HAL
open science

Meaning Metaphor for Visualizing Search Results

Nicolas Bonnel, Alexandre Cotarmanac'H, Annie Morin

To cite this version:

Nicolas Bonnel, Alexandre Cotarmanac'H, Annie Morin. Meaning Metaphor for Visualizing Search Results. 9th International Conference on Information Visualisation (IV'05), Jul 2005, Londres, Royaume-Uni, pp.467-472. inria-00001051

HAL Id: inria-00001051

https://inria.hal.science/inria-00001051v1

Submitted on 28 Mar 2006

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Meaning Metaphor for Visualizing Search Results

Nicolas Bonnel
France Telecom, Division R&D
4 rue du Clos-Courtel, BP 91226
35512 Cesson-Sévigné Cedex, France
nicolas.bonnel@francetelecom.com

Alexandre Cotarmanac'h
France Telecom, Division R&D
38-40 rue du Général Leclerc
92794 Issy les Moulineaux Cedex 9, France
alexandre.cotarmanach@francetelecom.com

Annie Morin
Irisa, Campus Universitaire de Baulieu
Avenue du Général Leclerc, 35042 Rennes Cedex, France
amorin@irisa.fr

Abstract

While searching the Web, the user is often confronted by a great number of results, generally sorted by their rank. These results are then displayed as a succession of ordered lists. Facing the limits of this approach, we propose a prototype to explore new organizations and presentations of search results, as well as new types of interactions with the results in order to make their exploration more intuitive and efficient. The main topic of this paper is the processing of the results coming from an information retrieval system. Although the relevance depends on the result quality, the effectiveness of the result processing represents an alternative way to improve the relevance for the user. Given the current expectations, this processing is composed by an organization step and a visualization step. Then the proposed prototype organizes the results according to their meaning using a Kohonen self-organizing map, and also visualizes them in a 3D scene to increase the representation space. The 3D metaphor proposed here is a city.

Keywords: *Search result visualization, 3D metaphors, Self-Organizing Maps, Adaptive interfaces.*

1. Introduction

Available data on the World Wide Web is constantly increasing, so it becomes more and more difficult to extract the relevant information for a given search. The search engines, which are a way to represent the web to the users, return a result number so great that it is necessary to search for new methods to process these results. These methods must be more adapted thanks to: a more relevant result organization, a richer visualization interface and an intuitive navigation in the result space.

This paper deals with the processing of search results. This processing, still neglected in some information retrieval systems, is becoming more and more important and essential. It can be considered as a solution for enriching the results. It is, in fact, complementary to the search process and also a mean to increase the result "relevance" for the user. If the result quality remains a major concern, the quality of the result restitution (classification and visualization) must be taken into account too. Facing the increase of search results, it seems obvious to want to organize and to visualize them in an effective and adapted way. That explains the goal of the prototype presented in this paper, which is to provide the user with a search interface enabling him to quickly find the relevant information.

The context of our work is web search, so the documents are the web pages returned by the query. We only use the textual information of the documents. This information enables to have a vectorial representation (word vectors) of the pages, which is frequently used in the Information Retrieval field. The number of results to process must also be specified because it is crucial for the organization and visualization choices. A recent study [5] shows that 81.7% of users will try a new search if they are not satisfied with the listings they find within the first 3 pages of results. But it would be too restrictive to only consider the first 30 results (10 results per page). Indeed this study has been done on search engines with linear result visualization (ordered lists) and users may want to see more results on 2D or 3D visualizations. That is why the number of results considered in this paper is more than 30 but less than a hundred (only 50 in the snapshots of this paper).

The two main points to reach our goal are a good data

Proceedings of the Ninth International Conference on Information Visualisation (IV'05)
1550-6037/05 $20.00 © 2005 **IEEE**

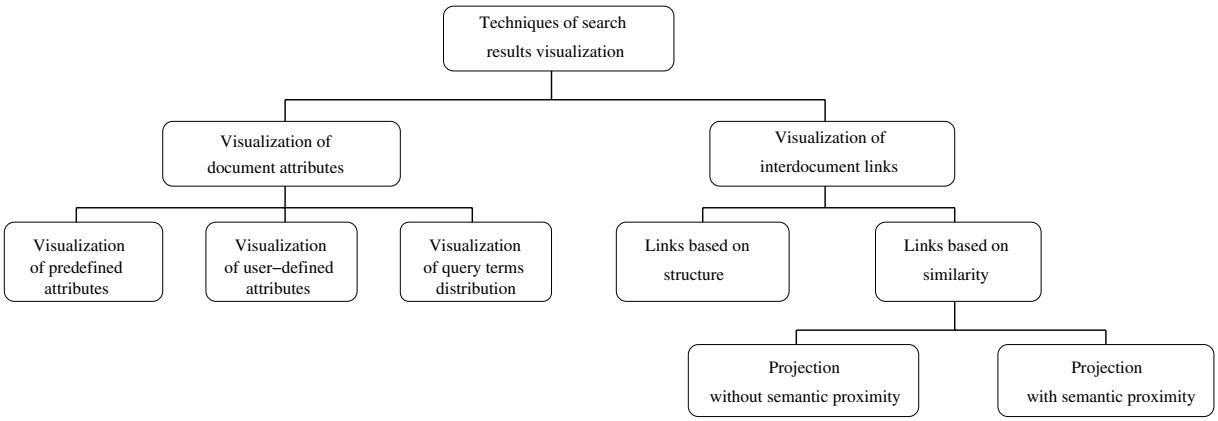

Figure 1. Taxonomy of search result visualization systems

classification and an effective visualization. Concerning these two aspects, our directions are a clustering method (the self-organizing maps) and a 3D visualization. The choice of a 3D visualization enables to increase the visualization space. Indeed, contrary to 2D, 3D offers the possibility to display a great number of results, which is not limited by the screen size but by the user's perception. 3D offers new interaction possibilities and allows to exploit the cognitive abilities of the users. So it enables us to bring a new point of view to the result visualization. However, considerable new problems appear, such as the navigation in such an environment. Finally we must remember that data visualization is strongly dependent on many criteria such as the search goal or the result number and type. One single solution for data visualization probably does not exist. That is why the prototype has many interfaces which adapt themselves according to the context.

This paper deals with the unsupervised organization of documents, the graphical representation of the results and the user interaction. It is structured as follows. The next section proposes a brief state of the art on search result visualization. Then the SmartWeb prototype is described in Section 3 and Section 4 is devoted to the prototype use. The last section allows us to conclude and gives an outlook on future work.

2. State of the art

Many works have been done on search result visualization in the last few years. The aim of this section is only to give a short overview of the main approaches in this field. As it exists many different techniques, many taxonomies of these techniques have already been proposed. But it appears that most of the classifications are inaccurate and depend on the used techniques. So we propose a taxonomy (see Figure 1) based on the previous works of [15, 2]. It is important to notice that the proposed classification only depends on the visualization goal (and also on the type of visualized objects or links). Techniques (like graphs, maps, 2D, 3D and so on) are not considered because they are particular implementations to reach the visualization goal.

The top level of the hierarchy differentiates the visualization of document attributes from the visualization of inter-document links. Attribute visualization represents the techniques used to only visualize some document attributes. Good examples are tilebars [4], visualization of the query terms distribution [15, 2], or systems based on many classical attribute visualizations [9]. But the problem with this approach is the lack of information on document similarities. So we are more interested in the other approach called link visualization between the documents. Indeed this approach gives the user information on the next document to select thanks to the inter-document links. This approach can be divided into two categories: the first one uses the links based on the document structure (hyperlinks) and the second one uses the links based on the content similarity. We focus on the second category: the visualization of inter-document similarities.

Visualization of inter-document similarities requires at least two dimensions to be effective. Indeed this kind of visualization must group similar documents together or explicitly show the inter-document links. So the techniques used in this approach have already given up the linear display of ordered lists. There are many techniques for visualizing the links between documents but the two main techniques used in our context are graphs and maps. In the graph approach, nodes generally represent web pages and arcs represent the links between web pages. Keywords are sometimes represented in these graphs as valued nodes or arcs. This approach is used in the meta search engine Kar-

1550-6037/05 $20.00 © 2005 IEEE

too [1] for which a drawback is the lack of a results overview. But with the increase of results and link complexity, graphs become more and more unreadable. In the map approach, we can generally take advantage of the cognitive aspect. Indeed the user is used to this kind of representation and therefore has *a priori* knowledge. That is why we often find geographic metaphors [12]. Among these cartographic techniques, some examples are the WEBSOM[2] project [7] which consists of computing self-organizing maps on textual data and Map.net[3] developed by Antarctica. Map.net gives a representation of ODP (Open Directory Project) using a 2D landscape metaphor. However we are still confronted by a readable problem when the results increase.

So one idea is to use 3D visualizations (which are less frequent and familiar for users) to increase the available space to represent information. The added dimension allows the display of complex graphs in a more readable way. We can also take advantage of this third dimension for replacing maps by 3D worlds: landscapes like VxInsight [1] or cities [11]. However the dimension increase makes navigation essential and especially more complex. We are facing another problem which is not obvious to solve. Finally the AVE method and its Periscope system [14] are the closest works to those ones presented in this paper. We have the use of X-VRML language and mixed interfaces (3D scene and 2D interface) in common, or also the use of many visualization metaphors which answer different goals. However our approach takes the problem of data organization in a "semantic" point of view into account. Indeed it is not sufficient in the context of web search to only order the pages according to some low-level descriptors.

3. SmartWeb prototype

SmartWeb is a prototype which is very close to a classical search engine according to the query and the database. The aim is to provide the user with the best organization and visualization of his query results, without soliciting him during the process. First we describe the global architecture of the prototype and then we focus on the two main modules which are the organization and the visualization of the documents.

3.1. Architecture

Figure 2 gives a simplified overview of the architecture of the prototype. On one hand we have the server side with the database and the different interface models. On the other hand we have the client side with the Java applet (2D part of the interface) and the VRML browser (3D scene of the interface). So the prototype access is easy because users only need to have a Java interpretor and a VRML browser.

A conceptual representation of the prototype is also proposed on Figure 2. So we can see the query and result progress between the various modules. The main steps are the organization, visualization and interaction parts. Documents (web pages) are indexed in the database whose descriptors are those frequently used by the search engines: URL, title, summary, snippet, word vector... User queries (simple or boolean) are interpreted and then sent to the database. The documents, corresponding to the query results coming from the database, successively go through two main modules: organization and visualization. These two modules are detailed in the following subsections. It is interesting to notice a certain independence between data organization and the selected metaphors for the visualization.

3.2. Document organization

In the context of this paper, it is essential to have an unsupervised method for organizing the documents. Among techniques which address this problem, one of them is particularly interesting: the Self-Organizing Maps (SOM) [6]. This choice is motivated by some properties of this method: it is a clustering method which organizes documents (or word vectors) on a map with predefined size, which guarantees a good use of space during the visualization. Moreover the obtained organization has a neighborhood concept. Indeed two neighboring documents on the map have similar word vectors. With the self-organizing maps it is also possible to have hierarchy levels or maps with dynamic size [3]. These two points can be interesting in our context but are not exploited at this time.

We detail some particularities of our implementation of the self-organizing maps. The first point concerns the word weighting. The *tf.idf* weighting [13] is used. This word weighting can be interpreted as a particular weighting of the Euclidean distance, which allows to increase or decrease the importance of some words. We use the LabelSOM method [10] for labeling the neurons. As the goal is to classify query results, an additional constraint is that the algorithm must be deterministic. Indeed the same query on the same corpus must always provide the same results. To do that, we use the batch SOM algorithm, a map with predefined size (5×5) and we always use the first data to initialize the neurons.

This SOM-based method (used for organizing the documents) is only based on word distribution and has the advantage to respect the "semantic" proximity of the documents. It also enables us to have a first abstraction level if we move on the neuron level. Then a hierarchical agglomerative clus-

[1] www.kartoo.com
[2] http://websom.hut.fi/websom
[3] http://maps.map.net

1550-6037/05 $20.00 © 2005 IEEE

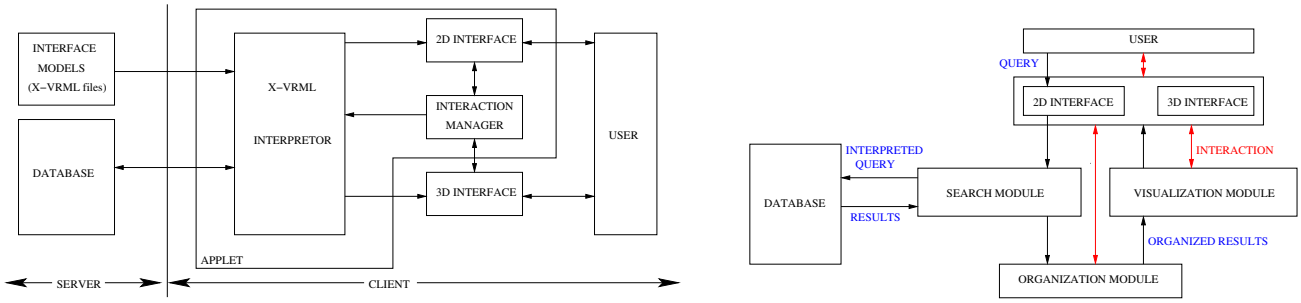

Figure 2. Simplified architecture (left) and conceptual schema (right) of the SmartWeb prototype.

tering algorithm is applied on the neurons and we only keep the classification in three clusters. So we have a second abstraction level which gives the user the three main topics of his search. These topics are defined by an arrangement of the neuron labels. These abstraction levels are not based on ontologies or document labeling. But we can consider crossing neuron labels with an ontology or using semantic data on documents during the algorithm. This approach is also open and close to other more semantic approaches like Topic Maps [8]. Indeed the self-organizing maps allow navigation in various abstraction levels like Topic Maps (topics are represented here by neurons or groups of neurons).

However one method which provides the best organization in all of the cases probably does not exist. That is why the aim of this module is to define many organizational methods and to select the most adapted to each case. For instance, it does not seem relevant to compute a self-organizing map when the number of query results is very weak.

3.3. Visualization module

The visualization module receives two input parameters: the previously organized results and an interface model (selected either by the user or automatically). It must provide output data for the applet (2D interface) and the VRML browser (3D scene). The model which defines the visualization metaphor and the used interactions, is selected in a list of interface models. However it is possible for the user to easily create his own model and thus to personalize the visualization interface and the interactions. The models are expressed in X-VRML[4] which is a meta-language (higher level than VRML) enabling to add many functions such as database interrogation or iterations. During the query processing, the model is interpreted in order to dynamically produce the VRML file which contains the 3D interface provided to the user. The visualization module also has an interesting characteristic: its adaptability. The possibility to choose the model makes the visualization adaptive. Indeed this choice can be done according to some criteria such as the result number, the search goal or the user category.

[4]X-VRML is an XML based language developed by France Telecom R&D and the Poznan University of Economics

4. SmartWeb use

First we define the metaphor word which is often used in this section: a metaphor is the realization of an association between graphical parameters of the presentation and information on the indexed documents. Then we present the prototype for a given interface: the city metaphor.

The choice of the city metaphor is mainly justified by the cognitive aspect of this metaphor. And it seems adapted to a 3D environment contrary to the map metaphor where two dimensions are enough. A first version of this metaphor was developed and a user test (on a panel of 50 users of various ages and backgrounds) was carried out on this metaphor. Upon the test results, the city metaphor has evolved. Figure 3 gives an overview of the new metaphor whose explanations can be classified in four classes: primitives, organization, visualization and navigation. This new version will also be tested to know if the users' expectations have been answered and to identify new issues.

4.1. Primitives

Each building of the city represents a web page. These buildings are grouped by districts which are placed on the ground according to a grid. Each district represents a neuron of the self-organizing map. The page relevance is represented by the building height, which enables to quickly see the best classified pages according to this criterion. As our mapping choice for the relevance does not allow to visually differentiate two successive ranks, an interval approach is adopted. So the first 5 ranks are associated to one building height and so on.

1550-6037/05 $20.00 © 2005 IEEE

Figure 3. SmartWeb prototype with the city visualization metaphor.

4.2. Organization

The districts are placed on the ground according to a 2D grid. This grid (which is square in our example) enables to carry out a classification of the search results. This classification is the result of the SOM algorithm previously computed by the organization module. Indeed the 2D grid on the ground is the same one as the SOM grid. This organization has two interesting properties: the unicity of the web pages (or buildings) in the city and the "semantic" neighborhood between the different pages and between the different districts. So the documents of the same district are close to each other and two neighboring districts correspond to two topics as close as possible.

4.3. Visualization

The choice of a 3D interface to visualize the search results pleased the users. Broadly the 3D visualization is not a problem because it corresponds to our natural vision. Moreover this 3D metaphor enables us to give an overview of a great number of results. The building texture represents the document content, which enables us to quickly have an overview of the results when hanging around in the 3D environment. Highlighting a building allows the user to see information (URL, snippet, keywords) about the associated document and about four neighboring documents which are obviously close to the chosen document. Three colors were chosen for representing the three clusters defined by the hierarchical clustering on the neurons. Each district is associated to one of these three colors (which are displayed on the ground). It enables us to show the three main topics of the search. With this metaphor, we can also interact with each document (represented by a building) in order to visualize all the images contained in the document. This image visualization is carried out by another 3D metaphor called gallery (see Figure 4). This gallery metaphor is very useful for an image search and was appreciated by the users. However, following the user test, we notice that the user seeks a compromise between the comfort and the effectiveness of the visualization more and more.

4.4. Navigation

The user test shows that the main drawback is the navigation in the city, which does not seem to be commonplace. We can add mouse sensibility problems or the loss of reference marks in the city. So certain displacements toward strategic places of the 3D scene were simplified. To do that the 2D map of the scene was made interactive in order to be able to move to any district in only one click. This modification makes navigation more comfortable but it must be coupled with other approaches. So solutions must be found in order to make navigation more familiar for the user (like navigation in a 2D interface). On the other hand, the gallery metaphor has the advantage of having a more constrained navigation and thus less tiresome for the user. Indeed the user can only move along one axis (the depth) and the side rotation angles are limited too. These navigation constraints avoid the user getting lost in the 3D environment and explain why the navigation in the gallery metaphor did not obstruct the users.

5. Conclusion

The main contributions of this paper are to propose a new taxonomy of search result visualization systems and to present an effective method for organizing and visualizing (in three dimensions) search results. The organization is

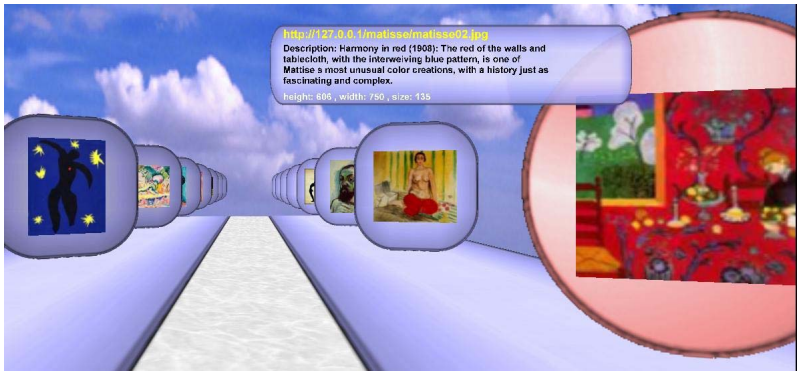

Figure 4. Gallery metaphor showing the images contained in a document.

based on a self-organizing map and the 3D visualization proposed here is based on a city metaphor which is very effective to represent organized documents. An essential point of the prototype is the coupling between the 3D scene and the 2D interface. The graphical interface is dynamically generated and interactive. With the proposed method, we provide the user with a three-level approach: low-level with document visualization, medium-level with neuron visualization (similar documents are grouped together), high-level with "topic" visualization (similar neurons are grouped together). The idea is to show the various topics of the query when going up in the hierarchy.

The search retrieval systems must be concerned with the quality of the returned results (not discussed in this paper) but with their organization, visualization and interaction too. So our goal is always to enrich the organization and visualization steps. For example, we must improve the interaction between the various hierarchical levels and keep on simplifying the navigation in the 3D scene. Other interesting outlooks are to make the classification algorithm interactive and to improve the interface adaptability. A last point concerns the decrease of the execution time for the algorithm and for generating the VRML code. Indeed this time is currently higher than ten seconds.

References

[1] K. W. Boyack, B. N. Wylie, and G. S. Davidson. Domain visualization using VxInsight for science and technology management. *Journal of the American Society for Information Science and Technology*, 53(9):764–774, 2002.

[2] M. Chevalier. *Interface adapative pour l'aide à la recherche d'information sur le web*. PhD Thesis, University Paul Sabatier of Toulouse, France, 2002.

[3] M. Dittenbach, D. Merkl, and A. Rauber. Using Growing Hierarchical Self-Organizing Map for Document Classification. In *Proceedings of European Symposium on Artificial Neural Networks*, pages 7–12, 2000.

[4] M. A. Hearst. TileBars: Visualization of Term Distribution Information in Full Text Information Access. In *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems*, pages 59–66, 1995.

[5] iProspect. iProspect's Search Engine User Attitudes Survey Results. White paper, 2004.

[6] T. Kohonen. *Self-Organizing Maps*. Springer, 1995.

[7] T. Kohonen, S. Kaski, K. Lagus, J. Salojärvi, J. Honkela, V. Paatero, and A. Saarela. Self Organization of a Massive Document Collection. *IEEE Transactions on Neural Networks, Special Issue on Neural Networks for Data Mining and Knowledge Discovery*, 11(3):574–585, 2000.

[8] B. Le Grand and M. Soto. Visualisation of the Semantic Web: Topic Maps Visualisation. In *Proceedings of the 6th Int. Conf. on Information Visualization*, pages 344–, 2002.

[9] T. M. Mann. Visualization of WWW-search Results. In *DEXA '99: Proceedings of the 10th International Workshop on Database & Expert Systems Applications*, pages 264–268. IEEE Computer Society, 1999.

[10] A. Rauber and D. Merkl. Automatic Labeling of Self-Organizing Maps for Information Retrieval. *Journal of Systems Research and Information Systems (JSRIS)*, 10(10):23–45, 2001.

[11] A. Rossi and M. Varga. Visualization of Massive Retrieved Newsfeeds in Interactive 3D. In *Proceedings of Int. Conf. on Information Visualization*, pages 12–17, 1999.

[12] A. Skupin and S. Fabrikant. Spatialization Methods: A Cartographic Research Agenda for Non-Geographic Information Visualization. *Cartography and Geographic Information Science*, 30(2):99–119, 2003.

[13] K. Sparck Jones. A statistical interpretation of term specificity and its application in retrieval. *Journal of Documentation*, 28(1):11–20, 1972.

[14] W. Wiza, K. Walczak, and W. Cellary. Periscope - A System for Adaptive 3D Visualization of Search Results. In *Proceedings of the Int. Conf. on 3D Web technology*, pages 29–40, 2004.

[15] O. Zamir. Visualization of search results in document retrieval systems. General Examination Report, 1998.

1550-6037/05 $20.00 © 2005 **IEEE**