Journal of Visual Languages and Computing 31 (2015) 87–103

Contents lists available at ScienceDirect

Journal of Visual Languages and Computing

journal homepage: www.elsevier.com/locate/jvlc

A map-like visualisation method based on liquid modelling

Robert P. Biuk-Aghai [a,*], Muye Yang [a], Patrick Cheong-Iao Pang [b], Wai Hou Ao [a], Simon Fong [a], Yain-Whar Si [a]

[a] Department of Computer and Information Science, Faculty of Science and Technology, University of Macau, Avenida da Universidade, Taipa, Macau S.A.R., China
[b] Department of Computing and Information Systems, The University of Melbourne, Victoria 3010, Australia

ARTICLE INFO

Article history:
Received 7 July 2015
Accepted 5 October 2015
This paper has been recommended for acceptance by Henry Duh
Available online 19 October 2015

Keywords:
Information visualisation
Polygon expansion
Liquid modelling
Map
Wikipedia
Category hierarchy

ABSTRACT

Many applications produce large amounts of data, and information visualisation has been successfully applied to help make sense of this data. Recently geographic maps have been used as a metaphor for visualisation, given that most people are familiar with reading maps, and several visualisation methods based on this metaphor have been developed. In this paper we present a new visualisation method that aims to improve on existing map-like visualisations. It is based on the metaphor of liquids poured onto a surface that expand outwards until they touch each other, forming larger areas. We present the design of our visualisation method and evaluations we have carried out to compare it with an existing visualisation. Our new visualisation has better usability, leading to higher accuracy and greater speed of task performance, as well as a lower error rate.

1. Introduction

Today's internet-scale applications produce huge amounts of data. Applications in the domains of social networking, collaborative filtering, online discussion, user-contributed content and others have up to hundreds of millions of users generating large amounts of data on an ongoing basis. There is much value to be gained in understanding this data and perceiving patterns in it. However, given the amount of data, doing so remains a challenge. Information visualisation can be employed to visually represent large amounts of data in a form that makes patterns and associations in the data appear as visual patterns, which aids assimilation of data [1].

Our research is framed within this context of perceiving patterns in large-scale user-contributed content. More specifically, we focus on visualising large hierarchies, namely the category data of Wikipedia, "the free encyclopedia that anyone can edit" (Wikipedia's own slogan). Since it was created in 2001 Wikipedia has experienced tremendous growth. Today (March 2015) there are over 4.8 million articles, over 24 million registered users, and over 760 million edits in the English Wikipedia alone,[1] and there are more than 280 other language editions. Moreover, the English Wikipedia keeps growing at a rate of around 800–900 new articles every day, and is edited about 3 million times per month.[2]

* Corresponding author.
E-mail addresses: robertba@gmail.com (R.P. Biuk-Aghai), muye.yang@gmail.com (M. Yang), cipang@unimelb.edu.au (P.-I. Pang), aowaihou@hotmail.com (W.H. Ao), ccfong@umac.mo (S. Fong), fstasp@umac.mo (Y.-W. Si).

[1] http://en.wikipedia.org/wiki/Special:Statistics.
[2] http://stats.wikimedia.org/EN/TablesWikipediaEN.htm.

http://dx.doi.org/10.1016/j.jvlc.2015.10.003
1045-926X/

Wikipedia articles are classified into categories. In English Wikipedia alone there are about 2 million categories, and it is common for an article to be assigned to multiple categories, in some cases dozens of categories. The classification of an article to its categories is displayed on each article page, but it is difficult to obtain a larger picture of the distribution of articles among categories. Such an overview would be useful to gain an understanding of the distribution of content among topic areas, which reflects their relative importance within the community of their contributors.

Previously we had devised a method for visualising the Wikipedia category hierarchy in the form of a geographical map, using an approach of tiling a plane of hexagons [2]. Experience with using that visualisation led us to identify several issues related to the readability and visual quality of the map, thus we set out to devise a new method that would produce a more readable visualisation of better visual quality. An early design of this new method, which places areas in the map by expanding polygons instead of tiling hexagons, was documented in a previous paper [3]. Our early design, however, also had some shortcomings, specifically it could not guarantee that areas in the map are displayed in the correct size since areas surrounded on all sides by other areas are prevented from expanding to their intended size. Since then we have significantly revised the design of our new visualisation method, solving its earlier shortcomings and overcoming the weaknesses of our previous hexagon-tiling method. In this paper we present the design and evaluation of our new method.

The remainder of this paper is organized as follows. Section 2 briefly reviews related work on map-like visualisation. Then we introduce the design of our visualisation method in Section 3, and present an evaluation of our visualisation in Section 4. Finally, we make conclusions in Section 5.

2. Related work

Structure and composition of large hierarchies are difficult to perceive, and various traditional visualisation techniques have been developed to try to visualise such data, including tree-maps [4], voronoi treemaps [5], mosaic plots [6], cone trees [7] and hyperbolic trees [8]. Tree-maps make efficient use of display space, but despite their name they do not actually look like (geographic) maps. Voronoi treemaps are an interesting variation of the traditional treemap, and can assume any shape, not only rectangular shapes, but have a very artificial visual appearance. Mosaic plots somewhat resemble tree-maps in that they have an overall rectangular shape, and rectangular sub-divisions. Cone trees and hyperbolic trees, on the other hand, are some of the techniques used to represent the tree structure of hierarchical data visually as nodes and edges, and as such make no attempt to resemble a map. In sum, these traditional techniques all visualise hierarchical data using a more or less abstract representation.

A more recent approach [2,9–12] is to visualise hierarchical and other relational data in the form of a geographic map. Maps consist of land and sea, of countries, provinces and counties that are separated by borders and that are labelled with their name. Maps are usually also coloured to represent a certain aspect of the terrain, such as elevation in the case of topographic maps, or another attribute such as economic output, unemployment rate, language spoken or others in the case of a thematic map.

The map form can be used to represent non-geographic data as well. Clusters of data can be represented as regions in the map, and attributes of this data can be mapped to visual attributes of map objects. The use of the map form has the significant advantage that most people understand geographic maps easily thanks to early exposure to maps in school. Even among pre-school children essential mapping abilities are well developed [13]. Elements such as mountains, valleys, land, sea, rivers, and cities, as well as the meaning of each, are readily recognized by people even without special training. Therefore visualising information structures in the form of a geographic map enables people to relate to such representations more easily without requiring prior instruction.

Skupin has used the map form to represent the geographic knowledge domain [9]. Taking 2200 conference abstracts, he applied the self-organizing map (SOM) method to produce a map-like representation of the topics contained in this document collection. Topics were clustered hierarchically, and this hierarchy was mapped to areas in the map. The final map output and labelling were produced by a GIS (geographic information system) software.

Hu et al. have devised another method for representing dynamic relational data in map form [10]. The approach, called GMap, transforms an arbitrary graph into a map representation. It represents clusters of nodes as areas in the map, and displays most areas fused together into one large continent, with only a few small separate islands, if any.

Gronemann and Jünger have visualised clustered graphs as topographic maps [11]. The input is a graph with an overlaid tree, expressing both relationships and hierarchies among nodes. This graph data is preprocessed to determine node placement, from which a triangle mesh is generated that assigns an elevation to each node. The result data is fed into a GIS software that outputs a map showing islands of areas whose elevation is above sea level, whereas other areas that have a negative elevation appear as submerged below the sea.

Auber et al. have proposed another method that produces visualisations resembling a geographic map, which they term GosperMap, as it relies on the use of a Gosper curve [12]. The input data has a tree structure, the leaves of which are projected to a 2D space-filling curve. Finally, regions containing nodes are filled to become areas corresponding to sub-trees in the input tree. The resulting maps consist of a single continent with countries that have highly irregular borders.

Biuk–Aghai and Pang have devised a different method to create map-like visualisations [2]. Their method takes a tree of hierarchical data and constructs a map that represents nodes at different levels as nested areas, reflecting the hierarchy. Areas are constructed by tiling hexagons in a hexagon matrix, starting from the area's centre and moving outwards in random order.

Examples of each of these five map-like visualisations are shown in Fig. 1. Our new visualisation method