TREE-MAP: A VISUALIZATION TOOL FOR LARGE DATA

Mahipal Jadeja
DA-IICT
Gandhinagar,Gujarat
India
Tel:+91-9173535506
mahipaljadeja5@gmail.com

Kesha Shah
DA-IICT
Gandhinagar,Gujarat
India
Tel:+91-7405217629
kesha.shah1106@gmail.com

ABSTRACT

Traditional approach to represent hierarchical data is to use directed tree. But it is impractical to display large (in terms of size as well complexity) trees in limited amount of space. In order to render large trees consisting of millions of nodes efficiently, the Tree-Map algorithm was developed. Even file system of UNIX can be represented using Tree-Map. Definition of Tree-Maps is recursive: allocate one box for parent node and children of node are drawn as boxes within it. Practically, it is possible to render any tree within predefined space using this technique. It has applications in many fields including bio-informatics, visualization of stock portfolio etc. This paper supports Tree-Map method for data integration aspect of knowledge graph. Social customer relationship management (CRM) tree-map example is briefly used to explain how data integration is supported by tree-maps. In this paper, key features of Tree-map are discussed briefly including expressive power of tree-map and types of queries supported by it. As an example of social network visualization, how twitter tree-maps can be used to answer dynamic queries interactively is also discussed in detail.

Categories and Subject Descriptors

H.3.3 [**Information Storage and Retrieval**]: Information Search and Retrieval: Search Process.

General Terms

Algorithms, Management, Measurement, Documentation, Performance, Design, Reliability, Experimentation, Human Factors,Theory

Keywords

Tree-Map, Large Data Visualization, NewsMap, Dynamic Query in Tree-Map, Social CRM Tree-Map,Twitter Tree-Map

SIGIR Workshop on Graph Search and Beyond '15 Santiago, Chile
Published on CEUR-WS: http://ceur-ws.org/Vol-1393/

1. INTRODUCTION

Tree-Maps are used to present hierarchical information on 2-D[1] (or 3-D [2]) displays. Tree-maps offer many features: based upon attribute values users can specify various categories, users can visualize as well as manipulate categorized information and saving of more than one hierarchy is also supported [3].

Various tiling algorithms are known for tree-maps namely: Binary tree, mixed treemaps, ordered, slice and dice, squarified and strip. Transition from traditional representation methods to Tree-Maps are shown below. In figure 1 given hierarchical data and equivalent tree representation of given data are shown. One can consider nodes as sets, children of nodes as subsets and therefore it is fairly easy to convert tree diagram into Venn diagram. Figure 2 represents Venn diagram and its equivalent representation as nested tree-map. Nested tree-map represents the nesting of rectangles. Finally in figure 3, tree-map representation of given hierarchical data is shown [4]. Tree-map is a comprehensive design in which a border is used to show nesting and it is more space efficient compared to nested version. Key advantages of tree-maps are easy identification of patterns and efficient usage of space.

Queries related to space can be answered easily with the use of tree-map visualization. Consider tree-map representation of operating system say UNIX. With the help of this representation, one can easily answer following queries: Identification of directory which is taking up most of space, how much amount of space is taken up by specific directories, types of files present in hierarchy etc. Tree-maps offer dynamic visualization. Key features of dynamic visualization are: immediate feedback mechanism, support for dynamic queries(queries which are incremental and reversible). In Section 5, types of queries supported by tree-map are explained in detail. Tree-maps can be used to represent complex social networks. Possible approaches to deal with dynamic queries during highly interactive sessions are discussed in subsection 6.1 for such types of tree-maps.

Intuitively, tree-map representation is better than simple manual list representation. Peet is a San Francisco Bay Area based famous coffee roaster as well as retailer since 1966. A marketing survey showed following result: For 92 out of 100 customers of peet, (who used the tree-map interface) online shopping was easy. Whereas for the manual lists users, this percentage was only 12. Tree-map interface of peet is shown in figure 4. Most of the other techniques of data visualization were invented in the absence of widely-available computational(computer) resources. Tree-maps were conceived as

a result of computerization and therefore they have crucial benefits from this more modern scenario.

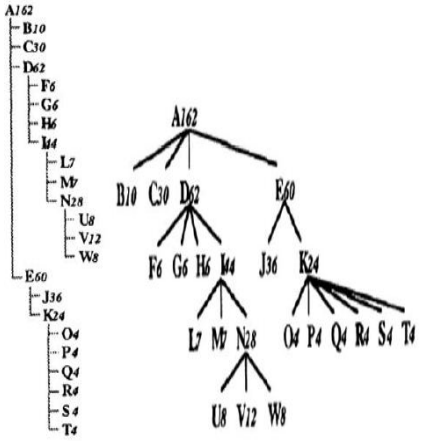

Figure 1: Hierarchical data and Corresponding Tree Representation

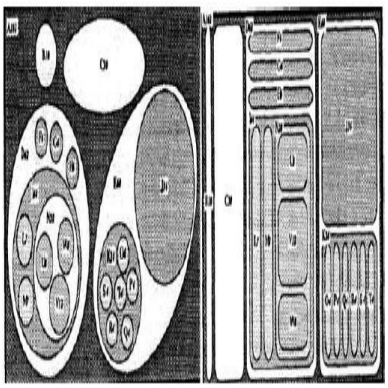

Figure 2: Venn Diagram and Nested Tree-Map

2. GUIDELINES FOR TREE MAP DESIGN

1. Every box of the tree-map can display two different measures namely size and color. Size should reflect quantity measure whereas color is used to display measure of performance and/or change. i.e. satisfaction of customer, growth rate etc.
2. In selection of tree-map layouts, extreme aspect ratios should be avoided [5].
3. Tree-maps are more suitable for high density data, for low density one can use bar charts.
4. Comparing non-leaf nodes is easier in tree-maps compared to bar charts.
5. Appropriate labels should be given and labels should be meaningful.
6. It is advisable to show labels only when user rolls over a tree-map box.

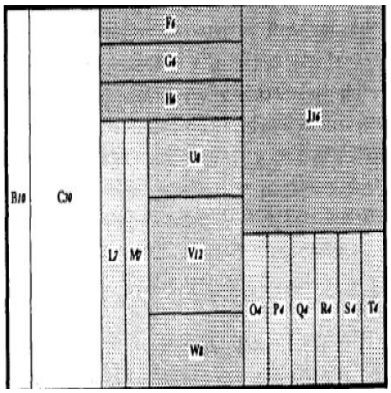

Figure 3: Tree-Map

7. Labels must be visible in multicolored background of tree-map.
8. Depending upon the nature of the color measure, one sided/two sided color range should be used.
9. In order to show correlation, highlighting should be used.
10. One can use animation in tree-maps to show change in the data.
11. Simple presentation method (Tooltip window/sidebar) can be used to show node detail.

Figure 4: www.peets.com

3. EXPRESSIVE POWER OF TREE-MAP

Tree-Maps are used to express a variety of nested as well as hierarchical data and data structures. In general, type of tree-map representation depends upon application and type of data hierarchy.

"Tree-map visualization generator" are used to display tree-maps for arbitrary hierarchical data. Tree-Maps can be provided as images in static forms or they can be used to provide interactive features (like zooming into small area of hierarchy) in applications. Tree-maps support browser as well as rich client applications. In one of the applications, tree-maps are incorporated with Windows Forms- Microsoft Corporation.

Tree-Maps are also famous amongst news designers. Examples are listed below.

1. NewsMap[6] (Newsmap.jp is developed by Marcos Weskamp and it represents current items of Google News using interactive Tree-map which is shown in figure 5.)
2. London 2012 Olympics and Tree- maps [7]
3. BBC News- SuperPower: Visualising the internet
4. The New York Times- Obama's budget proposal (Year 2011)
5. CNN Twitter buzz of South Africa (Year 2010)

Figure 5: www.newsmap.jp

4. SOCIAL NETWORK DATA AND TREE-MAP

For the promotion of brand, role of marketer is not significant in the modern era of social media. In the past, information was produced by marketers and consumed by customers. Currently more information is generated by customers about brands on social media including blogs, social media networks, online forums etc. Currently marketing teams are struggling in analysis of this online information, which is required for prediction of acceptance rates of products, patterns of purchase and level of satisfaction in customers. Marketers can use these new channels for promotion by developing customers as brand advocates.

For travel as well as hospitality industry, decisions related to purchase are mainly determined by online reviews as well as recommendations. Online customer data along with business functions information forms an integrated database. In order to study levels of customer loyalty, study of this integrated database is necessary.

It is possible to use customer tree-map for segmenting customers and generation of 'brand score' for customers and brand score depends upon 1) Brand engagement of customer-behavioral aspect and 2) Attitudes of customers.

Two different types of score namely spend value score and advocacy score are calculated using integrated database (traditional CRM and unstructured data). Social CRM tree-map can be created by plotting these scores (by integrating two data-sets) on a 2-D axis[9]. Example of social CRM tree-map is given in figure 6. Members without any spend value are defined as noncustomers. This tree-map is useful for calculation of overall "customer brand score".

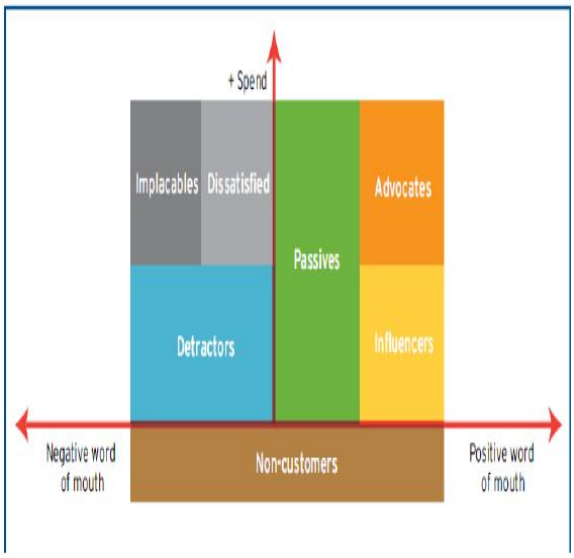

Figure 6: Social CRM Tree-Map

Advocates have following qualities: They have high values for spend value as advocacy score. They are brand evangelists and their behavior as well as attitude is very loyal to brand.

After successful development of Tree-map, organizations can take actions in order to cultivate advocates of brand.

5. TYPES OF QUERY SUPPORTED BY TREE-MAPS

Tree-Maps provide two important features by supporting dynamic queries:

1. Querying a large set of data.
2. To find out patterns in large data set. [10][11]

In tree-maps, dynamic queries are implemented using radio buttons, buttons and sliders. Tree-map follows principle of direct manipulation for searching in large database.
Key features of query processing of Tree-Map are listed below:

- Supports visual representation. (for components of query)
- Supports visual representation of query results
- Provides rapid, reversible and incremental control of query.

- Selection is done by just pointing, not by typing.
- Tree-map provides immediate as well as continuous mechanism of feedback

6. TREE-MAP FOR TWITTER DATA VISUALIZATION

Key requirements for visualization of any social network are listed below:

- Identification of the actors-members of the social network.
- Visualization should represent relationships of various types.
- Visualization should support aggregated as well as structured view of the complex social network.

Consider example of Twitter network with four sample actors namely Steve, John, Luke and Adam. Figure 7 represents this network as a Tree-map. Tree-map offers all the crucial features which are desirable for visualization tool. Here actors are represented by rectangles and size of each rectangle is proportional to the total number of tweets sent by that particular actor. The friendship relationship is represented by a common edge between two rectangles. In our example. rectangle corresponding to Luke has highest area which implies highest number of tweets amongst the four users. No common edge is present between Steve and Luke which implies that they are not friends in Twitter.

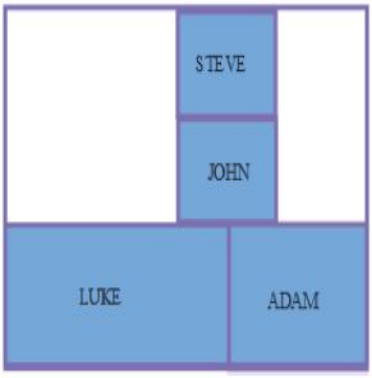

Figure 7: Basic Twitter Treemap

Other variants of Twitter tree-map are also shown in figure 8 and 9. Tree mapping is not as popular as other visualization techniques still recent survey results are encouraging for twitter tree-maps [12]. Better results are possible by improving current design of tree-maps as well as integration of tree-map with other visualization techniques.

6.1 Discussion on Interactivity of Tree-maps

Tree-map offers interactive feature which is distinctive. The main objective of this visualization tool is to provide interactive display on a computer screen. Because of this unique feature, one can explore the data hierarchy effortlessly and simultaneously decent level of estimation is also possible for quantitative aspects of the information. In order to provide element specific information in detail, various tree-map soft wares offers computer screen mouseovers

Figure 8: Twitter Treemap with additional information (Actor's interests)

Figure 9: Twitter Treemap integrated with Network diagram

using which the user can get specific information just by placing the computer mouse over the specific box. Because of these crucial interactive features, tree-map is emerging powerful visualization technique-also for large social datasets because real time feedback is essential in the case of complex social network. Due to this interactivity feature, the analyst has the ability to traverse the tree and he can also present categorical data view at every level.

Generally, queries on social network data focus more on relationships between different groups and size of particular category is often very common type of query. For example, which country has highest number of tweeter users? Now consider one complex query: Do white males in the North America use the twitter more than white females in the South? In order to answer this question one has to consider sub-questions for all data points. i.e. whether a particular person is black/white, whether he has twitter account or not and so on.

In order to answer these queries interactively for categorical social data, we propose the use of CatTrees.(enhancement of tree-maps) [13] It is possible to answer these types of question easily if the data has hierarchy because then for each possible answer pattern one can allocate leaf node with counter and to get final answer, the analyst can follow two different paths (depending upon query) from root to leaf nodes and give final result depending upon the comparison

of counters. So depending upon query, new hierarchy may be required every time. In short, dynamic hierarchies are required to support dynamic queries! Dynamic hierarchies are implemented by CatTrees.

All social data is not hierarchical in nature. Surprisingly tree-maps can be used to visualize non-hierarchical data too. In this case, imaginary hierarchy is provided as an input by the analyst [14].

7. CONCLUSION AND FUTURE WORK

Speed of data accessing is very crucial parameter for any visualization tool. Tree-maps should support hardware or parallel processing or grid computing approach for better results.

Overall design of tree-maps should be modified for offering better understandings of the data. Data accuracy is also equally important along with decent data accessing speed. Tree-maps should offer better meaningful results for various queries. Nowadays tree-maps are famous at the academy but they are not accepted as a general hierarchical tool. Tree-maps have various drawbacks: 1) Specific use 2) Lack of cognitive plausibility 3) Poor performance (Task-driven) 4) Average aesthetic qualities. Currently ongoing research in this area is trying to solve these issues.

Tree-maps are very useful tools for identification of extreme values in large database as well as primary trends. They are not meant for comparison of values precisely mainly because of two dimensional limited area and color encoding. Tree-maps are successful and can be understood easily by public.

8. ACKNOWLEDGMENTS

We would like to specially thank and acknowledge Dr. Jaap Kamps for motivating us and providing quality inputs. We would also like to convey our regard to SIGIR team for organizing GSB'15-the first international workshop for graph search and beyond, enabling us to participate and give us a chance to contribute to the community to the best of our abilities.

9. REFERENCES

1. Shneiderman, B. (1992). Tree visualization with tree-maps:2-d space-filling approach. ACM Transactions on graphics (TOG), 11(1),92-99.

2. Bladh, T., Carr, D. A., & Scholl, J. (2004, January). Extending tree-maps to three dimensions: A comparative study. In Computer Human Interaction (pp. 50-59). Springer Berlin Heidelberg.

3. Bederson, B. B., Shneiderman, B., & Wattenberg, M. (2002).Ordered and quantum tree-maps: Making effective use of 2D space to display hierarchies.AcM Transactions on Graphics (TOG), 21(4), 833-854.

4. Johnson, B., & Shneiderman, B. (1991, October). Tree-maps:A space-filling approach to the visualization of hierarchical information structures. InVisualization, 1991.Visualization'91, Proceedings., IEEE Conference on (pp.284 – 291). IEEE.

5. Kong, N., Heer, J., & Agrawala, M. (2010). Perceptual guidelines for creating rectangular tree-maps. Visualization and Computer Graphics, IEEE Transactions on, 16(6), 990 – 998.

6. Ong, T. H., Chen, H., Sung, W. K., & Zhu, B. (2005). Newsmap: a knowledge map for online news. Decision Support Systems, 39(4), 583-597.

7. Field, K. (2012). Mapping the London 2012 Olympics. The Cartographic Journal, 49(3), 281-296.

8. Cao, N., Lin, Y. R., Sun, X., Lazer, D., Liu, S., & Qu, H. (2012). Whisper: Tracing the spatiotemporal process of information diffusion in real time.Visualization and Computer Graphics, IEEE Transactions on, 18(12), 2649-2658.

9. Vijay Raghunathan(2014) Moving Beyond Social CRM with Customer Brand Score Available: http://www.cognizant.com/InsightsWhitepapers/Moving-Beyond-Social-CRM-with-the-Customer-Brand-Score.pdf

10. Deussen, O., Hansen, C., Keim, D. A., & Saupe, D. Interactive Tree-maps With Detail on Demand to Support Information Search in Documents.

11. Schlechtweg, S., Schulze-Wollgast, P., & Schumann, H.(2004, May). Interactive tree-mapswith detail on demand to support information search in documents. In Proceedings of the Sixth Joint Eurographics-IEEE TCVG conference on Visualization (pp. 121-128) .Eurographics Association.

12. Sathiyanarayanan, M., & Burlutskiy, N. Design and Evaluation of Euler Diagram and Treemap for Social Network Visualisation.

13. Kolatch, E., & Weinstein, B. (2001). Cattrees: Dynamic visualization of categorical data using treemaps. Project report.

14. Sirin, E., & Yaman, F. (2002). Visualizing dynamic hierarchies in treemaps.