Web LOG File Information Analysis and Visualization Inside Organizations

Florin Zamfir
Dep. of Electronics and Telecommunications
Dep. of Communication and Art
University of Aveiro
florin@ca.ua.pt

Óscar Mealha
Dep. of Communication and Art
University of Aveiro
oem@ca.ua.pt

José Nunes
Dep. of Communication and Art
University of Aveiro
jnunes@ca.ua.pt

Beatriz Sousa Santos
IEETA – Instituto de Engenharia Electrónica e Telemática de Aveiro
University of Aveiro
3810-193 Aveiro, Portugal
bss@ieeta.pt

Abstract

In the context of this work, organizations can be understood as organic systems composed of people, processes and procedures. The intranets are incorporated in organizations to support their mission and guarantee effective process delivery within an efficient framework. These technological platforms, with a web driven user front-end, are able to log several user interaction details. The main goal of this paper is to report how the analysis and visualization of interaction patterns inside an organization's internal web based system can be achieved, as mentioned by [Eick01] in a more general context and by [Healey01] in a specific e-commerce scenario.

In order to validate the research on conceptual models, information architecture and several visualization methods, a prototype is being developed, tested and permanently upgraded. At this stage of iterative development, the prototype can be considered as an experimental research platform [Nunes03], capable of integrating and testing specific visualization schemes [Mealha04] and visual correlation procedures. Besides this, the prototype incorporates features that can visually support answers for questions, such as:

i How is the site used?

ii Who is using the site?

iii What are the site areas / sectors / pages of interest?

iv What statistical information can be obtained from the log files?

v What are the problematic areas?

vi Which usability problems can be identified?

Some of the methodological stages, inherent to our research, will be described and exemplified in this paper, specifically, the conceptual model of the prototyped system, overall prototype design and unified user interface, and visual inspection methods/schemes. Examples will be presented as prototype screenshot images depicting visual inspection settings and results.

Keywords

Information visualization, information and communication management, usability, web logs.

1. INTRODUCTION

On the behalf of evolution of nowadays information systems, worldwide institutions developed and supported important informational frameworks, based on the latest hypermedia presentation systems. An emerging problem occurred, related to the management and understanding of these vast informational workspaces.

The latest evolution of Internet based technologies gave the institutions the means to develop large websites, meant mainly to support their strategic and operational

decisions. However, the management of such systems became more and more difficult, due to the unpredictable growth facilitated by these technologies. Starting with the role of technological support, the Internet became an efficient instrument to support business decisions, the prediction and understanding of its potential represents advantage for every institution.

In this context, partially reported by this paper, good analysis tools are needed, to support the managers in decision making, based on analysis results of their organizational information workspaces. However, this is not an easy task, from the computational point of view, specifically the development of applications that have quantitative and qualitative correlated representation and inspection features of the informational workspace.

The following sections describe methodological stages and usage examples of the prototyped system that represents a current research framework which provides answers for some strategic questions in the context of organizational web-based information and communication management.

2. SYSTEM MODEL

The prototype system model is organized to achieve two main goals focused on intranet usage. One is related to the site structure, the other to the interface design (visual workspace organization). It provides solutions to detect structural problems revealed by visual inspection of usage patterns with the possibility of having simultaneously qualitative and quantitative visual inspection methods. Another vital area, where severe system problems occur, is at interface level. The visual workspace hosts the visual component of the interface, i.e. an information area where the prototype user decides on major navigational issues. The registered interaction activity log is processed and the usage pattern is represented and analyzed.

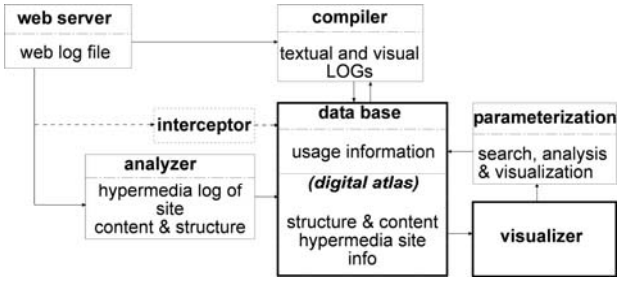

Figure 1 System Model

The global system architecture, as represented in figure 1, is user goal oriented and gathers raw information [Shneiderman96] from system log files, from controlled experience environment data (compiler) and from a copy of the pages loaded from the server and presented to the user (analyzer). A database is used to store the data and the visualizer to display it. The system conceptualization exercise and internal information structure uses, as an overall reference, the model proposed by [Card01, p. 554]. As a consequence of a state of the art exercise, it also complies with other raw information sources such as eye movement tracking technologies for a possible future system version.

3. UNIFIED INTERFACE–VISUAL CORRELATION STRATEGIES

Our system uses an interface strategy that integrates several visualization schemes, dialogues and feedback information units into one unified interface. It delivers an articulated and contextualized visual inspection method sustained on multiple simultaneously present views. The visual correlation of qualitative and quantitative network information properties is considered as shown in Figure 5.

An inspection exercise starts with the specification of a goal that may consider the homepage (or some other page) as a starting point. It ends with the specification of a final page, where the information or communication service entry is located, and is specified as a usability evaluation goal. An articulated and tightly coupled set of views are generated for deeper inspection analysis of inter-page related statistics and interface design parameters. All these schemes are triggered and redrawn by the same event and use the same raw information space to represent their visual features.

4. DATA VISUALIZATION SCHEMES

Visualization functions can be subdivided according to three main areas of concern, at this stage of conceptualization, development, use and evaluation:

- *Visual inspection of intranet structure and efficiency;*
- *Visual inspection of interface design coherence, specifically for Hot-spot and usage path overlapped representation;*
- *Representation of tree-structured information traversing in time.*

The following sections describe some of the authors' recent work in these areas. Some examples are presented based on real data to highlight the features inherent to the conceptualized and developed visualization schemes.

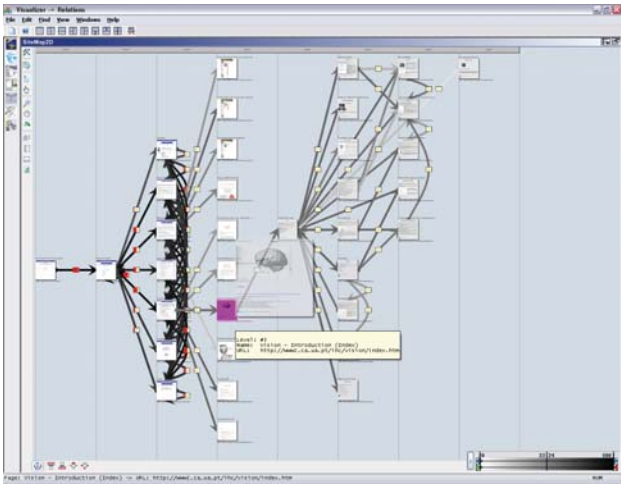

Figure 2 Tree representation with quantitative and qualitative usage data

4.1 Tree map based visual inspection features

Hierarchical tree based representation constitutes one of the visualization schemes. This scheme, Figure 2 and upper left quadrant of Figure 5, represents site page relational tree map [Card99], superimposed with usage data in qualitative and quantitative form with dynamic visual inspection capabilities.

The Colour Look Up Table (LUT), is used in this scheme, in greyscale for intensity representation purposes. It is a dynamic tool with multiple possibilities in its slider functions similar to slider applications as reported by [Eick99]. The following list describes some of the LUTs dynamic visual inspection capabilities, figure 3 and 4:

- *Remaps or masks usage behaviour into colours;*
- *Change between different colour schemes to increase contrast/visibility;*
- *Filter out representation objects to highlight specific usage information;*
- *Dynamically synchronize with the selected object's usage properties and remap them;*
- *Mask or hide various levels of details to allow highlighting others.*

This visualization scheme answers the first, third and fifth questions presented on the abstract, the site interconnections, pages and areas of interest being highlighted as the user explores the scheme, filters are helping distinguish between statistical usage levels.

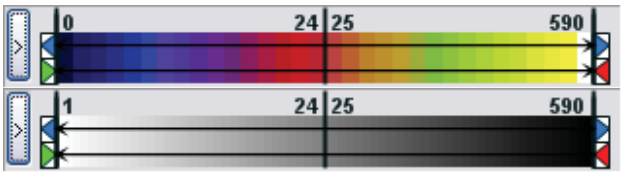

Figure 3 Colour LUT and greyscale LUT

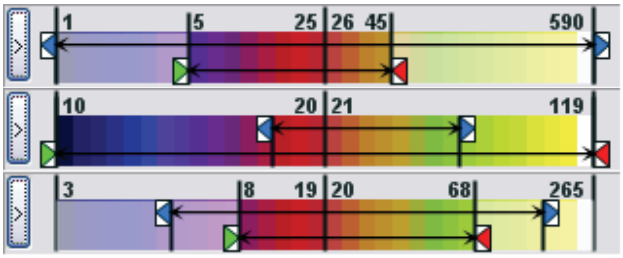

Figure 4 LUT dynamic filtering, masking and remapping of usage data

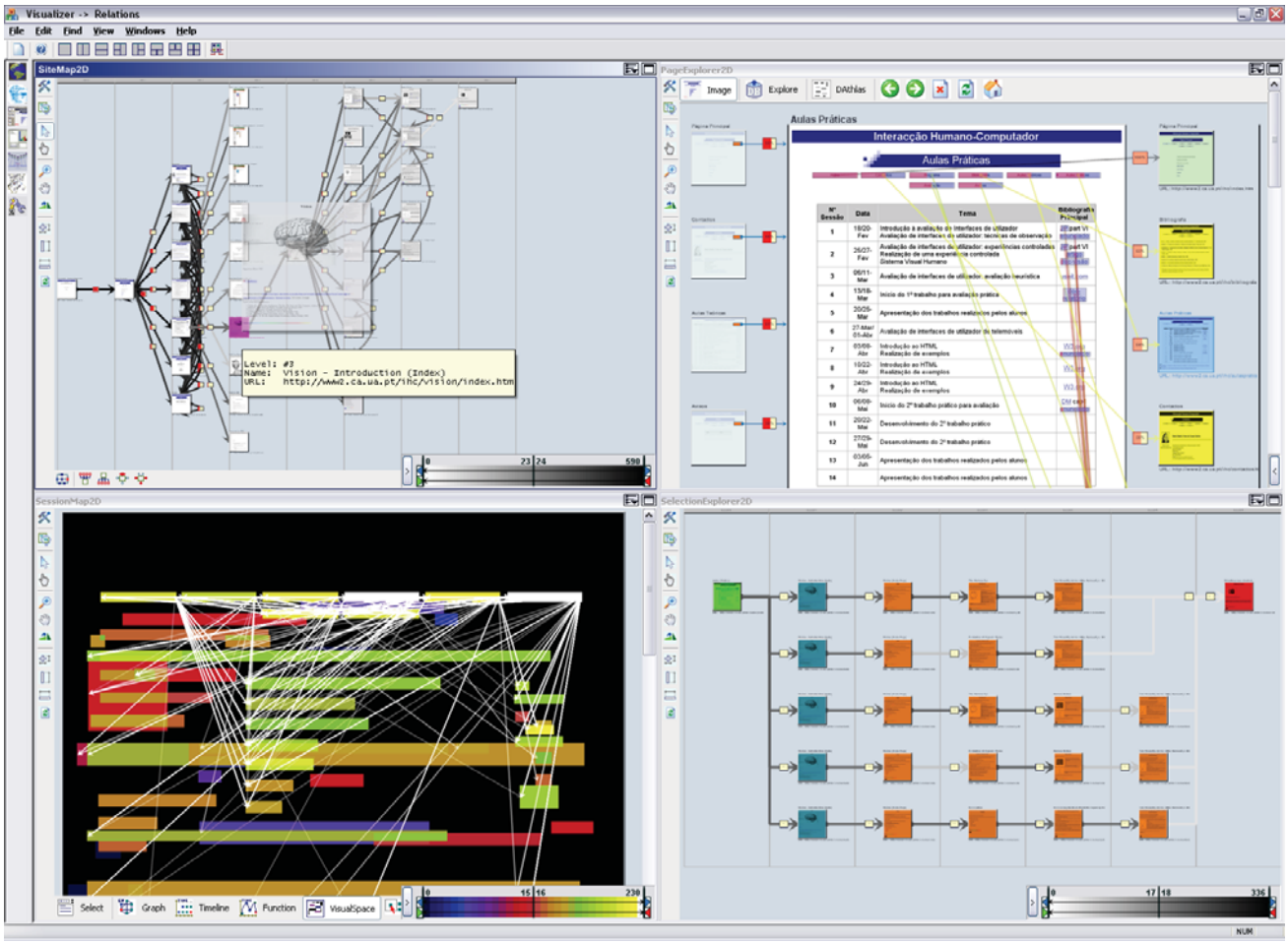

Figure 5 Unified interface with tightly coupled visual inspection methods

4.2 Page inspection tool

The site page inspection tool has a major goal of representing, in detail, all the hypermedia related information of a specific chosen site page, as seen in upper right quadrant of Figure 5 and Figure 6.

This scheme can:

- *Highlight statistical hotspots usage information and map it to the actual page position;*
- *Bind the selected page referrers using the physical position of their relational hotspots;*
- *Use the physical site interconnections mapped to the visual usage function for Referrer → Page, as for Page → Child page representations;*
- *Use masking and filters to highlight the connections Referrer → Page → Child page*
- *Interactive scheme exploration allowing page to page navigation, driven by statistical usage data and the spatial hotspot location on the visual field.*

This representation of the page helps the detection of usability problems by analysing the usage statistical information extracted from the log files applied to the structural UI of each page of the site. It answers to first, fourth and fifth questions presented on the abstract.

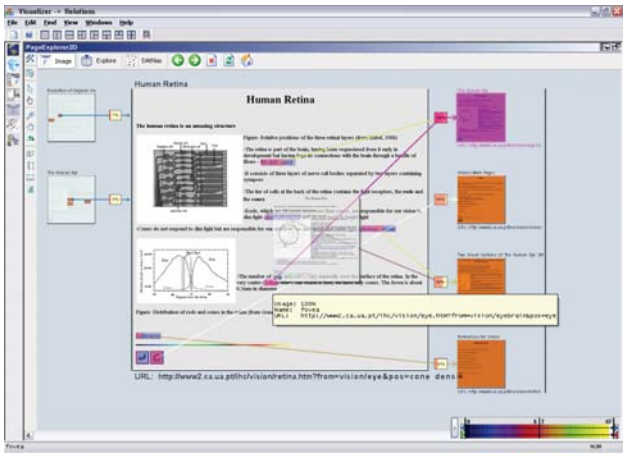

Figure 6 Hypermedia information superimposed to usage data of a specific site page

4.3 Visual workspace

The visual workspace (interface area on computer screen) inspection scheme delivers information on how the user interface of the site is being used. The scheme overlaps colour coded hotspot localization with usage data and hypermedia relations. As can be seen in Figure 7 and lower left quadrant of Figure 5, features can be reported as:

- *Overlap of hotspots of a session's usage to understand/validate the visual workspace organization;*
- *Direct mapping of session jumps to explore visual/interface relations between pages;*
- *Visual inspection of usage patterns and interface design layouts used to detect UI design errors.*

This visualization scheme helps viewing the hottest areas of the user interface, the visual workspace organization of the selected usage sessions being represented. It improves the detection of UI design errors by analyzing the concentration of hits on a specific visual area, one that does not comply with specific usability patterns. It answers the last three questions presented on the abstract.

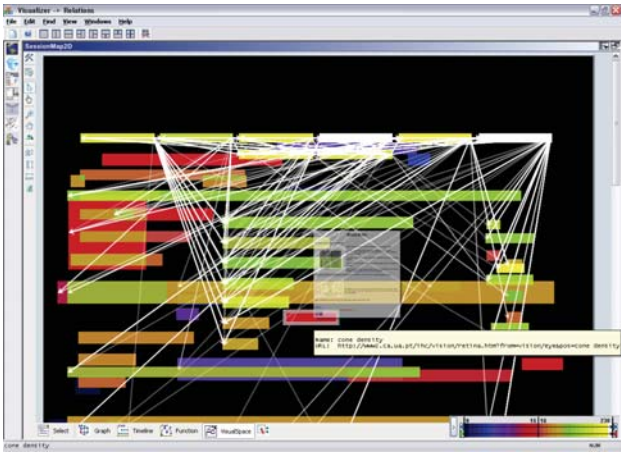

Figure 7 Overlap of hotspot representation superimposed to usage data on visual workspace

4.4 Behavioural function

The behavioural function B(x) represents the basis to calculate and identify a potential problem in the traversing of site data. Based on real usage data, this discrete function is defined with page position in hierarchical site tree (Y axis) and the time at which it was visited (X axis), Figure 8. The differential $\Delta B(x)/\Delta x$ of this function, specifically its high differential values, may depict a serious problem in page relation. Such a situation can be a synonym of information that is out of context in a specific page cluster, or simply just a bad page link established from a deep tree level to a near home page level. This function's major goal is to constitute an awareness feature, an instrument that marks a specific page hotspot/relation as needing inspection due to suspicious hyper-jump in the tree map representation.

The behavioural function answers the first three questions presented on the abstract. It allows the user to observe the site usage sessions, evaluate their comprehension of the inter-site relations and the mapping of the site structural content to the site navigation features.

The colour-codes and shape-codes used by the scheme are interpreted as follow:

- *Green and red shapes (filled circle or filled square) are used to highlight the beginning and, respectively, the end of a usage session;*
- *The blue filled circles are used to code regular jumps between site pages, forward exploration of neighbour site levels being considered regular jumps;*
- *Yellow filled squares represent user jumps out of context or jumps backward several levels,* Figure 9*;*

All these features help the automated detection of incoherent user jumps to backward located pages; the jump distance between site levels giving a good approximation of the user's effort in terms of site exploration.

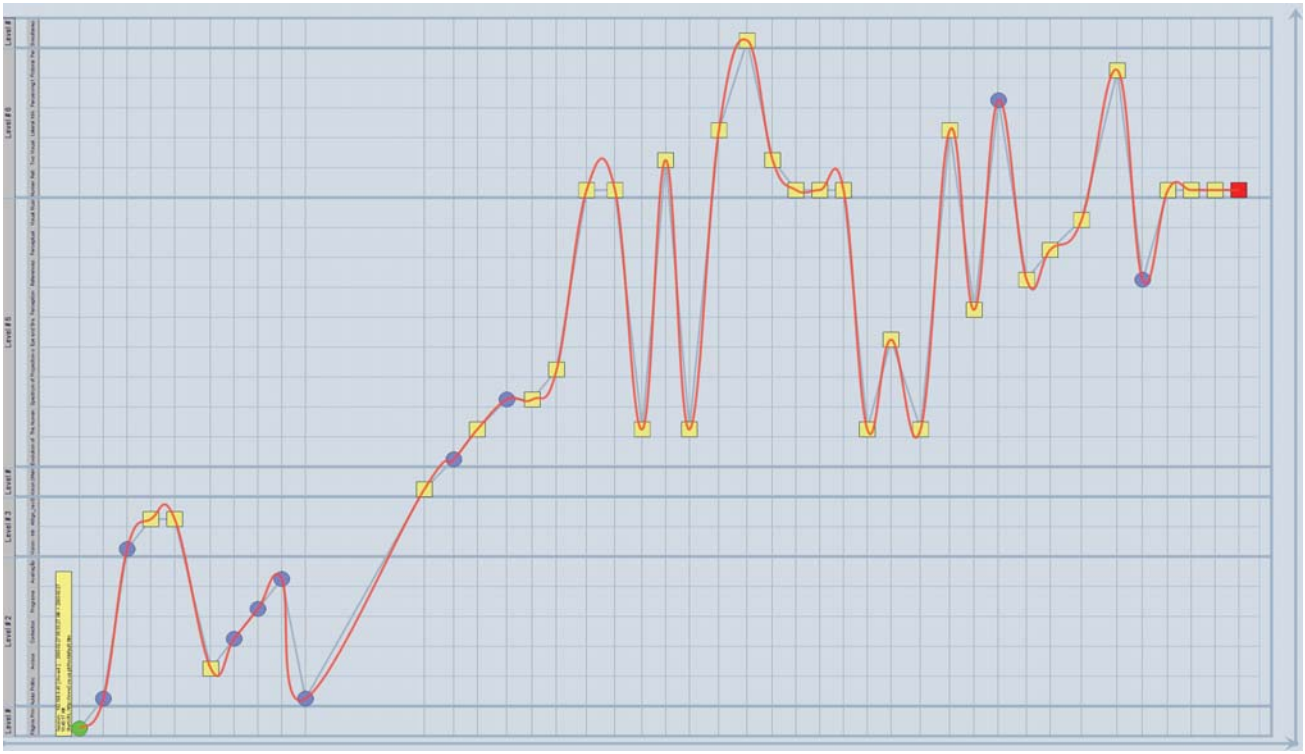

Figure 8 Behavioural function based on page traversing, session data

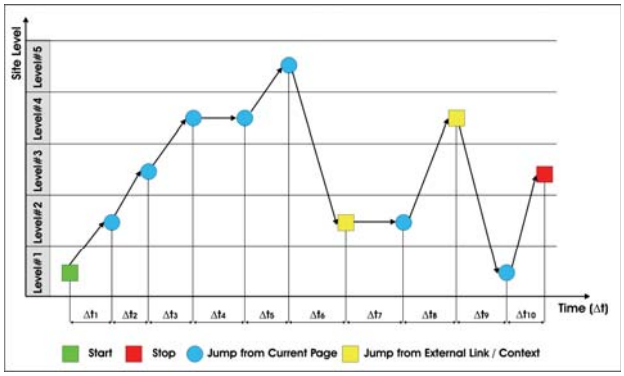

Figure 9 Behavioural function concept

5. CONCLUSIONS

The work reported in this paper represents a coherent research framework integrating visualization concerns in the analysis and representation of interaction / usage patterns inside web based technology / information repositories, specifically organization's intranets.

Besides the conceptual system model, specific details on the information architecture are presented alongside the visualization schemes that are used and the best visual correlation strategies to achieve efficient analysis procedures. Results are presented in the form of prototype screenshots depicting inspection conclusions applied to real data obtained inside University of Aveiro's (Portugal) web repositories. As a synthesis:

i Common visualization schemes, such as tree representations, are augmented with dynamic visual inspection tools based on LUTs and their specific slider functions to enhance inspection and analysis;

ii The visual workspace, or visual interface area on the screen, is inspected with a specific scheme that overlaps interaction hotspots and their hyper-relations between pages;

iii The behavioural function is introduced as an awareness feature that is used to capture form usage data, a potential information context problem in relations with deep tree pages with near home page level pages.

Future work is planned to enhance data visualization schemes directly related with alternative or augmented new website analysis and capture features.

Although ongoing evaluation is expected, the latest version of our system prototype has undergone an empirical and analytical evaluation reported at [Santos04]. The prototype's next evaluation phase will be developed as a real case study, considering common and expert evaluation subjects, and a specific organizational website. The purpose of this evaluation session will be to ask evaluation subjects to perform regular website management tasks, for an objective analysis of the proposed website. The evaluation's starting point is the capture of the website structure followed by the analysis of its usage data and website user's behaviour.

6. REFERENCES

[Card99] Card, S. K., Mackinlay, M. J. and Shneiderman, B., (editors). (1999). *Information Visualization: Using Vision to Think*. San Diego, CA, Morgan Kaufman, pp 493-502.

[Card01] Card, S. (2001). *Information Visualization*, in The Human-Computer Interaction Handbook Fundamentals, Evolving Technologies and Emerging Applications, ed. Jacko, J., Sears, A., Lawrence Erlbaum Associates, pp. 544-582.

[Eick01] Eick, S.G. (2001). *Visualizing Online Activity*. in Communications of the ACM, Vol. 44, No. 8, pp. 45-50.

[Eick99] Eick, S.G. (1999). *Data Visualization Sliders*. In Readings in Information Visualization: Using Vision to Think, (editors) Stuart K. Card, Jock D. Mackinlay and Ben Shneiderman., San Diego, CA, Morgan Kaufman, pp 251-252.

[Healey01] Healey, C.G., Amant, R. St. and Chang, J. (2001). *Assisted Visualization of E-commerce Auction Agent*, presented at Graphics Interface 2001 – Canadian Human-Computer Communications Society, pp. 201-208.

[Mealha04] Mealha, O., Santos, B.S., Nunes, J., Zamfir, F. (2004). *Integrated Visualization Schemes for an Information and Communication Web Log Based Management System*. 8th International Conference Information Visualization, (Eds.) Ebad Banissi, Katy Börner, Chaomei Chen, Muhammad Dastbaz, Gordon Clapworthy, Anthony Faiola, [Izquierdo] Ebroul Izquierdo, Carsten Maple, Jonathan Roberts, Chris Moore, Anna Ursyn and Jian Zhang, Published by the IEEE Computer Society, London, England, ISBN 0-7695-2177-0, ISSN 1093-9547, pp 295-301.

[Nunes03] Nunes, J., Zamfir, F., Mealha, Ó., & Santos, B. S. (2003). Web LogVisualizer: A Tool for Communication and Information Management. In *Proceedings of the 10th International Conference on Human-Computer Interaction – HCI International 2003*, Heraklion – Crete – Greece, Vol. 3, pp. 824-828.

[Santos04] Santos, B.S., Zamfir, F., Ferreira, C., Mealha, O., Nunes, J. (2004). *Visual Application for the Analysis of Web-Based Information System Usage: A Preliminary Usability Evaluation*, 8th International Conference Information Visualization, (Eds.) Ebad Banissi, Katy Börner, Chaomei Chen, Muhammad Dastbaz, Gordon Clapworthy, Anthony Faiola, Ebroul Izquierdo, Carsten Maple, Jonathan Roberts, Chris Moore, Anna Ursyn and Jian Zhang, Published by the IEEE Computer Society, London, England, ISBN 0-7695-2177-0, ISSN 1093-9547, pp 812-818.

[Shneiderman96] Shneiderman, B. (1996). The Eyes Have It: A Task by Data Type Taxonomy for Information Visualizations, In Proceedings of Visual Languages, (Boulder, CO, Sept, 3-6), IEEE Computer Science Press, Los Alamitos, CA, pp 336-343.