Luciana de Souza Gracioso, Letícia Reis da Silveira, Maria da Graça Simões, Luzia Sigoli Fernandes Costa

Image organization on the Web: an analysis from the perspective of cultural heritage of rural farms in Brazil

Abstract
The main objective of this work is to examine the organization methods of images in the field of information science and verify if the current imagery-related information systems available on the Web put into practice the guidelines on processing thematic photographs found in the scientific literature in the area. Specifically, the goal was to create, from such methodologies, a tool to analyze some of these systems, termed an “Observation and Analysis Guideline”. The work aimed to characterize the selected information systems, by applying the above guideline in the search and retrieval user interfaces, performing an analysis from the user’s perspective, and thus obtaining new parameters that were possibly not suggested in the methodologies studied. The specific study objective also includes applying the guideline to Virtual Memory software in order to validate the categories developed for the description of cultural heritage aspects (albeit focused on photos), verifying that they meet the best practices suggested by the literature and, if necessary, provide parameters that can assist in the processing and retrieval of images by the system.

Introduction
The study identifies to what degree certain online systems available for organizing and retrieving photos include some of the main categories related to thematic processing of photographs in their user search interface, as researched and reported in the literature of Brazilian Information Science, but which were developed from an international methodological framework. Among these systems, special attention was given to the analysis of Virtual Memory, a software developed within the Project “Criteria and Methodologies to perform the São Paulo Cultural Rural Heritage inventory”, financed by FAPESP, Federal University of São Carlos, Brazil, which proposes the construction of a standard description of information covering specific fields for the processes of indexing cultural heritage images. The objective of the Virtual Memory construction project is to protect and make available all types of knowledge related to historic farms, including photographs. In this context, parallel to the study of image organization systems, this research provides assistance for detailed guidance with regard to indexing historic photographs of Brazilian rural heritage. A photograph registration simulation was performed using the Virtual Memory software in order to define whether it includes guidelines for organizing photographs defined by information science, and thus validate it as an organization and retrieval information tool for this field. The main objective of this work is: (i) to create a tool, termed an “Observation and Analysis Guideline”, using image organization methods from the field of information science, in order to analyze some online image information and
retrieval systems; (ii) to characterize the selected information systems, applying the above guideline in the search and retrieval user interfaces, performing an analysis from the user’s perspective, and thus obtaining new parameters that were possibly not suggested in the methodologies studied; (iii) to apply the guideline to Virtual Memory software in order to validate the categories developed for the description of cultural heritage aspects (albeit focused on photos), verifying that they meet the best practices suggested by the literature and, if necessary, to provide parameters that can assist in the processing and retrieval of images by the system.

Methodology

This work is defined as exploratory research because in studying images, it seeks to analyze and define the best practices made available to users regarding the processing of this type of information system, while identifying such practices in the systems selected. Besides being exploratory research, it can also be defined as intervention research, because as previously mentioned, one of its objectives regards the analysis and suggestion of possible parameters for enhancing Virtual Memory. Methodologically, this is a bibliographic and documentary study of an exploratory nature, partly developed as intervention research and using participant observation to collect data. Following this introduction, we present the detailed methodological approach that was undertaken, considering the observation feedback of the guideline developed for systems analysis. We also consider the research outlined in the literature aimed at retrieving scientific knowledge on a particular topic in national and international databases, specifically for organization of information focused on images. Since the main aim was to study user interfaces of Web systems, we conducted participant observation, and given that these interfaces can be considered as documents, our work can thus also be considered as documentary research. It should be emphasized that this study was cross-sectioned, since the selection of the systems took place at a given point of the research.

It should also be clarified that to apply the observation guideline, the analysis was condensed into three main scopes: the academic scope (systems developed by higher education institutions), the institutional scope (systems developed by institutions dealing with cultural heritage information), and the social scope (systems developed for photo registration in Web environments). These institutions were all analyzed at global, national and local levels. Flowchart 1, shown below, was prepared to simplify its understanding and summarize the phases covered during this work.

The flowchart essentially explains the study of information science approaches to analyzing and organizing images, and thereafter parameters were set for preparing an observation and analysis guideline of previously selected systems. After constructing the guideline, it was applied in the selected systems, identifying new parameters that were later incorporated into the instrument. The guideline was then rewritten in order
to be applied to Virtual Memory software and represent search results.

**Methods, techniques and challenges for the organization and indexing of photos**

The photograph as a document is also a source of information, so that in addition to getting information from it, it is a fundamental object to assist people in forming knowledge, helping them to visualize, assimilate and understand certain concepts. For photographs to effectively convey information, they need to undergo a processing, which besides allowing them to be retrieved for use, also adds knowledge for the user. Regarding this thematic processing, its classification and indexing is outlined below.

When we address thematic processing of a photograph, it is necessary to take into account the different types of treatments from the various types of units that address such informational content, whether it be archives, museums, memorial centers, personal files, libraries or the Web itself. Each of these information units adopts a suitable methodology for processing its materials. Therefore, the objective is to address some of the image analysis techniques in the literature of information science, essentially focusing on the national literature, partly as a means of highlighting the progress made in this regard in Brazil and to encourage further studies, but also because we are analyzing a national scenario for processing information about historic farms.

Below there is a summary of the various methodologies identified, highlighting their key aspects concerning the representation and organization of image content. This served as a support for preparing the guideline tool for observing and analyzing user interfaces of systems that process Web photographs:

- **MANUAL OF THE NATIONAL LIBRARY (of Brazil 1998):** One of the first initiatives for representing images compiled by librarians from the National Library and used until the present day by professionals from various fields. It takes into account the historical aspect of illustrative materials, asking: Who photographed? When? Where? What and/or who was photographed? It also proposes that the image must pass through the librarian or historian to give a short summary and from this summary select the descriptive terms for later retrieval, always moving from broader to more specific terms. Use of controlled vocabulary and natural language.

• **MAIMONE (2007):** This approach covers mainly pictorial contents of an artistic and pictorial nature, but can also provide parameters for indexing other types of documents, such as photographs. The author relied primarily on the methodology developed by Agustín Lacruz (2006) and adapted it within a Brazilian context, in order to standardize the indexing of documents, with regard to the essential elements for effectively representing information and adding to user knowledge about museums and art galleries. The importance of this study thus lies in the measures it provides for processing materials in museums, especially pictorial information. The summary field is highlighted in this work, as well as the use of natural and controlled language for indexing.

• **TOREZAN (2007):** In this study, the author strives to approach photography as a source of information and understand the analysis of the photographic document within its visual and historical possibilities. It uses questions such as “Which elements are relevant in the description of the photographic document?” And “How should the relevant information of the photograph be organized?” to guide her work. Therefore, steps are set out to describe the image document, namely: Primary Analysis, Secondary Analysis, Supplementary Analysis, Support Analysis, Technical Production Analysis and Historical Research. These steps, according to the author, are sufficient to generate a new document about the photograph, which in turn constitutes a critical element for image indexing.

• **COSTA (2008):** Perhaps the most complete methodology for the analysis of photographs. Through a set of procedures from information science, literary and rhetoric narrative, this study formulates a model for the analysis and representation of historically-based advertising images. It uses content analysis of rhetorical discourse, based on concepts and terms related to Ranganathan’s essential categories (personality, matter, energy, space and time), extended to include each literary narrative category. Associated to this, the author also proposes that the image must be analyzed considering its denotative aspects – captured at a first-level of analysis, with visible and conceptually explicit signs – as well as its connotative aspects – noticeable at a second-level of analysis, with signs read as values, emotions and attitudes – thus taking into account all the general and specific historical context in which the image was produced. The rhetorical discourse analysis helps professionals to identify both denotative as well as connotative aspects of the image and its various meanings, which according to Costa (2008) can generate a differentiated documentary product, with higher added value. Finally, in literary theory, the historical and sociocultural context in which the work was produced is sought, contributing to the analysis of image content. It advocates the use of natural language for
indexing images.

- MARTINEZ (2009): This study focused on photographs as the study object of three areas of information science – museology, bibliothecology and archiving – analyzing the application of the Brazilian Standard for Archival Description (NOBRADE) to photograph description in these areas, enabling the exchange of information between different types of institutions. As a result of the comparison, the author prepared a table to simplify visualization of their similarities and differences, to allow clear identification of similar or non-similar elements. According to Martinez (2009), it can be concluded that NOBRADE encompasses all the elements in the research instruments of the institutions analyzed, and as such this standard can play an important role in the photographic description process, fulfilling desired goals.

From the guidance provided by these approaches, the most commonly processed categories were extracted, which resulted in obtaining parameters for how to include elements in the guideline. This tool was divided into three parts, in order to evaluate aspects relating to content analysis, information representation and information retrieval in the systems selected.

Results and final remarks

As a result, a guideline tool was developed to diagnose image representation practices, which was later validated and used to evaluate Virtual Memory. Table 1 summarizes the results of applying the guideline in the systems chosen for analysis, now including the results from its application to Virtual Memory. The categories are separated according to the three dimensions of analysis, using “S” and “N” which mean “Yes” and “No”, respectively:
The guideline prepared was applied in seven websystems that processed image content, in order to determine which categories have been developed and used by these systems for analyzing and organizing photographs, and also to verify whether these current systems include the techniques suggested by the scientific literature in the area of information science with regard to the organization and representation of information, as well as aspects regarding their search tools.

The results showed that that systems take into account most of the photograph analysis practices suggested by the methodologies studied; they provide users with

![Table 1: Guidelines’ application results in the systems](image-url)
advanced research tools in their interfaces, often innovating the categories to refine the search tools, such as the Corbis system that uses color filters, number of individuals in the photograph, similar photos, and other factors; they are concerned about providing the user with good information available on the interface and also a user-friendly system. Furthermore, the implementation of the guideline allowed us to diagnose a growing concern of the systems regarding social indexing, and four of the seven major systems examined had some type of access to this type of indexing – Corbis, Casa de Rui Barbosa Foundation, USP Integrated Search Portal and Flickr.

A second objective of this study was to examine Virtual Memory software, developed to group, organize and provide access to information. The specific focus was on how it dealt with Brazilian historical heritage in the context of 19th century coffee production farms, assessing whether this system also took into account the practices suggested by the scientific literature in the area regarding analysis and description of photographs, and to evaluate it in respect to content retrieval. Applying the guideline tool to analyze Virtual Memory software thus enabled us to conclude that it is a very complete system which broadly explores all the levels needed to thoroughly describe cultural heritage, including photographs, positively responding to the guideline parameters prepared for this work, with some exceptions, such as the expressive dimension field and social indexing.

In the case of Virtual Memory, aspects regarding expressive dimension can also be incorporated in the index field for the summary of the photograph. A further suggestion to optimize the preparation of the summary and to include expressive dimension aspects could be to incorporate a “template”, or guide, to direct the indexer when preparing the summary, for example, including fields with the expressions “Who? What? Where? How? When?”, as suggested by IC methodologies, to be filled in and also to guide the indexer to observe and identify issues related to the expressive dimension of the photograph. These could be ways to improve the quality of the summary, making it richer more detailed.

This research also argues that social indexing should be taken into account, because the user’s socio-cognitive perspective when indexing or searching for information can, in fact, contribute greatly to a more efficient retrieval of search results. One suggestion is to indirectly use the users’ perspective, as their search results could help to index terms. If the system is capable of storing the search terms used by users, these terms could be retrieved and subsequently analyzed by a professional cataloger, or reviewer, and added to the fields of topics and descriptors of the cultural heritage they relate to.

Another field to be discussed in relation to Virtual Memory, which the guideline analyzed, regards the system’s search tools. It is known that this system is still in development, and that it may be the most complete means of processing cultural heritage information. It is therefore crucial that it contain search tools that optimize the
users’ results. Through analysis of Web systems, the main trends regarding their search tools were observed, showing that with the exception of SACI, all systems had advanced search tools able to integrate their search terms through Boolean operators, and search by different categories, different databases, etc.

Thus, regarding the Virtual Memory proposal, we consider that an advanced search tool should be added, one that allows users to filter their searches by material types (Archaeological, Archival, Audiovisual, Bibliographic, Built, Mobile and Integrated or Landscape, according to the system’s categories) and also by institutions that have registered their cultural heritage assets, for a more specific search.

Besides the traditional fields available for advanced searches, such as authors, titles, locations, dates and subjects, the analysis of the Corbis system interface showed a multitude of search types that can further optimize the results for photographs in the system, such as filters that determine surveys by the color of the photos, number of people in the photo, gender, age, ethnicity, types of images (photographs, illustrations...), image orientation (horizontal, panoramic, square, vertical...), style (outdoors, indoors, silhouette...), visualization (head and shoulders, full body, empty space, cropped image...), panorama (aerial view, bottom view, top view, looking at the camera, looking away from the camera...). Other ways of further refining search results include looking for photographs similar to that found, or using keywords employed for indexing the image as a source for new searches. These are some of the interesting examples found throughout the study and which could enhance the search for historic photographs using Virtual Memory software, through the evaluation and definition of its needs. Essentially, these are just a few things to be considered, in addition to assessing their feasibility within the system.

As with most studies, the current work also encountered some limitations and methodological challenges, such as the difficulty in building research instruments and then defining the methodologies used, especially as this is a research-intervention study. Another difficulty regards the number of systems and photographs analyzed. Nowadays, many systems are developed for processing photographs, and there are thousands of pictures indexed by them. However, the relatively short time available for the research did not allow us to conduct a more in-depth analysis, analyzing more systems and more photographs. It is thus important to enable this research to continue and with it the development of new studies related to the theme, for example, applying the guideline tool in more systems in different contexts, to test their effectiveness.

Finally, as presented here, there are many forms of information, be they text, image, audio or sound, which need to be disseminated so that they make sense and can help form users’ knowledge for effective decision making, completion of informational gaps and also for the retrieval and preservation of memory. Image information, in turn, has increasingly gained ground as a source of information, transmitting knowledge as a
whole or to supplement other information sources. As demonstrated in this work, the increasing concern of institutions to form collections and make them available to the public proves this fact. Thus, it is vitally important to take into account the best practices for the processing of this information so that it can fully and satisfactorily reach the user.

With the efforts made in this work and the construction of a diagnostic tool based on observation and analysis, we believe that it can be useful to other institutions and studies regarding decision-making processes and for assistance in relevant practices for analyzing and organizing pictures, especially with regard to historical images.

References