

A Contribution for the Dissemination of Cultural Heritage Content to a Wider Public

Maristella Agosti¹, Lucio Benfante¹, and Nicola Orio²

¹ Department of Information Engineering – University of Padua
Via Gradenigo, 6/a – 35131 Padua, Italy
{agosti,benfante}@dei.unipd.it

² Department of Cultural Heritage – University of Padua
Piazza Capitaniato, 7 – 35139 Padua, Italy
orio@dei.unipd.it

Abstract. Digital resources are becoming an important tool for research in all the domains related to cultural heritage. Scholars have special requirements that need to be matched when developing digital library and digital archive systems that are to be used as tools to carry out scientific research. After having designed and developed a digital library application called IPSA as a system for researchers in illuminated manuscripts, we investigated how the digital library can be evaluated by non-domain users. Our goal was to highlight the overlaps and the differences in the user requirements between specialists, who use the digital archive to fulfill their research goal, and non-domain users, who interact with the digital library system because of a general interest about its content. The results have been used to re-engineer the digital library system and extend the functions of the digital library application in order to open up its use also to non specialists.

1 Motivations and Background

In past years, most systems able to manage specialised collections of cultural heritage documents have been envisaged and developed with one specific category of users in mind. In fact many systems have been created for managing collections to be used by researchers and scholars with specific requirements related to the research work carried out on the collections. More recently, many institutions have started to consider the possibility of opening up the use of those specialised collections and systems to other categories of users that may be interested in searching and navigating through cultural heritage resources. For instance, the DEBORA [9] and MonArch [12] projects involved different categories of users, from end-users to specialists in different domains, to develop collaborative access to cultural heritage content – Renaissance books and archeological sites, respectively.

The opening up of those collections and systems to new categories of users becomes a new challenge for the information communication technology specialists that have to address the generalization of systems previously designed for a

specific category of users. Cultural heritage applications can also be the starting motivation for the development of innovative tools to access multimedia content, for instance to annotate multimedia content [8] or to develop ad-hoc image processing techniques [11].

As an initial approach to this challenge, we considered the requirements that we gathered from a distinct category of domain specialised users to design a digital library application and tried to generalise them in order to re-design the system for its use by different categories of users. In particular, we focused on how the requirements gathered from domain specialized users (professional researchers) and used to envisage and design a digital library system can be considered to extend the functions of the system also to non-domain users.

A digital archive system of illuminated manuscripts was used as a case study for this investigation. The digital archive is called IPSA, which stands for *Imaginum Patavinae Scientiae Archivum* (archive of images of the Paduan science) [5]. IPSA was developed, from 2001 to 2005, with the main objective of being a scientific tool for the analysis of the role played by the Paduan school during the Middle Ages and the Renaissance in the spread of the new scientific method in different sciences, from medicine to astronomy and botany, through the study of illuminated manuscripts [10]. IPSA is a digital library system able to manage the description and the digital version of documents, dated especially from the late Middle Ages to the fifteenth century, that are of interest of botany, medicine, astronomy, and ancient astrology. The digital library software application [4] was envisaged and implemented by the University of Padua to study the medieval science and the scientific image in its tradition and evolution, in particular in relation to the studies conducted within the University. IPSA constitutes a valuable aid for people interested in the genesis of modern science also through the use of the visual transmission of knowledge.

The main goals of IPSA are: spreading the knowledge of ancient images both for their scientific and historical importance; creating links between images to relate them to different cultural areas of interest; and showing the importance that the University of Padua has played since the end of the Middle Ages in the spread and development of sciences and culture. Taking into account the main objectives of the project, it is clear that IPSA aimed at being used by professional researchers, i.e. scholars in history of medieval art specialised in history of illumination. It has to be noted that the text of an illuminated manuscript can be copied verbatim from older manuscripts, because the most relevant part of the illuminated manuscript is the iconographic part, so the text is accompanied by illustrations that can be copied from or inspired by older manuscripts, or taken directly from nature.

In the actual version, access to IPSA is given only to authorised users, because the manuscripts that are represented in the digital library application through metadata and digital version of pages are the property of different institutions spread throughout the world. In fact the archive includes 56 manuscripts belonging to some of the most important libraries in Europe and in the world. In order to grant access to a wider public, the research group responsible of maintaining the



Fig. 1. Examples of images from Egerton 2020, London, British Library

digital archive is in the process signing a license agreement with the institutions that own the manuscripts. A visible watermark has been added to each image, with a reference to the owner of the copyright.

A relevant example is the manuscript entitled *Liber Agregà de Serapion*, which is now property of the British Library (London, British Library, ms. Egerton 2020¹) and the digital version of the pages of the manuscript is inserted and managed by IPSA thanks to an agreement between those two institutions. Therefore, an example of the iconographic content of IPSA can be found in the British Library catalogue of illuminated manuscripts; some examples are reported in Figure 1 which shows the digital representation of three pages of Egerton 2020.

1.1 Long-Term Objectives for a New Digital Library Application

Taking into account that IPSA is a combination of digitised images of manuscripts and related metadata information, and that its content can be of interest to a much larger group of users in respect with the one that was the initial target of the work, the IPSA application was selected to contribute to the design, development and evaluation of the innovative research environment that is under design and development in the context of CULTURA (Cultivating Understanding and Research through Adaptivity)², a EU funded STREP project [7].

CULTURA aims at personalisation and community-aware adaptivity for digital humanities through the implementation of innovative adaptive services in an interactive environment. This goal is motivated by the desire to provide a fundamental change in the way digital cultural heritage is experienced, analysed and contributed to by communities of interested individuals. These communities typically comprise a diverse mixture of professional and apprentice researchers, informed users and interested members of the general public.

¹ The detailed record for Egerton 2020 is at the URL: <http://www.bl.uk/catalogues/illuminatedmanuscripts/record.asp?MSID=8320&CollID=28&NStart=2020>

² CULTURA Project Website, URL: <http://www.cultura-strep.eu/>

In line with the CULTURA objectives, an effort has been initiated to redesign the IPSA application to prepare an innovative digital library application able to face the challenge of supporting the different user groups of interest. In order to carry out an effective recollection of user requirements for a novel adaptive and interactive digital library system, we decided to carry out a first round of evaluations with students at the university level. The goal was to address the main problems that main arise while novel users were interactive with the digital archive. The presence of macroscopic issues, that can be due to a difference in the levels of expertise and motivation in using the system, may hide more subtle requirements that are more related to the development adaptive systems.

This paper reports the way this redesign has been addressed together with the initial results of user evaluation that are incorporated in the new version of the application now available³. This version will be the starting point for a second round of evaluation with a second cohort of students at the university level. Having addressed the most evident issues that arose during the initial evaluation, we are confident that a new recollection of user requirements will provide additional insights to achieve the aims of CULTURA project.

2 Requirements of Professional Users

As mentioned in the introduction, IPSA was developed as a tool for professional users, and, instead of limiting the requirements analysis to a number of interviews, the design approach was to create a research team where computer scientists and professional users (i.e. researchers in history of art specialized in history of illumination) collaborated together. Additional contributions from scholars in related disciplines, such as history of science, botany and astronomy, were integrated as well and formalized in a draft proposal that was presented and discussed with professional users. A similar approach was maintained during the development of the prototype system, because all the novel functions were directly tested by members of the research team.

The requirements for carrying out scientific research are in general more complex and articulated than the requirements of final users. Final users access an image digital archive to acquire information in a given field, researchers access the same archive to disclose knowledge and discover new relations between digital objects. IPSA was designed and developed taking into account the requirements of professional users in history of illumination. This means that IPSA is the outcome of the effort of producing an original and innovative system for a specialised group of professional users.

To understand the effort that has been recently started, in the context of the CULTURA project, aimed at re-designing IPSA to add new functions to the original ones to face the characteristics of interest of the new user groups of interest, it is necessary to know the inspiring requirements that pervade the

³ Authorised users can use the new IPSA digital library application at the URL:
<http://ipsa.ipsa-project.org/>

original IPSA system. For this reason, those relevant characteristics are briefly presented in the rest of this section.

2.1 Disclosure of Relations between Images

It is of primary importance for professional users in history of illumination to discover whether illustrations have been copied from images of other manuscripts, or they have been merely inspired by previous works, or if they are directly inspired by nature. A major IPSA function thus regards the possibility of enriching the digital archive by highlighting explicit relations that have been discovered by a domain professional user. In particular, the user should be able to create *links* that connect one image to another that is related to it in some way. The analysis of user requirements on link management highlighted a number of advisable features that needed to be implemented, these are link authorship, link typology, and paths [6].

The analysis of user requirements also suggested the use of typed annotations connecting two manuscripts, two images, or even two parts of different images. These annotations, which have been called *linking annotations*, have a type that describes the kind of relation between the two objects and provides a semantic to the link. For this reason, we proposed a taxonomy for linking annotations [2] which is divided in two classes, including annotations that express either hierarchical or relatedness links. Annotations have been developed and integrated within the digital archive according to the formal model described in [1].

Researchers did not show any interest towards content-based image retrieval tools to ease their work. Apart from a possible lack of trust on automatic tools, they motivated this choice considering that general visual similarity is not particularly useful for their research work, and in most cases the relation between images is due to stylistic reasons, like the way small details are drawn. For this reason, content-based image retrieval was not considered as a relevant feature.

2.2 Dynamic Records and Intellectual Rights

Almost every digital archive dynamically changes over the years, mainly because of new acquisitions that increase the number of documents that are stored and managed by the archive system. This is also true for a digital archive of illuminated manuscripts, but there are other reasons that produce changes on the archive over time. These include the creation of records describing the documents and the images of an illuminated manuscript, which is part of the scientific research itself as for any collection of historical works. Some examples of changes to records are that new relations with other works have been discovered, or that the attribution to a given author became less certain.

Because creating a new record or modifying an existing one is part of the scientific work of researchers, the data management has to deal with intellectual rights. A researcher may prefer that some of the newly created records are not accessible by other users, at least until the results of his research have been

checked and afterwards published. This situation implies that users may decide which information can be shared with other users and which cannot.

This novel knowledge, which is due to original results, should be stored in the digital archive at a different level than the information that is based on a general consensus. To this end, the use of annotations, both classical textual annotations and linking annotations, can be a viable tool providing that a user may state which annotations can be shared with the community or with his research group, and which ones have to remain private. Such a mechanism allows researchers both to use the digital archive as an advanced research tool and to protect their intellectual rights. Moreover, linking annotations add a hypertextual structure to the archive, which is different for each user and reflects his personal knowledge in the field.

2.3 Presentation of Digital Images

A digital archive of illuminated manuscripts has the double role of preserving cultural heritage and giving access to users in a networked environment. As always happens in this situation, there is a trade-off between the high quality required for preservation and the small size needed for transfer over the network of the image files. Moreover, it has to be considered that research users should be able to perform comparisons among images belonging to different manuscripts that, in principle, may differ in their original size. According to professional users involved in the original design of IPSA, the number of images that should be presented on the computer screen varies from one to a maximum of six.

This last requirement implies that the image size, and hence its resolution, can dynamically vary depending on the context, because in principle a link can be created between any pair of images. The image files transfer load can be reduced through the use of thumbnails, at least for the first presentation of images. Thumbnails may also be a viable solution when the comparison between images is not part of the scientific research but can be used for dissemination to students or, if future releases of IPSA will be available on the Web, to casual users without controlled access.

Image acquisition is another important issue, because researchers should be able to analyze even small details of images. At the same time, researchers also need to see the image of the complete page of a manuscript, because it gives the context in which a particular object is presented. Moreover, many manuscripts have more than a single image for each page, with images surrounding or overlapping with text. For these reasons, it is advisable to carry out multiple acquisitions of the same page, with different resolutions depending on the level of detail needed for the analysis by researchers.

3 IPSA Digital Library System

The IPSA prototype implementing the requirements briefly recalled in Section 2 was developed. The close collaboration within a single team of researchers and

scholars of all the disciplines involved allowed us to create a closed loop for evaluation, testing and refinement of the different functions of the evolving prototype. Once the underlying database structure had been designed and developed, the organization of the user interface and the development of the novel functions highlighted by the user requirements were done incrementally, with scholars in history of art starting to populate the archive with the initial collection of images while the refinement of the software tools was taking place.

The IPSA digital archive system was made available on its stabilised form in 2005 and from then on it has been used for research purposes by history of art researchers. Over the years the collection of manuscripts and images has been incremented. Due to the launch of the CULTURA project, the IPSA digital archive system has been reconsidered for use by different categories of users, and a re-engineering of the system has taken place to bring the system up-to-date with the new technologies that in the meanwhile have been made available, while the underlying model of content management has been kept. Taking into consideration that users mainly focus their attention on the graphical interface when interacting with a digital resource, the system interface has been re-designed to bring it more in line with recent advancements.

The new IPSA user interface aimed at simplicity and easy user accessibility. The main layout is designed for optimal visualisation with a screen resolution of 1024x768 pixels and up, horizontally centered and filling the vertical space. The layout contains three zones: the top header, the main area and the bottom footer.

The main header is as thin as possible. It contains the main starting points to the IPSA functionalities: a small IPSA logo which links to the home page, the login/logout button, a structured multi-level menu and a form for searching the IPSA illustrations. When users are logged in, their name is shown in the header, linked to their profile for editing, if necessary. Near the search form there is a link to the advanced search function. The menu adapts itself following the user permissions, and it guides the user in the navigation, showing the most common functionalities in its first or at maximum its second level. The footer is designed for containing secondary menus and non critical information for the user. At present it contains the copyright information and the language selectors. The IPSA user interface is fully localized in Italian and English.

Most of the screen is occupied by the main area. The layout of this zone is strictly related to each functionality, and is designed and implemented following the user needs of usability. It is designed for showing the main information on the left, with a small sidebar on the right containing the links to the operations on the currently displayed object, and the related information.

A screenshot of the present Web interface presenting an image and related metadata of the IPSA collection is shown in Figure 2. The small image on the left is a tool that allows user to zoom in relevant details, which are presented in the central part of the screen. The image on the right is a link to an images that has been considered in a relevant relation with the image under analysis by a researcher, and it can be directly accessed for further comparison.

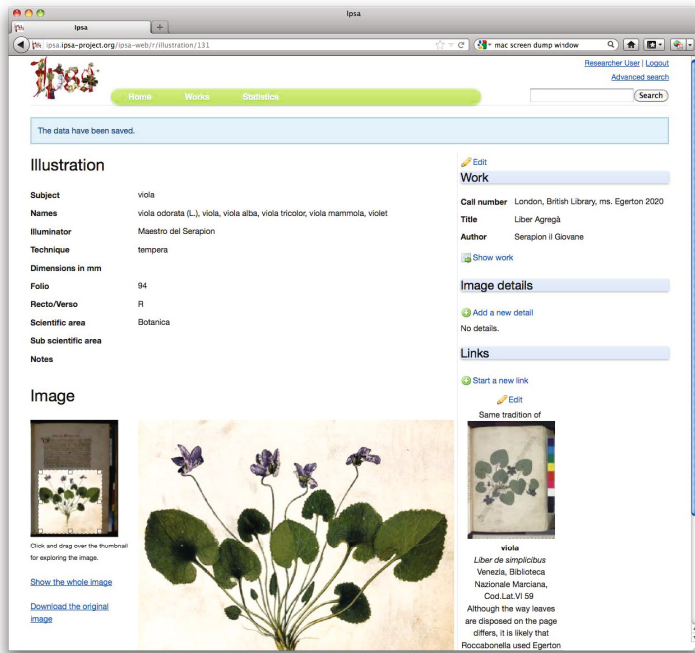


Fig. 2. Screenshot of the IPSA Web application

4 Accessing IPSA by Non-domain Users

Since 2005, IPSA has been used by professional users to carry out their studies on illuminated manuscripts. Starting from 2011, the new challenge is to investigate whether IPSA features can be of interest to non-domain users as well. So the goal is to use IPSA as a case study to compare the approach of different kinds of users to the same digital content. We conducted two subsequent evaluations: the initial one with different perspective users belonging to the class of non-domain professional research experts and to the class of the student community (in this case master students in archival science), the second one with the other groups that have to be taken into account in the context of the CULTURA environment and that were mentioned in Section 1.

The initial evaluation took place and completed with the main goal of highlighting possible overlaps between the requirements of domain professional users and the two considered groups. The main results were reported in [3] and for that are only summarised in the following. After having generalised the findings, the IPSA system has been re-engineered, as reported in Section 5. The second evaluation is still under development and possibly will give further useful insights.

Interaction with Digital Systems. As expected, computer skills play a central role in the way users interact with the system. This becomes particularly relevant in the case of specialists in other research areas who were asked to directly study multimedia content instead of bibliographic values. In contrast, master students had no problems interacting with multimedia content. Due to their habit of interacting with large multimedia collections, their requirements regarded search facilities, such as recommendations based on user-generated tags.

Hypertextual Structure. Probably because of its web interface, the archive was perceived as a hypertext and additional links towards external resources were considered an important improvement for IPSA. The possibility of using the digital archive as the starting point to retrieve other digital collections was considered highly relevant, maybe because users did not have a specific research interest towards the IPSA collection. The presence of navigation tools to browse the archive was considered important as well, because a lack of knowledge about archive content may prevent users from retrieving information when only direct search is made available. The presence of links between related images induces an hypertextual structure also to the collection of digital images and manuscripts. The exploitation of these links to improve research has already been proposed in [2], where an approach to mine the linking structure to discover novel relations is described. Moreover, the current evaluation highlighted that this feature will be useful also for non specialists as an alternative to direct search, in order to partially overcome the drawbacks of an imprecise knowledge of the domain.

Textual Descriptors. Although considered visually appealing, the digital images of the IPSA collections were not sufficient to raise interest when they are not paired by accompanying textual information. Analytic descriptions were suggested to improve user understanding of image characteristics, while additional bibliographic descriptors were suggested to make users aware of the cataloguing process. The approach to IPSA content depended on the particular field of interest of non-domain users.

5 Extending the Digital Library System

Although all non-domain users showed an interest towards IPSA digital archive, they highlighted a number of directions on how to improve interaction with the multimedia content. Part of these suggestions have already been implemented, in order to carry out a more effective evaluation with additional user groups. It is likely that the final outcomes of our evaluation will require a reengineering of the system.

First of all, a novel interface to display on screen a number of images has been developed. The interface now presents the images as a “wall” of thumbnails of the illustrations, with tools for incrementally loading additional slots of images at user request. A link can be followed from each image to its detailed description, where an image inspection tool is available to allow users to analyze its content in detail, and to follow a link to the manuscripts where it is contained.

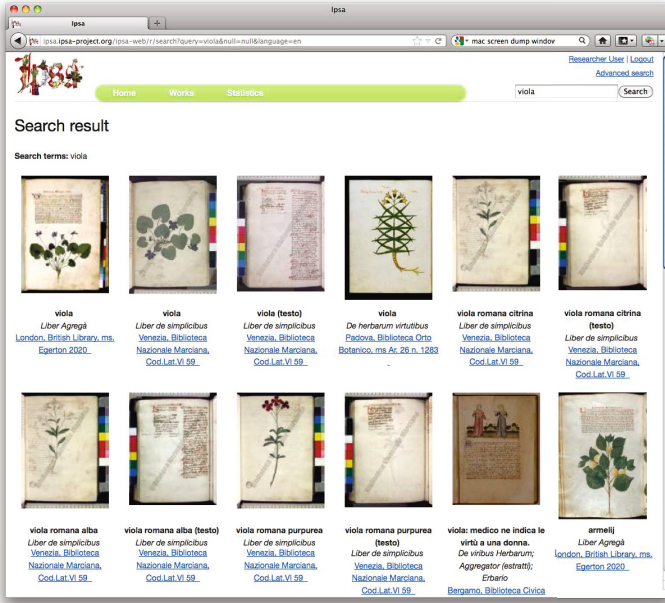


Fig. 3. Results of a search using the term “viola” (violet)

This novel interface, which is shown in Figure 3 for the results of an image search using the term “viola” (violet), is used consistently each time a number of images has to be presented on screen at the same time. In the particular case of images belonging to a given work, images are shown in the order as they appear in the manuscripts (given the focus on digital images, most of the pages that contain only text are not part of the IPSA collection) and the interface allows the user to select the central image of the wall of images.

Another improvement regarded the rendering of individual images, which was designed for expert users that might be interested in very small details. The initial version of the interface allowed a personalized rendering, because the full version of the image was processed at server-side each time a request was made by a user. The actual version is based on a pre-rendering of the magnified image, using a predetermined fixed-screen maximum resolution. The user can still interact with the image by dragging the mouse over its thumbnail, but the image is rescaled at client-side, obtaining a more fluid navigation and improving user experience.

The interface to create links between images has been re-designed as well. Now it is possible to always have available the thumbnail of the image that is used to start a link, and to select the second image from a wall of thumbnails, which is the results of an image search. The starting page of a link creation is shown in Figure 4.

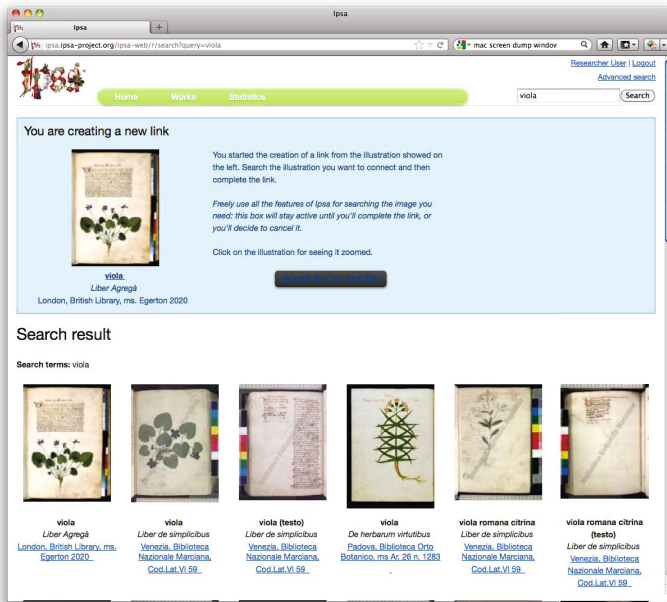


Fig. 4. Screenshot of the IPSA interface to create a link between two images

6 Conclusions and Future Developments

The IPSA digital library system was developed to help scholars carrying out their scientific research and is in the process of being extended to contribute to the dissemination of cultural heritage to a wider public. To this end, we carried out a first round of evaluation with non-domain users, in order to highlight a number of directions for improving the interaction with the digital collection of manuscripts. In this paper we describe how these requirements were translated in additional features of the IPSA digital archive.

We are organizing a second round of evaluation with other user groups, among the ones taken into account by the CULTURA project. It is likely that additional insights will be produced by the comments of new users, indicating that a digital library for cultural heritage should be a continuously growing system that has to evolve to adapt to new requirements in order to maintain its role of disseminating cultural heritage.

Acknowledgements. The authors would like to thank Giordana Mariani Canova and Chiara Ponchia of the Department of Cultural Heritage of the University of Padua and Nicola Ferro of the Department of Information Engineering for the useful discussions on aspects related to the reported effort.

The work reported has been partially supported by the CULTURA project, as part of the Seventh Framework Programme of the European Commission, Area “Digital Libraries and Digital Preservation” (ICT-2009.4.1), grant agreement no. 269973.

References

1. Agosti, M., Ferro, N.: A Formal Model of Annotations of Digital Content. *ACM Transactions on Information Systems (TOIS)* 26(1), 3–57 (2008)
2. Agosti, M., Ferro, N., Orio, N.: Graph-based Automatic Suggestion of Relationships among Images of Illuminated Manuscripts. In: Haddad, H. (ed.) *Proc. of the ACM Symposium on Applied Computing*, pp. 1063–1067. ACM Press, New York (2006)
3. Agosti, M., Orio, N.: To envisage and design the transition from a digital archive system developed for domain experts to one for non-domain users. In: Nelson, M., Van de Sompel, H., Sølvsberg, I. (eds.) *Proc. 12th ACM/IEEE Joint Conference on Digital Libraries (JCDL 2012)*. ACM Press, New York (2012)
4. Agosti, M.: Digital libraries. In: Melucci, M., Baeza-Yates, R., Croft, W.B. (eds.) *Advanced Topics in Information Retrieval. The Information Retrieval Series*, vol. 33, pp. 1–26. Springer, Heidelberg (2011)
5. Agosti, M., Benfante, L., Orio, N.: IPSA: A Digital Archive of Herbals to Support Scientific Research. In: Sembok, T.M.T., Zaman, H.B., Chen, H., Urs, S.R., Myaeng, S.H. (eds.) *ICADL 2003. LNCS*, vol. 2911, pp. 253–264. Springer, Heidelberg (2003)
6. Agosti, M., Ferro, N., Orio, N.: Annotating Illuminated Manuscripts: an Effective Tool for Research and Education. In: Marilino, M., Sumner, T., Shipman III, F.M. (eds.) *Proc. 5th ACM/IEEE Joint Conf. on Digital Libraries*, pp. 121–130. ACM Press, New York (2005)
7. Agosti, M., Orio, N.: The CULTURA project: CULTivating understanding and research through adaptivity. In: Agosti, M., Esposito, F., Meghini, C., Orio, N. (eds.) *IRCDL 2011. CCIS*, vol. 249, pp. 111–114. Springer, Heidelberg (2011)
8. Borghesani, D., Grana, C., Cucchiara, R.: Surfing on artistic documents with visually assisted tagging. In: *Proc. of ACM Multimedia*, pp. 1343–1352 (2011)
9. Lebourgeois, F., Emptoz, H.: DEBORA: Digital Access to Books of the Renaissance. *Int. Jour. on Document Analysis and Recognition* 9(2-4), 193–221 (2007)
10. Mariani Canova, G.: La cultura universitaria padovana e la nascita del realismo nell’immagine botanica. *Atti e memorie dell’Accademia di Storia della Farmacia* XX(3), 198–212 (2002)
11. Ogier, J.M., Tombre, K.: Madonne: Document Image Analysis Techniques for Cultural Heritage Documents. In: *Proc. of the Int. Conf. on Digital Cultural Heritage*, pp. 107–114 (2006)
12. Stenzer, A., Woller, C., Freitag, B.: MonArch: Digital Archives for Cultural Heritage. In: *Proc. of ACM Int. Conf. on Information Integration and Web-Based Applications & Services*, pp. 144–151 (2011)