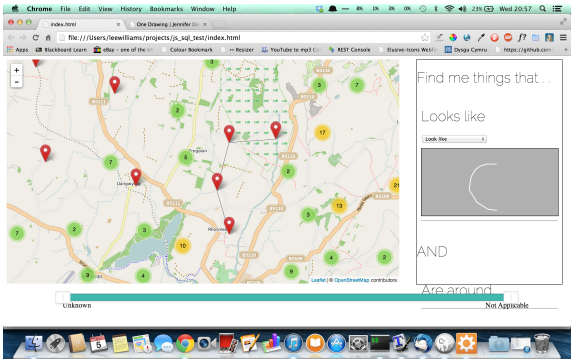

Figure 2: Figure showing the prototype visual query search. The drawn arc shape, drawn on the right, has selected four archaeology finds in the same shape.

4. Future Developments & Conclusions

The work to produce an interactive visual analytics system for the HER dataset is ongoing. The prototype system currently enable users to find different patterns and visualize the results.

Our query driven visualization method enables the user to explore the data in a natural and comprehensive way. The work is ongoing, with the domain scientists and computer scientists collaborating.

This work impacts the users of archaeological data in several ways, by providing a more expressive interface the users are able to understand their data better, and to hopefully discover new patterns.

There is much work to be achieved, but our prototype already demonstrates a way to find results that are 'similar' to others, and enables a more natural query interface for the heritage organizations.

References

[DMK05] DYKES J., MACEACHREN A. M., KRAAK M.-J.: *Exploring geovisualization*. Elsevier, 2005.

[MK01] MACEACHREN A. M., KRAAK M.-J.: Research challenges in geovisualization. *Cartography and Geographic Information Science 28*, 1 (2001), 3–12.

[Rob07] ROBERTS J. C.: State of the art: Coordinated & multiple views in exploratory visualization. In *Coordinated and Multiple Views in Exploratory Visualization, 2007. CMV'07. Fifth International Conference on* (2007), IEEE, pp. 61–71.

[YaKSJ07] YI J. S., AH KANG Y., STASKO J. T., JACKO J. A.: Toward a deeper understanding of the role of interaction in information visualization. *Visualization and Computer Graphics, IEEE Transactions on 13*, 6 (2007), 1224–1231.