# Analysis of Web-Based Communication Systems Usage Patterns: A Visual Application

Florin Zamfir, José Nunes[1], Leonor Teixeira[2], Óscar Mealha[1], Beatriz Sousa Santos

*1 - Dep. of Communication and Art, University of Aveiro, Portugal*
*2 - Dep. of Economics, Management and Industrial Engineering, University of Aveiro, Portugal*

***Abstract***

**In worldwide institutions, internal web sites have been developed in conformance with internal communication strategies. Efficient analysis and diagnosis tools must be designed to cope with those constantly growing sophisticated infrastructures, allowing the identification of problems at the human-computer communication level.**

**In this paper, we introduce an application meant to help usability specialists to organize the informational space of an institutional web site. This application is able to capture, compile and analyze the information related to the structure of the web site and the usage patterns. We explore the enormous capacities of the human visual system presenting that information through different visualization methods.**

**We briefly present the functionality we have identified as central to the application, an overview of the architecture, the main conceptual model conceptual model of the user interface and the visualization strategies used.**

## I. INTRODUCTION

The development of internet/web technologies in the last decades had an important impact on the organizational information flow inside institutions. Internal/external web sites have been developed in conformance with internal communication strategies, and reflect internal information management and workflow. In this context, an emerging problem concerns the management of these constantly growing intranets/extranets, which must serve the organization, adapt perfectly to its internal communication strategies, and sustain efficiently its information and workflow patterns. Good feedback instruments for problem identification are fundamental. Some commercial solutions do exist, however they tend to be biased by classical technical metrics for technical tuning, not for organizational communication and information analysis [18], [2]. Efficient analysis and diagnosis tools must be designed to cope with those sophisticated infrastructures, allowing the identification of user-system mismatch at the human-computer communication level; these problems should be located and handed on to the design or redesign team.

Our proposal is related to the representation of large quantities of information collected inside an institution mainly from the analysis of the web site structure and site usage logged information (obtained either during natural site usage or controlled experiments).

We try to provide help for answering the following general question:

- How is the site used?

Starting from this general question, we can detail a specific sub-set of questions, whose answers might provide useful information:

- Who is using the site?
- What are the site areas / sectors / pages of interest?
- What statistical information can be obtained from the log files?

Exploring the enormous capacities of the human visual system we present the information through different visualization methods, in order to provide help for answering the following additional question:

- Which are the areas with problems?
- What usability problems can be identified?

We provide an application able to capture, compile and analyze the information related to the structure of a web site and its usage patterns. It is meant mainly for helping usability specialists to organize the informational space of an institutional web site. However, this application does not provide direct solutions for the usability problems of the site; instead, it offers the possibility of taking decisions based on the usability problems discovered during the analysis process.

In this paper, we describe the overall philosophy we propose for such an application. While developing and testing a previous version of prototype, as a result of a user centered evaluation, we were able to better define the needed functionality and to improve the architecture of the application, as well as the conceptual model of the and user interface. In the following sections, we briefly present the functionality we have identified as central to the application, an overview of the architecture, the main aspects related to user interface and visualization strategies used.

## II. PROPOSED FUNCTIONALITY

As a result of a prior study concerning the goals that our application should meet and the corresponding functionality to include in it [10], we have identified as central to our application the following functionality:

- analyze the web site structure and catalog each page in terms of content and structural information;
- analyze the log files for the specified site. This process includes several steps like identification of users, sessions, statistical information, etc.;
- analyze and interpret the information collected from controlled experiments, like: mouse tracking, eye tracking, etc.;
- represent the information regarding the site structure;
- represent the information regarding the site user's information, the session's information and other communication additional information;
- represent statistical information collected from the log files and controlled experiments;
- manipulate various information visualization methods to make the presented information easier to perceive and interpret;
- analyze and process the structural and logged information, using predefined patterns and algorithms, in order to identify usability problems;
- offer a flexible and easy-to-use user interface, in order to allow an easy manipulation of the information representations and user interactions.

## III. ARCHITECTURE

We have used a layered architecture, which allows a greater independency from the development platform and the operating system, presented in Figure 1. The *Business Layer* (Framework) represents the link between the *Development Platform* and the *Presentation Layer* (PL). The PL includes all the components interacting directly with the user.

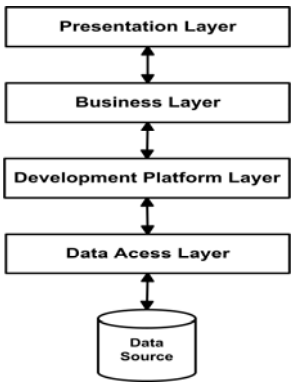

Figure 1 System architecture

Following a *Top Down* approach for the development process and using *UML* (*Unified Modeling Language*) [8], [11], [3] to represent the system, we were able to identify the architecture of our application. A simplified model of this architecture is presented in that corresponds to the presentation layer. This layer also encompasses all the other components for processing and manipulating the information.

According to the goals defined for our application, we were able to identify four major modules that interact and share data, having as final purpose to display the processed information to the communication and system manager, through specific information visualization methods.

The application is divided in four modules:

- *SiteAnalyzer* - meant to analyze the site in terms of structure and interconnections, to capture and classify the content of each page, to make a screen content capture of each page, to identify and classify the hotspot areas of each page and to store the resulting information in our database;
- *Interceptor* - meant to intercept the site user events at the client side, in a controlled experiment, to make a screen content capture for every context change and to send the collected information, via a Client-Server connection, to our server;
- *Compiler* - meant to analyze the log files stored on the site server, to interpret the information collected from these files and from controlled experiments, to organize the statistical information related to site usage, site users, page content, etc., and to store the resulting information in our database;
- *Visualizer* - meant to present the overall processed information in a way that is more adequate to the human perception and understanding. The used visualization methods are specialized on showing different types of information to the user and have different goals in terms of understanding. This module gives the user the possibility to choose different visualization methods for displaying the same information and switch among different representations. It does not modify our primary processed data; it only accesses it and translates numerical and textual data into visual representations.

This application has been developed for Windows Platform and as development platform we have used Microsoft Visual Studio .NET and Microsoft SQL Server for our relational database.

## IV. VISUALIZER

As mentioned earlier, our proposal involves the representation of a large amount of data. Based on our previous experience [9] and taking into account human-computer interaction and information visualization principles [4], [15], we have tried to find solutions for effectively representing those great amounts of collected data. In order to attain this goal we have not only visually represented data in an understandable way, but also offered means of interacting with it. This is the main role of the Visualizer.

Actually, the *Visualizer* is under a constant development and upgradeable phase. It is the most complex module of our application and its complexity comes mainly from the variety of visualization methods intended to be implemented. For this module, we were able to identify seven different components:

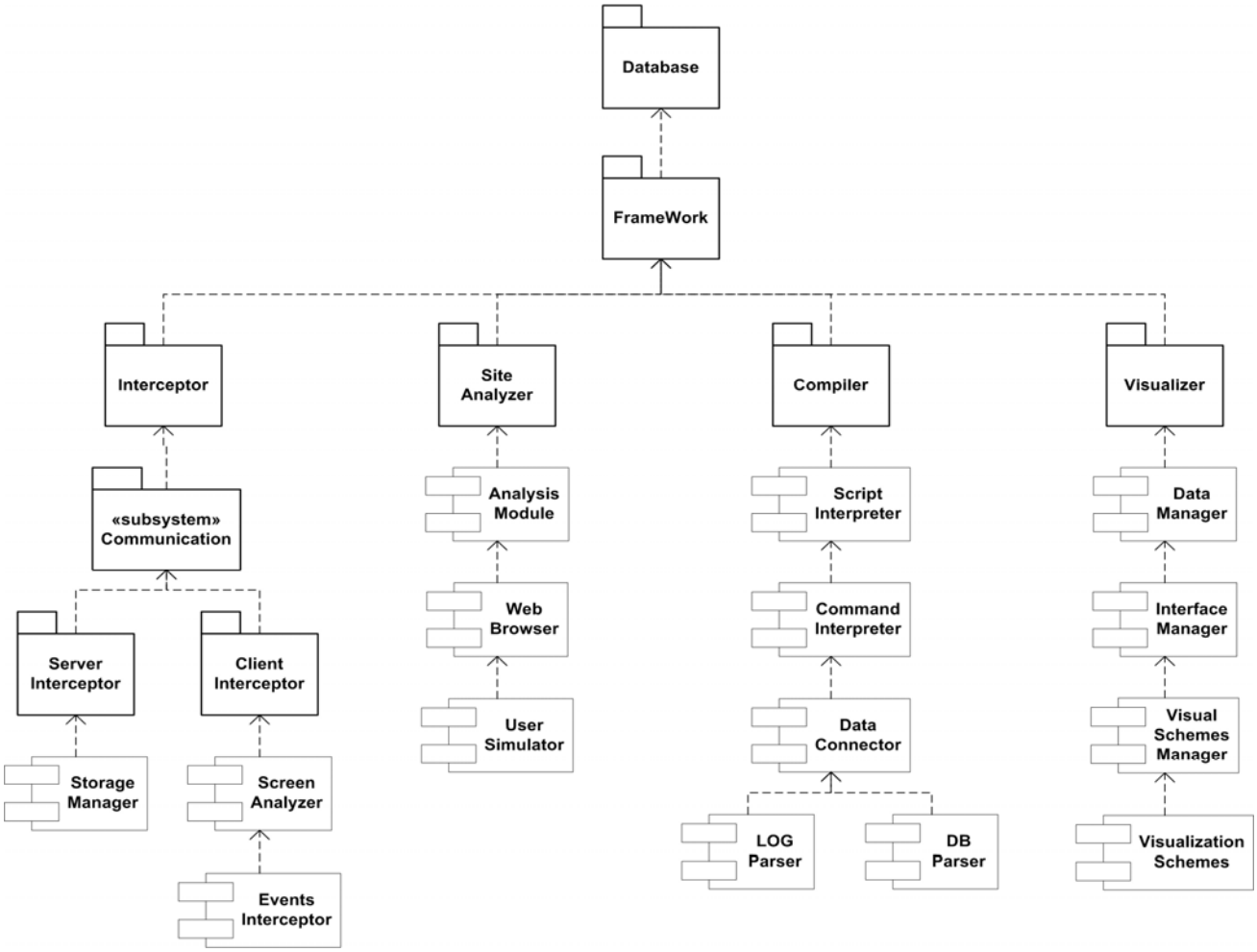


Figure 2 Application Simplified Model

- *Site Explorer* where the user has the opportunity to explore offline the structure of the selected site. Each page of the site is stored inside our database and every time the user selects a link from a page, the corresponding page is loaded from the database.
- *Site Relations* presents the information related to the site structure. Several representations of the information can be shown at a time, selected at any moment from the set of visualization methods available. Not only structural but also statistical information is presented, like usage patterns.
- *Sessions Exploration* where the session information can be observed using several representations of the information at a time. In addition, multiple sessions can be observed simultaneously. A overall usage statistical information is represented here.
- *General Statistics* is designed for showing the statistical information related to the site usage. Different visualization methods for representing the information are also available. In addition, this area offers a high level of interrogation of the database.
- *Problem/Solution Identifier* is designed with the purpose of showing information concerning usability problems related to the site structure (e.g. broken links) and the result of using metrics to assess the compliance with specific usability guidelines (ex. usage of colors, font size, overall clutter of the pages, etc.). By comparing the actual information with the paradigm associated with the site section, the tool is able to suggest (or not) a solution for the identified problem(s).
- *Usage Patterns* is designed for accessing the information related to user / user-group site usage, which patterns can be highlighted, related to the information flow or communication traffic.

One of the important aspects of Visualizer is that the user must have the possibility to observe the information, using different representations simultaneously. The synchronization between the representations offers good feedback to the application user, allowing a better observation and inspection of the selected information.

The graphical user interface is based on a multiple top-level window engine [North, C. and Shneiderman, B., 2000], which allows the manipulation of data representations, in an easy and practical way. The experience obtained from the previous prototype has proved that the manipulation of each different view, as a separate entity (window), is preferable. For a better flexibility while manipulating the representations, we decided to implement a set of controls (docking windows,

splitter windows, etc.) that allow the manipulation of the spatial position and size of the selected visualization representation window.

The division of the interface in several areas of interests, as presented in Figure 3**,** and the possibility to apply direct spatial manipulations to these areas, give a better visibility to the represented information. The context selection area allows changing the current context, which means to change the active window, each context corresponding to a separate window and a different type of information represented.

The application status window helps the user to have a better understanding about the actual state of the system; it provides information related to the description of the current context, the user interactions with the system and the effects of these actions. In addition, overall statistical information is presented on this window, information regarding the overall site usage, content and analysis period, for the selected site. Each visual and functional aspect of the application is configurable; the possibility of manipulating the user interface entities, and the interaction among these entities, also contributes to the overall flexibility.

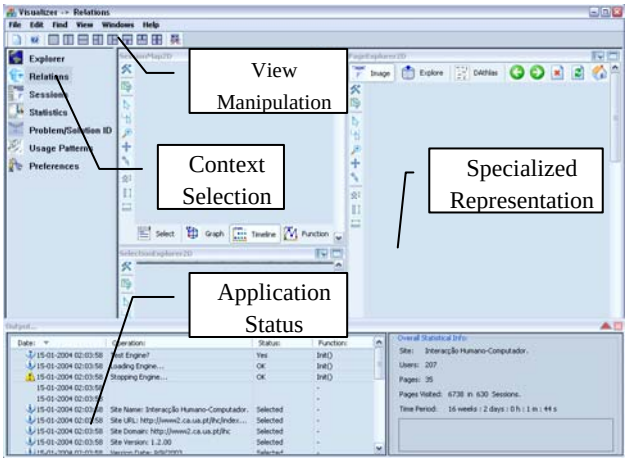


Figure 3 Visualizer: main aspects of the User Interface

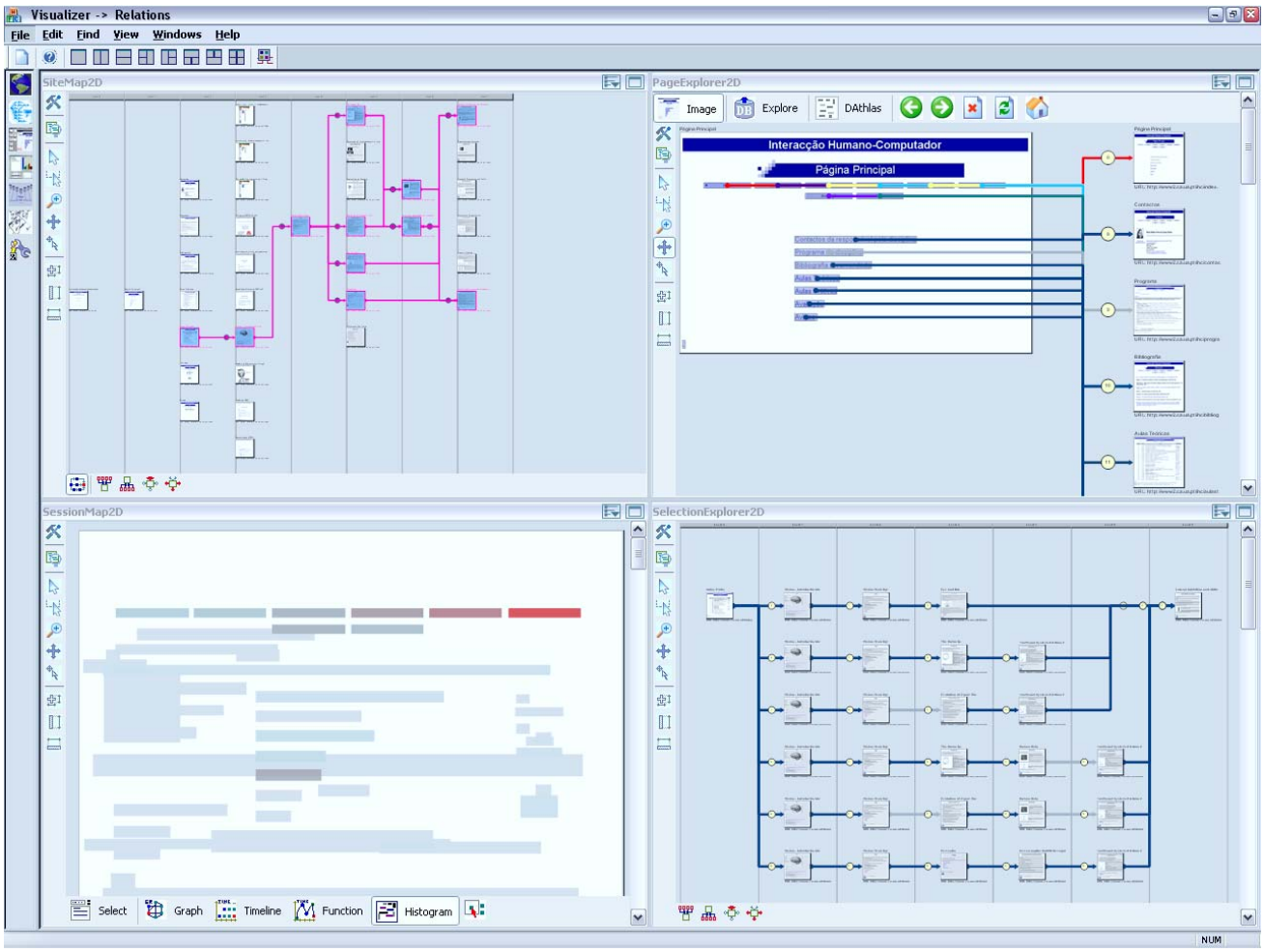


**Figure 4** Synchronous representation for information visualization and inspection.

In addition, the system integrates a Visual Logging system and error reporting, system that can be used for usability evaluation of the application and for observing how it responds to the user actions. This system allows recording the usage of the application, in terms of screenshot and event logging.

## V. VISUALIZATION METHODS

One of the major problems in dealing with large-scale hypermedia structures is to find a way for representing the information in an ideal scheme for efficient human understanding. Different authors have tried to solve this problem by developing various visualization methods [14, 16, 6, 7, 1, 17, 5]. The main concern while developing these methods is that the user needs a practical way to observe, explore and interact with the data in order to extract the information he/she needs. We have identified six visualization methods that could be of interest to our application. These visualization methods are complementary, each allowing a different way of representing the data that makes it more adequate to serve certain user goals; thus, they should be combined with each other and not used only one at a time. We offer the user the possibility to choose the number of visualization methods he/she wishes to use simultaneously, accordingly to which multiple view patterns and where on the screen the corresponding representation is going to be displayed.

The different views presented simultaneously are synchronized, i.e. any interaction with the objects of any view is reflected on the other views, with the result in a change of content and visual paradigm. This synchronization helps the user to better explore and understand the information (as shown in **Figure 4**).

The identified visualization methods and their advantages and disadvantages are briefly presented below.

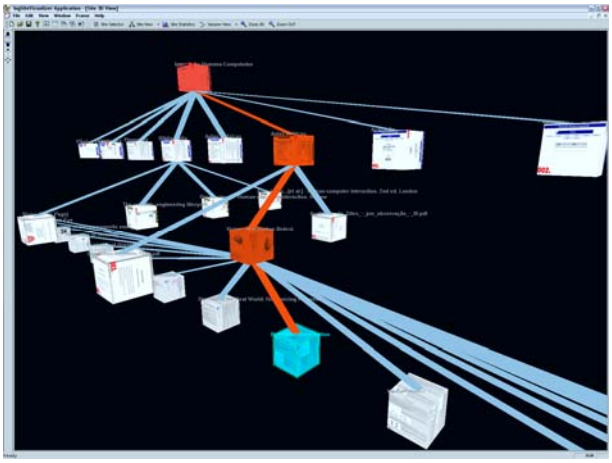

Figure 5 3D Site Map

**3D Site Map**. This scheme is based on a level based hierarchical representation, as observed in Figure 5. By default, consecutive levels are perpendicular to each other. Every page is considered a Parent and the links of the page are considered Children. It is possible to manipulate any object in the representation. Each Level position can also be dynamically manipulated. A page can be represented several times. This visualization seems particularly interesting to show the overall complex structure of large websites and highlight paths between pages. It may be used for representing complex web site structures, the 3D space offering the possibility of representing a large number of objects. However, navigation in 3D space may result difficult and cause disorientation to the user.

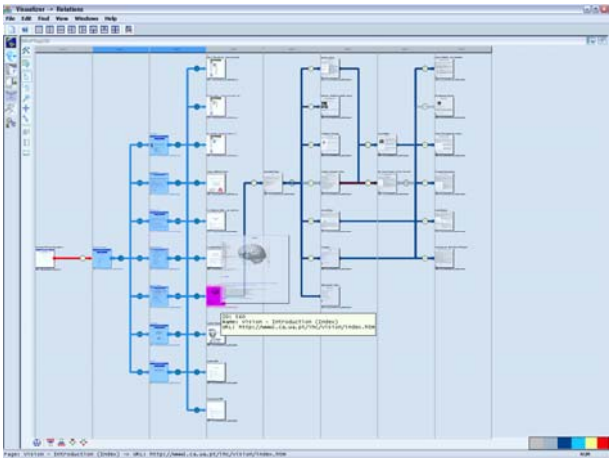

Figure 6 2D Site Map

**2D Site Map.** This scheme is based on a hierarchical representation. The levels are represented in columns, using a grid, as observed in **Figure 6**. Consecutive levels are represented in different columns. The links distribution is only in one direction – from the parent page to the children pages. Each page is uniquely represented. This visualization seems more adequate to represent small / medium sized sites and highlight large selected areas. Its main purpose is general exploration of the web site structure and content, also the interconnections between the site pages give good feedback. However, multiple interconnections between pages might unbalance the representation making difficult to represent large amount of data.

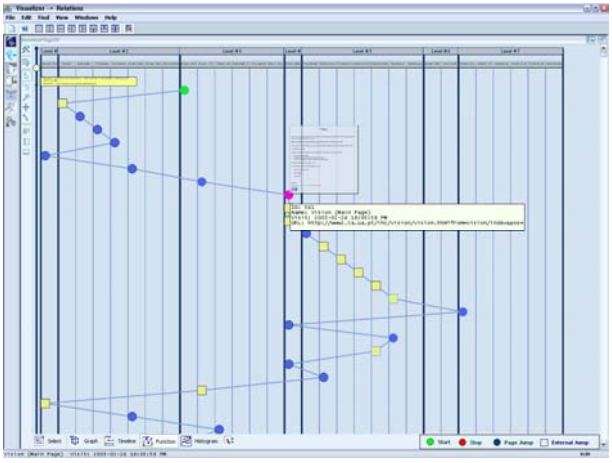

Figure 7 2D Session Map

**2D Session Map.** This scheme is based on a Time – Level representation in a Cartesian system. Visited pages are represented on a vertical timeline, according to the page level, within the steps followed by the site user. As can be observed in Figure 7, a user session can be represented by a discrete function. Work is being developed in order to understand this form of

representation and its usefulness for usage pattern parameterization. Additional information, color or/and shape coded, as in page/external jump between the pages, also gives good feedback about the site implementation or/and usage.

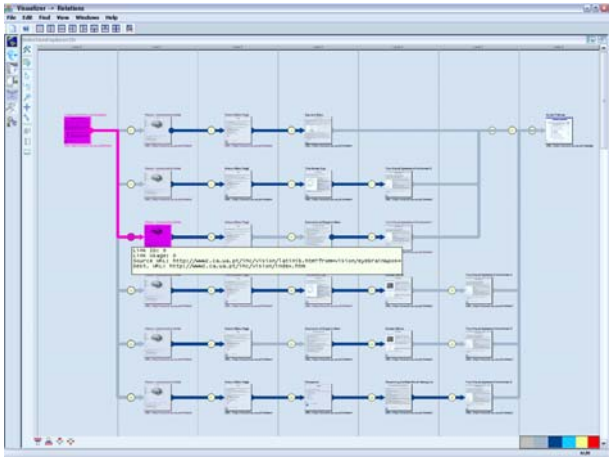

Figure 8 2D Selection/Paths Explorer

**2D Selection/Path Explorer.** This scheme can have two aspects. One for representing the selected objects (in other views), based on a graph representation**,** and another for representing the possible paths between two pages, based on a hierarchical representation, as it can be observed in Figure 8. Selected pages are represented uniquely and all the possible interconnections between the selected pages are shown. On the possible paths representation each page can be included several times. This scheme seems adequate for representing the selected pages and the interconnections among them or the possible paths between two pages. However, multiple interconnections among pages, or multiple possible paths, can make the representation very complex.

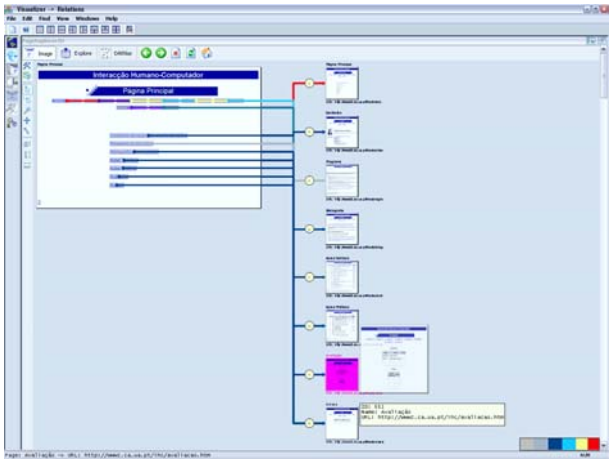

Figure 9 Page Explorer

**Page Explorer.** This scheme is based on the Referrers->Page->Page Links concept. Page Links are represented according to the corresponding Hotspot position on the page, as observed in Figure 9. Page Referrers can be represented on the opposite side of the Page Links and the links distribution is in only one direction. This scheme is adequate for representing the selected page links/referrers directly related to the physical position and the page statistical information (e.g. hotspot usage). However, complex pages with many links and referrers might be difficult to interpret and multiple interconnections between pages can unbalance the representation.

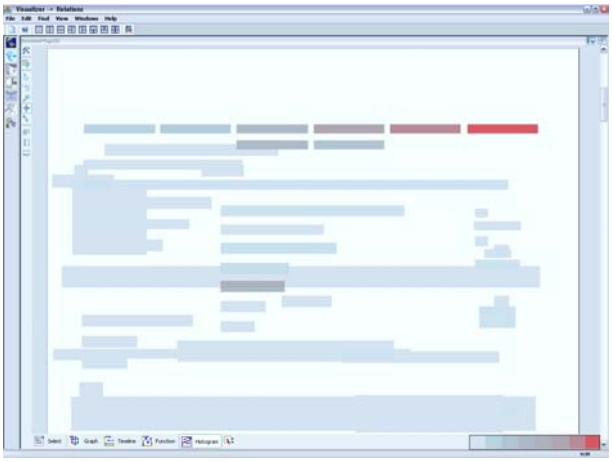

Figure 10 Hotspot Usage

**Hotspot Usage.** This scheme is based on a hotspot usage color coding representation. It is shown in Figure 10. This representation highlights the most visited hotspots by analyzing the selected user sessions. The color of a region gives the usage of the hotspots located in that region. Different color-coding techniques [19] can be used as shown in the lower right side of the following figures Figure 6, Figure 7, Figure 8, Figure 9 and Figure 10. In addition, the interaction with each region provides additional feedback. The scheme gives good information about the hottest areas of the site, uses the geometrical position of the most followed links as visual cues and gives visual feedback related to the page usage and structure.

## VI. CONCLUSIONS

We have introduced an application under development intended to help usability specialists to analyze usage patterns of an institutional web site. Our proposal is especially concerned to the representation of a large amount of data collected inside an institution mainly from the analysis of the web site structure and site usage logged information (obtained either during natural site usage or controlled experiments). We explore the enormous capacities of the human visual system and we present the information using specific visualization methods. The addition of user interaction and exploration capability of the visual representations improved the overall insight of the presented data. The application has been organized in four main modules: the site analyzer, the interceptor, the compiler and the Visualizer. The Visualizer is the most complex module and a third version of the prototype is currently under development; it offers the user a rich

functionality and a choice among several synchronized representations of the data obtained using different visualization methods.

We highlight as future work the following topics: evaluation of the visualization methods and implementation, further investigation of visualization methods and their relations, application to a case study.

## VII. REFERENCES


[1] Benford, S. et al.. Three Dimensional Visualization of the World Wide Web. *ACM Computing Surveys* 31(4), December 1999.

[2] Bieber, M. et al. Forth generation hypermedia: some missing links for the World Wide Web. *Int. J. Human-Computer Studies.* 47. 1997. pp 31-65.

[3] Booch, G. et al. *The Unified Modeling Language User Guide*. Reading (MA): Addison-Wesley, New York, USA. 1999.

[4] Dix, A. et al. *Human Computer Interaction*, 2nd. Prentice-Hall, London, England. 1998.

[5] Dodge, M. The Atlas of Cyberspace: Maps of Web Sites. http://www.cybergeography.org/atlas/web_sites.html. 2003.

[6] Durand, D. and Kahn, P. MAPA: A System for Inducing and Visualizing Hierarchy in Websites. *Proceedings of ACM Hypertext 1998*, Pittsburgh. 1998.

[7] Frècon, E. and Smith, G. WebPath, A Three,Dimensional Web History. *Proceedings of IEEE Symposium on Information Visualization (InfoVis 1998)*, Chapel Hill. 1998.

[8] Jacobson, I., Booch, G., Rumbaugh, J., 1999. *The Unified Software Development Process.* Addison-Wesley, New York, USA.

[9] Nunes, J., Zamfir, F., Mealha, O., Sousa-Santos, B. Web Log Visualizer for Communication and Information Management, *Actas da 3ª Conferência Portuguesa de Sistemas de Informação CAPSI 2002*, Coimbra, Portugal. 2002.

[10] Nunes, J., Zamfir, F., Mealha, Ó., Sousa-Santos, B. Web LogVisualizer: a Tool for Communication and Information Management. *Human Centered Computer: Cognitive, Social and Ergonomic Aspects, Vol. 3 of the proceedings of HCI International 2003*, pp. 854-858, Creta, Greece. 2003.

[11] Larman, C. *Applying UML and Patterns: An Introduction to Object-Oriented Analysis and Design.* Prentice-Hall. 1998.

[12] North, C. and Shneiderman, B. Snap-Together Visualization: Can Users Construct and Operate Coordinated Views. *International Journal Human-Computer Studies*, 2000. pp. 851-866.

[13] Richter, C. *Designing Flexible Object-Oriented Systems with UML*. MacMillan. 1999.

[14] Risch, S. et al. The STARLIGHT Information Visualization System. *Proceedings of the IEEE Conference on Information Visualization*, London, England. 1997.

[15] Spence, R., *Information Visualization.* Addison-Wesley. 2001.

[16] Shiozawa, H. and Matsushita, Y. WWW visualization giving meanings to interactive manipulations. *Advances in Human Factors/Ergonomics 21B (HCI International 1997)*, San Francisco, USA. 1997.

[17] Shneiderman, B. *Dynamic queries, starfield displays, and the path to Spotfire.* Feb 1999, http://www.cs.umd.edu/hcil/spotfire/. Visited at Nov 2003. 1999.

[18] Tauscher, L. and Greenberg, S. "How people revisit web pages: empirical findings and implications for the design of history systems". *Int. J. Human-Computer Studies*, 47. 1997. pp. 97-137.

[19] Tufte, E. Visual display of Quantitative Information. *Graphics Press*. 2001.