Bruno Azevedo. Universidade do Minho
Pedro Branco. Universidade do Minho
Oliveira e Sa Jorge Universidade do Minho
Rubén Tortosa Cuesta Universitat Poliècnica de València
Ana Alice Baptista Universidade do Minho

Interfaces for Science: Conceptualizing an Interactive Graphical Interface

Abstract

The exponential growth of Scientific Knowledge Objects (SKOs) on the Web, makes searches time-consuming. Access to the right and relevant SKOs is vital for research, which calls for several topics, including the visualization of science dynamics.
We present an interface model aimed to represent of the relations that emerge in the science social space dynamics, namely through the visualization and navigation of the relational structures between researchers, SKOs, knowledge domains, subdomains, and topics. This interface considers the relationship between the researcher who reads and shares the relevant articles and the researcher who wants to find the most
relevant SKOs within a subject matter. This article presents the first iteration of the conceptualization process of the interface layout, its interactivity and visualization structures. It is essential to consider the hierarchical and relational structures/algorithms to represent the science social space dynamics. These structures are not being used as analysis tools, because it is not objective to show the linkage properties of these relationships. Instead, they are used as a means of representing, navigating and exploring these relationships. To sum up, this article provides a framework and fundamental guidelines for an interface layout that explores the social science space dynamics between the researcher who seeks relevant SKOs and the researchers who read and share them.

Keywords: Information, visualization, Design, Interface

1 Introduction

Throughout history, the space of information flows supports, simultaneously, the dynamics of social practices and its relations [1] in all sectors and services of society. It equally supports the science dynamics [2] as it emerges in the intangible space infosphere and it is defined by flows of information [1]. Science dynamics is defined by a complex, self-organizing, and evolving relational multi-structure [2] of several axes including researchers, projects, scientific articles, ideas. Science dynamics reveals clutters, patterns and connections among scientific knowledge domains and sub domains, collaboration and citation networks [2]. Also, it is characterized by the merging of knowledge domains boundaries which in turn discloses the emergence of new fields of knowledge [2]. Therefore, science dynamics is characterized by a "complex self organizing and constantly evolving multiscale network" [2–4] of relations, which occurs in the space of information flows.

In this article it is argued that the science dynamics space is also defined by a social dimension, and it is not only characterized by network structures, but also by hierarchical structures. Although networks and hierarchies are distinct and opposing concepts, they coexist and complement each other [5]. These dynamics unveils the main architecture of science space, specifically composed of networks and hierarchies which, in turn, are the basic architecture of information [5–7]. In fact, the expression of the social science space is determined by complex interactions, among researchers and the research they perform. The social science space is the focus of this article, and the goal is to visualize the social dynamics in order to find the most relevant Scientific Knowledge Objects

(SKOs; e.g. articles, books, patents, software, disciplines) based on the reader's experience socially shared [8]. This shared experience functions as an organization/ranking mechanism. Thus, the researcher plays two interchangeable roles: First, there is the researcher (seeker), who could be new to some knowledge domain or subdomain, for instance; The second role is defined by the researcher that reads SKOs and shares his experience in the social space dimension. The objective is to design an interface layout that explores the social science space dynamics between the two researcher roles.

The main objective of this article is to present the first iterations of an interface conceptualization aimed to interactively depict the complex and dynamic social science space, based on the referenced interconnectedness and the researcher social experience.

This article is divided in five sections: the first one is this introduction to the problem; the section two describes the information deluge problematic in the context of science and digital libraries; the next section presents a brief description of related work; the section four provides the interface model and the interactivity conceptualization; and the article ends with the final considerations and some directions of future work are presented.

2 The Information Flood Problematics in Science

As a result of a number of factors, including the increased storage and processing capacity, the interconnection between different systems (internet, database) and development of new interfaces [9], accelerated the access, publication, and production of scientific content [5, 10]. 6,849.32 new research journal articles are published every day [11]. The current accelerated pace of scientific and technical discoveries and the emergence of new knowledge domains and subdomains in short periods contributed significantly to increase the number of scientific publications [9, 10].

Currently, the SKOs are stored in digital libraries that are the main repositories of scientific knowledge [9]. Digital libraries are fundamental services for accessing a broad typology of SKOs. Each document returned as a result of a query in a digital library platform is given a relevance score computed from a variety of factors (e.g. number of citations, the number of word hits). Even though the digital libraries engines allow filtering the results, in many cases the number of results is still unmanageable which calls for improvements at today's refinement and personalization algorithms.

Structuring, framing and filtering content represents an urgent and unceasing challenge [10]. In fact, if in previous times the main concern was to collect and store information or SKOs [5], one of today's digital information society main challenges is to devise and adopt strategies towards the reduction of the so called information deluge [10, 12]. In the scientific domain, the goal is to filter and reduce the volume of results and efficiently frame vast amounts of information in the researcher's cognitive and perceptual field [12, 13]. To efficiently frame a large body of information in the researcher (seeker) cognitive and perceptual field, it is fundamental to design a graphical language. One of the challenges underlying digital libraries is directly related with the information finding, filtering and visualization processes [14]. The application of advanced Information Visualization (InfoVis) techniques to access and depict the digital library as a social/interaction space is an unexplored challenging territory with few approaches. There are currently three approaches aimed at the visualization of science [9]: the first two are the static and interactive visualization which are aimed at the representation of science dynamics (e.g. citation, co-citation and collaboration (coauthorships) networks between researchers, articles, journals, and knowledge domains or subdomains); the third approach is related with exploratory interfaces, as already identified by [8] in the

next section. InfoVis constitutes a viable response to several tasks related with information structuring, filtering and finding processes, allowing a greater cognitive and perceptive efficiency [13, 15–17] [13–15, 18, 19]. Succinctly, InfoVis allows to use a graphical syntax to efficiently portray the science social space dimension and its inherent dynamics.

3 Related Work

Most of the former projects are related to the first and second approaches mentioned in the last paragraph, i.e., with static and interactive InfoVis. Some examples are the "Maps of Science", "Hypothetical Model of the Evolutions and Structure of Science", "The Structure of Science", "Map of Scientific Paradigms" or "Citespace" projects. Some projects related to the third approach, exploratory interfaces, are the GTOC, GRIDL, Envision, Antarcti.ca System Inc.'s Visual Map, Citiviz, Active Graph, Result Maps and VIDLS [8]. These projects were innovative and relevant in their time and provided the basis for new approaches more aligned with the approach described in this article. The work proposed by [8] provides a description and comparison of five significant projects with relevant use of information visualization techniques. The following four of these projects are applied to science: Well-Formed Eigenfactor, Apolo, Citeology and PaperQuest. In common these projects aim to develop exploratory interfaces using various techniques to represent and explore the dynamics of science. The remaining project, Mace, also uses a visual exploratory interface not for science but for architecture, it is relevant because it uses social taxonomies to classify topics.

These projects provide solutions for viewing patterns and trends, specifically interfaces aimed at the visualization of scientific network knowledge structures based on citations. Despite their objectives and their remarkable results, they are not well suited to explore SKOs when only a small subset of SKOs is relevant to the seeker researcher.

4 Designing the Interface Layout

The reading and publication of articles, books and other SKOs are part of a researcher's daily life. If a given number of SKOs in a given topic is handled by researchers, it is reasonable to infer that when each researcher reads, shares, comments, tweets, among other interactions, a related informal body of knowledge emerges. Our hypothesis is that this potentially emerging body of knowledge provides evidences about the existence of a social science space dynamics that plays a fundamental role in filtering the most relevant SKOs within a given topic.

In this sense, the equated hypothesis constitutes a new paradigm that determines a change in the focus of the approach, usually centered on the citation metrics and not on the researchers social experience.

The concept of our interface is based on the relationship between researchers, the articles they consider relevant and the subject interests of those who are accessing the interface by consulting this information. Therefore, in the next sections, we discuss the interface layout, its interactivity and its visualization structures. These structures are not being used as analysis tools, since we are not interested in showing the linkage properties of these relationships. Instead, they are used as a means of representing, navigating and exploring these relationships.

The great difference between our approach and other seemingly similar ones is that we relate to the researcher the articles he or she reads and share, while other projects are more focused on the authors and their production (e.g., citation/collaboration networks,

co-authorship networks). Our approach is reader centered while other projects are author-centered.

4.1 Interactivity and Graphical User Interface: Brief Considerations

In this section, a first iteration regarding the interface layout conceptualization, and underline a succinct explanation about the interactivity logic are presented. The interface is constituted by five main views: Global View, Researchers View, Researcher Tree View, Knowledge Subdomain View and Subdomain Researchers View. The graphical layout is subdivided in three main sections (Fig. 1): the top bar(main buttons and thesearch field box), the visualization section, and the reader researcher's and SKO's metadata section.

The Global View (Fig. 1) provides access to the most relevant SKOs within a knowledge domain. The objective is to visualize the most relevant SKO's that belongs to different knowledge subdomains within a knowledge domain. This view provides arious treemaps, where each treemap corresponds to a specific knowledge domain. Nominal information coding [20] is used to label knowledge domains, through the use of the variable color. Therefore, to each knowledge domain corresponds a fixed hue value. Gradients of the same hue provide the distinction between different knowledge Fig. 1. Global View: a color gradient identifies the SKOs knowledge subdomain (e.g. gradient of blues). By placing the mouse pointer over the SKOs areas, a descriptive tooltip about object metadata emerges in the metadata section. The bottom left rectangle provides the object relevance value (84.0 in this example). The directional buttons inside the visualization section provide the navigation menu, and plus and minus buttons provide the zoom interactivity. (Color figure online)

Fig. 1. Global View: a color gradient identifies the SKOs knowledge subdomain (e.g. gradient of blues). By placing the mouse pointer over the SKOs areas, a descriptive tooltip about object metadata emerges in the metadata section. The bottom left rectangle provides the object relevance value (84.0 in this example). The directional buttons inside the visualization section provide the navigation menu, and plus and minus buttons provide the zoom interactivity.(Color figure online)

The variable size provides the quantitative encoding (SKO's relevance). Each squarified area corresponds to a specific SKO. This means that it is possible to visualize in the

treemap the most relevant SKOs within a knowledge subdomain. The size of the squarified area will be correlated with a new metric to be developed in the future. A limited number of objects should be defined inside the treemap (e.g. the 25 most relevant SKOs). The hierarchy depth of the treemap will be a subject addressed in future work.

The Researchers View (Fig. 2), portrays graphical information about the most active research readers within a range of knowledge subdomains. This is graphically represented by a ranking of the most active readers within different knowledge subdomains (e.g. the 25 most active readers in Open Data). The circles and the graphical variable color represent different knowledge subdomains, and describe the total number of SKOs based on the relevance metric: the greater the relevance, the larger the diameter of the circumference. By clicking on the reader researcher name the user jumps to the Researcher's Tree View (Fig. 3).

The Researcher Tree View (Fig. 3) depicts the hierarchical relation between the research readers and the knowledge subdomains. The center node represents the researcher and the second layer nodes represent the knowledge subdomain fields (e.g. blue node). The third level of edges/nodes are the SKOs read and shared, which are children of the ancestor knowledge subdomain node (second layer nodes). The edges translate the readers relation between the knowledge subdomain and the various SKOs of that knowledge subdomain. The outer circle visually limits the number of SKOs organized by subdomain that a specific reader researcher read and shared. It should be noted that each SKO will act as a direct link to where the SKO is stored. The Knowledge Subdomain View (Fig. 4) provides the total number of read SKOs within a range of knowledge subdomains. The main objective of this view is to list knowledge sub-subdomains categories within a knowledge subdomain. By clicking on the knowledge subdomain name the user jumps to the Subdomain Researcher's View (Fig. 5).
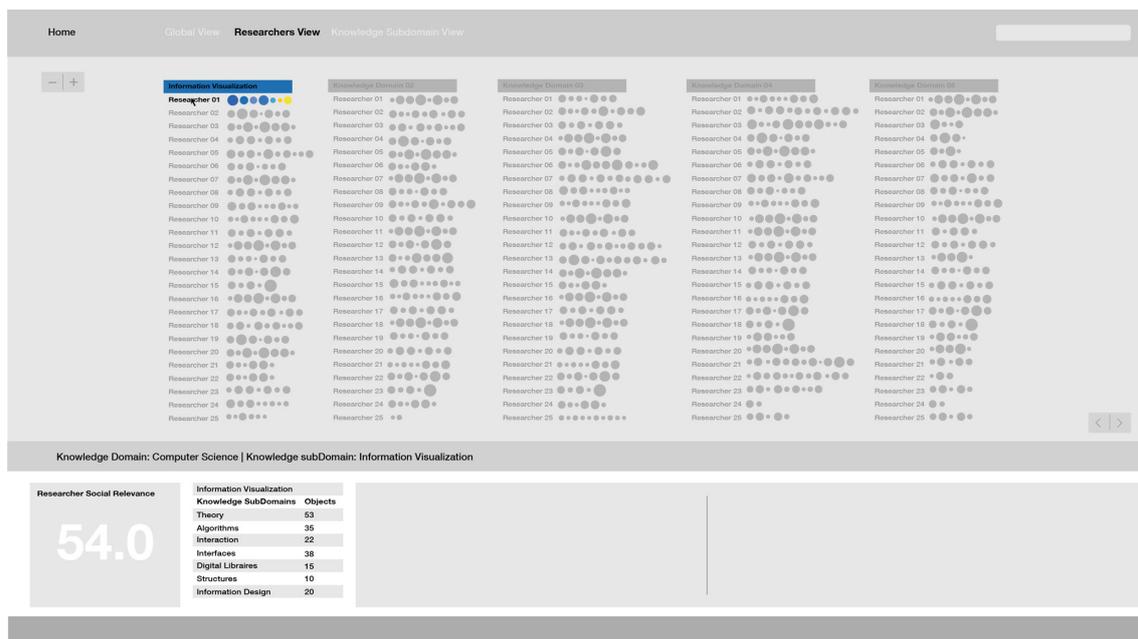
Fig. 2. Researchers View: by placing the mouse pointer over the reader name, a descriptive tooltip about research reader metadata is showed in the bottom section (e.g. ORCID number, researcher's knowledge domain and subdomains or total number of SKOs read and share by a given researcher). Making a rollover over the circles, a descriptive tooltip about SKO knowledge subdomain metadata is presented (bottom section) (e.g. type, total number of SKOs in a given knowledge subdomain). (Color figure online)
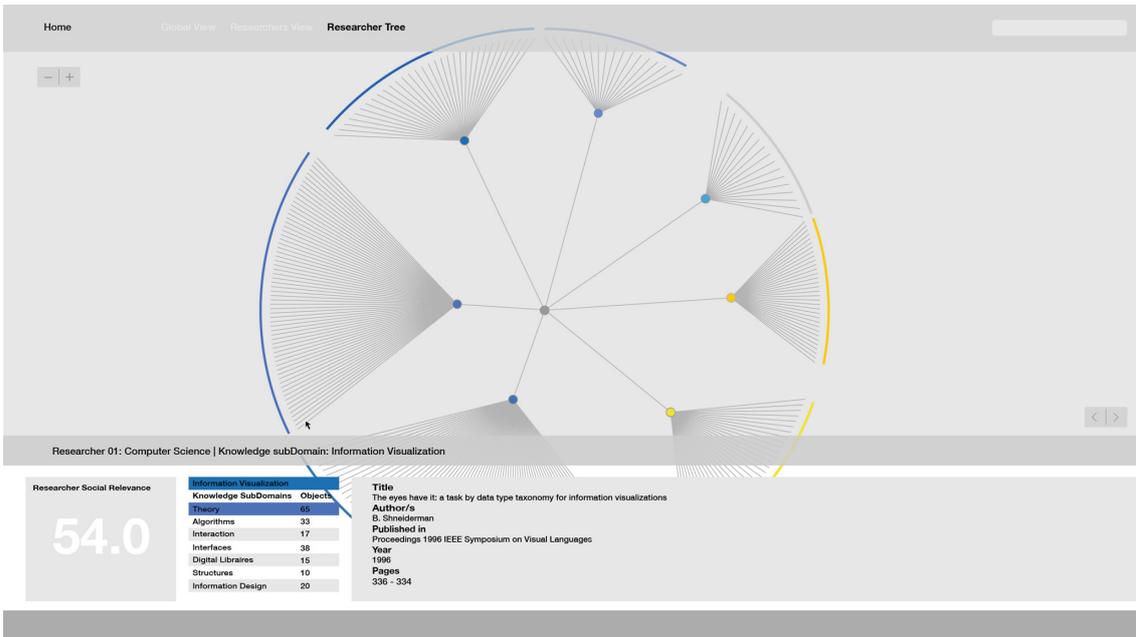
Fig. 3. Researcher Tree View: the metadata section provides a contextualization menu about the different knowledge subdomains related to the research reader. By placing the mouse pointer over the central node, a tooltip (bottom section) is displayed showing the reader metadata. By making a rollover over the second and third level of edges/nodes it is respectively displayed a tooltip (bottom section) describing the knowledge subdomain and the number of children ramification and SKOs metadata (e.g. knowledge domain, subdomain, title, DOI). (Color figure online)
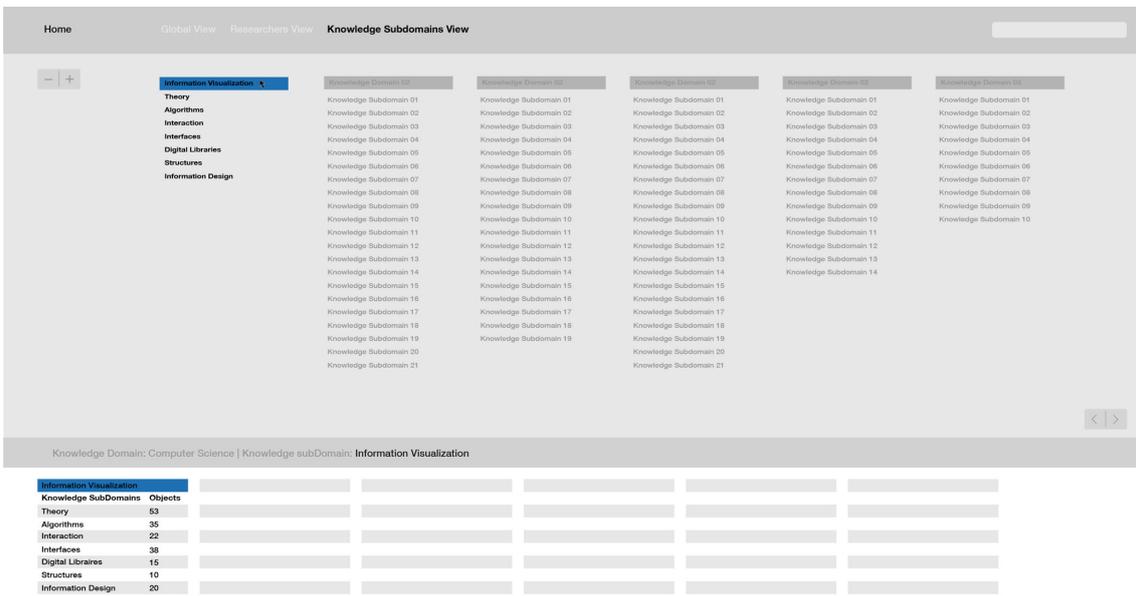
Fig. 4. The Knowledge Subdomain View provides a list of knowledge sub-subdomains categories within a knowledge domain (metadata section).

Fig. 5. Subdomain Researchers View: the represented layout reflects the interaction of a single reader. When placing the mouse pointer over the reader and SKO node, a tooltip (bottom section) provides information about the reader and SKO metadata. It also draws the relations between the reader and the SKO's.

The Subdomain Researchers View (Fig. 5) shows the most relevant research readers of a particular subdomain, as well as its relationship with several other subdomains taking into account the SKOs it reads and shares.

4.2 Objectives and Characteristics of Hierarchical and Relational

Structures The objective of the hierarchical containment structure (treemap algorithm) (Fig. 1) is to provide to the user an overview of the most relevant read/shared SKOs. Treemaps are an algorithm that allows the visualization of large hierarchies [21–24]. The treemap algorithm consists on a rectangular hierarchical structure of containment, aimed at the visualization of data in a compact layout. It is also characterized by an efficient use of the layout space when compared to horizontal or vertical structures, which are very extensive structures (Trees). The work presented in [23, 25] highlight the importance of ordering the adjacent data, providing more efficiency in the reading/finding process (readability) [26]. Another feature that is important to underline, is the rectangles with square aspects ratios [25]. This feature allows an easier distinction and selection of square shapes when compared to rectangular shapes, even when the shape proportions are similar. The advantage in using the ordered algorithm is that it preserves order while presenting small changes under dynamic updates [26]. According to [26] ordered treemaps with squarified rectangles provide legibility, usability and effortless updating. The latest increment in the squarify algorithm, and according to [27] provide rectangles with an improved aspect ratio and a higher consistency among nodes.

The objective of the hierarchical radial structure (Fig. 3) is to allow the user to visualize the researcher reader relational perspective. Specifically, the hierarchical social experience structure that emerges between the researchers and the read/shared SKOs, within various knowledge subdomains. The main characteristic of this algorithm is that it can represent large hierarchies [28] without losing context [28–30], in contrast to what happens in horizontal and vertical (linear) trees. Another feature that is important to highlight is that the nodes are arranged hierarchically in concentric circles around a central node (depth in the tree) [28, 29], which represents the selected research reader. The distribution of SKOs is determined by a specific metric to be developed and it is

hierarchical clockwise distributed. SKOs with analogous metric values are organized by year. Another issue is the number of SKOs (third layer nodes) within a knowledge subdomain node. The objective is to display a limited number of relevant nodes to reduce the visual clutter as, for instance, the 25 top SKOs (the determination of the adequate number needs further studies).

The circular structure purpose (Fig. 5) is to visualize the relational panorama between the researcher readers of a subdomain and the SKOs they read and share. The relational structure layout is composed by an outer circle that represents the researchers that have a relation with a subdomain, where the metric clockwise distributes the readers. The inner circle represents the SKOs that are clockwise arranged according to the metric. SKOs with analogous metric values will be organized by year. The edges provide a color gradation [31] between the reader subdomain and the SKO subdomains. The use of transparency is also considered: shorter edges have a higher opacity and, therefore, they are at a higher layer, and the long edges have less opacity and are at a lower layer [31]. The main characteristic of the circular layout algorithm is the reduction of visual clutter in case of a large number of edges. It is also an efficiently aesthetically technique to visualize adjacency relations in relational organized systems. It is important to underline that the work [31] provides important techniques such as the bundling strength and fundamental interaction & usability guidelines.

The next step will be the implementation of the algorithms in Processing programing language. Therefore, and as future work, it is fundamental to perform a literature review to implement both the radial and the relational algorithms.

5 Final Considerations and Future Work

This article conceptualizes an interface model/mock-up based on a previously defined architecture. This interface model/mock-up was designed to validate and guide the visualization and interaction process. The preliminary conceptualization provides a ramework of the relational structures that it will be considered in the next developments and implementations.

In what regards the visualization structures, a treemap programed in Processing language was already implemented in order to validate the treemap algorithm [32]. The future work will consist in the improvement of the squarification process based on the work of [27].

One issue that needs to be studied is the dimension of the circles of both the Hierarchical Relational View and the Circular Relational View. The greater the number of SKOs and researcher reader respectively, the greater the diameter of the circles need to be. In this sense, and conditioned by the size of the screen, issues related to managing the number of readers researchers and SKOs will have to be considered.

The next steps will consist in a deep literature review about the radial and circular layout to implement the algorithms in Processing language. It is important to underline that these algorithms are already implemented in D3 javaScript library but not open to the community. Another step will the study and implementation of interaction techniques such as Zoom, the interface validation by conducting usability tests and the validation of navigation steps.

References

1. Castells, M.: The Information Age Economy, Society, and Culture: The Rise of the Network Society, 2nd edn. Wiley-Blackwell Publishing, Hoboken (2010)
2. Fortunato, S., Bergstrom, C.T., Börner, K., et al.: Science of science. Science (80-) 359, eaao0185 (2018). https://doi.org/10.1126/science.aao0185
3. Barabasi, A., Albert, R.: Emergence of scaling in random networks. Science (80-) 286, 509– 512 (1999). https://doi.org/10.1126/science.286.5439.509
4. Albert, R., Barabasi, A.: Statistical mechanics of complex networks. Rev. Mod. Phys. 74,47–97 (2002). https://doi.org/10.1088/1478-3967/1/3/006
5. Wright, A.: Glut: Mastering Information Through the Ages. Cornell University Press, Ithaca (2008)
6. Simon, H.A.: The Sciences of the Artificial. MIT Press, Cambridge (1996)
7. Hidalgo, C.: Why Information Grows: The Evolution of Order, from Atoms to Economies. Basic Books, New York (2015)
8. Azevedo, B.M., e Sa, J.O., Baptista, A.A., Branco, P.: Information visualization: conceptualizing new paths for filtering and navigate in scientific knowledge objects. In:
2017 24o Encontro Português de Computação Gráfica e Interação (EPCGI), pp. 85–92. IEEE, Guimarães (2017)
9. Börner, K., Chen, C.: Visual Interfaces to Digital Libraries. Springer, Heidelberg (2002). https://doi.org/10.1007/3-540-36222-3
10. Thackara, J.: In the Bubble: Designing in a Complex World. MIT Press, Cambridge (2006)
11. Ware, M., Mabe, M.: The STM report: an overview of scientific and scholarly journal publishing (2015). https://www.stm-assoc.org/2015_02_20_STM_Report_2015.pdf. Accessed 14 Feb 2016
12. Wurman, R.S.: Information Anxiety 2, 2nd edn. QUE, London (2001)
13. Card, S., Mackinlay, J., Shneiderman, B.: Readings in Information Visualization: Using Vision to Think. Morgan Kaufmann, Burlington (1999)
14. Marks, L., Hussell, J.A.T., McMahon, T.M., Luce, R.E.: ActiveGraph: a digital library visualization tool. Int. J. Digit. Libr. 5, 57–69 (2005). https://doi.org/10.1007/s00799-004-0110-z
15. Mazza, R.: Introduction to Information Visualization. Springer, London (2009). https://doi. org/10.1007/978-1-84800-219-7
16. Meirelles, I.: Design for Information: An Introduction to the Histories, Theories, and Best Practices Behind Effective Information Visualizations. Rockport Publishers, Beverly (2013)
17. Börner, K., Polley, D.E.: Visual Insights. MIT Press, Cambridge (2014)
18. Börner, K., Chen, C.: Visual interfaces to digital libraries: motivation, utilization, and sociotechnical challenges. In: Börner, K., Chen, C. (eds.) Visual Interfaces to Digital Libraries. LNCS, vol. 2539, pp. 1–9. Springer, Heidelberg (2002). https://doi.org/10.1007/3-540-36222-3_1
19. Kim, B., Scott, J., Kim, S.: Exploring digital libraries through visual interfaces. In: Digital Libraries - Methods and Applications, pp 123–137. InTech (2011)
20. Ware, C.: Information Visualization: Perception for Design, 3rd edn. Morgan Kaufmann Publishers, Burlington (2012)
21. Johnson, B., Shneiderman, B.: Tree-maps: a space-filling approach to the visualization of hierarchical information structures. In: Proceeding Visualization 1991, pp 284–291. IEEE Computer Society Press (1991)
22. Shneiderman, B.: Tree visualization with tree-maps: 2-D space-filling approach. ACM Trans. Graph. 11, 92–99 (1992). https://doi.org/10.1145/102377.115768
23. Bederson, B.B., Shneiderman, B., Wattenberg, M.: Ordered and quantum treemaps: making effective use of 2D space to display hierarchies. ACM Trans. Graph. 21, 833–854 (2002). https://doi.org/10.1145/571647.571649
24. Long, L.K., Hui, L.C., Fook, G.Y., Wan Zainon, W.M.N.: A study on the effectiveness of tree-maps as tree visualization techniques. Procedia Comput. Sci. 124, 108–115 (2017). https://doi.org/10.1016/j.procs.2017.12.136
25. Bruls, M., Huizing, K., van Wijk, J.J.: Squarified treemaps. In: de Leeuw, W.C., van Liere, R. (eds.) Data Visualization 2000. EUROGRAPH, pp. 33–42. Springer, Vienna (2000). https://doi.org/10.1007/978-3-7091-6783-0_4
26. Shneiderman, B., Wattenberg, M.: Ordered treemap layouts. In: 2001 IEEE Symposium on Information Visualization. INFOVIS 2001, pp 73–78. IEEE (2001)
27. Cesarano, A., Ferrucci, F., Torre, M.: A heuristic extending the Squarified treemapping algorithm. In: CoRR (2016). https://arxiv.org/pdf/1609.00754.pdf. Accessed 6 Apr 2017
28. Book, G., Keshary, N.: Radial tree graph drawing algorithm for representing large hierarchies, University of Connecticut (2001). http://gbook.org/projects/
29. Yee, K.-P., Fisher, D., Dhamija, R., Hearst, M.: Animated exploration of dynamic graphs with radial layout. In: 2001 IEEE Symposium on Information Visualization. INFOVIS 2001, pp. 43–50. IEEE Computer Society Press (2001)
30. Sheth, N., Cai, Q.: Visualizing MeSH dataset using radial tree layout (2003)
31. Holten, D.: Hierarchical edge bundles: visualization of adjacency relations in hierarchical data. IEEE Trans. Vis. Comput. Graph. 12, 741–748 (2006)
32. Monteiro, A., Miguel, B.: Harnessing user's knowledge in the construction of rating flows: the design of a collaborative system applied to academic repositories. In: Ortuño, B.H. (ed.) 2016 6th International Forum

of Design as a Process: Systems & Design Beyond Processes and Thinking, pp. 780–792. Editorial Universitat Politècnica de València (2016)