Chapter 58
Visualizing the Bug Distribution Information Available in Software Bug Repositories

N. K. Nagwani
National Institute of Technology, India

S. Verma
National Institute of Technology, India

ABSTRACT

Software repositories contain a wealth of information that can be analyzed for knowledge extraction. Software bug repositories are one such repository that stores the information about the defects identified during the development of software. Information available in software bug repositories like number of bugs priority-wise, component-wise, status-wise, developers-wise, module-wise, summary-terms-wise, can be visualized with the help of two- or three-dimensional graphs. These visualizations help in understanding the bug distribution patterns, software matrices related to the software bugs, and developer information in the bug-fixing process. Visualization techniques are exploited with the help of open source technologies in this chapter to visualize the bug distribution information available in the software bug repositories. Two-dimensional and three-dimensional graphs are generated using java-based open source APIs, namely Jzy3d (Java Easy 3d) and JFreeChart. Android software bug repository is selected for the experimental demonstrations of graphs. The textual bug attribute information is also visualized using frequencies of frequent terms present in it.

There is a magic in graphs. The profile of a curve reveals in a flash a whole situation — the life history of an epidemic, a panic, or an era of prosperity. The curve informs the mind, awakens the imagination, convinces. - Henry D. Hubbard in Brinton

DOI: 10.4018/978-1-5225-1837-2.ch058

INTRODUCTION

Data visualization techniques use visual objects (images) to represent the data effectively. It helps in understanding the knowledge patterns, and is widely accepted in the field of data analysis for results representation. Visual data exploration supports integrating the human in the data exploration process, applying its perceptual abilities to the large data sets. The idea behind visual data exploration is to present the data in visual form. The visualized data can help the humans to get insight into the data and make useful conclusions. The users can directly interact with the data. Visual data mining techniques have proven to be of high value in exploratory data analysis and they also have a high potential for exploring large databases. The visualizations of the data allow the user to expand insight into the data and come up with new propositions.

With the help of visual data exploration the users can be directly associated with the data and at the same time get several advantages of it in terms of handling heterogeneous and noisy data. One of the major advantage is the visual data exploration requires no understanding of complex mathematical or statistical algorithms or parameters (Keim, 2002).

Visualization of data allows a faster data exploration and provides better results and their understanding. Visual data analysis provides higher degree of confidence in terms of findings the behavior of data and its variations. Visualizations along with the machine learning and statistical analysis can be combined together for developing advanced models for analyzing the bigger and complex data. Visualization techniques can also be used for getting overview of the data, and gives a choice to the users to explore interesting subsets. Recently, visualization techniques are popularly used for analyzing the large data sets. Data visualization has a high potential and many applications such as fraud detection and data mining will use information visualization technology for an enhanced data analysis.

VISUAL ANALYTICS AND VISUAL DATA MINING

Information is increasing drastically and is a well-known observable fact of the information age. Computing techniques are also becoming advanced day by day to collect and store the data. However, performing data analysis on this huge amount data is a critical problem and is the need of the day. It is one of the primary activity of knowledge workers is to extract the knowledge and useful patterns from the vast amount of data. A number of software tools exist for this purpose using which knowledge can be retrieved from the high volume of data. Although the numbers of software tools exist, still there is a major challenge of performing data analysis in handling the high volume of data. To bridge this gap visual analytics is introduced as one of the smart and quick way of analysis over the huge amount of data. Visual analytics uses graphs, charts and other visual objects to represent the information which allows an intelligent mechanism for direct interaction of users to perform data analysis and make fruitful conclusions in faster manners. The basic concept of visual analytics is information present in one graph is equivalent to information present in number of files (Keim et al., 2006).

In general, visual analytics can be described as “the science of analytical reasoning facilitated by interactive visual interfaces”. Visual analytics is consisting of a number of steps, some of the common steps in visual analytics are information collection, preprocessing of data, knowledge representation, user interaction, and decision making. Visual analytics is more than just visualization and can rather be seen as an integrated approach combining visualization, human factors and data analysis. Visual analytics