

**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

ABSTRACT	2
1. INTRODUCTION	8
1.1 BACKGROUND	8
1.2 NEW CONTRIBUTION	10
1.3 STAKEHOLDERS	11
1.4 AIMS	11
1.5 OBJECTIVES	11
1.7 SIGNIFICANCE	11
1.8 OUTLINE OF RESEARCH METHODS	12
1.9 ORGANISATION OF CONTENT	13
2. LITERATURE REVIEW	15
2.1 ACUTE LYMPHOBLASTIC LEUKAEMIA	15
2.2 HIGH-DIMENSIONAL LOW OBSERVATION DATA	17
2.3 DATA VISUALISATION	18
2.4 CAVE AUTOMATIC VIRTUAL ENVIRONMENT	19
2.5 HYBRID REALITY ENVIRONMENTS	22
2.6 VISUALISATION LIBRARIES	23
2.7 PARALLEL COORDINATES	25
2.8 RESEARCH GAPS	28
3. METHODS	29
3.1 DEVELOPMENT AND EVALUATION METHODS	29
3.1.1 DEVELOPMENT METHOD: FEATURE AND TIME DRIVEN METHODOLOGY	29
3.1.2 EVALUATION METHOD	31
3.2 TOOLS	31
3.2.1 INSTALLATION OF OMEGALIB	32
3.2.2 VIRTUAL BOX	33
3.2.3 REQUIRED LIBRARIES	35
3.2.4 ACUTE LYMPHOBLASTIC LEUKAEMIA DATA AND OTHER DATA USED IN THE DEVELOPMENT PROCESS	36
3.3 FAILED SOLUTIONS	36
3.3.1 QT AND OMEGALIB SOLUTION	37
3.3.2 OSG SVG READER SOLUTION	37
3.4 WEBVIEW SOLUTION	38

3.5 SVG PARSER SOLUTION: A WORK IN PROGRESS	39
3.6 TEST PLAN	40
3.6.1 FEATURES TO TEST	40
3.6.2 TEST CASES	41
4. RESULTS	46
4.1 WEBVIEW SOLUTION	46
4.1.1 PARALLEL COORDINATES IN THE DATA ARENA	46
4.1.2 TEST REPORT	47
4.2 THE SVG PARSER SOLUTION: A WORK IN PROGRESS	65
4.2.1 PARALLEL COORDINATES IN THE DATA ARENA	65
4.2.2 TEST REPORT	66
5. CONCLUSION	69
5.1 FUTURE WORK	71
6. REFERENCES	73
7. APPENDIX	80
7.1 SOURCE CODE	80
7.1.1 WEBVIEW SOLUTION	80
7.1.2 SVG PARSER SOLUTION	85
7.2 EXAMPLE EMPIRICAL METHOD: EVALUATING THE IMPLEMENTED PARALLEL CO- ORDINATES IN A DATA ARENA WITH PARALLEL COORDINATES ON A PC USING A FACTORIAL DESIGN	114

Table of Figures

FIGURE A – A PARALLEL COORDINATE GRAPH SURROUNDED BY INTERACTIVE SOFTWARE AND DIAGRAMMATIC CUES (SIIRTOLA ET AL. 2009).....	26
FIGURE B - PROGRESSION THROUGH FEATURE EVALUATION	30
FIGURE C - UBUNTU VIRTUAL BOX SETTINGS	34
FIGURE D - GENTOO VIRTUAL BOX SETTINGS	35
FIGURE E - PARALLEL COORDINATES WEBPAGE IN THE DATA ARENA	47
FIGURE F - CLOSE-UP OF PARALLEL COORDINATES WEBPAGE IN THE DATA ARENA.....	47
FIGURE G - MULTIPLE SELECTED AXES IN THE DATA ARENA.....	49
FIGURE H - FORCE DIRECTED NETWORK DIAGRAM IN THE DATA ARENA	51
FIGURE I – ‘WHO DO THEY SERVE’ WEBSITE IN THE DATA ARENA	52
FIGURE J - TREE DIAGRAM IN THE DATA ARENA WITH A NODE SELECTED FOR EXPANSION.....	53
FIGURE K - UK TEMPERATURE 1910-2014 SCATTERPLOT	54
FIGURE L - SCATTERPLOT MATRIX IN THE DATA ARENA WITH SMALL BRUSHED AREA.....	55
FIGURE M - LES MISÉRABLES CO-OCCURRENCE HEATMAP IN THE DATA ARENA.....	56
FIGURE N - DAYS-HOURS HEATMAP IN THE DATA ARENA	57
FIGURE O - CHERNOFF FACES IN THE DATA ARENA.....	58
FIGURE P - SMALL MULTIPLES IN THE DATA ARENA	59
FIGURE Q - INTERACTIVE SMALL MULTIPLES IN THE DATA ARENA.....	60
FIGURE R - STAR PLOTS IN THE DATA ARENA.....	61
FIGURE S - DASHBOARD IN THE DATA ARENA	62
FIGURE T - CROSSFILTER IN THE DATA ARENA.....	63
FIGURE U - BASIC PARALLEL COORDINATES WITH TRANSPARENT BACKGROUND.....	64
FIGURE V - BASIC PARALLEL COORDINATES WITH WHITE BACKGROUND.....	64
FIGURE W - SVG PARSED PARALLEL COORDINATES WITH 3D GLASSES	65
FIGURE X - SVG PARSED PARALLEL COORDINATES WITHOUT 3D GLASSES.....	66
FIGURE Y - CLOSE-UP OF SVG PARSED PARALLEL COORDINATES	66
FIGURE Z - STUDY DESIGN FOR INDIVIDUAL TESTING	115
FIGURE AA - STUDY DESIGN FOR GROUP TESTING.....	116

Table of Tables

TABLE 1 - TEST CASE TEMPLATE	31
TABLE 2 - TEST CASE 01	41
TABLE 3 - TEST CASE 02	41
TABLE 4 - TEST CASE 03	41
TABLE 5 - TEST CASE 04	42
TABLE 6 - TEST CASE 05	42
TABLE 7 - TEST CASE 06	43
TABLE 8 - TEST CASE 07	45
TABLE 9 - TEST CASE 08	45
TABLE 10 - TEST CASE 09	45
TABLE 11 - TEST CASE 01 RESULTS	47
TABLE 12 - TEST CASE 02 RESULTS	48
TABLE 13 - TEST CASE 03 RESULTS	49
TABLE 14 - TEST CASE 04 RESULTS	49
TABLE 15 - TEST CASE 05 RESULTS	50
TABLE 16 - TEST CASE 06 RESULTS	51
TABLE 17 - TEST CASE 07 RESULTS	67
TABLE 18 - TEST CASE 08 RESULTS	67
TABLE 19 - TEST CASE 09 RESULTS	67