

Semantic Cartography in Information Retrieval Systems

Férihane Kboubi, Anja Habacha Chabi and Mohamed BenAhmed

RIADI-ENSI, Compus Universitaire de Manouba, 2010, Tunisia

Ferihane.kboubi@riadi.rnu.tn, Anja.habacha@ensi.rnu.tn,

Mohamed.BenAhmed@riadi.rnu.tn

Abstract

In this paper we examine the state of research into the domain of information visualization. We start by describing the spectrum of current paradigms of information visualization. We address the problem of their integration in the web applications and precisely in information search systems. Then, we deal about the challenges and directives for a good conception of semantic visualization. Then, we discuss about the existing methods of information visualization evaluation and present the most relevant challenges facing both information visualization construction and evaluation tasks.

Keywords: *Information visualization, Representation paradigm, Interaction and navigation, Cartography paradigms, Visualization challenges, Visualization evaluation*

1. Introduction

If search engines play a dominating role for the search of information on the web, their graphic interfaces are rarely revolutionary: a single-line text field for the query, a button to validate it, and a list sometimes ordered by pertinence to display the results. These classical methods have been criticized for many reasons.

The query based search engines offer to users only one search type “*the precise search*” which supposes that the user knows exactly for what he looks for: *a precise paper knowing its title, authors and major theme*. It is not unusual for users to input search terms that are different from the index terms used by the information search engines. Various methods have been proposed to help users choosing their search terms and articulating their queries. One widely used approach is to incorporate into the information system a thesaurus-like component that represents both the important concepts in a particular subject area and the semantic relationships among those concepts [6]. Unfortunately, the development and use of thesauri is not without its own problems. The thesaurus development process, if done manually, is both time consuming and labor intensive and the produced thesaurus has often been developed for a general subject area and needs significant enhancement to be tailored to the information system where it is to be used. Some other research works tried to resolve a part of this problem by proposing methods for query expansions [14]. These methods suppose that users know exactly their needs and assist them to define their query in order to reduce the problem of noise and confusion. However, users may have the need to make search without knowing exactly to what they look for, at least at the beginning. In these cases, it will be very interesting to offer to users other search types which assist them in their searches.

The second great limit of existing information search systems is the means used for displaying the search results: *result lists*. Indeed, the result lists return an enormous quantity of information which leads to a cognitive overload of users who cannot, in the majority of the cases, consult all the returned documents [30]. An innovative idea to guide users in their searches is to propose them a method of location centered on the navigation in the informational spaces. This type of interaction benefits from an important characteristic of the human cognition: *it is easier to users to discover or to locate for what they look, than to produce formal descriptions of information which they do not have* [16]. So, navigation within maps can replace advantageously writing of queries as far as semantics, being more explicit in maps, limits the problems of confusion and ambiguity often met in the query based systems. As a primarily step to find solutions for the previously evoked problems we propose in this paper a study of the state of the art concerning the existing paradigms, challenges and evaluation methods in the domain of information visualization.

The remaining of this paper is organized as following. In section 2, we present a survey of existing cartography paradigms. Based on the study of the state of the art, we propose in section 3 some directives that must be respected for a good conception of semantic maps. In section 4, we present a survey of the existing evaluation methods of information visualization based systems and discuss about the relevant challenges relative to this visualization task. Then, in section 5, we study the real integration of cartography paradigms in existing information search systems.

2. Survey of Cartography Paradigms

The field of Information Visualization is influenced by many different research domains including psychology, semiotics, graphic design, and art. The goal of information visualization is generally defined as providing useful tools and techniques for gaining insight and understanding in a dataset, or more generally to amplify cognition [8]. In the last few years, several paradigms of knowledge and information cartography were proposed in the literature. These paradigms were classified in three categories [45]: the representation paradigms, the visualization paradigms and the interaction paradigms. In the following we present a synthesis of these three categories of paradigms.

2.1. Representation Paradigms

The representation paradigms allow representing the structure of the information. Several and different representation paradigms were proposed in the literature, we classified them into five categories according to the type of information structure they support (See Figure 1):

- The representation paradigms supporting the tabular structure are used to represent the characteristics and the attributes of the entities [21], [33][70][71][72].
- The representation paradigms supporting the agglomerative structure (Themescape [50]) are used to represent agglomeration of entities which are grouped together according to certain criteria.
- The representation paradigms supporting the treelike structure are used to represent the hierarchical relations between the different entities in a single tree or in multiple tree visualization [60]. Treelike structure are commonly represented using a node link

approach (cone tree [35], bifocal tree [5], radial tree [45], Ring tree [42], Botanic tree [69]), an indented list approach [59] or by containment approach (Beamtrees [48], DocuBurst [55] [74], Sunburst [58]).

- The representation paradigms supporting the graph structure are used to represent different type of relations between the entities (associations, connections) [75]. There are two main approaches for representing graph structures: the matrix representation [56] and the node-link representation (VisuGraph [29], Hypergraph (<http://hypergraph.sourceforge.net/>) and TouchGraph systems).
- The representation paradigms supporting the temporal structure are used to represent the time dimension which could be modeled linearly (ThemeRiver [19], Linear representation [73]) or periodically (spiral representation [7])

In Table 1 we give some strengths and weaknesses of the presented representation paradigms.

2.2. Visualization Paradigms

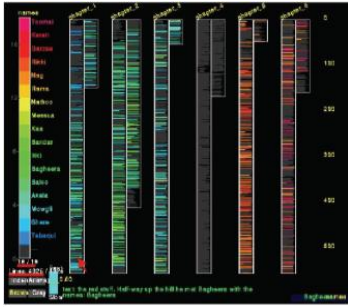
The visualization paradigms concern the means of displaying the information representations in a clear and coherent way on a limited space of visualization so that a person can become aware quickly of the presented information. The visualization techniques are classified in two groups (see Figure 2):

- The uniform visualization techniques which are based on geometric transformations of the map like zoom, translation, rotation [23].
- The not uniform visualization techniques which display the elements on the map with a variable level of details according to the interest of the user. These techniques include bifocal visualizations like document lens [34] and the elusive walls [32], and polyfocal visualizations like fisheye visualizations [28] [36] [2]).

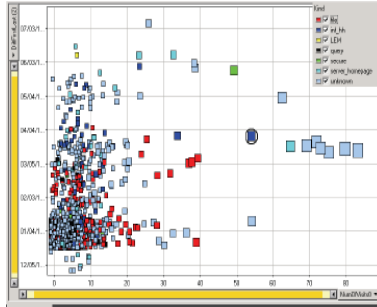
In Table 2 we give some strengths and weaknesses of the presented visualization paradigms.

2.3. Interaction Paradigms

The interaction paradigms concern techniques put at the disposal of users to interact with the produced visualizations. Interaction paradigms could be classified in three types of techniques according to the goal of the user: Overview techniques (Reducing Data Quantity [9], Miniaturizing Visual Glyphs [25]), Navigation techniques (Zoom+Pan, Overview+Detail [61], Focus+Context, semantic zoom [4][17]) and Interaction techniques (Selecting, Linking, Rearranging and Remapping, Filtering[20]). Some interaction paradigms are closely related to the representation paradigm used in the visualization. For example, in information disk representation paradigm we can find specific interaction techniques like angular zoom navigation, external detail navigation, internal detail navigation [41], circular and radial distortion [51]. Examples of interaction paradigms related to treelike and graph representations are node centering, pruning and expansion [5] [42].



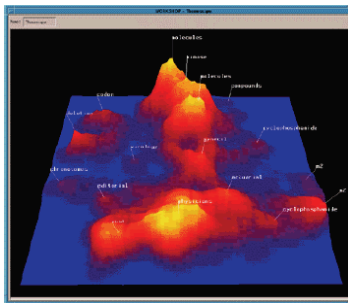
Seesoft : Unidimensional space structure representation [72]



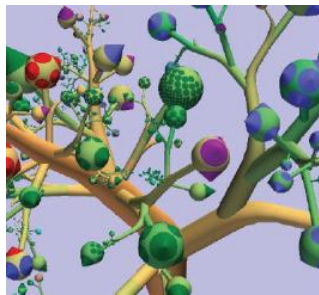
SpotFire : Bidimensional space structure representation [70]



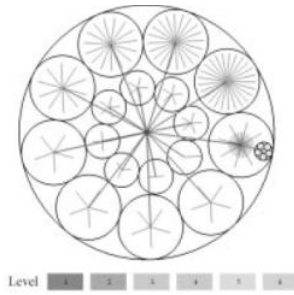
Star diagram :
 Multidimensional space structure representation [71]



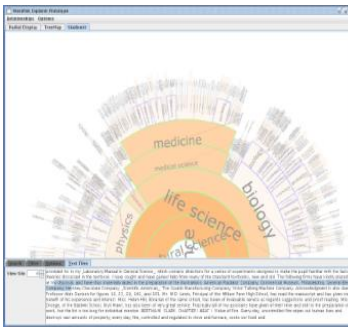
Themascape [50]



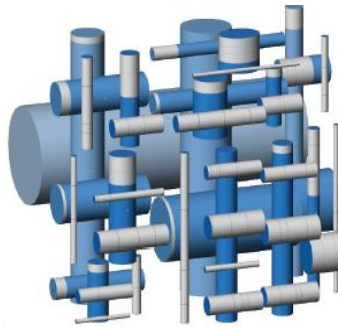
Botanic tree [69]



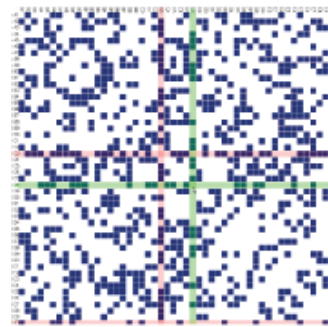
Ring System [42]



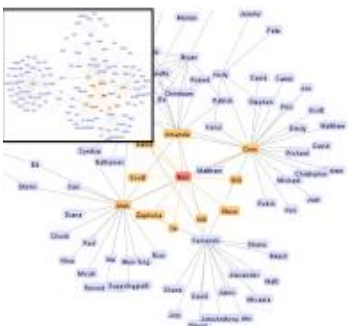
DocuBurst [74]



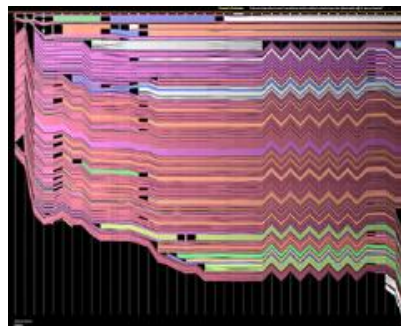
Beamtrees [48]



Matrix representation [56]



[75]



Linear representation [73]



Spiral representation [7]

Figure 1. Examples of Representation Paradigms

Table 1. Strengths and Weaknesses of the Representation Paradigms

Representation paradigms	Characteristics (+) strengths (-) weaknesses
TablesLens[33]	+Apply deformations in order to magnify certain ranges of cells and emphasize them - Do not reveal the underlying structure in the data
Parallel Coordinates [21]	+Allow to discover underlying structures in the informative space - It is difficult to understand every case in detail independently of the others
Chernoff Faces[57]	+ It exploits the habit and the strong capacity of the human to perceive very light changes in the facial expressions - The number of variables which we can represent is limited - Interpretation of the meaning is not intuitive; requires an important cognitive effort
Matrix Representations [56]	+ Avoid the problems of occlusion - It is very delicate to represent several not binary relations with a single matrix. - Do not allow to understand every entity independently of the others
Node-Link Representations [5][29 [35]	+ Reveal underlying structures in the data and more particularly the related and strongly related constituents - Problem of occlusion due to the tangle of the links. As soon as the size of the graph or the density of the links increases, it becomes very difficult for the user to visually explore the graph and to interact with its elements
Containment Representations [48][55][58]	+ Allow to visualize big trees +Avoid the problems of occlusion -The representation of the structure is not explicit - Cognitive overload (visual exploring of the structure requires a detailed reading)

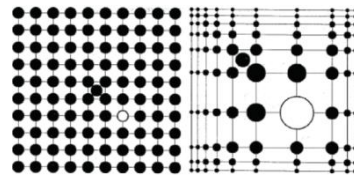
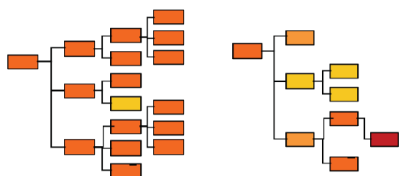
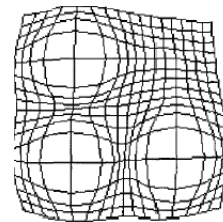
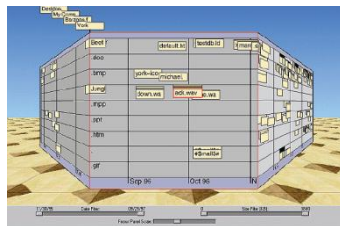
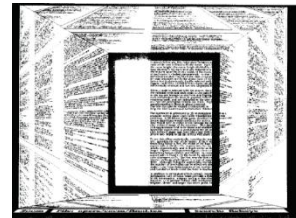
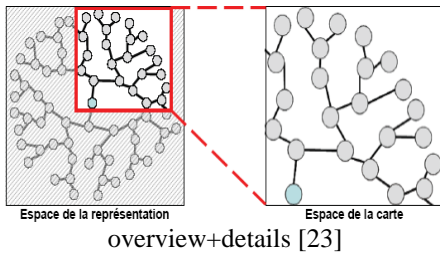


Figure 2. Examples of Visualization Paradigms

Table 2. Strengths and Weaknesses of the Visualization Paradigms

Visualization paradigms	Characteristics (+) strengths (-) weaknesses
Zooming in a Single View [23]	- The user is forced to change permanently of point of view what engenders the cognitive overload of the user
Multiples view [23]	+ Avoid the inconvenience of the cognitive overload caused by the change of zoom in a single view - Create a spatial discontinuity with several points of focus, what leads to another problem of cognitive overload for users
Elusive walls [32]	- Allows the visualization of one-dimensional vector spaces only
Polyfocal transformation [28]	+ Allows to obtain a visualization with several points of focus
Fisheye [2]	+ Filter the display of the information with regard to their degree of interest + The user can concentrate on his center of interest and use his peripheral vision to watch the remaining space of the map - Due to the effects of the deformation, visual components are not aligned and sometimes overlap - Unpredictable Movement of elements during the manipulation of the structure. This engenders the cognitive overload and the disorientation of users by losing temporarily the perception of the context

3. Challenges for the Conception of Visualizations

The conception of semantic visualizations is a very challenging task. One of the most important problems, during the conception of semantic maps, is not to associate a visual structure to every type of objects (for example a rectangle or a circle) but to distribute them in the space of the map knowing that this distribution has an impact on the meaning perceived. Besides, a good paradigm of cartography has to find a balance to the compromise of the quantity of information to be visualized and the legibility of the map. Indeed an excess of precision filled the cognitive environment with noise and forces the user to supply a big cognitive effort to filter the information from the useless details. A map must be conceived so that it represent and reflect the maximum of information with the minimum of cognitive effort supplied in a minimum of time with least ink onto the smallest surface [3] [46]. A map has to be also multivariate, showing simultaneously several variables.

There are many other challenges which make difficult the conception of an efficient visualization such as the complexity and the quantity of the concerned information [24]. Indeed, the quantity of the treated information could be enormous and these informations can have several types and different structures.

But, the biggest challenge of the conception of visualization is that there is no strategy of “ideal” visualization; the conception is always specific to the application. Different systems are efficient for users having different backgrounds and needs (expert or novice, scientist or general information). A universal model is difficult to be generalized. However there is certain number of criteria which, when they are present, improve sharply the quality of the map. Below we give some general directives for a good conception of maps. These directives, we classified them according to their natures in three categories: semantic directives, cognitive directives and technical directives.

The *semantic directives* concern especially the analysis of the informative space and are interested in the following criteria:

- *Consideration of the domain semantics*: it would be necessary to conceive a graphic representation which is intuitive and which represents semantics of the domain and the structures of the knowledge.
- *Multidimensional visualization*: often we cannot visualize on the same map all the knowledge required by the user. So, organizing the visualizations according to several axes would be doubtless very useful. Every axis has to display a facet or a point of view different from the others to satisfy a specific need of users.
- *Management of the big quantities of information*: how to display a very big quantity of information on a limited space? This is one of the fundamental problems of the researchers in the field of the information visualization. To answer this question, several solutions can be considered. These solutions share the same principle which is to build a global view of the space. Two approaches are possible. The first one consists of reducing the quantity of information to be visualized by aggregation or by filtering. The second consists of reducing the size of the visual used representations (miniaturization).

The goal of the *cognitive directives* is to minimize the cognitive effort made by the readers. These directives are relative to two criteria:

- *User familiarization with the produced visualizations*: this is dependent to the knowledge of the readers about the reading-keys of the visualizations.
- *Consideration of user's needs and profiles*: Humbert et al [18] assert that the efficiency of a representation is defined according to its use: “*from a set of data shared by several persons, each of them can require a different and fragmentary point of view on these data, according to the objective at which it aims and its need in information.*” Besides, each type of user must be represented by a predefined set of needs. But the user builds gradually and individually his knowledge by the collected information, so that his needs evolve [10]. The problem is then to find a means to enrich this model by taking into account the evolution of the individual needs.

The *technical directives* concern the process of visualization and are interested in the following criteria:

- *Management of the space of the screen*: during the conception of a map it is necessary to think to exploit the maximum and most adequately possible the vacuum of the screen without cluttering the interface. So, it is necessary to eliminate from the screen the elements which have no informative or operational utility and to choose one paradigm of representation which exploits well the space of the map.
- *Optimization of the interaction modes*: a static map is either too much cluttered (if we want to integrate in it the maximum of information), or too restrictive (if we want to have a clear and readable map). To keep the interface at an optimal level of complexity, the user must have the possibility to move interactively between the various layers of information and to reduce dynamically the number of shown variables. It is also necessary to use an interaction language which is intuitive and substantial which supports and increases the cognitive flow. To optimize the interaction modes it would be necessary to provide a variety of tools for the navigation, the search and the exploration. The interaction has to offer the maximum of features with the minimum of operational difficulties.

– *Management of the contextual stability*: It is necessary to make sure that every presented information is contextualized in term of whom, what, when, where and how it is produced [13]. The contextual stability requires, in the first place, the authorization of the user to choose the information to be shown to satisfy its needs. Secondly, visualize most details possible (defined by the user) as long as the map remains readable. So, it is necessary to make sure that the visualization of the context remains stable to promote the understanding and the familiarization. For this it is necessary to keep easily available views of the global context and the details of the center of interest, for example hidden in the background, in the neighboring window or integrated directly into the map.

4. Evaluation Methods of Information Visualizations

The task of information visualization evaluation aims at determining the performance of visualizations by estimating some criteria such as: functionality, effectiveness, efficiency, usability and usefulness. This allows answering questions like: To what extent the system provides the functionalities required by users? Does the visualization provide value? Do they provide new insight? To what extent the visualization may help users in achieving a better performance? How easily users interact with the system? Are the information provided in clear and understandable format? Is the visualization useful? How may benefit from it?

Current evaluation practices are based on methodologies established in Human Computer Interaction (HCI) which fall into two types: analytic evaluations and empirical evaluations [31].

The Analytic evaluation methods come from psychological models of human information processing and are based on studies of human cognition and behaviour. They are performed with expert-based methods such as *heuristics evaluations* (where an expert evaluates an interface and judge its compliance with recognized usability principles called “heuristics” [54]), or *cognitive walkthroughs* (where an expert walks through a specific task using a prototype system, thinking carefully about potential problems that could occur at each step) [43]. They are also used to evaluate usability and accessibility issues. Analytic evaluations usually occur during the system's design and are oriented to identify problems and guide modifications during the development of a system. In [54] the authors used three sets of previously published heuristics to assess a visual decision support system that is used to examine simulation data. The meta-analysis shows that the evaluation process and results have a high dependency on the heuristics and the types of evaluators chosen. Zuk et al describe issues related to interpretation, redundancy, and conflict in heuristics and provide a discussion of generalizability and categorization of these heuristics. We can distinguish from this work between three categories of heuristics: **Perception heuristics** [49] [44] like (Ensure visual variable has sufficient length, Don't expect a reading order from color, Color perception varies with size of colored item, Consider people with color blindness, Preserve data to graphic dimensionality, Put the most data in the least space, Remove the extraneous (ink), Provide multiple levels of detail, Integrate text wherever relevant), **Usability heuristics** [38] like: (Zoom and filter, Overview first, Details on demand, Relate, Extract, History) and **Discovery process heuristics** [1] like: (Expose uncertainty, Concretize relationships, Determination of Domain Parameters, Multivariate Explanation, Formulate cause & effect, Confirm Hypotheses).

The Empirical evaluation methods (also known as *user studies*) involve real users in the study and allow designers to obtain qualitative and quantitative data [43] [39]. Usually they are performed with system already implemented (in form of prototypes or demonstrators), as they are suitable to make formal claims. Empirical evaluation can be done by quantitative studies and/or by qualitative studies. *The Quantitative studies* consist of an analysis of determinate hypotheses tested through direct measurements [11]. This requires the definition of one or more variables related to the hypotheses examined and a metric associate to each of them (time required to learn the system, time required to achieve a goal, error rates, retention of the use the interface over the time). The evaluation is carried out usually by the means of *controlled experiments* (also known as *experimental studies*) [22]. They consist of asking the user testers to run a task and performing some measurements using observation, and completing the study with questionnaires or interviews.

The Qualitative studies involves the analysis of qualitative data, which may be obtained through questionnaires, interviews and observing users using the system, to understand and explain social phenomena. They are opposite to quantitative methods used in experimental studies for their ability to analyse phenomenon from the point of view of the participants that it is largely lost when textual or analytical data are quantified [26]. The combination of both qualitative (e.g. focus group [31] and individual interviews) and quantitative methodologies (e.g. experimental study) is known as *cross examination* or *triangulation* [27] and it is appropriate to InfoVis applications because it allows to examine data gathered with different and complementary ways, establish commonalities or differences and to provide rigour to the study [12].

The evaluation of information visualization is a very problematic task [37] [40][47]. Several challenges rise when researchers conduct an infoViz evaluation. These challenges can be related to many factors: the context of use, participant gathering, data collection, existence of evaluation environment (standard, reference tool for comparison, etc.). In the following we synthesize some of the most important challenges.

- *Integrating Tools in Daily Work Processes*: Tools have to be stable, robust to changing data sets and tasks, and – if they replace previous tools – should support the functionalities of the tools being replaced.
- *Choosing an Evaluation Context*: There may be many teams with similar data analysis tasks and data types that can collaborate in the evaluation process but the qualitative results collected during the evaluation can be vastly different.
- *Finding Domain Expert Participants*: Getting domain experts for studies is generally difficult.
- *Attachment to Conventional Techniques*: By working with their traditional tools over a long period of time people may be very accustomed to them. This may lead to certain reluctance to learn a new system. In addition, some domain experts may have learned to master complex tools and data analysis tasks over the years. If the designed tool significantly simplifies a specific data analysis compared to a previous tool these experts may be stripped from their respected expert status so that other peoples can also conduct the same tasks [21]. These issues can complicate acquiring participants for the evaluation studies.

- *Getting the Data*: To evaluate a visualization tool the evaluator may have to deal with issues of interoperability between different data sources on different machines and within different work groups. Additional challenges can be raised by different data versions, different or inappropriate format, unmaintained sources, and most importantly security restrictions.
- *Confidentiality of Information*: large companies often have confidentiality guidelines and restriction policies (Intellectual Property Rights (IPR) security requirements) that might forbid certain recording techniques and the publication of the results.
- *Complex Work Processes*: One important goal in information visualization is to support people in solving complex tasks. For this purpose, an important first step is to understand current data analysis problems with pre-design evaluation [14] which may be difficult for an outsider.
- *Existence of comparable (reference) systems*: Relatively few systems in the area are designed to address the exact same problem. When two systems have differing goals and objectives, even if only slightly different, it is difficult to perform direct comparative evaluations of them.
- *Technical performance*: It is difficult to separate a technique from its specific system implementation. That is, the accompanying user interface and set of basic user interaction capabilities of a system can strongly influence the utility of said system and people's perceptions of the value of the system.
- *The lack of standard*, accepted evaluation tests and techniques in the field of information visualization is yet another challenge. Such tests and techniques would help researchers who are not so experienced in evaluation to assess and evaluate their new systems. The variety and diversity of information visualization techniques may make standardized evaluations unlikely and impractical, however.
- *Formalization of outcomes*: When a software tool addresses a problem with concrete, easily-quantifiable outcomes, then evaluation of the tool is relatively straightforward. For instance, when we can clearly determine if the user of a system has achieved a desired outcome, then evaluation of the system is made easier. In the field of information visualization, however, outcomes of system use are difficult to clearly articulate and quantify.

5. Information Retrieval Visualizations

In a traditional information retrieval system, information retrieval is primarily keyword-based search and the search process is discontinuous because users have no control over the internal matching process. The internal matching process is not transparent to users, search result list presentation is linear and has a limited display capacity, relationships and connections among documents are rarely illustrated, and the retrieval environment lacks an interactive mechanism for users to browse. These inherent weaknesses of traditional information retrieval systems prevent them from coping with the sheer complexity of information needs and the multitude of data dimensionality [63].

To remedy to these problems a new emerging trend is to integrate information visualization techniques in the information search process. According to Gershon and

Page [16] the visualization *amplifies the cognition* and it allows to users and readers to observe, to understand and to make sense of these information.

Information retrieval visualization refers to a process that transforms the invisible abstract data and their semantic relationships in a data collection into a visible display and visualizes the internal retrieval processes for users. Basically, information retrieval visualization is comprised of two components: visual information presentation and visual information retrieval. The visual information presentation provides a platform where visual information retrieval is performed or conducted.

According to Zhang there are three information retrieval visualization paradigms [52]:

- The *QB* paradigm (Query searching and Browsing). An initial regular query is required to submit to an information retrieval system to narrow things down to a limited search results set, then the search results set is visualized in a visualization environment. Finally, users may follow up with browsing to concentrate the visual space for more specific information.
- The *BQ* paradigm (Browsing and Query searching). For the *BQ* paradigm, a visual presentation of a data set is first established for browsing. Then users submit their search queries to the visualization environment and corresponding search results are highlighted or presented within the visual presentation contexts.
- The *BO* paradigm (Browsing Only). This paradigm does not integrate any query searching components.

These three paradigms presuppose the integration of the information visualization paradigms in the information search systems. However, in spite of the variety of the cartographic paradigms proposed in the literature, their concrete integration on web applications remains however very limited and this for two main considerations. In the first place, from the user point of view, numerous are the ones who are not familiarized yet with these new paradigms [30]. Secondly, as regards to material and software configurations, a big part of equipments connected on the net are not adapted to this type of applications.

Nevertheless, the evolution of hardware performance and the considerable development in the domain of interactive information visualization for years, allowed the emergence of new systems integrating techniques of information visualization with varied levels. Among these systems let us quote Kartoo (<http://www.kartoo.com>), Toolnet (<http://www.toolenet.com>), Ujiko (<http://www.Ujiko.com>) and Grokker (<http://www.grokker.com>).

All these systems are based on query definition as a search mode, but as output they offer to users a graphical result maps. The common limit to all these systems is that the means given to users to interact with the produced maps remain elementary (selection, zoom). There are no mean of semantic interaction and navigation in the informational space visualized by the maps.

Based only on a query search mode, these systems support only a single search type which is the precise search (where the user knows exactly for what he looks). It would be very useful to propose to users other search mode [62] guiding and assisting them in their searches and allowing them to navigate in the produced maps to refine their searches and to discover new knowledge.

It would be very interesting to integrate in a visual context search types such as “*thematic search*” (allowing the user to navigate in the corpus according to a particular

theme [15][45]), or “*connotative search*” (allowing the user to discover the associated and similar concepts of his interest concept [53][47]) or “*explorative search*” (allowing the user to make an idea about the content of the corpus; and after a preliminary consultation that he will exactly define his needs).

In [67][68] the authors present an *explorative and thematic search platform based on the semantic mapping paradigms*. The idea is to offer to users graphic maps in which they navigate to look for the information which interest them. This allows helping users to go beyond their difficulties for formulating their needs using queries. The architecture of this platform is divided into two constituents. The first one concerns the descriptive, conceptual and thematic annotation and indexation of the textual corpus. The second constituent, concerns the problem of exploring textual documentary space using graphical visualization and semantic navigation techniques.

Several recent research works are undertaken in information visualization domain. The study presented by Masquillier and Cuxac [64] aims to analyzing in a synthetic way a scientific domain. It offers a representation mode of the scientific literature using thematic maps. These maps allows an expert, in this study an ecologist, to visualize the landscape of the scientific ecology such as it appears in 2010 and ten years ago in the scientific publications and to identify the relations between the search themes.

A similar research was realized by Collignon and Cuxac [65] within the framework of the valorization of the medical scientific production of African countries. The results of these two research works are visualized as maps developed using Neurodoc module of Stanalyst platform <http://stanalyst.inist.fr/>.

In [66] Samuel Szoniecky discuss about what makes the symbolic languages limited to face the knowledge management stakes in particular in front of the information ecosystems vitalisme. Szoniecky develops a proposition of allegorical language to show how an analogical approach is better adapted to the knowledge management through Web. He develops an application which uses an allegorical agent to put in relation two catalogs of bookmarks resulting from the application of management of folksonomies www.delicious.com.

6. Conclusion

In this paper we present a survey of the domain of information visualization. We talk about the representation, visualization and interaction paradigms and the challenges of information visualization conception. We propose some guidelines to enhance the conception of visualizations, and we summarize the principle of the existing evaluation methods. The study presented in this paper could be considered as a first step toward the conception of new solutions for information retrieval visualizations.

This review of current work demonstrates that information retrieval visualization is still an open research topic. We can identify several research directions in this domain. However, it is in the techniques of interaction that we can expect for the most significant developments. Indeed it is at the level of the interaction that we can improve the coupling between techniques of information visualization and information retrieval.

References

- [1] R. Amar and J. Stasko. (2004) A Knowledge Task-Based Framework for Design and Evaluation of Information Visualizations. In Proc. of IEEE InfoVis, pages 143–149, Los Alamitos, USA, IEEE Press.
- [2] Bederson, B.B. (2000), Fisheye menus, Proceedings of the 13th annual ACM symposium on User interface software and technology, San Diego, California, United States, ACM Press

- [3] Bertin J. (1967) *La sémiologie graphique*. Paris : Gauthier-Villars.
- [4] Lyn Bartram, Albert Hot, John Dill, Frank Henigman, (1995) The Continuous Zoom: A Constrained Fisheye Technique for Viewing and Navigating Large Information Spaces, Proc. ACM Symposium on User Interface Software and Technology (UIST'95), pp. 207-215.
- [5] Riccardo A. Cava, Paulo R. G. Luzzardi, Carla M. D. S. Freitas, (2002) The Bifocal Tree: a Technique for the Visualization of Hierarchical Information Structures, IHC'2002.
- [6] Chen, H., Lynch, K.J., Basu, K., & Ng, T. D. (1993). Generating, integrating, and activating thesauri for concept-based document retrieval. *IEEE Expert*, 8(2), 25-34.
- [7] John V. Carlis and Joseph A. Konstan. (1998) Interactive visualization of serial periodic data. In Proceedings of the 11th Annual ACM Symposium on User Interface Software and Technology, pages 29–38. ACM Press
- [8] S. K. Card, J. D. Mackinlay, and B. Shneiderman. (1999) *Readings in Information Visualization: Using Vision to Think*. Morgan Kaufmann, San Francisco, CA, USA.
- [9] Conklin, N., Prabhakar, S., North, C. (2002) Multiple Foci Drill-Down through Tuple and Attribute Aggregation Polyarchies in Tabular Data. Proc. IEEE Symposium on Information Visualization, pp.131-134.
- [10] David, A., (1999) *Modélisation de l'utilisateur et recherche coopérative d'information dans les systèmes de recherche d'informations multimedia en vue de la personnalisation des réponses*. Mémoire HDR, France.
- [11] Dix, A., Finlay, J., Abowd, G., Beale, R. (1998) *Human-Computer Interaction*. Prentice Hall, second edition.
- [12] Denzin, N. K. and Lincoln, Y. S. (1998) *Collecting and interpreting qualitative materials*. Sage, Thousand Oaks, CA.
- [13] Eppler, Martin, Burkhard, Remo, (2004): *Knowledge Visualization*, <http://www.knowledgemia.org/modules/pub/view.php/knowledgemia-67>, accessed 24 December 2009
- [14] El AyeB Bilel (2009) SARIPOD: Système multi-agent de recherché intelligente POSSibiliste de Documents Web, PhD Thesis, ENSI, Manouba, Tunisia
- [15] Dominic Forest (2006) *Application de Techniques de Forage de Textes de Nature Prédictive et Exploratoire à des Fins de Gestion et d'Analyse Thématique de Documents Textuels non Structurés*, PhD thesis, Université du québec à montréal, Canada
- [16] Gershon N., Page W. (2001) What Storytelling can do for information visualization, In : *Communication of ACM*, Vol.44, n°8.
- [17] Hascoët, M. et Beaudouin-Lafon, M. (2001), 'Visualisation Interactive d'Information', *I3: Information, Interaction, Intelligence*, Vol. 1, n° 1, p.77-108.
- [18] Pierre Humbert, Claire François, Pascal Cuxac, Amos David (2007) *La visualisation des connaissances scientifiques : intégration des besoins des utilisateurs*, Congrès de l'ACSI / CAIS (Canadian Association for Information Science), Montréal (Québec)
- [19] Susan Havre, Beth Hetzler, Lucy Nowell (2002) ThemeRiverTM*: In Search of Trends, Patterns, and Relationships, In *IEEE Transactions on Visualization and Computer Graphics*, volume 8, pages 9–20. IEEE Computer Society Press.
- [20] Hascoët-Zizi, M. Pediotakis, N. (1996) Visual Relevance Analysis. Proc. ACM Conference on Digital Libraries (DL'96). pp. 54-62.
- [21] Inselberg, A. et Dimsdale, B. (1990) 'Parallel coordinates: a tool for visualizing multi-dimensional geometry', Proceedings of the 1st conference on Visualization '90, San Francisco, California, IEEE Computer Society Press
- [22] Johnson, P. (1992) *Human-computer interaction: psychology, task analysis and software engineering*. McGraw-Hill, London.
- [23] Jerding, D.F. et Stasko, J.T. (1995) The information mural: a technique for displaying and navigating large information spaces', Proceedings of the 1995 IEEE Symposium on Information Visualization, Atlanta, Georgia, IEEE Computer Society
- [24] Kerren, A.; Stasko, J.; Fekete, J.-D.; North, C. (2008) (Eds.), *Information Visualization : Human-Centered Issues and Perspectives, IX*, 177 p. 15 illus. in color., Softcover, ISBN: 978-3-540-70955-8
- [25] Keim, D., Hao, M., Dayal, U., Hsu, M. (2002) Pixel bar charts: a visualization technique for very large multi-attribute data sets. *Information Visualization*, 1(1), pp.20-34.
- [26] Kaplan, B. and Maxwell, J. A. (1994) Qualitative research methods for evaluating computer information systems. In Anderson, J. G., Aydin, C., and Jay, S. J., editors, *Valuating Health Care Information Systems: Methods and Applications*, pages 45 68. Sage, Thousand Oaks, CA.
- [27] Krueger, R. A. (2000) *Focus Groups: A Practical Guide for Applied Research*. Third Edition. Sage Publishing, Newbury Park, CA.

- [28] Leung, Y.K. et Apperley, M.D. (1994), A review and taxonomy of distortion-oriented presentation techniques', *ACM Trans. Comput.-Hum. Interact.*, 1, 2, p.126-160.
- [29] Loubier E., Bahoun W., Dousset B. (2007) Visualisation de l'évolution des informations relationnelles par morphing de graphe. In : *Journées Francophones Extraction et Gestion de connaissances (EGC 2007)*, Namur, Belgique, 23-26 Janvier, Cépaduès Editions, p.43-54.
- [30] Wingyan Chung, Hsinchun Chen, Jay F.Nunamaker Jr, (2003) *Business Intelligence Explorer : A Knowledge Map Framework for Discovering Business Intelligence on the Web*, *Proceedings of the 36th Hawaii International Conference on System Sciences (HICSS'03)*
- [31] Riccardo Mazza (2006) *Evaluating Information Visualization Applications with Focus Groups: the CourseVis experience*, *BELIV'06 - BEyond time and errors: novel evaluation methods for Information Visualization*. Venice May 23, ACM Press
- [32] Mackinlay, J.D., Robertson, G.G. et Card, S.K. (1991), *The perspective wall: detail and context smoothly integrated*, *Proceedings of the SIGCHI conference on Human factors in computing systems: Reaching through technology*, New Orleans, Louisiana, United States, ACM Press
- [33] Rao, R. et Card, S.K. (1994) *The Table Lens: Merging Graphical and Symbolic Representations in an Interactive Focus+Context Visualization for Tabular Information*, in Beth, A., Susan, D. et Judith, S.O. (eds.), ACM Press, p.318-322.
- [34] Robertson, G.G. et Mackinlay, J.D. (1993) *The document lens*, *Proceedings of the 6th annual ACM symposium on User interface software and technology*, Atlanta, Georgia, United States, ACM Press
- [35] Robertson, G.G., Mackinlay, J.D. et Card, S.K. (1991) *Cone Trees: animated 3D visualizations of hierarchical information*, *Proceedings of the SIGCHI conference on Human factors in computing systems: Reaching through technology*, New Orleans, Louisiana, United States, ACM Press
- [36] Spence, R. et Apperley, M. (1999), *Data base navigation: an office environment for the professional*, *Readings in information visualization: using vision to think*, Morgan Kaufmann Publishers Inc., p.333-340.
- [37] Michael Sedlmair, Petra Isenberg, Dominikus Baur, Andreas Butz, (2010) *Evaluating Information Visualization in Large Companies: Challenges, Experiences and Recommendations*, *Proceedings of the CHI Workshop Beyond Time and Errors: Novel Evaluation Methods for Information Visualization (BELIV)*. Atlanta, USA.
- [38] B. Shneiderman (1987) *Designing the User Interface: Strategies for Effective Human-Computer Interaction*. Addison-Wesley, Reading, MA.
- [39] Ben Shneiderman, Catherine Plaisant (2006) *Strategies for Evaluating Information Visualization Tools: Multi-dimensional In-depth Long-term Case Studies*, *Proceedings of the BELIV'06 workshop Advanced Visual Interfaces Conference*, Venice
- [40] John Stasko (2006) *Evaluating Information Visualizations: Issues and Opportunities*, *BELIV'06 - BEyond time and errors: novel evaluation methods for Information Visualization*. Venice May 23. ACM Press
- [41] Stasko, J. et Zhang, E. (2000), *Focus+Context Display and Navigation Techniques for Enhancing Radial, Space-Filling Hierarchy Visualizations*, *Proceedings of the IEEE Symposium on Information Visualization* p. 57-65
- [42] Soon Tee Teoh, Kwan-Liu Ma (2002) *RINGS: A Technique for Visualizing Large Hierarchies*, In *Proceedings of Graph Drawing'2002*. pp.268-275
- [43] Tory M. and Möller T. (2004) *Human Factors in Visualization Research*. *IEEE Transaction of Visualization and Computer Graphics*. Vol. 10 N. 1.
- [44] M. Tory and T. Moller (2005) *Evaluating Visualizations: Do Expert Reviews Work*. *IEEE Computer Graphics and Applications*, 25(5):8-11, September/October.
- [45] Christophe Tricot (2006) *Cartographie sémantique, des connaissances à la carte*, thèse de doctorat, Université de Savoie, France.
- [46] Tufte, Edward R. (1997). *Visual Explanations: Images and Quantities, Evidence and Narrative*. Cheshire, CT: Graphics Press. ISBN 0961392126.
- [47] Jarkko Venna, Jaakko Peltonen, Kristian Nybo, Helena Aidos, Samuel Kaski, (2010) *Information Retrieval Perspective to Nonlinear Dimensionality Reduction for Data Visualization*, *Journal of Machine Learning Research* 11, 451-490
- [48] van Ham, F. et van Wijk, J.J. (2003), *Beamtrees: compact visualization of large hierarchies*, *Information Visualization*, 2, 1, p.31-39.
- [49] C. Ware. (2004) *Information Visualization: Perception for Design*. Morgan Kaufmann, 2nd edition.

- [50] Wise, J.A., Thomas, J.J., Pennock, K., Lantrip, D., Pottier, M., Schur, A. et Crow, V. (1995), 'Visualizing the non-visual: spatial analysis and interaction with information from text documents', Proceedings of the 1995 IEEE Symposium on Information Visualization, Atlanta, Georgia, IEEE Computer Society
- [51] Y. Yang, M.O. Ward and E.A. (2002) Rundensteiner. InterRing: An interactive tool for visually navigating and manipulating hierarchical structures. Proc. Of the IEEE Symposium of Information Visualization 2002, p. 77-84.
- [52] Jin Zhang (2008) Visualization for Information Retrieval, ISBN: 978-3-540-75147-2 e-ISBN: 978-3-540-75148-9, Springer-Verlag Berlin Heidelberg
- [53] Junliang Zhang, Javed Mostafa, Himansu Tripathy (2002) Information Retrieval by Semantic Analysis and Visualization of the Concept Space of D-Lib® Magazine, D-Lib Magazine, October 2002, Volume 8 Number 10, ISSN 1082-9873
- [54] Torre Zuk, Lothar Schlesier, Petra Neumann, Mark S. Hancock, Sheelagh Carpendale, (2006) Heuristics for Information Visualization Evaluation, In the Proceedings of BELIV '06, Proceedings of the Workshop Beyond Time and Errors: Novel Evaluation Methods for Information Visualization, held in conjunction with the Working Conference on Advanced Visual Interfaces, Venice, Italy, ACM Press, May 23--26.
- [55] Collins, Christopher (2006) DocuBurst: Document Content Visualization Using Language Structure. Proceedings of IEEE Symposium on Information Visualization, Poster Session. Baltimore.
- [56] Ghoniem, M., Fekete, J.-D. et Castagliola, P. (2004), 'A Comparison of the Readability of Graphs Using Node-Link and Matrix-Based Representations', Proceedings of the IEEE Symposium on Information Visualization (INFOVIS'04) - Volume 00, IEEE Computer Society
- [57] Chernoff, H. (1973), 'The Use of Faces to Represent Points in k-Dimensional Space Graphically', Journal of the American Statistical Association, 68, 342, p.361\368.
- [58] Stasko, J. et Zhang, E. (2000), 'Focus+Context Display and Navigation Techniques for Enhancing Radial, Space-Filling Hierarchy Visualizations', Proceedings of the IEEE Symposium on Information Visualization 2000, IEEE Computer Society
- [59] Nation, D.A. (1998), 'WebTOC: a tool to visualize and quantify Web sites using a hierarchical table of contents browser', CHI 98 conference summary on Human factors in computing systems, Los Angeles, California, United States, ACM Press
- [60] Graham, M., Kennedy, J., (2010) A survey of multiple tree visualisation, Information Visualization Vol. 9, 4, 235–252
- [61] Card, S.K., Mackinlay, J.D. et Shneiderman, B. (1999), Overview + detail, Readings in information visualization: using vision to think, Morgan Kaufmann Publishers Inc., p.285-286.
- [62] M. Cluzeau-Ciry, Typologie des utilisateurs et des utilisations d'une banque d'images, Le documentaliste, 25(3):155-120, 1998.
- [63] G.Madhu, A.Govardhan, T.V.Rajinikanth, Intelligent Semantic Web Search Engines: A Brief Survey, International journal of Web & Semantic Technology (IJWesT) Vol.2, No.1, January 2011
- [64] Masquillier M.-L., Cuxac P., Cartographie de l'écologie scientifique et analyse diachronique, ISKO Maghreb, 13-14 Mai, Hammamet, 2011
- [65] Alain Collignon, Pascal Cuxac, Les bases de données bibliographiques vecteur de connaissances de la production scientifique : étude des sciences médicales en Afrique, ISKO Maghreb, 13-14 Mai, Hammamet, 2011
- [66] Samuel Szoniecky, Le langage du Web du symbolique à l'allégorique vers une représentation de la connaissance en train de se faire, ISKO Maghreb, 13-14 Mai, Hammamet, 2011
- [67] Férihane Kboubi, Anja Habacha Chabi, Mohamed BenAhmed, Plateforme de Navigation Sémantique dans un Corpus Textuel, ISKO Maghreb, 13-14 Mai, Hammamet, 2011
- [68] Férihane Kboubi, Anja Habacha Chabi, Mohamed BenAhmed, Semantic Visualization and Navigation in Textual Corpus, DAS in Press
- [69] Kleiberg, E., van de Wetering, H. et van Wijk, J.J. (2001), 'Botanical Visualization of Huge Hierarchies', Proceedings of the IEEE Symposium on Information Visualization 2001 (INFOVIS'01), IEEE Computer Society
- [70] Ahlberg, C. (1996), Spotfire: an information exploration environment', SIGMOD Rec., 25, 4, p.25-29
- [71] Chambers, J.M., Cleveland, W.S. et Tukey, P.A. (1983), Graphical Methods for Data Analysis, New York, Chapman and Hall.
- [72] Eick, S.G., Steffen, J.L. et Jr., E.E.S. (1992), 'Seesoft-A Tool For Visualizing Line Oriented Software Statistics.' IEEE Trans. Software Eng., 18, 11, p.957-968.

- [73] Fernanda B. Viegas, Martin Wattenberg, and Kushal Dave. Studying cooperation and conflict between authors with history flow visualizations. In Proceedings of the 2004 conference on Human factors in computing systems, volume 6, pages 575–583. ACM Press, 2004.
- [74] Collins, Christopher; Carpendale, Sheelagh; and Penn, Gerald. DocuBurst: Visualizing Document Content using Language Structure. Computer Graphics Forum (Proceedings of Eurographics/IEEE-VGTC Symposium on Visualization (EuroVis '09)), 28(3): pp. 1039-1046, June, 2009.
- [75] Jeffrey Heer, Stuart K. Card, James A. Landay, prefuse: a toolkit for interactive information visualization, CHI '05 Proceedings of the SIGCHI conference on Human factors in computing systems, 2005