Visual Informatics 2 (2018) 50–59

Contents lists available at ScienceDirect

Visual Informatics

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

Metro-Wordle: An Interactive Visualization for Urban Text Distributions Based on Wordle

Chenlu Li[a], Xiaoju Dong[a,*], Xiaoru Yuan[b]

[a] Department of Computer Science and Engineering, Shanghai Jiao Tong University, 800 Dongchuan Road, Shanghai, P.R. China, 200240
[b] Key Laboratory of Machine Perception (Ministry of Education), and School of EECS, Peking University, Beijing, P.R. China, 100871

ARTICLE INFO

Article history:
Received 11 December 2017
Received in final form 23 February 2018
Accepted 12 March 2018

Keywords: Text visualization, Location retrieval, Urban data, Metro map, Word cloud

ABSTRACT

With the development of cities and the explosion of information, vast amounts of geo-tagged textural data about Points of Interests (POIs) have been generated. Extracting useful information and discovering text spatial distributions from the data are challenging and meaningful. Also, the huge numbers of POIs in modern cities make it important to have efficient approaches to retrieve and choose a destination. This paper provides a visual design combing metro map and wordles to meet the needs. In this visualization, metro lines serve as the divider lines splitting the city into several subareas and the boundaries to constrain wordles within each subarea. The wordles are generated from keywords extracted from the text about POIs (including reviews, descriptions, etc.) and embedded into the subareas based on their geographical locations. By generating intuitive results and providing an interactive visualization to support exploring text distribution patterns, our strategy can guide the users to explore urban spatial characteristics and retrieve a location efficiently. Finally, we implement a visual analysis of the restaurants data in Shanghai, China as a case study to evaluate our strategy.

1. Introduction

Rapid urban growth enriches people's life and gives people more choices about where to go, but it also becomes more difficult for people to select one from numerous POIs. At present, many websites provide discussions, reviews and descriptions on POIs to help users make a choice, therefore a large number of textual data has been generated. However, people can be really at lost in confront of the huge data. It is in great need efficiently finding the wanted location. For the text data on POIs has a close connection with locations, it is meaningful to visualize the texts combining with their geographic information. In addition, metro has been a popular urban transportation choice, and subway lines are like the skeleton of a city, so we use the metro lines to separate urban space into several areas, then we study how the texts distributed across the areas and try to find their spatial characteristics.

Traditional methods to observe and retrieve locations include maps and search engines. As for maps, designers have made many kinds of colorful maps for tourists to explore the city, such as chinamaps.org and vidiani.com. These maps are well designed and attractive, but are mostly static with a small amount of information and take a lot of manpower. Some online web mapping services like Google Map support interactive searching, however, they are lack of detailed descriptions. As for search engines, the results obtained by queries are usually listed as text documents which are short of geo information. It's difficult to compare a large number of POIs using the textual and geo information at the same time through the existed methods. At present, there is much research aiming at analyzing geo-tagged data of POIs. Lee et al. (2011), Yin et al. (2011) and Xu et al. (2017) analyze text data combining their locations to discover hidden spatial patterns, however, in these works, the text view and map view are arranged side by side which adverses to observe their relations directly. There are also some

*Corresponding author: Xiaoju Dong (tel: +86-21-34205060; fax: +86-21-34204728;)
e-mail: chenluuli@gmail.com (Chenlu Li), xjdong@sjtu.edu.cn (Xiaoju Dong), xiaoru.yuan@pku.edu.cn (Xiaoru Yuan)

https://doi.org/10.1016/j.visinf.2018.04.006

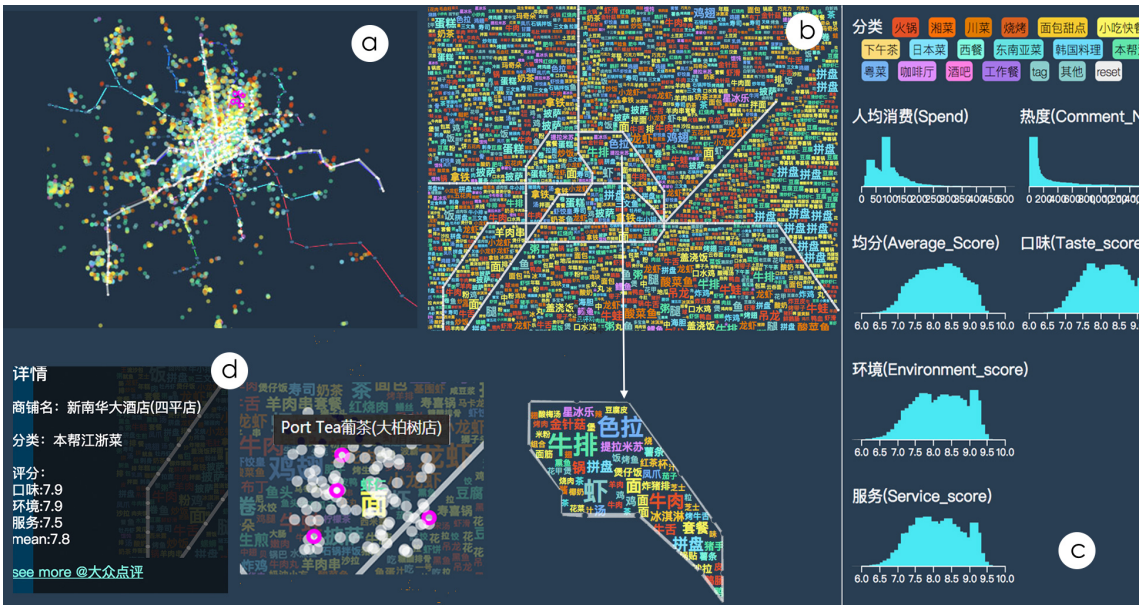

Fig. 1. System overview. (a) Geographic View. Each point represents a POI and is placed according to its location. (b) Metro-Wordle View. Wordles contain the keywords in the subarea constrained by the metro line segments. (c) Filter and Statistics View. Statistics data of POIs and filters are provided in this view. (d) POIs Detail View.

works aiming to generate a more intuitive overview of how semantic information distribute across space. Wu et al. (2013), Afzal et al. (2012), BRA and Meitzler are some examples of visualizations that merge text and spatial data into one graph. They considered aesthetics and user-friendliness, but they are more suitable for displaying rather than exploring data because of their small amount of information and insufficient interactions.

As a result, we want to design a visualization that achieves the following objects:

- All POIs are accessible through keywords. In the main view, we only provide a certain number of keywords due to the space limitation, but via these words and with the help of the filters, users can retrieve any POIs in the city.
- Textural and geographical information can be observed in one visual representation. So that their spatial relations and characteristics can be discovered.
- Multiple dimensions data of POIs are presented to give users comprehensive information.

To achieve the goals we proposed above, this paper introduces a visual approach to integrate text and metro map. First, we used the semantic analysis of POIs records and obtained some keywords after extracting and clustering the text documents. Then, we automatically calculate the layout of metro map based on Mixed-Integer Programming (MIP) algorithm. After getting the keywords of each area, we visualized the words as wordles and embedded them into the metro map using a greedy algorithm. Finally, we design a user interface to support rich interactions. We use the metro-wordle graph as the main view with several other visual models and a variety of operations for users to explore in the system.

The main contributions of this article are as follows:

- We introduce and implement a new visual strategy to combine text and map. We use the metro lines as the boundaries for wordles, and in each region, we place each word considering its actual geographic location. This method supports efficiently discovering the pattern of text spatial distributions.
- We propose new requirements for automatic drawing metro map. We take the amount of infomation in each region into consideration when schematizing a metro map.
- We present an interactive visualization to support urban geo-tagged text analysis. This work helps people retrieve and select POIs from large amount data efficiently. It also allows users to explore the patterns of urban topic spatial distributions.

The sections are organized as follows. This paper discusses related works on the automatic layout technique for metro maps, wordle embedding, textual and urban data visualization in Section 2. Section 3 gives an overview of this work. Section 4 explains the methods of data analysis including words extraction methods, metro map layout algorithm and wordle embedding. Section 5 presents the visual design of the system. Then we use the metro-wordle map for restaurants in Shanghai as a case study in Section 6. In the end, we discuss and conclude our work.

2. Related Work

2.1. Wordle

Numerous text-based visualization methods have been proposed. We choose wordle as the main method to visualize text in our work. Wordle was launched in 2009 (Viégas et al.