Effective Organization and Visualization of Web Search Results

Nicolas Bonnel, Vincent Lemaire, Alexandre Cotarmanac'H, Annie Morin

Proceedings of the 24th IASTED International Multi-Conference
INTERNET AND MULTIMEDIA SYSTEMS AND APPLICATIONS
February 13-15, 2006, Innsbruck, Austria

EFFECTIVE ORGANIZATION AND VISUALIZATION OF WEB SEARCH RESULTS

Nicolas Bonnel
IRISA
Rennes, France
nicolas.bonnel@irisa.fr

Vincent Lemaire and Alexandre Cotarmanac'h
France Telecom, Division R&D
Lannion and Issy-Les-Moulineaux, France
vincent.lemaire@francetelecom.com
alexandre.cotarmanach@francetelecom.com

Annie Morin
IRISA
Rennes, France
amorin@irisa.fr

ABSTRACT

While searching the web, the user is often confronted by a great number of results, generally displayed in a list which is sorted according to the relevance of the results. Facing the limits of this approach, we propose to explore new organizations and presentations of search results, as well as new types of interactions with the results 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 results quality, the effectiveness of the results 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 approach organizes the results according to their meaning using a Kohonen Self-Organizing Map (SOM), and visualizes them in a 3D scene to increase the representation space. The 3D metaphor proposed here is a city.

KEY WORDS

Search Results Visualization, 3D Metaphors, Human-Computer Interfaces, Web Mining, Evaluation.

1 Introduction

Searching the web is one of the most frequent tasks, but often one of the most frustrating too. Search engines which are a way to represent the web to the users, are mainly used for web searches. However they are as easy to use as their results are difficult to interpret, which shows the web search contradiction. Indeed it becomes more and more difficult to extract the relevant information for a given search since available data on the World Wide Web is constantly increasing. The search engines return a number of results so great as 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 query 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 is also a way to increase the result "relevance" for the user. If the result quality remains a major concern, the quality of the result restitution (organization and visualization) must be taken into account too. Without an effective organization of the results, the user has to process manually the huge amount of results or refine the query in order to limit the results. This last solution can be compared to use a search engine for searching into the results! So these alternative solutions require efforts from the users.

Facing the increase of query results, it seems natural to want to **organize** and **visualize** them in an effective and adapted way. That explains the goal of the presented approach, which is to offer the user a search interface enabling him to quickly find the relevant information. The two main points to reach this goal are a good document organization 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 exploit cognitive metaphors (and more specially the spatial metaphors) in a more intuitive way. 3D offers new interaction possibilities too. 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.

This paper deals with the unsupervised organization of documents and the graphical representation of the results. A previous work [1] already deals with the same processing of web search results (*i.e.* a SOM-based organization of the results and a 3D city-based visualization), but it is mainly intended to give a description of our prototype. In this paper we focus more precisely on the SOM-based organization and the 3D visualization based on the city metaphor, as two independent tasks and without taking the prototype implementation into account. And an evaluation of this 3D metaphor is also proposed. Considering the above introduction, this paper is structured as follows. The next section proposes a short overview of the main issues in search results processing. Then Section 3 explains the self-organization method and Section 4 is devoted to the visualization. The last section allows us to conclude and gives an outlook on future work.

2 Two Issues in Search Results Processing

The search results processing has two main issues: results clustering and results visualization. For the first one, the goal is to find an effective method which allows to group similar results together and to organize the various clusters. The second one is to find an effective visualization of the organized results. These two points are briefly discussed in the following subsections.

2.1 Search Results Visualization

Many works have been done on search results visualization in the last few years (some examples can be found in [2] [3]). The goal of this subsection is not to present an exhaustive overview of the various approaches in this field, but only to give some information to locate our visualization approach (proposed in Section 4) in the literature and to explain our choices.

In this paper, one particularity is that the visualization is done on organized (or clustered) search results. Therefore the considered methods are only those which show the content-based links between the documents. Another constraint on the visualization is that the document projection shows the semantic proximity. The taxonomy of search results visualization systems proposed in [1] enables to show where the methods of visualization which address this problem are located in the literature. As this kind of visualization groups similar documents together, it gives the user information on the next document to select.

Visualization of inter-document similarities requires at least two dimensions to be efficient. So the techniques used in this approach have already given up the linear display of ordered lists. Indeed the two main techniques are graphs and maps. The meta search engine KARTOO[1] proposes a cartography of search results but a drawback is the lack of an overview of these results. Another well-known example is the WEBSOM[2] project [4]. The map approach can take advantage of the cognitive aspect. That is why geographic metaphors are often used such as in MAP.NET[3] (developed by ANTARCTICA) or in [5]. However with the increase of the results and the links complexity, graphs and maps become more and more unreadable.

So one idea is to exploit 3D visualizations to increase the available space to represent information. The added dimension allows the display of complex graphs in a more readable way. This third dimension can also be used for replacing maps by 3D worlds: landscapes [6] or cities [7] [8]. Other works are the AVE method [9] and its PERISCOPE system which are the closest works to those ones presented in this paper. We have the use of mixed interfaces (3D scene and 2D interface) in common, or the use of many visualization metaphors which answer different goals. However the approach proposed in this paper 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. However the dimension increase makes navigation essential and especially more complex. We are facing another problem which is not obvious to solve.

2.2 Search Results Clustering

There are many works in the area of text classification which is not discussed in this paper. Indeed our goal is more to find a good projection of documents and/or clusters than the classification itself.

Clustering can be used to improve retrieval results, which has been investigated in many previous works [10] [11]. The two main possibilities are static clustering (pre-clustering on the entire corpus) and on-the-fly clustering which can be considered as a post-retrieval document browsing technique (*e.g.* the clustering engine VIVÍSIMO[4]). In this paper, we are only interested in the second solution. Moreover the context of this paper imposes to have an unsupervised method for organizing the documents. Among techniques which address this problem, one of them is particularly interesting: the Kohonen self-organizing maps [12]. Indeed this method enables to cluster **and** to project documents onto an output space (generally a 2D space). In other words, 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. Privileged application areas of the SOM are visualization and cluster analysis [13]. With the SOM, it is also possible to have hierarchy levels or a map with dynamic size [14]. These two points can be interesting in our context but are not exploited at this time. Self-organizing maps have already been used for textual data clustering such as in the WEBSOM project [4] or in [15].

However the organization of web search results implies a particular SOM application whose adaptation is described in subsection 3.1. For example, web search results are special textual documents due to their various size, con-

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

tent, vocabulary or reliability. And they can be badly defined according to the number of words used to describe them.

3 Self-organization of Web Search Results

In the context of web search, documents are web pages returned by the query. In this paper, only the textual information of the documents is used. 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 [16] 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. However 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 results visualization (ordered lists) and users may want to see more results on 2D or 3D visualizations. Facing the lack of studies for 2D and 3D visualizations, this number is currently fixed to the first 50 results, which is more than the 30 results with which the users are satisfied. This weak number of results is not really a problem to compute the SOM, because the result organization is more important than the clustering itself.

3.1 SOM-based Organization

The SOM-based organization can be divided in three main steps: a pre-processing step to define the documents representation, a computation step with the algorithm, and a post-processing step to group similar clusters.

Pre-processing step. This first step deals with the **documents representation**, and more precisely how the word vectors are defined from the document content. The terms contained in the indexed pages[5] are extracted for representing the documents. With this term extraction, we are confronted by a great number of words. So our first processing consists of removing the stop words thanks to a dictionary. With the remaining words, two lists are created and organized according to a decreasing order. The first list contains word frequency in the corpus, and the second list contains the number of documents in which words appear. So the selected words for results organization are those which belong to the intersection of the first two hundred values of both lists. This choice is sufficient for a good representation of the documents and is compatible with the runtime efficiency. Each selected word is then associated with a variable of the self-organizing map. So the maximum number of dimensions for the input space is two hundred.

All the variables of the input space do not have the same importance for the results clustering. So a term weighting is used, which allows to increase or decrease the importance of some words. The term weighting used here is the *tf.idf* [17] one. This term weighting can also be interpreted as a particular weighting of the Euclidean distance. So the weighting of word w_i in document d_k is

$$tf.idf(w_i, d_k) = tf(w_i, d_k) \times idf(w_i)$$

$$idf(w_i) = \log \frac{\text{corpus size} + 1}{n_i + 1}$$

where tf means term frequency, idf means inverse document frequency and n_i is the number of documents which contain the word w_i.

The variable selection makes it possible to have documents just represented by a weak number of variables, which means that some documents can be badly represented. In data mining, these badly represented documents are generally removed before the analysis. However, in the context of web search, these documents can not be removed. Indeed document suppression could be assimilated to hide information to the user. So a weight is given to each document. This weight is high for well-defined documents and low for badly defined documents. The document weighting used in this paper is defined as follows: each document has a weight which is proportional to its number of non-zero variables.

Computation step: modified algorithm. This second step corresponds to the execution of the modified algorithm described in this part (as mentioned in subsection 2.2). First some notations are introduced: $\mathbf{x_k}$ is the k^{th} document of the corpus and its weight is w_k. $\mathbf{m_i}$ is the prototype vector of the i^{th} neuron of the map. N is the number of documents, M is the number of map units and h is the neighborhood function. The prototype vectors define a tessellation of the input space into Voronoi sets. And W_i is the weight sum of the documents which belong to the Voronoi set associated to the i^{th} neuron. Then Algorithm 1 gives a detailed description of this computation step.

In this SOM-based approach using a document weighting, particular values of the weights can be interpreted as follows. If all the document weights are set to 1, this algorithm is the classical batch SOM algorithm. Setting one document weight to zero means that this document is only projected on the computed map (not used during the computation). As the aim is to classify web search results, a constraint on the algorithm has to be added: it must be deterministic. The same query on the same corpus must always provide the same results. To do this, the batch algorithm is computed on a map with predefined size (5 $\times$

[5] An indexed page is only the web page associated to the URL, not the full site.

Algorithm 1 *modified batch SOM algorithm*

1: Initialization

2: Vector quantization:

Computation of the BMU[6] (for all the documents): $\forall k \quad BMU(\mathbf{x_k}) = \arg\min_i ||\mathbf{x_k} - \mathbf{m_i}||$

Computation of the new value (for all the prototypes): $\forall i \quad \bar{\mathbf{m}}_\mathbf{i} = \frac{1}{W_i} \sum_{k=1, BMU(\mathbf{x_k})=i}^{N} w_k \mathbf{x_k}$

3: Smoothing step: $\forall i \quad \mathbf{m_i} = \frac{\sum_{j=1}^{M} W_j h_{ij} \bar{\mathbf{m}}_\mathbf{j}}{\sum_{j=1}^{M} W_j h_{ij}}$

4: Go to step 2 as long as the convergence is not reached

5) and the first data is always used to initialize the neurons. The LabelSOM [18] method is used for labeling the neurons.

Post-processing step. This last step is the **neuron clustering**. The goal is to group similar neurons in order to obtain the various topics of the search. Many methods address this problem such as k-means or hierarchical agglomerative clustering (HAC). However the context of web search needs a deterministic solution, so it seems most appropriate to use HAC. For this clustering step, the input data is the neurons whose document number is not null. Then the HAC algorithm is computed on those neurons and the classification in three clusters is currently kept. The idea is to extract the three main topics of the search. This choice of the number of clusters allows to give the user a first overview of the main topics of the search. It is also interface-oriented because each cluster is associated with one color. However data can have more or less than three natural clusters, that is why this parameter can be modified by the user or it is also possible to move in the hierarchy (computed by the HAC algorithm). Each cluster is labeled with the most frequent labels of its neurons. Although this method is basic, it is currently sufficient to give the user a word based labeling of the topics.

3.2 Discussion

The documents organization presented here is based on the Kohonen self-organizing maps. This method, which is only based on word distribution, has the advantage to respect the “semantic” proximity of the data. It also enables us to have a first abstraction level if we move on the neuron level. Then a HAC algorithm is applied on the neurons using a Ward criterion which is fixed to only have three clusters. So it provides a second abstraction level which gives the user the three main topics of his search. These topics are defined by an arrangement of the neurons labels. These abstract levels are not based on ontologies or document labeling. However crossing neuron labels with an ontology or using semantic data on documents during the algorithm can be considered. This approach is also open and close to other more semantic approaches like Topic Maps [19]. 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 many organization methods must be defined in order 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.

4 Visualization of Self-organized Results

A mixed interface for visualizing self-organized results is proposed in this section. This interface is composed of a 2D part (a Java applet) and a 3D scene which represents the metaphor. A good definition of the metaphor word is: the realization of an association between graphical parameters of the presentation and information on the indexed documents.

4.1 Metaphor Context

The visualization interface proposed in this paper is based on the city metaphor. This choice 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 (detailed in subsection 4.3) was carried out on this metaphor. Upon the test results, the city metaphor has evolved. The new version will also be tested in order to know if the modifications answered what the users are waiting for and to identify new issues. Figure 1 gives an overview of the new metaphor whose explanations can be divided in four categories: basic elements, cluster representation, visualization and navigation. The last three categories are described in the following subsection. Now the basic elements of the city metaphor are introduced. Each building of the city represents a web page, and the buildings are grouped by districts which are placed

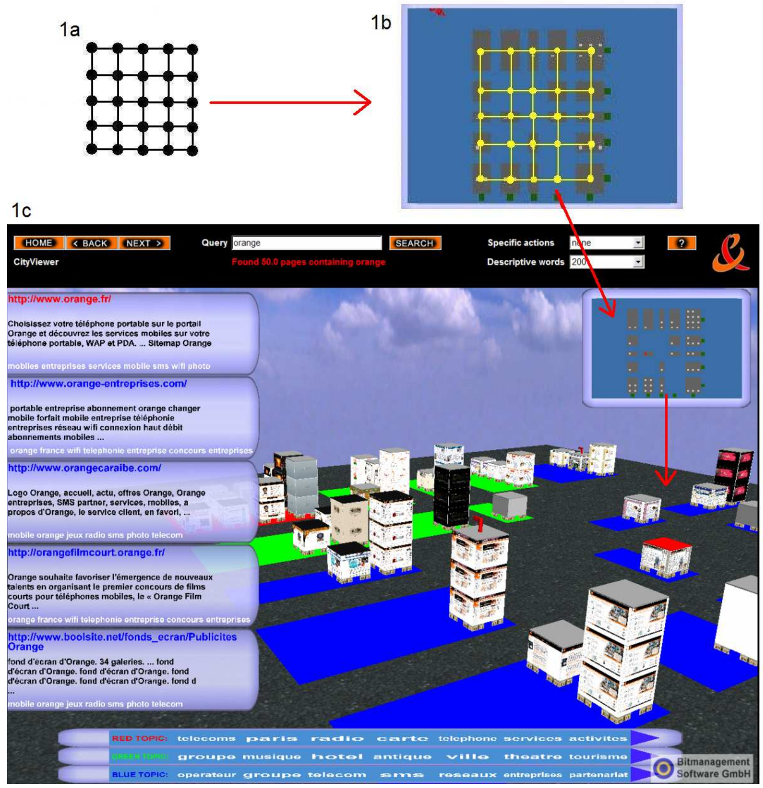


Figure 1. Visualization based on the city metaphor. The computed SOM (1a) is displayed on the ground, which can be seen on the top view (1b) or directly on the 3D scene (1c). The SOM grid is superimposed on the top view (1b) to show the mapping.

on the ground according to a grid. The building height represents the page relevance, 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 10 ranks (*i.e.* the first result page in traditional search engines) are associated to a high building height. The following 20 ranks are represented by a medium building height. And the other ranks have a small building height. These interval choices are motivated by a study on user attitudes [16].

4.2 Application of the City Metaphor on Self-organized Results

Cluster representation. The districts are placed on the ground according to a 2D grid and each district represents a neuron of the self-organizing map. So this grid (which is square in our example) enables to map the search results classification obtained with the self-organizing algorithm presented in the previous section. Indeed the 2D grid on the ground is the same one as the SOM output grid. This organization has two interesting properties: the web pages (or building) unicity 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 neighboring as possible. Three colors were chosen for representing the three clusters defined by the hierarchical clustering on the neurons. Each district is associated with one of these three colors (which are displayed on the ground). It enables us to show the three main topics of the search.

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 information about four neighboring documents which are obviously close to the chosen document. However, the user test reveals that the user seeks a compromise between the comfort and the effectiveness of the visualization more and more.

Navigation. A user study was carried out (*c.f.* subsection 4.3) and shows that the main drawback is the navigation in the city, which does not seem to be commonplace.

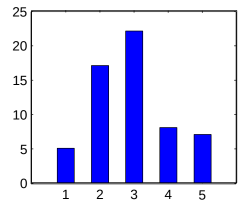
Other navigation problems are mouse sensibility or the loss of reference marks in the city. So certain displacements toward strategic places of the 3D scene were simplified. To do that the 2D map of the scene was made interactive in order to be able to move to any district in only one click. This modification 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). A more constrained navigation (and thus less tiresome for the user) must be proposed to avoid the user getting lost in the 3D environment. Another point concerns the used distance. In real life, to go from one district to another in a city, we follow the streets in a block like fashion, which corresponds to the Manhattan distance. This is the case in this metaphor only if the user uses the walking mode. However one interest of this metaphor is the use of the flying mode, and then the displacements are based on the Euclidean distance (the same as in the organization algorithm). So the user is not constrained to use the streets.

4.3 Evaluation

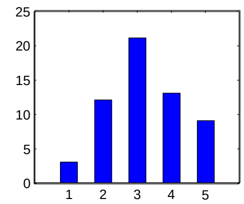
The evaluation task is something very important in the visualization process. In our case, the evaluation needs to take the organization and the visualization metaphor into account. The user study proposed in this paper is mainly oriented to evaluate the visualization metaphor. It is based on the well-known propositions of [20] for evaluating graphical interfaces. Table 1 and Figure 2 give an extract of this user study. For each question, the user can give a mark between 1 and 5, or decide not to answer. However the interpretation of the results is hard because the users prefer the average marks and then avoid the extreme marks. The average mark for each question is generally higher than 3, so it can be said that there is no drawback to use our approach for searching the web (compared to the existing ones). Another interesting remark concerns the last question which shows that the users are ready to use visualizations of clustered web pages (see Figure 2(h)). Concerning the runtime efficiency, it seems that organization and visualization of 50 results take almost 10 seconds (there is no code optimization and organization of the results is more time-consuming than 3D visualization).

4.4 Discussion

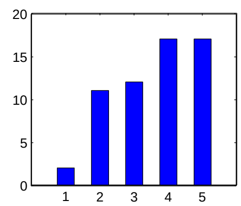
The topics separation makes it possible to consider a more adapted ranking algorithm. Indeed, it is not really relevant to compare the rank of two pages which do not deal with the same topic. Unfortunately, it is the case in most of the search engines. In our approach, an individualized ranking for each of the three main topics (resulting from the HAC algorithm) can be considered. To do that, the document ordering within a topic (cluster) is kept but the values are modified. So this approach allows us to have various doc-



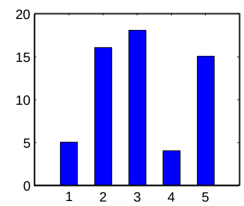
(a) Is the prototype clear to use?



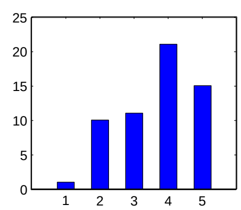
(b) Is the prototype easy to use?



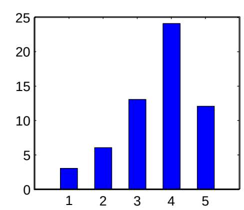
(c) Learning how to use the prototype



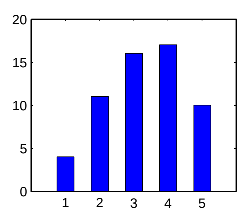
(d) Intended for all the user categories



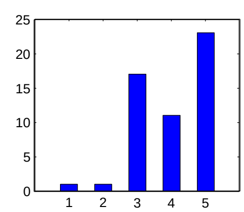
(e) Understanding of the city metaphor



(f) Relevance of the city metaphor



(g) Navigation in the city metaphor



(h) Usefulness of a visualization with clusters of web pages

Figure 2. Graphical representation of the extract of the user study proposed in Table 1.

	MARK (m)						
	1	2	3	4	5	NA	$\bar{m}$
QUESTION							
Is the prototype clear to use?	5	17	22	8	7	1	2.9
Is the prototype easy to use?	3	12	21	13	9	2	3.2
Learning how to use the prototype	2	11	12	17	17	1	3.6
Intended for all the user categories	5	16	18	4	15	2	3.1
Understanding of the city metaphor	1	10	11	21	15	2	3.7
Relevance of the city metaphor	3	6	13	24	12	2	3.6
Navigation in the city metaphor	4	11	16	17	10	2	3.3
Usefulness of a visualization with clusters of web pages	1	1	17	11	23	7	4

Table 1. Extract of the user study carried out on a panel of 60 people of various ages and backgrounds. The worst mark is **1** and the best mark is **5**. The column **NA** represents the not applicable answers and the column $\bar{m}$ corresponds to the average mark.

uments (which deal with various topics) as top results of the search. That avoids supporting one topic for a query and that makes it possible to let the user choose. Another point concerns data visualization which is strongly dependent on many criteria such as the results number and type, the search goal or the user category. So one single solution for data visualization probably does not exist. That is why an interesting characteristic consists of making the visualization adaptive. To do that many interfaces must be defined in order to choose the most adapted according to the context.

5 Conclusion and Outlooks

In this paper we propose an effective method for organizing and visualizing search results. The results organization is based on a self-organizing map which is adapted to the context of web search results. As the input space is reduced to optimize the computation, some documents can be badly represented in the chosen subspace. So a weighting scheme is applied on the documents to take their representation quality into account. Concerning the visualization, we propose a new 3D approach based on a city metaphor which is very effective for representing organized documents. The graphical interface is dynamically generated, interactive and based on a compromise between the 3D scene and the 2D interface. With the proposed method, we provide the user with a three levels 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. An evaluation of this 3D interface was carried out and a short extract is proposed in this paper. This study shows that users are globally satisfied with the 3D metaphor of the city, although they identified some points which have to be improved such as the navigation or the cluster labeling.

The information retrieval systems must be concerned with the quality of the results returned by the system (not discussed in this paper) but with their organization, visualization and interaction too. So our aim is always to enrich the organization and visualization steps. Concerning the organization, interesting outlooks are to make the organization algorithm interactive and to generalize the preprocessing rules (by testing them on more searches with various results). We can also include semantic analysis (not discussed in this paper) for visualization purposes, which will be a good way to improve organization and visualization of search results. Some other perspectives are the interaction improvement between the various hierarchical levels, the navigation simplification and the reduction of the response time. The general approach presented in this paper can be adapted to other exciting issues. For example, one idea is to compute a personalized representation of the web through web search results. The web searches of the user make the representation growing according to a criterion such as the visited pages or the first results. Another interesting issue is to use a SOM computed on the user's knowledge to display the search results. It is also possible to use the 3D metaphor of the city for visualizing other data such as RSS feeds, bookmarks or news.

Prototype

The works presented in this paper have been integrated in 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 their query results, without soliciting them during the process. A simplified overview of the prototype can be described as follows. On one hand the server side is composed by the database and the different interface models. Documents (web pages) are indexed in the database whose descriptors are those frequently used by the search engines such as URL, title, summary, snippet, word vectors. On the other hand the client side is composed by the Java applet (2D part of the interface) and the VRML browser (3D scene of the interface). The main parts of the prototype are the organization and the visualization, and the prototype implementation allows a certain independence between data organization and the selected metaphors for the visualization. The two input parameters of the visualization module of the prototype are: the previously organized results and an interface model (selected either by the user or automatically). The model which defines the visualization metaphor and the interactions used, 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[7] which is a meta-language (higher level than VRML) enabling to add many functions (database interrogation or iterations). During the query processing, the model is interpreted in order to dynamically produce the 2D interface and the VRML file which contains the 3D interface provided to the user. More details about this prototype can be found in [1].

References

[1] N. Bonnel, A. Cotarmanac'h, and A. Morin. Meaning Metaphor for Visualizing Search Results. In *Proc. of Int. Conf. on Information Visualisation*, pages 467–472. IEEE Computer Society, 2005.

[2] O. Zamir. Visualization of Search Results in Document Retrieval Systems. General Examination Report, 1998.

[3] T.M. Mann. *Visualization of Search Results from the World Wide Web*. PhD Thesis, University of Konstanz, Germany, 2002.

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

[5] 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.

[6] K.W. Boyack, B.N. Wylie, and G.S. Davidson. Domain Visualization Using VxInsight for Science and Technology Management. *JASIST*, 53(9):764–774, 2002.

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

[8] F. Sparacino, C. Wren, A. Azarbayejani, and A. Pentland. Browsing 3-D spaces with 3-D vision: body-driven navigation through the Internet city. In *Proc. of Int. Symp. on 3D Data Processing Visualization and Transmission*, pages 224–233, 2002.

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

[10] M.A. Hearst and J.O. Pedersen. Reexamining the Cluster Hypothesis: Scatter/Gather on Retrieval Results. In *Proc. of Int. ACM SIGIR Conf. on Research and Development in Information Retrieval*, pages 76–84. ACM, 1996.

[11] O. Zamir and O. Etzioni. Web Document Clustering: A Feasibility Demonstration. In *Proc. of Int. ACM SIGIR Conf. on Research and Development in Information Retrieval*, pages 46–54. ACM Press, 1998.

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

[13] J. Vesanto. *Data Exploration Process Based on the Self-Organizing Map*. PhD Thesis, Helsinki University of Technology, Finland, 2002.

[14] M. Dittenbach, D. Merkl, and A. Rauber. Using Growing Hierarchical Self-Organizing Maps for Document Classification. In *Proc. of ESANN*, pages 7–12, 2000.

[15] A. Rauber and D. Merkl. The SOMLib Digital Library System. In *Proc. of ECDL'99*, pages 323–342. Springer-Verlag, 1999.

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

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

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

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

[20] B. Shneiderman. *Designing the User Interface*. Addison-Wesley, 1998.

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