GitHub Viz: An Interactive Visualization to Acquire Knowledge from Authoritative Developers

Chanhee Park* Sungjun Do† Eunjeong Lee‡ Hanna Jang§ Sungchan Jeong**
Hyunwoo Han†† Kyungwon Lee¶
Ajou University

ABSTRACT

'GitHub Viz' is helpful for developers to examine the work of major developers in areas similar to their interests for making decisions [1] [2]. Our Visualization is a web application that can visually explore the rapid trend changes in computer science and data science. Through this application, users can explore other technologies that are relevant to a specific technology, research the key developers through specific development areas and keywords, and view the changes in key developers technology interests over time. To solve these problems, we have developed a visualization application that provides three sub-views with RadViz, which combines several features.

Keywords: RadViz, GitHub, Visualization, Decision Making

Index Terms: H.5.2 User Interfaces: Graphical user interfaces.

1 INTRODUCTION

Computer science and data science are changing rapidly. Identifying the changing trends is extremely important for developers in these areas. Developers require information on the recent technological issues in a development field, the influential developers in each field, and how their keywords of interest have changed over time. However, it is not easy for ordinary developers to quickly understand trends. Hence, we propose a solution to this problem by visualizing an open source project shared in GitHub.

2 OBJECTIVES

We set up the following 5 objectives to help the users of this application explore multiangle: 1) Find key developers related to specific fields and technologies. 2) Explore the core technology of fields. 3) Find other technologies that are relevant to a specific technology. 4) Explore trends in the areas of interest of key developers over time. 5) Explore changes in the technology, language, and libraries that are of interest to key developers.

* { ch13p, kand148, dmswjd2340, jhn9592, tmdcks6628, ainatsumi, kwlee }@ajou.ac.kr

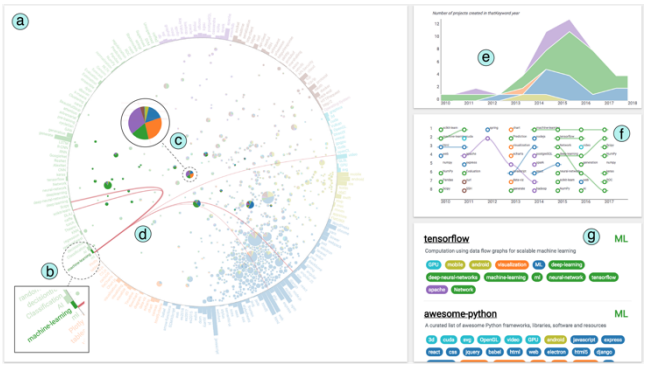

Figure 1: (a) RadViz with keywords as DAs and users as instances. (b) DAs are indicated by different colors for each field and have a bar chart where bars are proportional to the number of related projects. (c) The color of the nodes is expressed as a pie chart according to the proportion of the field. (d) Two keywords appearing in the same repository. (e) A graph showing the change in the number of generated repositories.(f) A graph showing the keyword ranking changes in the repositories. (g) List of repositories comprising the above two sub-views.

3 METHOD

3.1 Data

Data are collected and refined through the GitHub API. We gather information about repositories and the users contributing to these repositories. The 1000 most popular repositories are selected based on the number of stars. The title, description, tags, and readme.md of a repository are used to determine which keywords and development areas are associated with the repository. Keywords are assigned to all users according to participating repositories and their related keywords.

3.2 RadViz

RadViz is a circular visualization based on the spring paradigm, which projects the data of n dimensions into a plane [3]. In RadViz, n dimensions are placed on a circle in the form of axes and connected with a spring to the inner instance of the circle. This point is referred to as a dimension anchor (DA). All user data are represented as individual instances located in circles. The location of the instances within the visualization is determined by the strength of the pull among the connected keywords.

RadViz shows the technologies and the relationship between the developers related to a specific field. Development-related keywords are located outside RadViz, and user instances are located inside it. An instance is a developer, and it is positioned based on the detailed technology associated with all repositories

that he or she participates in. The size of the instance is proportional to the number of stars in all repositories created by the developer. This makes it easier to view who is more popular. We redesign the RadViz to solve specific problems such as 'key developer detection by field' and 'relevance analysis between keywords'. Our RadViz provides three distinctive features.

The first feature is color. Each DA (keyword text) and instance (user) is associated with a color. Different colors are assigned to each development field; this makes them easier to explore [4]. Instances are represented on a pie chart such that a user can view the percentage of each field in which a developer participates [5]. In addition, the pie chart shows each developer's main field and other fields in which major developers from any field are involved.

The second feature is the combination of DAs and bar charts [2]. Bars are assigned to the keywords constituting the DAs of RadViz, and the heights of the bars represent how many times a keyword appears in a repository. Through this, a user can view at a glance which technologies are the core of each field.

The last feature is to link DAs and instances [2][6]. This technique is commonly used in network visualization to link two related instances. We conclude that if two keywords appear in the same repository, they are related to each other, and we link the two DAs with red color line. If two users are simultaneously participating in more than one repository, it is determined that there is an association between the two users. Through this process, users can view other keywords related to the keywords of interest.

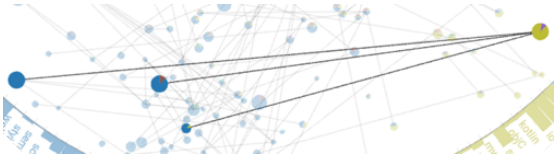

Figure 2: A Link between Web (blue nodes) and Mobile (a yellow node) developers.

3.3 Stacked Area Chart

When one selects a user (instance) or keyword (DA), he or she can view a summary of all repositories that meet the selected criteria [7][8]. This visualization is a stacked area chart. The horizontal axis represents time, and the vertical axis represents the number of generated repositories. Thus, one can view the change in the number of generated repositories in each field over time.

3.4 Keywords Ranking Visualization

When one selects a user or keyword, he or she can view the changes in keyword rankings in all repositories that meet the selected criteria [7]. The horizontal axis represents time, and the vertical axis represents the ranking of the frequency of use. Ranking visualization is structured using these axes. Through this visualization, one can explore the trends in technology, language, and library over time.

3.5 Repository List View

If one selects the area of a field in the stacked area chart or a keyword in the keyword ranking visualization, all repositories that satisfy the condition are displayed, sorted by the number of stars. This view shows the actual repositories that comprise the visualization so that a user may not only trust the visualization but also navigate to the GitHub links in the repositories to further explore them.

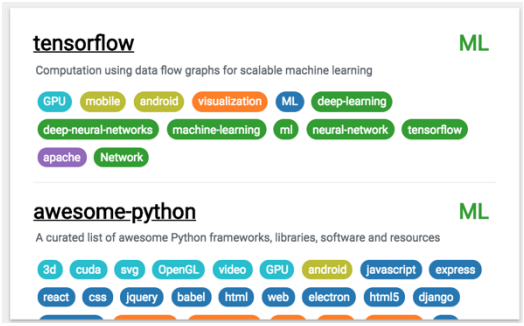

Figure 3: A repository list view when select a keyword 'machine-learning'.

4 Conclusion

When a beginner developer starts studying development, our visualization application can help provide information on the core technologies in each field, the recent trends in the technologies, and which technologies have been used consistently in each field. In addition, when experienced developers want to expand their areas of interest, they can use our visualization to identify areas that are relevant to their area of expertise.

The sub-views on the right have time flow, while the main view on the left has no time flow. So, additional researches will include two. First, we will deal with instance movement in RadViz over time. Second, we will give time weight to the RadViz.

References

[1] F. Du, C. Fan, et al, "Finding similar people to guide life choices: Challenge, design, and evaluation", Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems. ACM, 2017.

[2] L. Liu, D. Silver, and K. Bemis, "JobViz: Interactive Visualization of Majors & Jobs", IEEE VIS, 2016.

[3] A. G. Gee, M. Yu, and G. G. Grinstein, "Dynamic and interactive dimensional anchors for spring-based visualizations", Technical report, computer science, University of Massachussetts Lowell, 2005.

[4] Jihye Lee, Geon Hur, Yongkyun Lee, Yerin Ga, LeeKyung Hong, Kyungwon Lee, "Research Trend Case Study: For Understanding Interdisciplinary Keywords in South Korea", IEEE VIS, 2015.

[5] T. Kurosawa, and Y. Takama, "Co-authorship networks visualization system for supporting survey of researchers' future activities", Journal of Emerging Technologies in Web Intelligence 4.1, 2012.

[6] S. Liu, X. Wang, et al, "Topicpanorama: A full picture of relevant topics" Visual Analytics Science and Technology (VAST), 2014 IEEE Conference on. IEEE, 2014.

[7] F. Heimerl, Q. Han et al, "CiteRivers: Visual analytics of citation patterns", IEEE transactions on visualization and computer graphics 22.1, 190-199, 2016.

[8] Y. Yang, Q. Yao, and H. Qu, "VISTopic: A visual analytics system for making sense of large document collections using hierarchical topic modeling", Visual Informatics 1.1, 40-47, 2017.