FACTA UNIVERSITATIS
Series: **Automatic Control and Robotics** Vol. 20, Nº 2, 2021, pp. 111 - 121
https://doi.org/10.22190/FUACR210715009F

Regular Paper

TOOL FOR INTERACTIVE VISUAL ANALYSIS OF LARGE HIERARCHICAL DATA STRUCTURES

UDC ((681.883.75:303.442.3):004.3)

Milena Frtunić Gligorijević, Miloš Bogdanović, Nataša Veljković, Leonid Stoimenov

University of Niš, Faculty of Electronic Engineering, Department of Computer Science, Republic of Serbia

Abstract. *In the Big Data era data visualization and exploration systems, as means for data perception and manipulation are facing major challenges. One of the challenges for modern visualization systems is to ensure adequate visual presentation and interaction. Therefore, within this paper, we present a tool for interactive visualization of data with a hierarchical structure. It is a general-purpose tool that uses a graph-based approach. However, its main focus is on the visual analysis of concept lattices generated as the output of the Formal Concept Analysis algorithm. As the data grow, concept lattice can become complex and hard for visualization and analysis. In order to address this issue, functionalities important for the exploration of the large concept lattices are applied within this tool. The usage of the tool is presented in the example of visualization of concept lattices generated based on the available data on the Canadas open data portal and can be used for exploring the usage of tags within datasets.*

Key words: *Visualization, visual analytics, visual exploration, visualization tool, data exploration*

1. INTRODUCTION

The era of Big Data has made available large amounts of diverse data for analysis and drawing conclusions for further processes, knowledge, or results. Due to the volume and heterogeneity of the available information, data exploration and visualization systems have a great importance in decision making process in various sectors. Both data exploration and visualization are means for extracting knowledge and making sense of data [24][25]. Their goal is to support information perception and manipulation. The

Received July15, 2021 / Accepted September 23, 2021
Corresponding author: Milena Frtunić Gligorijević
University of Niš, Faculty of Electronic Engineering, Department of Computer Science
Aleksandra Medvedeva 14, 18000 Niš, Republic of Serbia
E-mail: milena.frtunic.gligorijevic@elfak.ni.ac.rs

© 2021 by University of Niš, Serbia | Creative Commons License: CC BY-NC-ND

visual analysis involves users developing insights about significant relationships, domain-specific contextual influences, correlations and causal patterns [25].

Most traditional visualization systems have reached their limits when visualizing a very large amount of data. Authors in [26] have emphasized that the "bigness" of large data sets and their complexity in terms of heterogeneity contribute to complicate the representation of data and that effective representation of all the information at the same time is really challenging. Therefore, exploring and visualizing large datasets has become a major research challenge [2]. Furthermore, one of the main requirements that should be addressed by the modern visualization systems is visual presentation and interaction that will provide a clear preview and an easy exploration, handle a large number of objects and prevent information overloading.

Within this research, we would like to present an interactive visualization tool for the visual analysis and data exploration of data with a hierarchical structure. This tool relies on the graph-based approach and supports both an overview of the whole hierarchical structure and options for a detailed exploration of the data. For a detailed exploration the tool supports options such as zooming, filtering the data by their keywords, and exploring only parts of the graph and relations that are of interest for the analysis.

The visualization tool that is presented within this paper is a general purpose tool that can be used for the visual analysis of any datasets with node-link structure. However, its main advantage and focus is on the visualization of concept lattices that hold the hierarchical order of formal context generated as the output of the Formal Concept Analysis algorithm. Therefore, the tool supports functionalities vital for visualizing large concept lattices and securing their easy analysis.

The visualization of concept lattice is very important for the analysis when using Formal Concept Analysis. As reported in [27][27], the visualization of the hierarchical order of concept lattice structure is an important concern for practical applications of this algorithm. The main problem for the visualization can be the size of the concept lattice due to the large number of the formal concepts which can lead to the very complex and impractical concept lattice. Furthermore, there are not many tools that can overcome this challenge and successfully visualize large concept lattices. Thus, our focus was on creating a tool that applies a graph-based approach with interactive mechanisms to overcome the visualization problem of concept lattices and enable a clear and informative preview of the data hierarchy that can be used for the decision-making processes.

The rest of the paper is organized as follows. Section 2 covers state of the art for graph-based visualization systems. Section 3 presents the developed tool for the interactive visual analyses and data exploration along with the example of using this tool for analyzing concept lattices generated for all open data portal categories available on the Canadas open data portal. Section 4 presents the conclusion of this paper.

2. State of the Art Graph-Based Visualization Systems

As we find ourselves a part of the Big Data era, information volume, variety, veracity, and velocity became a significant problem for each person when trying to manually explore and analyze data. In such a situation, we all rely on data exploration and visualization systems when performing visual information seeking mantra: "overview first, zoom and filter, then details on demand" [1].

Contemporary data visualization and exploration tools are adjusted to fit the needs of Web of Data and Linked Data. This adjustment was necessary due to inability of most traditional systems to handle the large size of many contemporary datasets. As proven throughout literature [2], [3], [4], [5], [6], [7], the development of tools and systems capable of handling current data velocity has become a major challenge. Researchers and developers worldwide are struggling to overcome performance issues, along with visualization and interaction challenges, with the same goal – provide users with a tool capable of easily exploring various datasets.

In the recent years, a significant effort has been put into resolving visual exploration and analysis problems [8], [9], [10]. As a result, a large number of tools and systems have been developed whereas only a part of them resemble the tool we present in this paper. These systems are usually referred to as Graph-based visualization systems. The majority of Graph-based visualization systems implement the node-link approach [11], where nodes are used to represent entities, and edges (links) exist between the nodes if there is a relation between the represented entities. The graph-based visualization can be very valuable for analyzing data from various fields of research since it reveals the structure and relations between the analyzed data and can be used as a knowledge base in the decision-making processes. The authors of [12] have performed a very good overview and comparison of graph-based visualization systems by analyzing their functionalities. Among all, the following characteristics are vital for graph-based visualization system usability: keyword search, data filtering, sampling techniques, aggregation techniques (for example clustering), incremental techniques in terms of result computation, the use of external memory, information domain, and application type.

RelFinder [12] is implemented using AdobeFlex. It is a Web application which runs in all Web browsers with Flash Player support. Its main functionality is relationship discovery. Algorithm implemented within RelFinder utilizes connections between selected data sources and extracts as many relationships as possible, according to basic search constraints – relationship length and directionality. LODeX [14] a tool that produces a representative summary of a Linked open Data source. As an input, LODeX uses the URL of a SPARQL endpoint and sends a set of predefined SPARQL queries. Query results are used to form a visual report containing statistical and structural information of the LOD dataset. IsaViz [15] is a visual environment for browsing and authoring RDF models represented as graphs. It features a 2.5D user interface allowing smooth zooming and navigation in the graph, creation and editing of graphs by drawing ellipses, boxes and arcs, RDF/XML, Notation 3 and N-Triple import, RDF/XML, Notation 3 and N-Triple, SVG and PNG export. graphVizdb [16] is a platform for the interactive visualization of very large graphs. Interactions with the visualized graph are similar to exploration of maps due to indexing graph node coordinates within a spatial database. User operations are translated into simple spatial operations on the backend side of the system which leads to lower latency.

Fenfire [17] is an open-source RDF browser and editor. It uses a graph view and focuses on an engaging and interactive browsing experience. Similarly, LodLive [18] uses RDF and SPARQL standards to enable RDF resources exploration. ZoomRDF [19] is a framework based on Gene Ontology, implemented with the purpose of RDF data exploration and editing. It adapts a space-optimized visualization algorithm for RDF, combines the visualization with a fisheye zooming concept, and considers both the importance of resources and the user interaction on them. Gephi [20] uses 3D render engine to display large graphs in real-time. It is an open-source graph exploration software that allows for spatializing, filtering, navigating, manipulating, and clustering. Trisolda RDF visualizer [21] is based on the clustered navigation technique. This tool offers users with optional access to some detailed information without unnecessarily occupying space in the basic view of the data. Paged Graph Visualization (PGV) [22], is a semi-autonomous tool for RDF data exploration and visualization. It consists of the "PGV explorer" and the "RDF pager" modules, whereas the later utilizes BRAHMS main-memory RDF storage system. The system described in [23] is a graphical tool for visualizing arbitrary RDF graphs. It incorporates a search tool capable of finding a portion of the graph likely to be of user's interest while the system focuses on providing a clear graphical visualization of that portion using a node-centric view.

3. Interactive Visual Analysis

Our focus in this research is the development of a web-based visualization tool for the interactive analysis of datasets with a hierarchical structure. This is a general-purpose tool that can be used for visual analysis of any kind of data that contains a node-link structure. The tool uses a graph-based approach for visualization and relies on the d3.js library for the visual presentation of data.

Although it is a general tool, its main purpose is to support the visualization of concept lattices generated as the output of the Formal Concept Analysis algorithm. Concept lattice represents a collection of formal concepts logically organized into a hierarchy of concepts interconnected using subconcept-superconcept relations [28]. Described in a less formal manner, if there is a dataset where each object in the dataset is described with a particular set of attributes, based on the relationship between those objects and attributes, concept lattice contains the hierarchical order of all attributes that are used for describing all objects in the particular dataset.

With the growth of the dataset and combinations of attributes that describe objects in them, the complexity of concept lattice increases, and the analysis of the hierarchy becomes harder. In order to overcome that issue and its complexity, visualization of concept lattice has to be interactive and offer the possibility of both overviews of the whole hierarchy and detailed relations between the attributes in the lattice. Therefore, the visualization tool that we present in this paper supports several key functionalities that provide a clear preview and analysis of the lattice (Fig. 1).

Since the tool visualizes the concept lattice with a graph-based approach, nodes are used to represent certain combinations of attributes, and links between the nodes represent hierarchical relations between the nodes for the particular data. In order to provide exploration of the hierarchy between the attributes, the tool uses the *gravity approach* and positions nodes in the way that the nodes that are at the top in the graph are higher in the hierarchy, and nodes at the bottom of the graph are lower in the hierarchy. Consequentially, attributes covered by the nodes in the upper part of the concept lattice are attributes with higher hierarchy then those that are lower in the concept lattice.

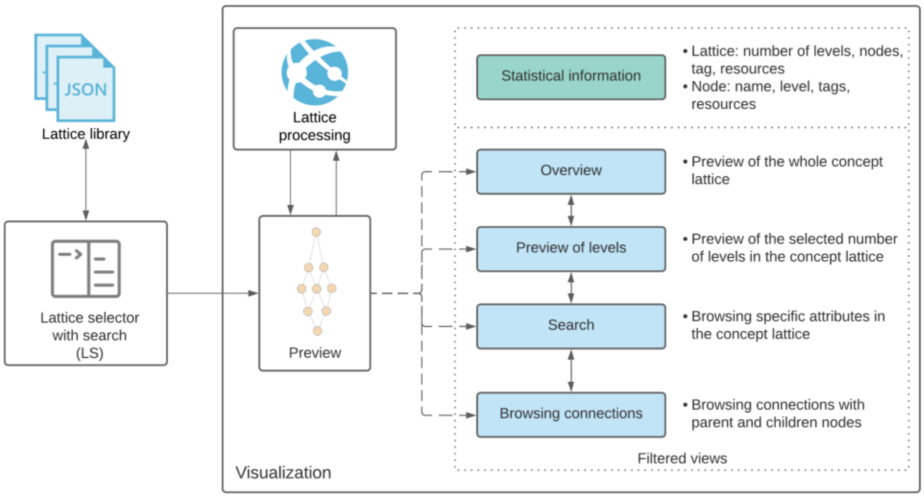

Fig. 1 Visualization tool functionalities

Our tool scans concept lattices that are available in the application library (Lattice library in Fig.1) in JSON format and offers them to the user. The user is given two options: to select a particular concept lattice which will be visualized or to search all available concept lattices (LS component in Fig. 1). At this point, the search mechanism receives a list of attributes that are of interest for the user and marks concept lattices that contain at least one searched attribute. In that way, the user can narrow the concept lattices that are of interest for the particular user. This possibility can be a valuable option for analysis of concept lattices from the same domain. The visualization of the particular concept lattice supports several options for the analysis:

- The preview of the whole concept lattice
- Preview of only a selected number of levels in the lattice
- Browsing positions of particular attribute in the lattice
- Browsing connections to all top and all bottom nodes from the selected node
- Statistical information

The preview of the whole concept lattice – gives an overview of all nodes in the concept lattice and all links between the nodes. This preview is valuable for gaining a general impression of the whole structure of the hierarchy.

Preview of only a selected number of levels in the lattice – gives an overview of the part of the concept lattice where only nodes that belong to the selected number of levels in the graph are presented. This option is especially useful when the concept lattice is complex because the user has an option to analyze only the upper part of the lattice which contains attributes with a higher hierarchy and thus greater importance for a particular dataset.

Browsing positions of particular tags in the lattice – this is a search option available to the user. The user is given an opportunity to browse specific attributes in the lattice. The search results will be shown in the concept lattice by highlighting nodes using different colors based on the subset of the searched attributes those nodes cover. The tool does not highlight all the nodes that cover some of the searched attributes but only the most significant ones. Node's significance is determined based on the number of searched attributes covered by the particular node and the number of additional attributes. Based on the significance, nodes are colored with different colors: red, purple, green, and blue. The red color is used to mark nodes that contain all searched attributes and no additional attributes. Purple is used to mark nodes that contain a part of the searched attributes and no additional attributes. Green is used to mark nodes that contain all of the searched attributes and the minimum possible number of additional attributes. The blue color is used to mark nodes that contain part of the searched attributes and the minimum possible number of additional attributes. This option is especially useful for fast discovery of particular attributes.

Browsing connections to all top and all bottom nodes from the selected node – this option allows the user to select a node, and by doing so, all parent and children nodes as well as the links between them will be highlighted. This allows users to easily and clearly analyze connections between a particular subset of nodes.

Statistical information – along with all visualization options, the user is given the statistical information regarding the lattice. This functionality visualizes information such as total number of nodes in the lattice, total number of datasets in the lattice, total number of attributes in the lattice, and total number of levels. Further, the tool supports options for the selection of a particular node which gives a preview of all the information regarding the selected node: node name, level at which the node is positioned, number and list of attributes covered by the selected node, list of objects covered by the selected node and the number of objects that are represented by the covered attributes.

Additionally, presented visualization options can be combined which further facilitates analysis of the position, usage and importance of attributes in one concept lattice, especially if the concept lattice contains huge number of nodes, attributes and datasets.

3.1. Example of concept lattice visual analysis

For the example of the usage of the presented tool, we are using concept lattices generated from the available datasets on Canada's open data portal. This portal contains over 80000 datasets organized into 19 categories where some of the categories contain a significant number of datasets and tags used for describing those datasets. For each category on the portal, we created one concept lattice based on all datasets belonging to that category. The aim of these concept lattices is to reveal the usage of tags that are used for describing datasets and hierarchical order among tags within one category.

For the purpose of demonstration of the visualization, we have used the category *Law* from the Canada's open data portal. This category is one of the smaller categories on this portal in terms of both the number of datasets and the number of different tags used for describing the datasets belonging to this category. Although it is one of the smaller categories, the concept lattice generated for this category is rather complex – it contains 325 nodes organized in 8 levels. The preview of the summary information about the whole concept lattice for category *Law* and an example of the summary information for the single selected node in the concept lattice is presented in Fig. 2.

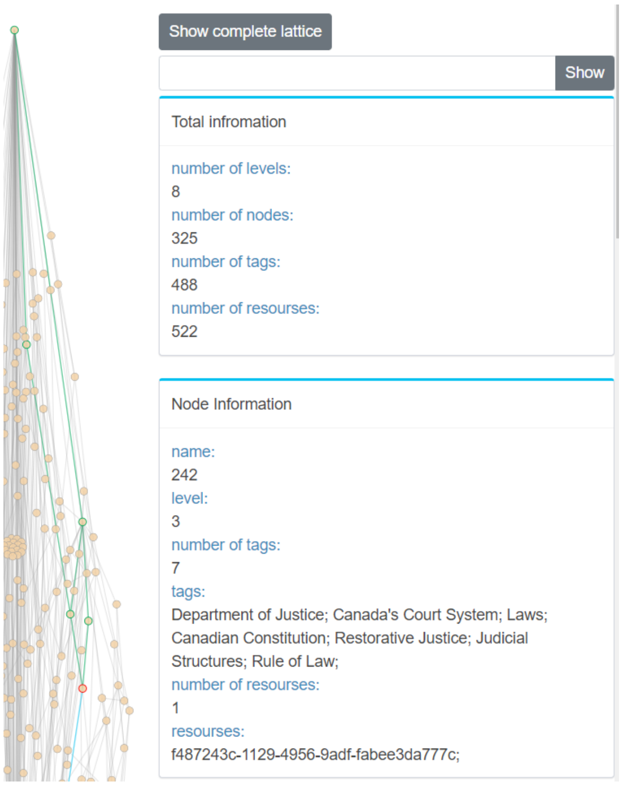

Fig. 2 Preview of the summary information about the whole concept lattice and example of the single node summary information

The visualization option for analyzing the complete lattice, especially complex as this one, can be used for some initial analysis of the overall positions of some nodes and their position in the lattice. In Fig. 3a we have demonstrated how this can be done. We have used the possibility of previewing the whole concept lattice together with browsing for nodes that contain some tags. In this example we searched for tags: *Research*, *Annual reports*. As shown in Fig. 3a, the concept lattice does not contain a node that covers both searched tags and no additional tags. Additionally, there are nodes in the lattice that cover only searched tags separately (purple nodes), and by placing the cursor on them it can be noted that both nodes are on the first level in the lattice. Also, there are several nodes that contain part of the searched tags and the same minimum number of additional tags (in this case one additional tag). Finally, in the results, there is one green node that contains both searched tags and some additional tags (in this case three additional tags). All detailed information about the nodes can be acquired by selecting the node of interest.

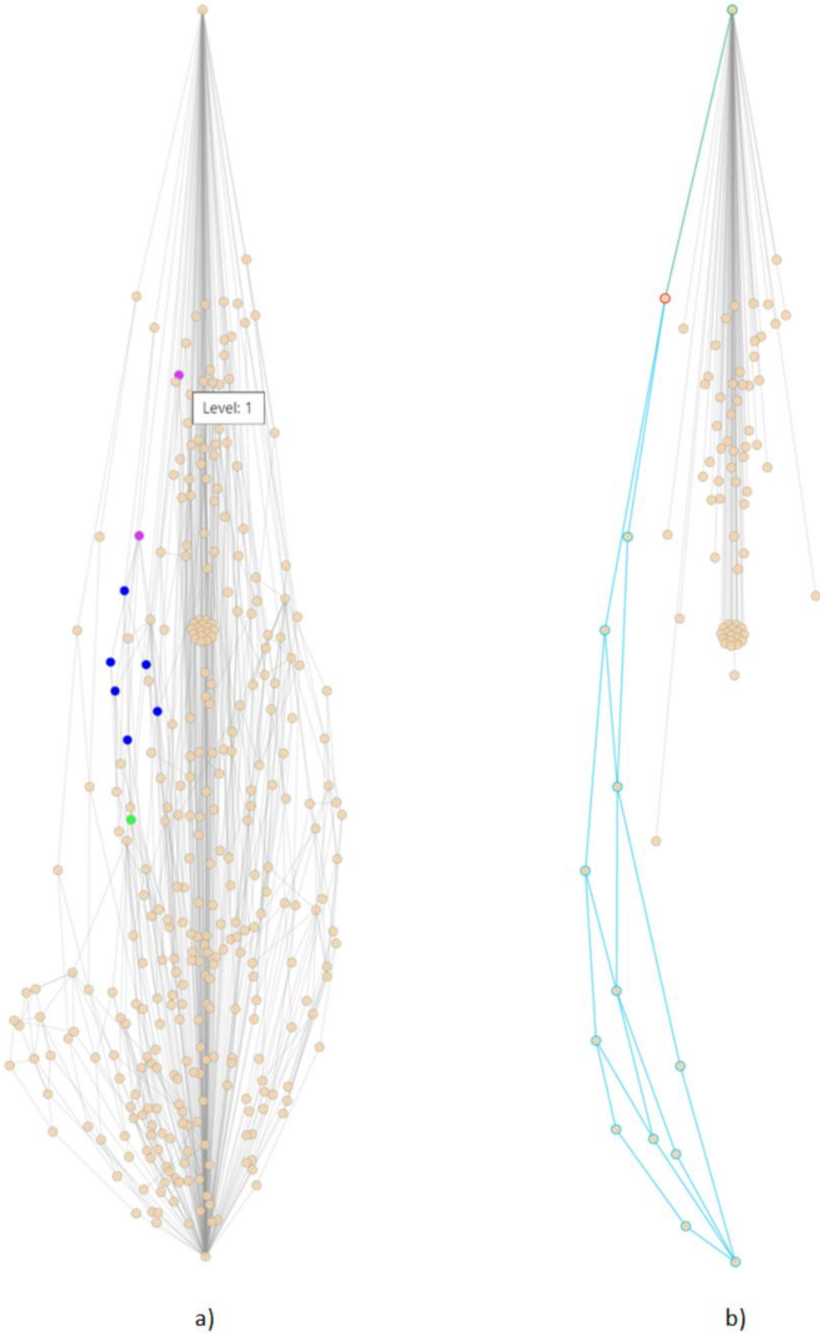

Fig. 3 a) Complete concept lattice for category Law from Canada's open data portal; b) Combination of visualization options for category Law from Canada's open data portal

However, for a deeper analysis such as analyzing connections between the nodes, a huge number of nodes and links may cause confusion and problems. For that reason, to enable a clearer preview, our tool offers a possibility for hiding some of the nodes trough an option of selecting only few levels to be shown.

One example of analysis with some of the nodes being hidden is shown in Fig. 3b. In this figure, a preview of the selected number of levels in the lattice is combined with the preview for browsing connections to all top and all bottom nodes from the selected node. The selected number of levels is set to one, which significantly reduced the number of nodes. After that, browsing the connections can be done with a simple click on the node of interest. After the selection, the chosen node is marked with a red border. Links from the selected node to nodes higher in the hierarchy, as well as the upper nodes' borders are colored green. Additionally, links from the selected node to nodes lower in the hierarchy and lower nodes' border are colored blue.

As it can be seen, such preview is much clearer comparing to the one that can be seen in Fig. 3a. This type of preview can be used for a detailed analysis of the relations between subsets of interrelated tags within one category. Our tool provides several modes for concept lattice preview and analysis. These modes can be combined to enable an easy overview analysis depending on the user's needs.

4. CONCLUSION

Due to the growth of the volume and variety of data in recent years, modern data exploration and visualization systems have to adopt new techniques that will provide clear and suitable preview of data. For graph-based data this means introducing adequate interactive options that will provide easy exploration of the connections that exist between the nodes in the graph and structure of the visualized dataset.

In this paper we have presented a general tool for visualization of data with hierarchical structure that provides these options. Our tool is focused on the visual analysis of the large concept lattices that are the output of the Formal Concept Analysis algorithm. Concept lattices are proven to be very useful way for the analysis of hierarchical data as they reveal a shared conceptualization originating from the usage of attributes that describe particular set of data. However, one of the downsides of the visual exploration of concept lattices is that their structure becomes very complex as data grows.

Therefore, in order to overcome this issue, we have developed a tool that provides functionalities that facilitate visual analysis of large concept lattices. We implemented methods that can be used to filter only nodes that are of interest for the user. Further, we provide search mechanism that marks only important nodes that correspond to the searched attributes and we use different colors to mark them differently depending on their fitness to the searched criteria. Moreover, for exploring the hierarchy we implemented possibility of selecting particular node and exploring its hierarchical relations with nodes that are both upper and lower in the hierarchy. With possibility of previewing only part of the concept lattice we provide a streamlined preview which is valuable for a successful analysis.

With these functionalities we have provided a tool that can be successfully used for the visual exploration and analysis of large concept lattices. At the end of the paper we gave an example of the usage of our tool on data available on the Canada's open data portal. In our example we are analyzing the usage of tags within different categories on the portal with a goal to determine the importance of different tags within different categories and reveal the hierarchy among them.

Although this is a general tool, it is adjusted for the visualization of concept lattices. Therefore it does not offer a lot of flexibility for the node structure. At this moment, if the node structure varies from the initial setting, it can be changed manually. Implementation of the automatic change of the node structure could be a valuable feature for the analysis of different kinds of datasets with a hierarchical structure. For that reason, in the future, this tool will be extended with the possibility for automatic setting of the inner node structure that could be further used for statistics, filtering, and visual exploration.

REFERENCES

[1] B. Shneiderman. The Eyes Have It: A Task by Data Type Taxonomy for Information Visualizations. *In IEEE Symposium on Visual Languages*, 1996. [Online]. Available: https://doi.org/10.1109/VL.1996.545307

[2] N. Bikakis and G. Papastefanatos. Visual Exploration and Analytics of Big Data: Challenges and Approaches, 2016.

[3] B. Shneiderman. Extreme Visualization: Squeezing a Billion Records into a Million Pixels. In SIGMOD, 2008. [Online]. Available: http://dx.doi.org/10.1145/1376616.1376618

[4] J. Heer and S. Kandel. Interactive Analysis of Big Data. ACM Crossroads, 19(1), 2012. [Online]. Available: https://doi.org/10.1145/2331042.2331058

[5] K. Morton, M. Balazinska, D. Grossman, and J. D. Mackinlay. Support the Data Enthusiast: Challenges for Next-Generation Data-Analysis Systems. PVLDB, 7(6), 2014. [Online]. Available: https://doi.org/10.14778/2732279.2732282

[6] E. Wu, L. Battle, and S. R. Madden. The Case for Data Visualization Management Systems. PVLDB, 7(10), 2014. [Online]. Available: https://doi.org/10.14778/2732951.2732964

[7] P. Godfrey, J. Gryz, and P. Lasek. "Interactive Visualization of Large Data Sets." *IEEE Transactions on Knowledge and Data Engineering*, vol. 28(8), pp. 2142-2157, 2016. [Online]. Available: https://doi.org/10.1109/TKDE.2016.2557324

[8] A. Dadzie and M. Rowe. Approaches to visualising Linked Data: A survey. Semantic Web, 2(2), 2011. [Online]. Available: http://dx.doi.org/10.3233/SW-2011-0037

[9] N. Marie and F. L. Gandon. Survey of Linked Data Based Exploration Systems. In IESD, 2014.

[10] F. Alahmari, J. A. Thom, L. Magee, and W. Wong. Evaluating Semantic Browsers for Consuming Linked Data. In ADC, 2012.

[11] S. Mazumdar, D. Petrelli, K. Elbedweihy, V. Lanfranchi, and F. Ciravegna. Affective graphs: The visual appeal of Linked Data. Semantic Web, 6(3), 2015. [Online]. Available: http://dx.doi.org/10.3233/SW-140162

[12] N. Bikakis and T. Sellis. "Exploration and Visualization in the Web of Big Linked Data: A Survey of the State of the Art." *LWDM '16,* 2016.

[13] P. Heim, S. Lohmann, and T. Stegemann. Interactive Relationship Discovery via the Semantic Web. In ESWC 2010. Lecture Notes in Computer Science, vol. 6088. Springer. 2010. [Online]. Available: https://doi.org/10.1007/978-3-642-13486-9_21

[14] F. Benedetti, L. Po, and S. Bergamaschi. A Visual Summary for Linked Open Data sources. In ISWC, 2014.

[15] E. Pietriga. IsaViz: a Visual Environment for Browsing and Authoring RDF Models. In WWW, 2002.

[16] N. Bikakis, J. Liagouris, M. Krommyda, G. Papastefanatos, and T. Sellis. graphVizdb: A Scalable Platform for Interactive Large Graph Visualization. In ICDE, 2016. [Online]. Available: https://doi.org/10.1109/ICDE.2016.7498340

[17] T. Hastrup, R. Cyganiak, and U. Bojars. Browsing Linked Data with Fenfire. In WWW, 2008.

[18] D. V. Camarda, S. Mazzini, and A. Antonuccio. LodLive, exploring the web of data. In I-SEMANTICS, pp. 197-200, 2012. [Online]. Available: https://doi.org/10.1145/2362499.2362532

[19] K. Zhang, H. Wang, D. T. Tran, and Y. Yu. ZoomRDF: semantic fisheye zooming on RDF data. In WWW, 2010

[20] M. Bastian, S. Heymann, and M. Jacomy. Gephi: An Open Source Software for Exploring and Manipulating Networks. In ICWSM, 2009.

[21] J. Dokulil and J. Katreniaková. Using Clusters in RDF Visualization. In Advances in Semantic Processing, 2009.

[22] L. Deligiannidis, K. Kochut, and A. P. Sheth. RDF data exploration and visualization. In *Workshop on CyberInfrastructure: Information Management in eScience*, pp. 39-46, 2007. [Online]. Available: https://doi.org/10.1145/1317353.1317362

[23] C. Sayers. Node-centric RDF Graph Visualization, 2004. Technical Report HP Laboratories

[24] S. Idreos, O. Papaemmanouil, and S. Chaudhuri. Overview of Data Exploration Techniques. In SIGMOD, 2015. [Online]. Available: https://doi.org/10.1145/2723372.2731084

[25] J. Heer and B. Shneiderman. Interactive Dynamics for Visual Analysis. Commun. ACM, 55(4), 2012. [Online]. Available: https://doi.org/10.1145/2133806.2133821

[26] E. G. Caldarola, and A. M. Rinaldi. "Big Data Visualization Tools: A Survey - The New Paradigms, Methodologies and Tools for Large Data Sets Visualization." *DATA* (2017).

[27] P.K. Singh, C. Aswani Kumar, G. Abdullah, "A comprehensive survey on formal concept analysis, its research trends and applications," Int J Appl Math Comput Sci, vol. 26, no. 2, pp. 495–516, 2016. [Online]. Available: https://doi.org/10.1515/amcs-2016-0035

[28] R. Belohlavek, "Introduction to formal concept analysis," 2008. [Online]. Available: https://phoenix.inf.upol.cz/esf/ucebni/formal.pdf [Accessed on November 2020].