Development of Parallel Coordinates Using D3 for Hybrid Reality Environments:

Visualising Acute Lymphoblastic Leukaemia Data

Author: Samantha Crome

Date of Submission: 11 December 2015

Keywords: Parallel Coordinates, Cancer, Biomedical Data

Abstract

Visualisation is the process of creating a visual representation of data. Visualisation is a tool that assists the analysis of data, the finding of patterns and meaning, the exploration of relationships and the development of hypotheses. Increasing amounts of medical data are being captured systematically as part of routine clinical care. In addition, medical research is identifying new biomarkers of disease diagnosis, prognosis and response to treatment such as gene expression profiles. Visualisation of biomedical data can assist in finding patterns in a particular disease to optimise treatment and prognosis. This is important, for example, for cancer patients, in which the opportunity of personalised treatment can make a large difference in the probability of a cure or remission. The primary objective of this research is to develop code to extend the 2D visualisation functionality of a newly built Data Arena (a CAVE-like Hybrid Reality Environment) by enabling a parallel co-ordinates data visualisation tool. This data visualisation tool will be used to evaluate multivariate high dimensional biomedical data of children with acute lymphoblastic leukaemia, with the aim of developing better predictive risk stratification models to guide therapeutic decisions, which in turn, may reduce comorbidity and improve survival.

Visualisation is an established exploratory tool for use with multivariate high dimensional data. It could be used in computer generated virtual environments. Virtual environments are three-dimensional computer generated environments that simulate sensory information. A recent technology is Hybrid Reality Environments (HREs) which combine the immersion of CAVE with the high resolution of large ultra-resolution displays. HREs are promising for the analysis of complex data. The University of Technology Sydney has built a CAVE-like HRE called the Data Arena. There are many visualisation libraries that could be integrated into the Data Arena for data visualisation tools. D3, a JavaScript library, has several components which improve interoperability, documentation, expressiveness, compatibility, and performance. D3 can also create Scalable Vector Graphics (SVG) images, which is useful when performing interactions of the images and rendering. Of the many multivariate data visualisations tools that could be used to explore the relationships within the acute lymphoblastic leukaemia data using D3 in the Data Arena, parallel coordinates has been shown to be useful because of its ability to find patterns and exceptions.

There are two aims for this thesis. The first aim is to demonstrate that parallel coordinates visualisation of a paediatric acute lymphoblastic leukaemia dataset can be implemented in the Data Arena. A secondary, minor aim is to be able to implement other 2D multidimensional

visualisations in the Data Arena. Through iterative development and evaluation two potential solutions emerged. The first solution is to use the WebView Omegalib module and use it to view a webpage containing a D3 parallel coordinates visualisation. Omegalib is the application framework used in the Data Arena. This solution has the benefit of having easy interaction, but is heavily dependent on the progress of the module for increase in quality. The second proposed solution is a work in progress proof of concept. It uses an XML parser to parse an SVG created using the D3 library that is first extracted by the user for a webpage using D3. By analysing the parsed elements, the elements are translated into an Omegalib Cyclops module objects.

An evaluation of these two solutions demonstrates that the first aim of the thesis is successful, however the second aim is not fulfilled. The parallel coordinates visualisations of acute lymphoblastic leukaemia data perform well in the Data Arena passing all test cases. The WebView solution succeeds in adapting some other types of 2D visualisations from external sources, but fails to visualise others. Additionally, as the SVG Parser solution is in the proof of concept stage there is still development needed to make it truly adaptable, but the potential is there. Overall, the thesis benefited the acute lymphoblastic leukaemia project by demonstrating that acute lymphoblastic leukaemia data can be visualised using parallel coordinates in the Data Arena with high quality and interactivity. Additionally, it assisted Data Arena developers because the process of using Omegalib and communicating with the developers gave feedback and identified some difficulties with Omegalib.

TABLE OF CONTENTS

Table of Figures

Table of Tables