TOBIE - Web User Interface

Zhe Wang, Zhiyan Lu, Akmal Khikmatullaev

University of Bonn
s6wazhee,s6znluuu,s6akkhik@uni-bonn.de

Abstract. At present, it has to admit that the employment situations are becoming more and more serious. Therefore, it requires the data visualizations to provide the comprehensive information for job seekers. At the same time, the Internet is playing an increasingly important role in everyone's daily life. So it comes that the online information about job demand is available massively, and it will be convenient for job seekers online.
We focus on visualization technique to visualize the sophisticated data. Our project involves building a Trend Analysis Ontology-based Information Extraction (TOBIE) System, discovering how to use the TOBIE system as raw data for finding changes and detecting trends on the skills to trend analysis of skills. The TOBIE system is committed to creating interactive graphs which contain all information and visualizing trend results on web-based user interactive interface. The contribution of our project improves employment problem by visualizing result more intuitively for users.

Keywords: Employment Problem, Data visualization, Trend Analysis

1 Introduction

A current simple fact is that more young people including fresh graduates are meeting problems when finding new jobs. For example, job seekers can't understand the specific content of jobs very well, or the related information about new jobs is out of date. At the same time, the Internet plays an increasingly important role in everyone's daily lives, which certainly including the role of information sharing. As a consequence, online information about job demand is massively available on the Internet.

For the employment information, comprehensiveness and validity are two most important aspects [1]. The comprehensiveness of information ensures that job seekers are able to obtain a large amount of employment information which helps them find a new job. The validity of information ensures that employment information on websites is always the latest. However, a large part of websites which provide the employment information performs not well enough on these two aspects. Some of them show the graphs which contain limited information and can't be changed by visitors, and some of them show graphs which are out of update. The aim of our project is to use data visualization technique to improve user experience on both two aspects.

Data visualization technique is a perfect tool to visualize sophisticated data [2] [3]. Our project involves building a Trend Analysis Ontology-based Information Extraction (TOBIE) System, discovering how to use TOBIE system as raw data for finding changes and detecting trends on skills to trend analysis of skills. The TOBIE system relies on a vocabulary to guide the information extraction process and employs the co-word analysis to dynamically map skills' demands and structures. Finally, the TOBIE system helps to create interactive graphs which contain all information about skills on the web-based interactive interface to visualize results.

The most important feature of our project is interaction. For a user who is visiting our website, it is free to see anything that he is really interested. He can not only choose different time periods, but also set different variables. It is no exaggeration to say that every user can customize his own graph by himself! What's more, when a user is selecting nodes or edges on graphs, it will display detail information or zoom into another related graph. Therefore, the comprehensiveness of information is guaranteed. Meanwhile, there is always a text-label showing the current time and country of graphs to remind users. Moreover, administers can upload csv files which contain latest employment information to the database. Thus, it can guarantee the validity of information.

The contribution of our project are two-folders: (1) Use TOBIE system as data visualization technique for analyzing; (2) Create interaction graphs on web-based interactive interface to visualize results.

Based on our project, users can see the entire and latest trends of the skills on our website directly, and it will help users to select the most useful and popular skills during their job seeking. We are sure that our project must improves employment problem by visualizing the trends of skills for users in a more intuitive way.

The remainder of this paper is divided into five sections. In section 2, we introduce concepts of background which are required to understand the problems in our project. In section 3, we present existing approaches that address the same situation. In section 4, we define the problem addressed in our project, the solution to problems and the implementation of proposed solutions. The results and use cases of our proposed solution which can be applied in details are described in section 5. Finally, section 6 summarizes conclusions of our work with an outlook. We also explain the limitations of current solution and how it can be extended in the future.

2 Background

2.1 Data Visualization

Data visualization has been viewed by many disciplines as a modern equivalent of visual communication. In some fields, it means "information that has been abstracted in some schematic form, including attributes or variables for the units of information". Data visualization is a study of the visual representation of

data. The visual representation of the data is defined as a kind of summary which summarizes the form of the information extraction [4].

The most important goal of data visualization is to communicate the information about the data clearly and efficiently by encoding it as the visual objects (e.g., points, lines or bars) contained in graphics to users. It is one of the most important steps in data analysis or data science. Data visualization is closely related to information graphics, information visualization, scientific visualization, and statistical graphics.

2.2 Trend Analysis

Trend Analysis is a method of using the time series data (Ex. The information in sequence over time) to analyze the involving comparison of the same item over a significantly long period. Trend Analysis is aimed at projecting both the current and the future movement of events through the use of time series data analysis which involves comparison of data over a sequential period of time to spot a pattern or trend. In one word, the trend analysis is taking the past data and using it to project the future results [5].

Trend analysis involves collection the information from multiple time periods and drawing the information on the horizon for the further review. The purpose of this analysis is to find an operational model in the information which is provided. When the trend analysis is used to predict the future, keep in mind that factors which previously affected data points may not be done under the same situation. This means that the inference of the historical time series does not necessarily produce an effective prediction for the future.

2.3 Ontology-Based Information Extraction (OBIE)

Information extraction (IE) is designed to automatically retrieve the natural language text to retrieve the certain types of information. Ontology-based information extraction (OBIE) has recently become a sub-domain of the information extraction. The general idea behind Information Extraction is automatically retrieving certain types of information from the natural language text. it aims to process natural language text and to retrieve the occurrences of a particular class of objects or events [6] [7].

In the field of ontology-based information extraction, the information extraction process always uses the ontology, and the output of the process is usually rendered through the ontology. It should be noted that the ontology is defined as an explicit formal and specification of the shared conceptualization. Since the information extraction essentially involves the task of retrieving the information of a particular domain, it is helpful to formally and explicitly specify the concept of the domain through the ontology itself. In general, this area will grow significantly in the future.

3 Related Work

Data visualization is the study of the visual representation of data, meaning "information which has been abstracted in some schematic form, including attributes or variables for the units of information". The main goal of data visualization is to convey information clearly and efficiently with the help of graphic technologies.

Our information age more often feels like an era of information overload. Excess amounts of information are overwhelming; raw data becomes useful only when we apply methods of deriving insight from it [8]. As we humans can easily detect the patterns among charts rather than a series of raw data, for this reason, data visualization becomes a powerful tool to communicate between different information systems.

A number of visualization techniques have been developed in last few years due to representation of huge massive information and to analyze it. They are: table, line chart, area chart, column and bar chart, pie chart, scatter and bubble chart, dynamic chart, 3D chart, etc [9]. Besides, the developers could add many features, such as usability, interactivity, which make these methods easy to use and easy to understand. In addition, there are some data visualization methods, compared to above methods, which are less known. Force-directed graph is one of these methods.

Force-directed graph drawing is used to assign a set of nodes, with a set of edges, which represent the links between above nodes. It positions the nodes in two-dimensional space, and put all the edges to connect the nodes, with more or less equal length. By assigning forces among the set of edges and the set of nodes, based on their relative positions, it then use these forces either to simulate the motion of the edges and nodes or to minimize their energy [10]. Force-directed graph is very useful when displaying nested data with complex relations in a simple natural way. In addition, data visualizations should be not only static, but also interactive. Force-directed graph has the advantage of interactivity. With the property of interactivity, the users can follow how the graph evolves, transforming it from a mess to a good-looking configuration. Besides, users can also directly manipulate with the nodes and links, pulling them out of their state and getting the detailed information of a specific node or link from tooltips which will display when hovering on the node or the link. Another advantage of force-directed graph drawing is the flexible aspect. This algorithm can be easily adapted and extended to fulfill additional requirements and functionalities. Developers can extends the graph to directed graphs, 3D graph drawing, cluster graph drawing and dynamic graph drawing. Besides, we can find that force-directed graph has good-quality displaying when the number of nodes is in medium and small size. The length of every link is uniform and the distribution of every node seems very comfortable for users. The graph can also be performed through adding various features such as zoom in and zoom out, overview and detail, zoom and pan, fisheyes, for aesthetic purposes.

The main disadvantage of force-directed graph is that the running time of this algorithm is in general considered very high. Since in every iteration, the

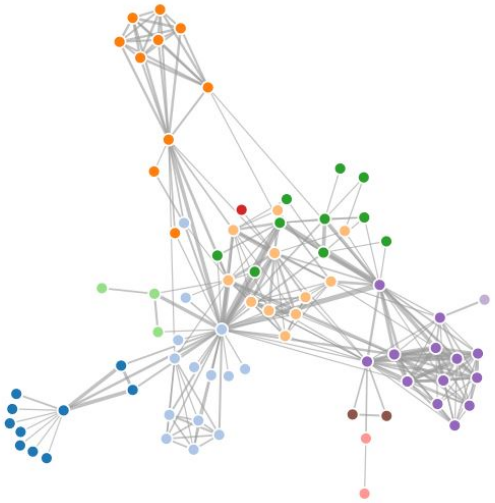

Fig. 1. Example of force-directed graph

algorithm needs to visit all pairs of nodes and compute their mutual force. This deficiency restricts number of nodes we can put in one graph and also the performance of graph displaying. Approaches to the drawing of graphs by high dimensional embedding can improve the running time to a level that is linear to the graph size [11], while force-directed algorithms running time is a polynomial function of graph size.

To display force-directed graph as interactive components on HTML, there are several technologies can be considered to build applications. Scalable Vector Graphics (SVG), which is used define vector-based graphics for the web, has matured into a widely-implemented specification for interactive graphics, used in viewers, authoring tools, and other standards [12]. Drawing with divs and other native HTML elements is possible, but a bit clunky and subject to the usual inconsistencies across different browsers. Using SVG is more reliable, visually consistent, and faster [8]. D3 is most useful when used to generate and manipulate visuals as Scalable Vector Graphics (SVG). Although there are many javascript libraries to animate SVG, such as Vivus, Bonsai, Velocity, Raphael, Snap, Lazy Line Painter, SVG.js, Walkway [13], D3 is always chosen as the best technology to present force-directed graph on web, not only because it is very popular and has bulletproof core, but also it has an amazing physical tool named d3-force [14] to support all features of this graph.

4 Proposed Approach

The ontology-based information extraction and integration system TOBIE generates csv files which consist of a list of data to be analyzed. For this case, the

main focus of this work is to present the above data as a front-end visualization of the data analysis filtered based on a particular time-range. We want to enable the users to efficiently analyze the trend of different variables as time goes on and clearly see the inter-relationship between skills. As a result, we build a website consists of several web pages for the target users visiting.

4.1 Problem Statement

The input for data visualization is two kinds of csv files, as one contains information about skill occurrence, and the other contains information about connections between different skills. Each csv file is concerned about information mentioned above in one specific month. So the first problem of this project is how to transfer data into which can be manipulated by web browsers, because some browsers dont support to load local files, e.g. Chrome.

After that, as we begin to show interactive data visualization, we need to choose some proper visualization techniques to realize all functionalities to fulfill users requirements. There are many kinds of graphs or charts, such as line charts, area charts, column charts, pie charts, bubble charts, etc. We should find the best one or more to enable users to easily detect the most useful information.

4.2 Proposed Solution

To build a good website, we implement yii2, which is a generic web programming framework for developing all kinds of web applications using PHP. Because of its component-based architecture and sophisticated caching support, it is especially suitable for developing large-scale applications [15]. In additional, this framework provides Database Access Objects for accessing relational database, supporting MySQL, which is selected as database to store data from csv files. Following the target users requirements, we design an uploading web page, to provide a fast channel for the web managers to submit two csv files in different kinds of a single month, so that the managers can efficiently input data to MySQL. In that page, the manager should also select the proper information, as month, year, and threshold (co-occurrence, external links and internal links) for their submitted files.

While we have already detailed the approach for uploading files to transfer data to our DB, we want to introduce how to visualize the data as interactive graphs. According to the data structure, we decide to make three individual charts to separately display the information. First of all, a keyword structure is created, in which skills are represented as nodes, whose radius is in accordance skills occurrence, and connections are represented as links, whose thick is in accordance with strength between two skills. For this case, a force-directed graph is the most suitable one.

For page strategic diagram, we display the centrality and density of each cluster, in which skills has strong strength and similar properties with each other. The centrality and density are calculated in the back-end through particular formulas. For the time being, the graph we create in this page is in the form

of bubble chart. At the end, on the page skill occurrence, we use a variant of stacked bar chart to display top 10 skills with most skill occurrence.

4.3 Implementation

For the page keyword structure, our basic skeleton of force-directed graph is based on Mike Bostocks original example [16] which is created through d3. An element SVG is appended to the html body and then we append circle elements displaying the nodes and line elements displaying links to the SVG element. Then also drawing text elements to display the name of each node. As every node belongs to one specific cluster, which is represented as a number, we assign node color according to its cluster, so that nodes in same cluster have same color. D3 also provides an element tip to enable us to add tooltip that inform the detailed data of each node and each link. The tooltips will show when mouseover on the nodes or links.

The second web page, structure diagram, is embedded with a bubble chart based on a Highcharts demo bubble chart [17]. In this case, one bubble represents on cluster and the x-axis value is its centrality, while the y-axis value is its density. The whole chart is divided into four quarter by drawing lines in both x-axis and y-axis with the value as medians of centrality and density. We provides a legend for the users to display bubbles in one quarter and hide all the other bubbles. We realized that users should have an easy access to page keyword structure in order to see the target cluster has what skills which can be used for analysis. For this reason, we add a click event to each bubble. By making an active of the event, the page will jump into keyword structure and it will display nodes and links of the target cluster, while other nodes and links will be hidden.

The final important web page for graph is drawing a stacked bar chart to display the trend of skill occurrence. As there are hundreds of skills retrieved and calculated the number of occurrence, it is not possible to input every skill as a bar in the chart. We extract ten skills with most number of occurrence and put them in the bar chart which help the users to find which skills are most popular. The chart is generated by implementing a Highcharts demo stacked bar [18].

5 Use Cases

Here are several basic use cases in our project. These use cases are divided into two different parts according to different roles of visitors. One part is for users and another part is for administrators.

5.1 Users' use case 1: Enable users changing time through time-span

Users can use the time-span bar to change time period in order to see different information in different time period.

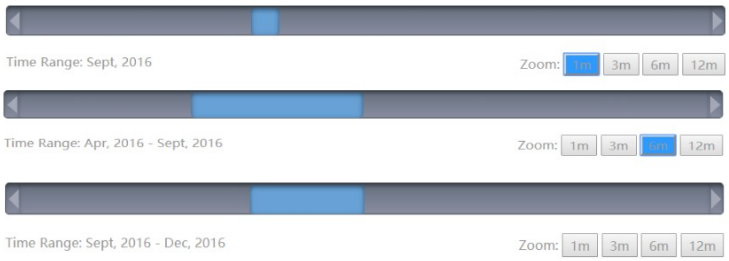

Fig. 2. Time Slider

On the bottom of website page, there is an interactive bar to control time period. The entire bar means whole time period. The blue part means the current time period now. At the bottom left corner of interactive bar, the Time Range explicitly shows current time period by text. At the bottom right corner of interactive bar, the Zoom part explicitly shows predefined settings of fixed time periods. Users can click the Zoom part to choose a corresponding fixed time period, then move blue bar horizontally on interactive bar to change time period. Users can change the length of the blue column and move it to choose a flexible time period. Users can drag blue bar directly by using right/left arrow button to set the current time period forward or backwards one month.

5.2 Users Use case 2: Select the nodes/edges to display more detail information

Users can select the specific nodes or edges on graphs by themselves to show more detail information.

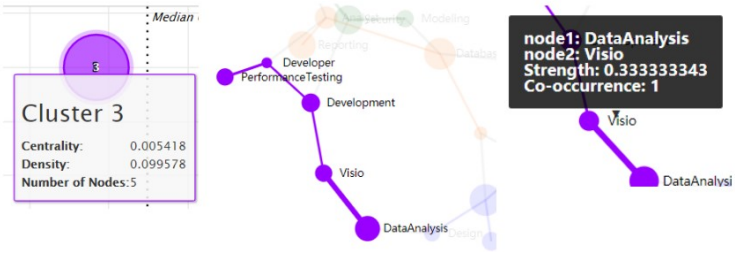

Fig. 3. Result by selecting nodes/edges

In the graphs on website, there are nodes or edges. Each node or edge has its own relevant details. For a node in Strategic Diagram graph, it has all information such as centrality, density and number of nodes about the specific cluster. For an edge in Keyword Structure graph, it has all information such as node1,

node2, strength and co-occurrence about the connections between two connecting skills. For a column in Occurrence of Skills graph, it has all information such as skill name and a total number about the specific skill. Users can select nodes or edges by hovering on them to see more detail information. Users can click one specific node in Strategic Diagram graph, then the page will zoom into relevant nodes which are highlighted with the same color in Keyword Structure graph to show more details about connections.

5.3 Administrators Use case 3: Upload csv files

Administrators can upload the new csv files which contain the latest information into database to update the graphs on the website.

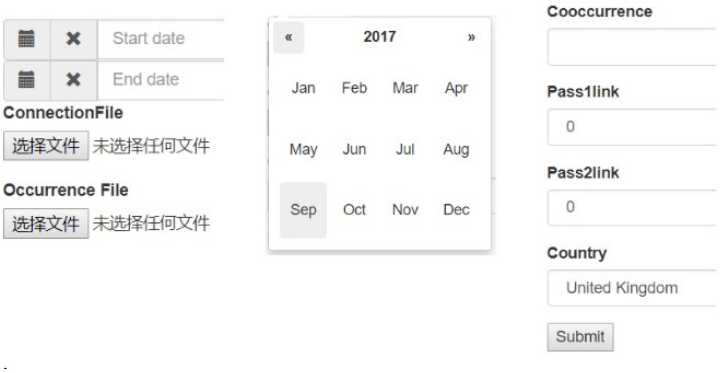

Fig. 4. Processing of uploading csv files

When uploading a new csv file, administrators can choose start date and end date which csv file covers, and set different values of variables such as cooccurrence, pass1link, pass2link and country for the csv file by themselves.

6 Conclusions and Future Work

This paper focus on analyzing the approach to build an interactive front-end Web user interface that making a visualization with data by extracting from the content of job advertisements to identify the specific skills and characteristics that are demand in EU countries labor market. We contribute to works that emphasize the usability and performance of data visualization and enable users to have a good trend analysis in a very efficient way. Therefore we complete the work of conducting the last stage of TOBIE system C trend analysis phase.

In fact, the user interface can be extended as TOBIE system grows. Therefore, in future work, we need to create more pages to fulfill new requirements. In addition, for trend analysis, currently we enable users to detect the changes

through manipulating time slider. While it has a disadvantage that users have no intuitive knowledge about the trend. Therefore we will create charts with time line to display the specific skills occurrence and which cluster a skill belongs to over all time.

References

1. Tze-Haw Huang, Mao Lin Huang, Quang Vinh Nguyen, Laiping Zhao, Weidong Huang, and Jinjun Chen. A space-filling multidimensional visualization (sfmdvis) for exploratory data analysis. *Proceedings of the 7th International Symposium on Visual Information Communication and Interaction*, 2014.
2. Guangchen Ruan and Hui Zhang. Closed-loop big data analysis with visualization and scalable computing. *Big Data Research*, 8:12–26, July 2017.
3. Cristian Felix, Anshul Vikram Pandey, and Enrico Bertini. Texttile: An interactive visualization tool for seamless exploratory analysis of structured data and unstructured text. *IEEE Trans Vis Comput Graph*, 23(1):161–170, Jan 2017.
4. Michelle Hoda Wilkerson and Vasiliki Laina. Youth reasoning with interactive data visualizations: A preliminary study. *Proceedings of the 2017 Conference on Interaction Design and Children*, pages 411–416, June 2017.
5. Joseph F. Grcar. Research trends in msc 65 numerical analysis. *Communications in Statistics - Simulation and Computation*, 2017.
6. Elisa Sibarani, Simon Scerri, Camilo Morales, and Diego Collarana. Ontology-guided job market demand analysis: A cross-sectional study for the data science field. Sept 2017.
7. Daya C. Wimalasuriya and Dejing Dou. Ontology-based information extraction: An introduction and a survey of current approaches. *Journal of Information Science*, 36, March 2010.
8. Scott Murray. *Interactive Data Visualization for the Web*. 2013.
9. Muzammil Khan and Sarwar Shah Khan. Data and information visualization methods and interactive mechanisms: A survey. *International Journal of Computer Applications*, 34(2), 11 2011.
10. Stephen G. Kobourov. Spring embedders and force directed graph drawing algorithms. 2012.
11. David Harel and Yehuda Koren. Graph drawing by high-dimensional embedding. *Journal of Graph Algorithms and Applications*, 8(2):195–214, 2004.
12. Scalable vector graphics (svg) 1.1 (second edition). `https://www.w3.org/TR/2011/REC-SVG11-20110816/`, Aug 2011.
13. Agus. 8 javascript libraries to animate svg. `http://www.hongkiat.com/blog/javascript-libraries-animate-svg/`.
14. d3. d3/d3-force. `https://github.com/d3/d3-force`, Sep 2017.
15. What is yii. `http://www.yiiframework.com/doc-2.0/guide-intro-yii.html`.
16. Mike Bostock. Force-directed graph. `https://bl.ocks.org/mbostock/4062045`, Aug 2017.
17. Bubble chart — highcharts. `https://www.highcharts.com/demo/bubble`.
18. Stacked bar — highcharts. `https://www.highcharts.com/demo/bar-stacked`.