

Visual Articulation of Navigation and Search Systems for Digital Libraries

Mario Pérez-Montoro
(University of Barcelona)
Jaume Nualart
(University of Canberra)

ABSTRACT

Journal and digital library portals are the information systems that researchers turn to most frequently for undertaking and disseminating their academic work. However, unlike other information systems, such as e-commerce websites, their interfaces have not been improved on the basis of the findings provided by user studies, nor have the advances developed in specific disciplines, such as information architecture, or those derived more generally from User Experience (UX), been applied to them. In an attempt at overcoming these limitations, from the late eighties onward, a series of prototypes have been developed that seek to improve the visualization of results from journal and digital library portals. Yet, these prototypes and advances in visualization have not been widely adopted for both practical and methodological reasons. To overcome these limitations, new solutions and low-cost tools that can be readily implemented, and which can improve user-experience and user-satisfaction with these information systems, need to be identified. One possible alternative is the articulation of the navigation and search systems in a single visual solution that would allow the simultaneous exploration and interrogation of the information system. *Area* is a new, low-cost visualization tool that is easy to implement, and which can be used with large collections of documents. Moreover, it has a short

learning curve that articulates the two systems using a two-dimensional structure which enhances both user-experience and user-satisfaction with journal and digital library websites.

1. Introduction

When designing a digital information system, the first objective that has to be met is that of facilitating the most intuitive means for users of locating information. To satisfy this objective, the systems of organization, labeling, navigation and searching have to be properly designed, as do the controlled vocabularies that articulate this digital environment (Morville and Rosenfeld, 2007).

For a web page, for example, this means that the organizational systems must serve to structure and organize website content. They are usually constructed by using a classification, based on one or more specific criteria of the content housed on that page (for example, the subject that is being dealt with, the date of creation or the audience being targeted). The labeling system consistently and efficiently defines and determines the terms used to name the categories, options and links used on the web in a user-friendly language. The navigation system allows users to move comfortably around the different sections that make up the website. It provides a method of orientation for users to move in a controlled way from one point of the website to another and to ensure that at all times they know where they are and where they can go within the structure of the web. Based on a previous indexing strategy, the search system allows the user to formulate queries and to retrieve information from within the website. Controlled vocabularies or languages are documentary resources (thesauri, taxonomies, synonym rings, etc.) that facilitate, by articulating the other elements of the architectural structure, the search and retrieval of information on the site (Pérez-Montoro, 2010).

While all these elements form part of the architectural anatomy of a digital information system, the two elements used most frequently by users when seeking information are the search and navigation systems. These two systems tend to be clearly identified in the system interface using the search box and the navigation bar, respectively. Users are typically well versed in their use and, to improve their performance, they are usually articulated via the labeling system (i.e., the navigation system labels are used as indexing terms in the search engine).

In the case of journals and digital libraries, in common with other digital information systems, architectural elements are usually employed to facilitate user location of the information they manage.

Among these elements, the most frequently used are typically their navigation and search systems. In this case, the navigation system is usually quite simple, allowing an exploration of the resources filtered through such criteria as author, year of publication, journal or publisher and, in the best of cases, subject. The results of this navigation appear as a list of clickable labels that lead the user to the set of resources, listed alphabetically, corresponding to these criteria. Search systems usually allow the formulation of queries (e.g., Any Word, All Words or Exact Phrase) by field (e.g., title, description, keywords or anywhere). The result of the query is a list of resources, normally sorted alphabetically too, which corresponds to the criteria in the search interface.

These architectural systems and their interfaces are typically adapted to the nature of the documents managed by these systems and to the metadata used. The documents are static, non-dynamic, resources as far as their content is concerned, and they do not change over time. Moreover, their metadata describe the contents stored (based on qualitative, ordinal, nominal or hierarchical data) (Hearst 2009; van Hoek and Mayr, 2014a).

These systems are the direct heirs of the classical interfaces of the document databases on CD-ROM developed in the eighties and which have barely evolved since. In contrast with other information systems, such as e-commerce websites, their interfaces have not been improved on the basis of the findings provided by user studies, nor have the advances developed in specific disciplines, such as information architecture, or those derived more generally from User Experience (UX), been applied to them.

2. Visualization of information in digital libraries

One of the options for improving classical interfaces is the introduction of new visual solutions in the search process that improve user-experience and user-satisfaction with these digital systems of scientific information.

Traditionally, following on from the initial query, the search systems implemented in information systems of this type offer a very simple representation of the results retrieved. They usually only provide a vertical list of results sorted alphabetically, and, for each result, they give additional information about the retrieved item, such as its author, the title or date of publication of the document, among others.

This strategy of traditional representation has significant limitations. On the one hand, it does not always provide sufficient information about the content of the document to enable the user to accept it or dismiss it without having to read or interact with it first (Baeza-Yates, 2011; Nualart et al., 2014). And, on the other, it does not allow the user to deploy techniques of berrypicking in the search process (Bates, 1989), which could refine the results obtained so as to propose subsequent, more efficient searches in keeping with the user's changing information needs following interaction with the results.

In an attempt at overcoming these limitations, from the late eighties onward, a series of prototypes have been developed that seek to improve the visualization of results from journal and digital library portals. Some have focused on the representation of the content of the retrieved documents (Hearst, 1995; Egan et al., 1989; Weiss-Lijn, McDonnell, & James, 2001; Woodruff et al., 2001; Lam and Baudisch, 2005; Hoerber and Yang, 2006; Nualart and Pérez-Montoro, 2013); while others have contributed new interactive visualizations of the set of results after formulating the search query.

If we focus on the second group of prototypes, we can identify two main types of strategy, some of which are interactive: first, those that provide support for query creation and refinement and, second, those that offer visual support for the presentation of results (van Hoek and Mayr, 2014a).

The earliest techniques were designed to help the user in formulating the query, facilitating the use of Boolean operators (Jones, 1988; Wong et al., 2011) or supplying and suggesting possible terms to the user for building their queries (Schatz et al., 1996).

Those focusing on the visual presentation of results include different alternatives. Some offer two-dimensional visualizations of the relationships between the retrieved documents by using maps or clusters (Chalmers and Chitson, 1992; Andrews et al., 2001; Andrews et al., 2002) or by using two-dimensional tables or grids (Fox et al., 1993; Shneiderman et al., 2000; Kim et al., 2011). Others present strategies based on three-dimensional visualizations of the retrieved results (Robertson et al., 1991; Hearst and Karadi, 1997; Börner et al., 2002; Cugini et al., n.d.; Hienert, 2012). These visual prototypes made a series of significant improvements to the classical interfaces of journal and digital library portals. Thus, on the one hand, they provided more rapid search times compared to those of traditional non-visual methods (Hienert, 2012) and, on the other, they permitted a more efficient formulation of queries in

a way that was tailored to the information needs of users. And, finally, they provided additional information to users, information that was not available on a page of more conventional results. This extra information, which shows different semantic relationships between the documents retrieved, provides a better interaction with the results and facilitates the refinement of subsequent queries (Bauer, 2014).

Yet, even with these advantages, these prototypes and advances in visualization have not been widely implemented in the portals or websites of journals or digital libraries. The reasons for this are varied, but they can be classified into two main groups: reasons of a practical nature and methodological reasons.

In the case of the practical reasons, in resources of this type these tools are implemented as separate pages from the basic search interfaces, which means users perceive them as being secondary tools. Furthermore, these solutions, especially those that visualize the results, involve a high level of abstraction and conceptualization that means they are not very intuitive for users (Hienert, 2012). And, perhaps more importantly, implementing these techniques, unlike traditional interfaces, does not offer any clear commercial or economic benefits in the world of digital systems of scientific information of this type (van Hoek and Mayr, 2014a and 2014b).

If we focus on the methodological reasons, it can be seen that very few of the proposed techniques have been tested and evaluated with end users, which makes it difficult to draw any clear conclusions about their efficiency. Moreover, the prototypes have only been used with small collections of documents, and so their efficient use with large collections has not been demonstrated to users. Likewise, the paucity of the quantitative results reported in these studies of visual prototypes fails to demonstrate whether they are any better than the classical versions of the interfaces. As such, experiments are needed that analyze a period of

widespread use over a broader period of time before it can be concluded whether or not the difficulty in using them stems from the users' learning curve and their degree of familiarity with the system. Similarly, when these prototypes are constructed by articulating different techniques it becomes more difficult to compare them, because it is not possible to attribute unequivocally the success or failure of the system to one or more of the techniques implemented. And, in this sense, these tools do not share a methodological design that would allow us to compare the results of each proposal and to analyze them jointly (van Hoek and Mayr, 2014a and 2014b).

3. *Area*: an alternative visualization proposal

To overcome these practical and methodological limitations, new solutions and low-cost tools that can be readily implemented, and which can improve user-experience and user-satisfaction with these information systems, need to be identified. One possible alternative is the articulation of the navigation and search systems in a single visual solution that would allow the simultaneous exploration and interrogation of the information system.

Area is a new, low-cost visualization tool that is easy to implement, and which can be used with large collections of documents. Moreover, it has a short learning curve that articulates the two systems using a two-dimensional structure that can enhance both user-experience and user-satisfaction with journal and digital library websites.

Although the idea for *Area* originated in 2006, it has evolved since then with the development of versions in several computer languages for a range of different uses and purposes. However, for the experiment reported here *Area* has been completely rewritten. Today it is a simpler

version that runs completely on the client side from a standard browser. *Area* is free software.

In presenting this alternative visualization proposal, we have selected the contents of the journal *Information Research* to serve as our corpus of texts on which we demonstrate the tool's visualization and exploration capacities. To do so, we replicated these contents on a standalone server, where *Area* is presented as an alternative interface to that of the *Information Research* journal, yet emulating all its capabilities and adding additional ones (see Table 4) (Nualart, 2014a).

We have chosen the contents of *Information Research* for two reasons. On the one hand, it serves as a good example of an open access journal with the collection being published online under a Creative Commons license and, on the other, academic papers represent a controlled collection of texts with a similar language register, structure and length, which gives the collection a homogeneous shape. All the codes related to this experiment, as well as the *Area* software itself, can be downloaded from the GitHub repository (Nualart, 2014c).

3.1. *Area*'s Visualization Capacity

Generally speaking, *Area* is an architectural proposal in which we articulate, in a single structure, the two main systems facilitating the location of information in digital contexts, i.e., the navigation system and the search system. These two systems are present in most contexts, which is a guarantee that users are fully familiar with them and that additional specific instructions are not needed for them to use *Area* efficiently and comfortably.

Area represents two of the eligible properties simultaneously. The first property is represented graphically as blocks. These form a grid of blocks

that contain the items in the collection, depicted as small squares. The second property is the color representation of each item (see Figure 4). This particular architectural structure provides the tool with a series of capabilities for locating and visualizing the information contained in the collection that makes up the web page of the journal or digital library.

First, the system can browse the collection and simultaneously select two of the attributes of each document in the collection: the year of publication, the volume in which it appears, the issue in which it was published, the number of references per article and the language in which the article is written. The application of this double selection process generates a two-dimensional representation in which all (not just part) of the collection of documents managed on the web page of the journal or in the digital library is depicted, unlike classical navigation and search systems. This presentation allows us to visualize information about the collection, such as the volume of the collection referred to, the way in which the volume or issues are distributed throughout the year, the annual variation in the number of references included in the documents and the distribution of articles by languages. These indicators are not available in classical systems.

By clicking on one of the rectangles (representing a document) in the grid, a central window opens showing all the available bibliographic information (title, author, volume, number of references, etc.) about that selected paper. The system allows 25 combinations of “eligible properties” (5x5), of which twenty relate two different properties (bivariates) and five represent the collection in terms of a single attribute (univariates), where blocks and colors coincide. Figure 4 shows the entire collection of documents from *Information Research* using as our criteria the Year and the Number of References. Each block corresponds to a Year and each rectangle corresponds to a document colored according to the number of references that it includes.

Second, once the collection has been presented in terms of the combination of criteria or parameters, the system allows us to apply a series of filters to locate documents that can help the user meet her information needs. The documents corresponding to the filters are highlighted in black. Specifically, three different types of filter are available: author, subject and manual with field selection.

If we click the tab marked “by Author” tab (top left), we can write the name of the author of the documents we seek or choose the author from the list of all authors that have published in the journal. This second option should be understood as a system query-builder (Figure 1).

If we click the tab marked “by Subject” (top left), we can write the subject of the documents we seek or choose the subject from the list of all subjects dealt with by the documents in the collection. This second option should, once more, be understood as a system query-builder (Figure 2).

The manual filter – “filter paper” (in the right-hand column, mid-zone) – allows a text to be filtered by the attributes or parameters of the document, namely, Title, Citation, Year, Authors or Subject/s (Figure 3). The user can choose which of these fields they want to filter for. If more than one filter is selected, the OR operator functions between them. If the Author filter is selected, we can also filter by the university to which the author is affiliated or the city in which the author lives.

It should be pointed out that the first two filters (“by Author” and “by Subject”) are not cumulative, so that every time we write something in the corresponding box this overrides the previous filter. Once one of these two filters is completed, if we change the attributes, the filter is maintained. Moreover, once we have used one of those two filters, we can see what we have typed using the query builder (clicking) as it will appear written in the manual filter box.

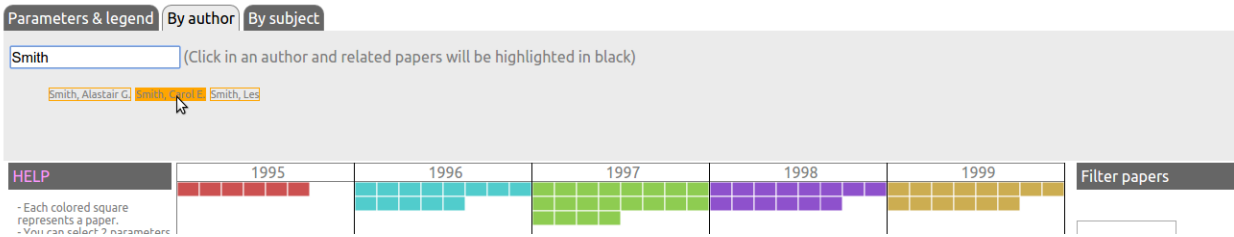


Figure 1. Detail of filter “by Author”.

Third, *Area* allows us to customize the visualization by giving users the possibility of varying the colors (fixed, random or gradient mode) and thus overcome any potential problems of color-blindness that users might suffer from. It should also be stressed that it incorporates (left-hand column, mid-zone) a support text which explains how to use the tool and an overview of the data in the collection making up the journal or digital library web page.

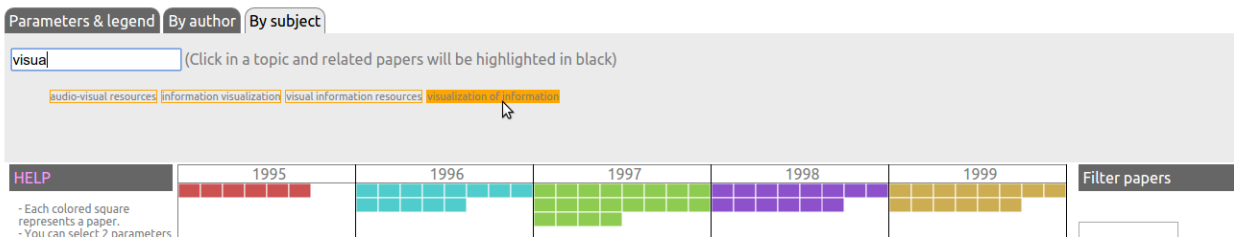


Figure 2. Detail of Filter “by Subject”.

Fourth, *Area* also includes the original location systems available on the *Information Research* website. Thus, the filters offered by the tool can be understood as a complement to the Google and Atomz searches offered by the *Information Research* website.

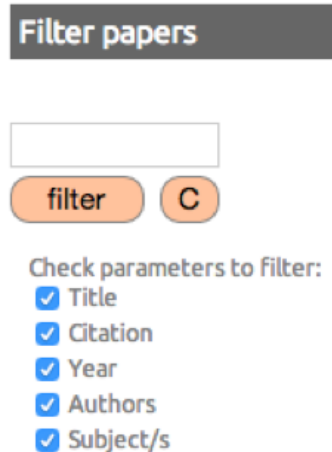


Figure 3. Detail of manual filter.

Finally, unlike alternative visualization systems, the possibilities offered by *Area* are not visually affected by the size of the collection represented. By incorporating a grid that grows in function of the size of the collection, and not depending on other systems such as 3-D or clusters, it avoids the potential visual overlapping of information and the production of visual noise when representing large quantities of documents. *Area*, as specified in the technical description, is recommended for collections of up to 50,000 items.

3.2 Technical description

Area is a simple, small application coded in Javascript, which uses the libraries jquery and D3, HTML, and CSS. The data files are stored in JSON format and the application is accessible with a modern browser. When visiting the *Area* website the client can download all the necessary files to run the application entirely on the client side.

The implementation of this application faces two main constraints: the number of items in the dataset and the dataset size. The first of these is related to screen resolution while the second is related to the size of RAM memory available on the client side. Performance tests conducted

(Nualart, 2015c) suggest the use of collections that do not exceed fifty thousand items.

Area represents the metadata of a collection of items, allowing filtering and the exploration of the contents of each item. Each time the application and the data files are downloaded, the properties from the metadata schema are analyzed. In those cases in which the number of possible values of a property is not greater than a configured value, then the group of eligible properties is added.

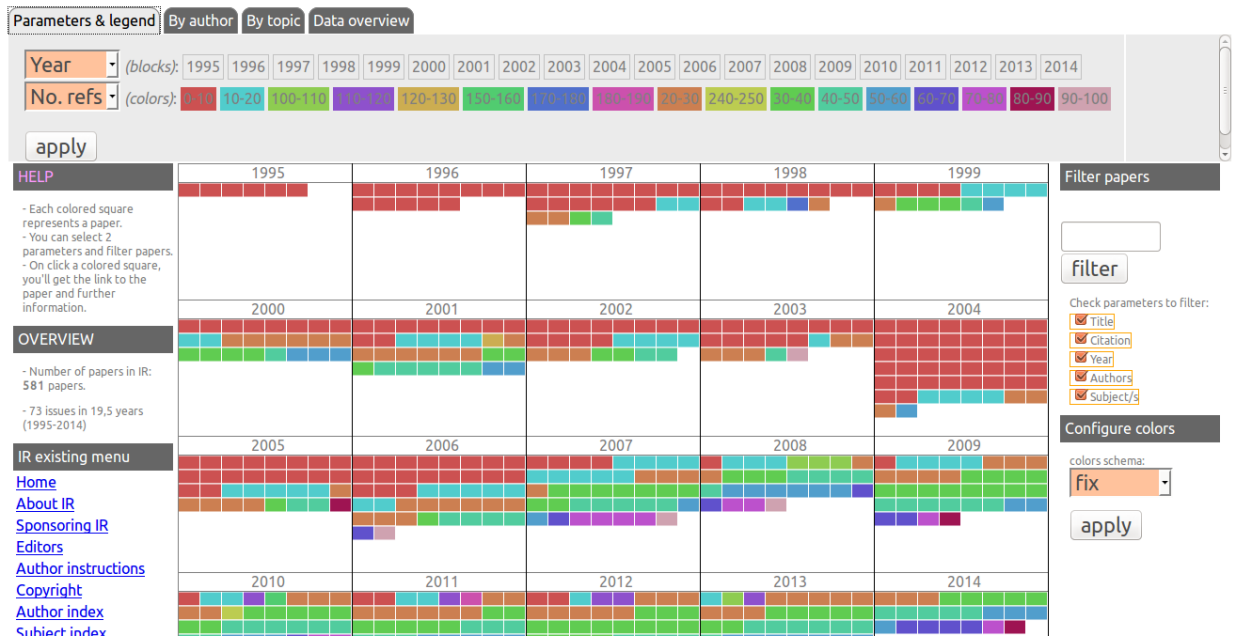


Figure 4. Screenshot of Area interface. The first eligible property is “Year” of publication, represented by twenty-five blocks. The second eligible property is “no. of references/papers”, grouped in seventeen categories and is represented by a different color.

Area was tested in 2014 on desktops, laptops, mobile phones and tablets. All were found to offer good interface responsiveness. However, small screens need to use scroll and zoom in order to provide the same experience as that on larger screens.

4. User evaluation test

To gain a better understanding of the potential of the visual exploration and search of text collections with the *Area* tool, we undertook a web-based survey.

The aim of the survey was to compare the text-based website with *Area* for the presentation of collections of texts, specifically, scientific papers. To this end, we addressed the following questions: Are users able to detect the new features? Do users still prefer or require access to the existing presentation? Are users able to understand the new features? Do users feel confident and positive about using the new features?

The design of the experiment is based on the established technology acceptance model (TAM) (Davis et al., 1989), and the task technology fit (TTF) (Goodhue 1995). TAM seeks to understand why people accept or reject information technologies, whereas TTF says that technologies will be used if, and only if, their available functionalities support the user's activities. As such, the focus is on the match between the user's task needs and the available functionalities of a given technology. The questions have been designed following Taylor-Powell and Marshall (1996).

In the rest of this section we explain the data collection process: choice, download and storage. Then we describe the demographics of the participants. Finally, we explain in detail the content of the questionnaire administered to the users. In the section the follows we discuss the results of the evaluation.

4.1. Data Collection

We used the collection of papers in *Information Research* (IR), edited by Prof. T.D. Wilson (<http://www.informationr.net/ir>). It has been published since 1995, and as of November 2014 the journal has published 592 papers, in 74 quarterly issues, and 19 yearly volumes.

We selected the contents of *Information Research* to provide the corpus of texts for this experiment for the two main reasons discussed above, namely its status as an open access journal, published online under a Creative Commons license and, because its academic papers constitute a controlled collection of texts with a similar language register, structure and length, giving the collection a homogeneous shape.

In designing *Area* we sought to provide most of the features that the existing *Information Research* website offers. Indeed, for some features *Area* redirects the user to the existing services on the website. This is the case of Atomz search and the domain-restricted Google search. Other features have been improved in *Area*, specifically, exploring the collection by year, by language, by number of references per paper, by issue and by volume, and exploring by subject and by author.

To obtain the data collection we harvested the contents from the journal's website. Papers have been published in different versions of HTML, reflecting the evolution of the markup language since 1995 and changes dictated by the publishers in the structure of the pages. We customized the spiders to the non-homogeneous HTML structure of the corpus.

After cleaning the data and adding HTML entities for all special characters, above all for authors' names, we selected several metadata properties for each paper. See Table 1 for a detailed list.

#	metadata properties	type	In which page of the journal is this data?	Function (no. of different values)
1	Volume	integer	paper page + issue page	Eligible property (nineteen volumes)
2	Issue	integer	paper page + issue page	Eligible property (seventy-four issues and four values)
3	Year	integer	paper page + issue page	Eligible property (twenty years)
4	Number of references (grouped)	integer	paper page	Eligible property (seventeen groups)
5	Language	string	paper page + issue page	Eligible property & Searchable (three languages: English, Spanish, Portuguese)
6	Title	string	paper page + issue page	Searchable (592 values)
7	Authors, institution/country	string	paper page	Searchable (582 values)
8	Citation	string	paper page	Searchable (592 values)
9	Paper URL	URL string	paper page + issue page	External link (592 values)
10	Issue URL	URL string	paper page + issue page	External link (74 values)
11	Number of references	integer	paper page	Property (101 values)
12	Paper subjects	string	by-subject page	Property (400 subjects)
13	Individual author names	string	by-author page	Property (895 authors)

TABLE 1: Metadata properties: list and details.

In line with the conditions described above, five properties were labeled as being eligible: Volume, Issue number, Year, Grouped number of references/papers, and Language. The remaining properties (eight) were: Title, Authors with institutions and countries, Citation, Paper URL, Issue URL, Number of references, Paper subjects, and Individual author names.

This valuable metadata from the contents of *Information Research* were stored in JSON files, and like the rest of the code, have been published under free licenses to allow others to reuse them.

4.2. Questionnaire description

Online questionnaires are the most frequently employed method for collecting quantitative data from users for statistical analysis. Questionnaires allow the participation of an unlimited number of people and can be used to gather data about users' knowledge, beliefs, attitudes, and behaviors (Taylor-Powell and Marshall, 1996). Online questionnaires also make it easier to protect the privacy of participants.

The questionnaire comprised fourteen questions. Seven demographic questions and seven specific questions compare the tasks and features of the *Information Research* website and the *Area* website.

Eligible respondents of the questionnaire were any potential visitors of an academic journal. Initially, participants were invited to visit the existing *Information Research* website and the *Area* website in order to familiarize themselves with them and so as to be able to answer the questions. In order to find participants, open calls were sent out using mailing lists of PhD and Master's students.

5. Results

The questionnaire was answered by forty-four respondents, with thirty-seven completing all the questions. One out of three respondents were women and seven out of ten were between thirty and fifty years of age. All the participants said they had either a good, very good or expert technical knowledge of computers in approximately equal proportions. In

line with this, seven out of ten of the participants use web browsers several times a day (see Table 2).

<p>Gender</p> <table border="1"> <thead> <tr> <th>Answer</th> <th>Count</th> <th>Percentage</th> </tr> </thead> <tbody> <tr> <td>Female</td> <td>13</td> <td>29.55%</td> </tr> <tr> <td>Male</td> <td>28</td> <td>63.64%</td> </tr> <tr> <td>Other</td> <td>1</td> <td>2.27%</td> </tr> <tr> <td>No answer</td> <td>2</td> <td>4.55%</td> </tr> </tbody> </table>	Answer	Count	Percentage	Female	13	29.55%	Male	28	63.64%	Other	1	2.27%	No answer	2	4.55%	<p>How old are you?</p> <table border="1"> <thead> <tr> <th>Answer</th> <th>Count</th> <th>Percentage</th> </tr> </thead> <tbody> <tr> <td>less than 20</td> <td>0</td> <td>0.00%</td> </tr> <tr> <td>20 to 30</td> <td>6</td> <td>13.64%</td> </tr> <tr> <td>30 to 40</td> <td>14</td> <td>31.82%</td> </tr> <tr> <td>40 to 50</td> <td>16</td> <td>36.36%</td> </tr> <tr> <td>50 to 60</td> <td>3</td> <td>6.82%</td> </tr> <tr> <td>more than 60</td> <td>3</td> <td>6.82%</td> </tr> <tr> <td>No answer</td> <td>2</td> <td>4.55%</td> </tr> </tbody> </table>	Answer	Count	Percentage	less than 20	0	0.00%	20 to 30	6	13.64%	30 to 40	14	31.82%	40 to 50	16	36.36%	50 to 60	3	6.82%	more than 60	3	6.82%	No answer	2	4.55%			
Answer	Count	Percentage																																									
Female	13	29.55%																																									
Male	28	63.64%																																									
Other	1	2.27%																																									
No answer	2	4.55%																																									
Answer	Count	Percentage																																									
less than 20	0	0.00%																																									
20 to 30	6	13.64%																																									
30 to 40	14	31.82%																																									
40 to 50	16	36.36%																																									
50 to 60	3	6.82%																																									
more than 60	3	6.82%																																									
No answer	2	4.55%																																									
<p>How would you rate your technical knowledge of computers and the web?</p> <table border="1"> <thead> <tr> <th>Answer</th> <th>Count</th> <th>Percentage</th> </tr> </thead> <tbody> <tr> <td>None</td> <td>0</td> <td>0.00%</td> </tr> <tr> <td>Some</td> <td>0</td> <td>0.00%</td> </tr> <tr> <td>Good</td> <td>16</td> <td>36.36%</td> </tr> <tr> <td>Very good</td> <td>14</td> <td>31.82%</td> </tr> <tr> <td>Expert</td> <td>12</td> <td>27.27%</td> </tr> <tr> <td>No answer</td> <td>2</td> <td>4.55%</td> </tr> </tbody> </table>	Answer	Count	Percentage	None	0	0.00%	Some	0	0.00%	Good	16	36.36%	Very good	14	31.82%	Expert	12	27.27%	No answer	2	4.55%	<p>How often do you use web browsers?</p> <table border="1"> <thead> <tr> <th>Answer</th> <th>Count</th> <th>Percentage</th> </tr> </thead> <tbody> <tr> <td>Once a month or less</td> <td>0</td> <td>0.00%</td> </tr> <tr> <td>Once per week</td> <td>0</td> <td>0.00%</td> </tr> <tr> <td>Several times a week</td> <td>3</td> <td>6.82%</td> </tr> <tr> <td>Every day</td> <td>7</td> <td>15.91%</td> </tr> <tr> <td>Several times a day</td> <td>32</td> <td>72.73%</td> </tr> <tr> <td>No answer</td> <td>2</td> <td>4.55%</td> </tr> </tbody> </table>	Answer	Count	Percentage	Once a month or less	0	0.00%	Once per week	0	0.00%	Several times a week	3	6.82%	Every day	7	15.91%	Several times a day	32	72.73%	No answer	2	4.55%
Answer	Count	Percentage																																									
None	0	0.00%																																									
Some	0	0.00%																																									
Good	16	36.36%																																									
Very good	14	31.82%																																									
Expert	12	27.27%																																									
No answer	2	4.55%																																									
Answer	Count	Percentage																																									
Once a month or less	0	0.00%																																									
Once per week	0	0.00%																																									
Several times a week	3	6.82%																																									
Every day	7	15.91%																																									
Several times a day	32	72.73%																																									
No answer	2	4.55%																																									
<p>Do you like to encounter new features in the pages you visit the most?</p> <table border="1"> <thead> <tr> <th>Answer</th> <th>Count</th> <th>Percentage</th> </tr> </thead> <tbody> <tr> <td>Not at all</td> <td>0</td> <td>0.00%</td> </tr> <tr> <td>Rarely</td> <td>2</td> <td>4.55%</td> </tr> <tr> <td>No opinion</td> <td>8</td> <td>18.18%</td> </tr> <tr> <td>Sometimes</td> <td>25</td> <td>56.82%</td> </tr> <tr> <td>Often</td> <td>7</td> <td>15.91%</td> </tr> <tr> <td>No answer</td> <td>2</td> <td>4.55%</td> </tr> </tbody> </table>	Answer	Count	Percentage	Not at all	0	0.00%	Rarely	2	4.55%	No opinion	8	18.18%	Sometimes	25	56.82%	Often	7	15.91%	No answer	2	4.55%	<p>Are you happy with the information tools and interfaces that you use?</p> <table border="1"> <thead> <tr> <th>Answer</th> <th>Count</th> <th>Percentage</th> </tr> </thead> <tbody> <tr> <td>Very happy</td> <td>1</td> <td>2.27%</td> </tr> <tr> <td>Happy</td> <td>20</td> <td>45.45%</td> </tr> <tr> <td>Indifferent</td> <td>6</td> <td>13.64%</td> </tr> <tr> <td>Unhappy</td> <td>14</td> <td>31.82%</td> </tr> <tr> <td>Very unhappy</td> <td>1</td> <td>2.27%</td> </tr> <tr> <td>No answer</td> <td>2</td> <td>4.55%</td> </tr> </tbody> </table>	Answer	Count	Percentage	Very happy	1	2.27%	Happy	20	45.45%	Indifferent	6	13.64%	Unhappy	14	31.82%	Very unhappy	1	2.27%	No answer	2	4.55%
Answer	Count	Percentage																																									
Not at all	0	0.00%																																									
Rarely	2	4.55%																																									
No opinion	8	18.18%																																									
Sometimes	25	56.82%																																									
Often	7	15.91%																																									
No answer	2	4.55%																																									
Answer	Count	Percentage																																									
Very happy	1	2.27%																																									
Happy	20	45.45%																																									
Indifferent	6	13.64%																																									
Unhappy	14	31.82%																																									
Very unhappy	1	2.27%																																									
No answer	2	4.55%																																									
<p>Have you ever published a paper in an academic conference?</p> <table border="1"> <thead> <tr> <th>Answer</th> <th>Count</th> <th>Percentage</th> </tr> </thead> <tbody> <tr> <td>Yes</td> <td>22</td> <td>56.41%</td> </tr> <tr> <td>No</td> <td>17</td> <td>43.59%</td> </tr> <tr> <td>No answer</td> <td>0</td> <td>0.00%</td> </tr> </tbody> </table>	Answer	Count	Percentage	Yes	22	56.41%	No	17	43.59%	No answer	0	0.00%	<p>How often do you read papers from academic journals?</p> <table border="1"> <thead> <tr> <th>Answer</th> <th>Count</th> <th>Percentage</th> </tr> </thead> <tbody> <tr> <td>More than once per week)</td> <td>11</td> <td>29.73%</td> </tr> <tr> <td>Once per month</td> <td>11</td> <td>29.73%</td> </tr> <tr> <td>Once per year</td> <td>6</td> <td>16.22%</td> </tr> <tr> <td>Never</td> <td>9</td> <td>24.32%</td> </tr> <tr> <td>No answer</td> <td>0</td> <td>0.00%</td> </tr> </tbody> </table>	Answer	Count	Percentage	More than once per week)	11	29.73%	Once per month	11	29.73%	Once per year	6	16.22%	Never	9	24.32%	No answer	0	0.00%												
Answer	Count	Percentage																																									
Yes	22	56.41%																																									
No	17	43.59%																																									
No answer	0	0.00%																																									
Answer	Count	Percentage																																									
More than once per week)	11	29.73%																																									
Once per month	11	29.73%																																									
Once per year	6	16.22%																																									
Never	9	24.32%																																									
No answer	0	0.00%																																									

Table 2: Questionnaire demographics

The attitude of the participants to the new features found on the websites was positive: they like to find new features sometimes (56.82%) or often

(15.91%). Other answers were: no opinion (18.18%), and rarely (4.55%). In contrast, almost half of the participants said they were happy (45.45%) with the information tools and interfaces they use. Finally, more than half of the participants (56.41%) have published scientific papers, and three out of four read scientific papers on quite a regular basis.

<p>How many papers have been published in the journal since the first issue?</p> <table border="1"> <thead> <tr> <th>Answer</th> <th>Count</th> <th>Percentage</th> </tr> </thead> <tbody> <tr> <td>IR existing interface</td> <td>2</td> <td>5.41%</td> </tr> <tr> <td>IR Area interface</td> <td>31</td> <td>83.78%</td> </tr> <tr> <td>No difference</td> <td>4</td> <td>10.81%</td> </tr> <tr> <td>No answer</td> <td>0</td> <td>0.00%</td> </tr> </tbody> </table>	Answer	Count	Percentage	IR existing interface	2	5.41%	IR Area interface	31	83.78%	No difference	4	10.81%	No answer	0	0.00%	<p>How is the term "visualization" distributed in the history of the journal?</p> <table border="1"> <thead> <tr> <th>Answer</th> <th>Count</th> <th>Percentage</th> </tr> </thead> <tbody> <tr> <td>IR existing interface</td> <td>1</td> <td>2.70%</td> </tr> <tr> <td>IR Area interface</td> <td>31</td> <td>83.78%</td> </tr> <tr> <td>No difference</td> <td>5</td> <td>13.51%</td> </tr> <tr> <td>No answer</td> <td>0</td> <td>0.00%</td> </tr> </tbody> </table>	Answer	Count	Percentage	IR existing interface	1	2.70%	IR Area interface	31	83.78%	No difference	5	13.51%	No answer	0	0.00%
Answer	Count	Percentage																													
IR existing interface	2	5.41%																													
IR Area interface	31	83.78%																													
No difference	4	10.81%																													
No answer	0	0.00%																													
Answer	Count	Percentage																													
IR existing interface	1	2.70%																													
IR Area interface	31	83.78%																													
No difference	5	13.51%																													
No answer	0	0.00%																													
<p>How many papers talk about visualization?</p> <table border="1"> <thead> <tr> <th>Answer</th> <th>Count</th> <th>Percentage</th> </tr> </thead> <tbody> <tr> <td>IR existing interface</td> <td>3</td> <td>8.11%</td> </tr> <tr> <td>IR Area interface</td> <td>24</td> <td>64.86%</td> </tr> <tr> <td>No difference</td> <td>10</td> <td>27.03%</td> </tr> <tr> <td>No answer</td> <td>0</td> <td>0.00%</td> </tr> </tbody> </table>	Answer	Count	Percentage	IR existing interface	3	8.11%	IR Area interface	24	64.86%	No difference	10	27.03%	No answer	0	0.00%	<p>When exploring papers of the journal website: do you have a better overview of the journal using the existing interface or the Area interface?</p> <table border="1"> <thead> <tr> <th>Answer</th> <th>Count</th> <th>Percentage</th> </tr> </thead> <tbody> <tr> <td>IR existing interface</td> <td>6</td> <td>16.22%</td> </tr> <tr> <td>IR Area interface</td> <td>30</td> <td>81.08%</td> </tr> <tr> <td>No difference</td> <td>1</td> <td>2.70%</td> </tr> <tr> <td>No answer</td> <td>0</td> <td>0.00%</td> </tr> </tbody> </table>	Answer	Count	Percentage	IR existing interface	6	16.22%	IR Area interface	30	81.08%	No difference	1	2.70%	No answer	0	0.00%
Answer	Count	Percentage																													
IR existing interface	3	8.11%																													
IR Area interface	24	64.86%																													
No difference	10	27.03%																													
No answer	0	0.00%																													
Answer	Count	Percentage																													
IR existing interface	6	16.22%																													
IR Area interface	30	81.08%																													
No difference	1	2.70%																													
No answer	0	0.00%																													
<p>Understanding the topics and themes of the journal</p> <table border="1"> <thead> <tr> <th>Answer</th> <th>Count</th> <th>Percentage</th> </tr> </thead> <tbody> <tr> <td>IR existing interface</td> <td>4</td> <td>10.81%</td> </tr> <tr> <td>IR Area interface</td> <td>31</td> <td>83.78%</td> </tr> <tr> <td>No difference</td> <td>2</td> <td>5.41%</td> </tr> <tr> <td>No answer</td> <td>0</td> <td>0.00%</td> </tr> </tbody> </table>	Answer	Count	Percentage	IR existing interface	4	10.81%	IR Area interface	31	83.78%	No difference	2	5.41%	No answer	0	0.00%	<p>Finding papers related to your personal interests</p> <table border="1"> <thead> <tr> <th>Answer</th> <th>Count</th> <th>Percentage</th> </tr> </thead> <tbody> <tr> <td>IR existing interface</td> <td>3</td> <td>8.11%</td> </tr> <tr> <td>IR Area interface</td> <td>30</td> <td>81.08%</td> </tr> <tr> <td>No difference</td> <td>4</td> <td>10.81%</td> </tr> <tr> <td>No answer</td> <td>0</td> <td>0.00%</td> </tr> </tbody> </table>	Answer	Count	Percentage	IR existing interface	3	8.11%	IR Area interface	30	81.08%	No difference	4	10.81%	No answer	0	0.00%
Answer	Count	Percentage																													
IR existing interface	4	10.81%																													
IR Area interface	31	83.78%																													
No difference	2	5.41%																													
No answer	0	0.00%																													
Answer	Count	Percentage																													
IR existing interface	3	8.11%																													
IR Area interface	30	81.08%																													
No difference	4	10.81%																													
No answer	0	0.00%																													
<p>Exploring new topics and discovering new research in this field</p> <table border="1"> <thead> <tr> <th>Answer</th> <th>Count</th> <th>Percentage</th> </tr> </thead> <tbody> <tr> <td>IR existing interface</td> <td>5</td> <td>13.51%</td> </tr> <tr> <td>IR Area interface</td> <td>29</td> <td>78.38%</td> </tr> <tr> <td>No difference</td> <td>3</td> <td>8.11%</td> </tr> <tr> <td>No answer</td> <td>0</td> <td>0.00%</td> </tr> </tbody> </table>	Answer	Count	Percentage	IR existing interface	5	13.51%	IR Area interface	29	78.38%	No difference	3	8.11%	No answer	0	0.00%																
Answer	Count	Percentage																													
IR existing interface	5	13.51%																													
IR Area interface	29	78.38%																													
No difference	3	8.11%																													
No answer	0	0.00%																													

Table 3: Task questionnaire

We asked participants to compare several tasks completed with the journal's existing interface, on the one hand, and with Area's interface, on the other. To answer the questions we encouraged participants to visit both sites and to familiarize themselves with their interfaces before they

started to complete the questionnaire. For all seven tasks, users preferred the new interface. In six out of the seven, participants preferred *Area* for solving the proposed tasks in 80% of cases.

6. Discussion

As is apparent from the results obtained, the *Area* visualization prototype is, in the opinion of users, better than the conventional tool available on the website of the *Information Research* journal.

In response to all the questions asked, *Area* obtained a more positive response than that given to the classical visualization tool (in six of them there was a roughly 80% preference and in the remaining task a 64.86% preference).

Thus, *Area* was preferred by 80% of the users for completing the following tasks:

- Verifying the number of papers making up the collection.
- Identifying the number of papers addressing a specific subject and their distribution in time.
- Obtaining an overview of the collection.
- Understanding the subjects addressed by the journal.
- Finding papers related to a user's interests.

And 64.86% of users preferred it for:

- Exploring new topics and discovering new research in a specific field.

These results can be attributed to the enhanced capacity of visualization provided by *Area* compared to that provided by the classical resources of information presentation included on the journal's website. In the case of the following functions: *Explore by issue as a list of papers*, *Search with*

Atomz, and *Search with Google*, *Area* redirects users to the resources on the journal's web page.

However, *Area* betters the classical visualization tool (included in the interface of most journals and digital libraries) in several of its features (Table 3). On the one hand, it incorporates new visualization features that are not available in the classical proposal. For example, it allows the user to visualize the whole collection in different ways depending on the two properties and filters selected, and not just as a subset of the whole as in classical systems. We have named this new function: *Multiple overviews of the collection*. *Area* also provides rapid access to the quantitative characteristics of the collection (Numbers of papers, Issues, Volumes, Years, etc.), a function named: *Numerical overview*. *Area* also shows the user how a subject is distributed during the history of the journal as it allows filtered papers to be marked during exploration. A feature we have named *Topic distribution*. And, finally, *Area* allows the user to explore papers by language and to see the evolution in this language, since language is an eligible property and it can be represented in combination with the other properties. We have named this new function: *Explore by language*.

Features	Information Research	Area	Comment
<i>Explore by issue as a list of papers</i>	✓	✓	<i>No changes: Area redirects to the existing issue page</i>
<i>Search with Atomz, and Search with Google</i>	✓	✓	<i>No changes: It redirects to the IR search page</i>
<i>Multiples overviews of the collection</i>	✗	✓	<i>New feature: (no. of eligible properties)² This is 5² = 25 combinations of eligible properties</i>















<i>Numerical overview</i>			<i>New feature: Area shows an overview of the main numbers of the collection</i>
<i>Topic distribution</i>			<i>New feature: filter papers are marked during exploration.</i>
<i>Explore by Language</i>			<i>New feature: Language is an eligible property. So it can be represented in combination with the other of properties.</i>
<i>Explore by Year, Issue and Volume</i>			<i>Improved feature: multiple representations and evolution visualization</i>
<i>Explore by Subject</i>			<i>Improved feature: TAB “by topic”, allows filter by typing</i>
<i>Explore by Author (authors can have more than one paper)</i>			<i>Improved feature: TAB “by author”, allows filter by typing</i>
<i>How many papers talk about a subject?</i>			<i>Improved: Area shows the papers and their context. A better visualization of the group of results</i>

Table 4. Comparison of features of existing Information Research site and Area

Area also improves certain functions that already exist in the classical version. For example, it improves the *Explore by Year, Issue and Volume* function by allowing multiple representations and evolution visualization. It also improves the functions of *Explore “by author”* and *Explore “by subject”* by allowing filter-by-typing. Finally, *Area* improves the function of identifying *How many papers talk about a subject?* by showing the papers and their context.

7. Conclusions

These new visualization functions and the outcomes recorded allow us to draw a number of conclusions.

The test conducted on *Area* provided positive responses to the questions that we set out to answer: users detect and understand new features; users prefer or require new ways of presenting information; and users feel confident and positive about using the new features.

The simplicity and economy of the *Area* prototype should pave the way for the widespread introduction of these visualization tools in the portals and websites of journals and digital libraries. The fact that *Area* is not implemented as a page which is independent of the basic search interface means that it is not perceived by users as a secondary tool; nor does the prototype present a high level of abstraction and conceptualization that means its use is not very intuitive for users. Similarly, *Area*, by basing its visualization power on the metadata file, is a non-intrusive system that only needs to be accessible from any point in the network and, once downloaded locally, it allows interaction without an Internet connection. Unlike other prototypes that have been implemented only with small collections of documents and in highly controlled experimental conditions, *Area* has been implemented in a real world context with the entire collection of documents from a journal (not just with a subset of retrieved documents). Therefore, the user-satisfaction results reported here cannot be dismissed on the grounds of their having been obtained with a limited collection or a limited number of documents. Finally, it should be stressed that *Area* is a free licensed tool that is readily implemented which, unlike other more abstract and expensive prototypes, facilitates its introduction in journal and digital library sites.

8. References

- Andrews, K., Gutl, C., Moser, J., Sabol, V., & Lackner, W. (2001). "Search result visualisation with xFIND". *Proceedings of the Second International Workshop on User Interfaces to Data Intensive Systems*. p. 50–58. IEEE Computer Society Press.
- Andrews, Keith, Kienreich, W., Sabol, V., Becker, J., Droschl, G., Kappe, F., Tochtermann, K. (2002). "The InfoSky visual explorer: exploiting hierarchical structure and document similarities". *Information Visualization*, 1(3-4), p. 166–181.
- Baeza-Yates, R. (2011). "Tendencias en recuperación de información en la web". [Trends in information retrieval on the Web.] *BiD: textos universitaris de biblioteconomia i documentació*, desembre, núm. 27. Retrieved from <http://www.ub.edu/bid/27/baeza2.htm> on 22-10-2014.
- Bates, Marcia J. (1989). "The Design of Browsing and Berrypicking Techniques for the Online Search Interface". *Online Review*, n. 13, p. 407-424.
- Bauer, Sabine (2014). *Interactive Visualizations for Search Processes*. PhD Dissertation.
- Chalmers, M., and Chitson, P. (1992). "Bead: explorations in information visualization". p. 330-337). ACM Press.
- Cugini, J., Laskowski, S., & Sebrechts, M. (n.d.). *NIRVE paper. Design of 3-D Visualization of Search Results: Evolution and Evaluation*. Retrieved February from <http://zing.ncsl.nist.gov/cugini/uicd/nirve-paper.html>, 20-10-2014.
- Davis, F.D., Bagozzi RP, Warshaw PR (1989). User acceptance of computer technology: a comparison of two theoretical models. *Management Science* 35(8):982-1003.
- Egan, D.E., Remde, J. R., Gomez, L.M., Landauer, T.K., Eberhardt, J. & Lochbaum, C.C. (1989). Formative design evaluation of superbook. *ACM Transactions on Information Systems (TOIS)*, 7(1), p. 30– 57.
- Fox, E.A., Hix, D., Nowell, L. T., Brueni, D. J., Wake, W. C., Heath, L. S., & Rao, D. (1993). "Users, user interfaces, and objects: Envision, a digital library". *Journal of the American Society for Information Science*, 44(8), p. 480–491.
- Frayling, Christopher. *Research in art and design*, volume 1. Royal College of Art London, 1993.
- Goodhue, D.L. (1995). Understanding user evaluations of information systems. *Management Science* 41(12):1982-1844.

- Hearst, Marti (1995). TileBars: visualization of term distribution information in full text information access. *CHI '95: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, (p. 59- 66). New York, NY:ACM Press/Addison-Wesley Publishing Co.
- Hearst, Marti (2009). *Search User Interfaces*. Cambridge; New York: Cambridge University Press.
- Hearst, M., and Karadi, C. (1997). "Cat-a-Cone: an interactive interface for specifying searches and viewing retrieval results using a large category hierarchy". *ACM SIGIR Forum*, 31(SI), p. 246–255.
- Hienert, Daniel (et al.) (2012). "Integrating Interactive Visualizations in the Search Process of Digital Libraries and IR Systems". [*Advances in Information Retrieval Lecture Notes in Computer Science*](#) Volume 7224, p. 447-450.
- Hoerber, O. and Yang, X. D. (2006). A comparative user study of web search interfaces: HotMap, Concept Highlighter, and Google. In *WI '06: Proceedings of the 2006 IEEE/WIC/ACM International Conference on Web Intelligence*, p. 866-874. Washington, DC: IEEE Computer Society.
- Jones, S. (1998). "Graphical query specification and dynamic result previews for a digital library". *Proceedings of the 11th annual ACM symposium on User interface software and technology*. p. 143–151. ACM Press.
- Kim, B., Scott, J., & Kim, S. E. (2011). "Exploring Digital Libraries through Visual Interfaces". *Digital Libraries - Methods and Applications*, p- 123– 136.
- Kitchin, R., & Tate, N. (2013). *Conducting research in human geography: theory, methodology and practice*. Routledge.
- Lam, H., & Baudisch, P. (2005). "Summary thumbnails: readable overviews for small screen web browsers". *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. p. 681–690. ACM Press
- Morville, Peter and Rosenfeld, Louis (2007). *Information Architecture for the World Wide Web*. 3rd Edition. Sebastopol (CA): O'Reilly Media Inc.
- Nielsen, J., & Molich, R. (1990, March). Heuristic evaluation of user interfaces. In *Proceedings of the SIGCHI conference on Human factors in computing systems* (pp. 249-256). ACM.
- Niles, R. (2006). Robert Niles' journalism help: Statistics every writer should know. RobertNiles.com [accessed 2009] <http://www.robertniles.com/stats>.
- Nualart, Jaume (2014a). Retrieved January 3, 2015, from <http://research.nualart.cat/area-ir/>

- Nualart, Jaume (2014b). Area repository. Retrieved January 3, 2015, from <https://github.com/jaumet/Area>
- Nualart, Jaume (2014c). Area stress. Retrieved January 3, 2015, from <http://research.nualart.cat/area-stress/>
- Nualart, Jaume and Pérez-Montoro, Mario (2013). "Texty, a Visualization Tool to Aid Selection of Texts from Search Outputs". *Information Research*, v. 18, nº 2, paper 581. ISSN 1368-1613.
- Nualart, J.; Pérez-Montoro, M. and Whitelaw, M. (2014). "How We Draw Texts, a Review of Approaches to Text Visualization and Exploration". *El profesional de la información*, v. 23, n. 3, p. 221-235.
- Pérez-Montoro Gutiérrez, Mario (2010). *Arquitectura de la Información en entornos web*. Gijón: Trea.
- Robertson, G. G., Mackinlay, J. D., & Card, S. K. (1991). "Cone Trees: animated 3D visualizations of hierarchical information". *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. p. 189–194. ACM Press.
- Schatz, B. R., Johnson, E. H., Cochrane, P. A., & Chen, H. (1996). "Interactive term suggestion for users of digital libraries: using subject thesauri and co-occurrence lists for information retrieval". *Proceedings of the first ACM international conference on Digital libraries*. p. 126–133. ACM Press.
- Shneiderman, B., Feldman, D., Rose, A., & Grau, X. F. (2000). "Visualizing digital library search results with categorical and hierarchical axes". p. 57–66. ACM Press.
- Taylor-Powell, E., & Marshall, M. G. (1996). *Questionnaire Design: Asking questions with a purpose*. University of Wisconsin-Extension Cooperative Extension Service.
- van Hoek, Wilko and Mayr, Philipp (2014a). "Assessing Visualization Techniques for the Search Process in Digital Libraries". Keller, S. A., Schneider, R., Volk, B. (eds.) (2014). *Wissensorganisation und repräsentation mit digitalen Technologien*, p. 63–85. DeGruyter Saur.
- van Hoek, Wilko and Mayr, Philipp (2014b). "Is Evaluating Visual Search Interfaces in Digital Libraries Still an Issue?". T. Seidl, M. Hassani, C. Beecks (Eds.) (2014). *Proceedings of the LWA 2014 Workshops: KDML, IR, FGWM, Aachen, Germany, 8-10 September 2014*, published at <http://ceur-ws.org>.

Weiss-Lijn, M., McDonnell, J., & James, L. (2001). "Supporting Document Use Through Interactive Visualization of Metadata". *Proceedings of the ACM/IEEE Joint Conference on Digital Libraries*.

Wilson, T. (Ed.). (n.d.). *Information Research*. Retrieved January 3, 2015, from <http://www.informationr.net/ir/>

Wong, W., Chen, R., Kodagoda, N., Rooney, C., & Xu, K. (2011). "IN- VISQUE: intuitive information exploration through interactive visualization". *CHI '11 Extended Abstracts on Human Factors in Computing Systems*. p. 311. ACM Press.

Woodruff, A., Faulring, A., Rosenholtz, R., Morrisson, J., & Pirolli, P. (2001). "Using thumbnails to search the Web". *CHI '01: Proceedings of the SIGCHI conference on Human factors in computing systems*. p. 198–205). ACM Press.