Enhancing End User Access to Cultural Heritage Systems: Tailored Narratives and Human-Centered Computing

Maristella Agosti[1], Marta Manfioletti[1], Nicola Orio[2], and Chiara Ponchia[2]

[1] Department of Information Engineering, University of Padua, Italy
{agosti,manfioletti}@dei.unipd.it
[2] Department of Cultural Heritage, University of Padua, Italy
nicola.orio@unipd.it, chiara.ponchia.1@studenti.unipd.it

Abstract. This paper reports on the results of a study that aims to support end users of a multimedia system that manages a digital cultural heritage collection. The system is provided with automatic tools that simulate the behavior of the research method adopted by professional users when they interact with the multimedia application. The experimental results have been obtained using a multimedia application that manages the digital representation of historical botanical manuscripts.

Keywords: Cultural heritage systems, multimedia systems, human-centered computing, history of art research method, digital archives, illuminated manuscripts, historical botanical manuscripts, recommendation.

1 Introduction

Advanced cultural heritage systems are often digital applications that manage multimedia content for diversified categories of users, ranging from professional researchers to interested members of the general public. To this end, these systems have to be adapted to the user needs of such a diversified user range. One major challenge for system designers is thus to allow the general public to take advantage of the automatic tools created to support professional research while at the same time exploiting scholars' knowledge about the digital collections. The dissemination of cultural heritage can be fostered when end users are able to actively interact with the system, discover the professional knowledge and have a rewarding experience that would not otherwise be possible.

This paper presents the findings of a study that aims to support end users of a multimedia system that manages the digital representation of historical botanical manuscripts. The system is provided with automatic methods that are based on the research method of the professional users of history of art.

The paper is structured as follows. An initial section presents the context that gave rise to the study. This is followed by a section that presents the basic characteristics of the research method used by history of art professional users when they identify

A. Petrosino, L. Maddalena, P. Pala (Eds.): ICIAP 2013 Workshops, LNCS 8158, pp. 278–287, 2013.
© Springer-Verlag Berlin Heidelberg 2013

correspondences and similarities between images in manuscripts that may have been written over the centuries and/or in different geographical areas. After having identified and presented the essential characteristics of the methodological tools used by history of art experts, we present the process adopted to define and enrich the digital representations of botanical manuscripts with useful metadata representations that are used with a proposed method for automatically discovering similarity between images. At the end we make some final remarks together with some suggestions for future work.

2 Context

The findings presented in the paper originated in the CULTURA project, a European project that aims at increasing user engagement with cultural heritage digital collections through the development of a new adaptive and dynamic environment[1]. To fulfill the challenging goal of the project, continuous interaction with different categories of end users is required to understand the needs, wishes and desires CULTURA has to address. As an exemplary cultural heritage digital collection, we used the digital archive containing the digital representations of illuminated manuscripts identified and chosen in the context of the IPSA (*Imaginum Patavinae Scientiae Archivum*) project.

The IPSA digital archive was created at the University of Padua for professional researchers in History of Illumination to allow them to compare the illuminated images held in the collection and to examine in depth certain aspects of scientific illustration in Italy[2]. The approach adopted in designing the IPSA multimedia system was focused on the user with the aim of developing a real human-centered multimedia application [1].

Due to involvement in the CULTURA project, we had to face the stimulating task of opening up such a specialist collection to new user categories with different interests toward the IPSA collection and History of Art in general. In fact, a subset of the IPSA collection was selected to be imported in the CULTURA environment for use as a case study to test the new environment and its functions. This new environment was named IPSA@CULTURA to underline that the IPSA content was being used with tools and services of the CULTURA environment[3]. CULTURA provides a service-oriented architecture, where the user can interact with a number of functions that have been developed and are maintained by partners of the project; the portal was developed using Drupal by the research group of the Trinity College Dublin [2].

With our involvement in the design and development of IPSA@CULTURA, we have come to the conviction that a valuable way to catch the attention of new users is to give them the basic critical tools and information to approach the collection and to involve them in the research paths of professional users. This would then help them to develop their own research method and have a more satisfactory interaction with the system.

Therefore, we prepared a *narrative* to guide users through the collection, and we presented them with both the IPSA@CULTURA contents and the main functions.

[1] http://www.cultura-strep.eu/
[2] http://ipsa.dei.unipd.it/en_GB
[3] http://kdeg.cs.tcd.ie/ipsa/

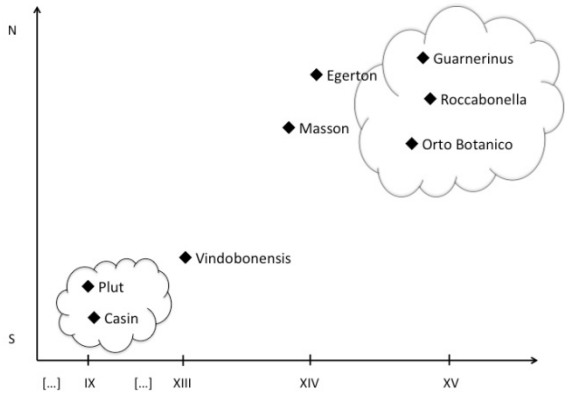

Fig. 1. Graphical representation of the three groups of manuscripts used in the study

Up to now art historians have developed the recommendations to create the narrative, but we decided to investigate how they are related to features that can be directly computed from metadata descriptors and multimedia content.

3 Art Historical Research Method Recommendations

The main functions of IPSA@CULTURA include enabling the final user to follow different paths through the collection thanks to the use of narratives. With the term narrative we mean short lessons that allow the user to become acquainted with the collection and to obtain a clearer idea of the research that can be carried out in IPSA@CULTURA. Illuminations are presented to the user according to their interest toward the collection: a beginner or user with a low level of interest would only be shown the basic resources of the narrative. A mid-level user would be shown more information, whereas a person with a high level of interest is shown all the documents. However, it is possible for a beginner to see additional resources from the higher levels if they are so interested.

Let us see a narrative in detail. Since the IPSA collection includes a rich group of illuminated herbals, it was considered very significant to create a narrative on the development of botanical illustrations in Italy from the early centuries of the Middle Ages until the XV century. Therefore, not only can the user follow the history of this particular kind of illustration, but they are also given a quick and simple overview of some of the main medieval periods that are relevant from an art-historical point of view.

Art historians employ a number of methods in their research into the qualities, nature and history of objects. They examine works of art through an analysis of form (i.e. the use of line, shape, color, texture and composition, and iconographic analysis) that takes account of the particular design elements. Art historians are able to assign a date to different works of art by studying their style and making comparisons with

other images (frescos, illuminations, jewelry decoration and so on) whose realization time is unknown. For example, if an illumination presents the same stylistic and iconographic feature of the Scrovegni Chapel frescoes in Padua, which were painted by Giotto between 1303 and 1305[4], it can reasonably be assumed that the illumination was painted at the beginning of the XIV century, maybe by a Paduan illuminator or by an illuminator that visited Padua.

Table 1. Catalog information on the IPSA manuscripts used in the study

Call Number	Label	Century	Provenance
Montecassino, Archivio della Badia, Casin. 97	Casin	IX	Montecassino
Firenze, Biblioteca Medicea Laurenziana, ms. Plut. 73.41	Plut	IX	Montecassino
Wien, Osterreichische Nationalbibliothek, Codex Vindobonensis 93	Vindobonensis	XIII	Southern Italy
Paris, Bibliothèque de l'Ecole des Beaux-Arts, ms. Masson 116	Masson	XIV (1370-1380)	Northern Italy
London, British Library, ms. Egerton 2020	Egerton	XIV (1390-1404)	Padua
Venezia, Biblioteca Nazionale Marciana, Cod.Lat.VI 59	Roccabonella	XV (1445-1448)	Venice
Padova, Biblioteca Orto Botanico, ms Ar. 26 n. 1283	Orto Botanico	XV	Veneto
Bergamo, Biblioteca Civica Angelo Mai, ms MA 592 (già Lambda 1.3)	Guarnerinus	XV (1441)	Feltre

According to this methodology, the IPSA herbal collection can be divided into three main groups of manuscripts that were produced roughly at the same time and that present illuminations with the same stylistic features and sometimes with similar iconographic elements. These three groups constitute three main steps in the development of botanical illustrations: (1) the Montecassino group, which includes very early botanical manuscripts produced in Southern Italy between the VIII and IX centuries [3], (2) the Federico II group, constituted by manuscripts produced at the court of Federico II in Southern Italy during the XIII century, and (3) the Veneto group, which consists of herbals produced in the Veneto region during the XIV and XV centuries and characterized by very realistic images. In Figure 1 the three mains groups of manuscripts are represented in a graph where the vertical axis represents the geographic areas of provenance of the manuscripts, where N stands for North of Italy and

[4] http://www.cappelladegliscrovegni.it/index.php/en/

S stands for South of Italy, and the horizontal axis represents the time, from the IX to the XV century.

Once users start to access the narrative, they are shown an introductive page with instructions on how to use the lesson functions. In one lesson they can use the "Lesson Block", which is located on the left hand side of the screen of the IPSA@CULTURA application, to read their tasks and to navigate through the lesson. While in the middle of course, they can move from one page to the next by simply clicking on the "Next" button, or in some cases they can choose between the "Next" button and the "See extra resource related to this one" button to access more detailed information on the issue.

In this narrative lesson, users are firstly shown an illumination of the Rosmarinum plant belonging to a manuscript of the Montecassino group. Then, users can decide whether to view the illumination of another manuscript of this group ("See extra resource related to this"), thus obtaining more knowledge of the manuscript production and botanical illustration of this period, or whether to skip to the second important moment of botanical illustration in Italy, the production of the Federico II manuscripts ("Next").

All the illuminations are accompanied by detailed explanations of the subject of the manuscript they belong to as well as their historical and art-historical context, as can be seen in Figure 2.

The narrative proceeds by guiding users through the centuries until the period of Veneto herbal production, a highly important point in the history of botanical illustration in Italy. In fact, whereas illuminated herbals during the Middle Ages presented very simple and highly schematic illustrations, illuminations of the Veneto herbal group have in common a marked search for realism. The first herbals with highly realistic illustrations of plants were produced in Padua in the XIV century, thanks to the scientific research undertaken in the University of Padua, particularly by Pietro d'Abano. The new scientific studies had a deep influence on the development of contemporary painting, which became increasingly more faithful to nature [4]. Users can verify this sea-change by comparing two illuminations of the rosemary plant and two illuminations of the "Mandragora". The Mandragora is an imaginary plant, thought to be half human and half vegetable. Users are invited to view its different representations to gain a better understanding of how the style developed, as it is easier to verify this evolution by comparing images of the human body rather than images of plants.

4 Thematic Paths Starting from Metadata

Originally the IPSA archive was designed for expert users, scholars and researchers with a particular interest in history of illumination. Users of such a sample typically have a broad knowledge of the archive content and interested scholars usually already have detailed knowledge of the whole collection of images. With the opening of the IPSA collection to the general public, in the context of the CULTURA project, the system needed to be redesigned to fulfill not only professional requirements, but also the needs of the general public. One of the most important aspect which needed improvements to satisfy user requirements is the metadata corpus, and for this reason a challenging task was undertaken on the metadata that describe the collection.

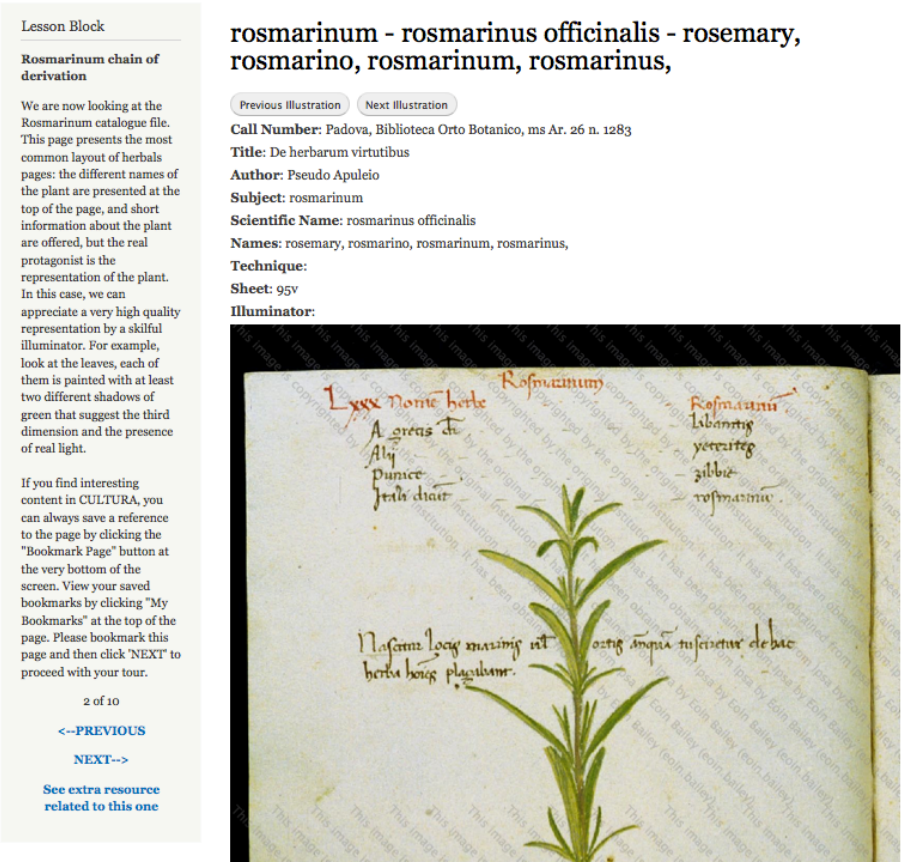

Fig. 2. Screen shot of a step of a narrative lesson in IPSA@CULTURA

IPSA@CULTURA contains nine digitized manuscripts for a total of more than 3000 digital images representing about 1000 different plants, since it is common that the same plant is contained in more than one manuscript. Each manuscript is described by 25 fields which specify physical description of the manuscript (dimension, binding, etc.), historical information (century, dating, provenance, etc.) and side notes (bibliography, observations, etc.). In addition, each image contained in a manuscript has a specific description that consists of 10 fields.

Since IPSA gathers illuminations of botanical manuscripts, the main issue regarding metadata concerns plant names. In medieval times plants did not have a standard classification as they do today, nor had an orthographic standard been established. As a consequence, the very same plant could be called by different names, and at the same time, these names could be written using different orthographic variations. Therefore, normalization of the plant names was critically needed to guarantee a correct and complete functioning of the research tool. In fact, when a user looks for a specific plant, the system should present all the images in the result list which represent that specific plant, and not only those which present the identical written shape of the query.

This outcome is possible only if the metadata that describe the name of the plants are rendered uniform and normalized.

The work of normalization was entrusted to a linguist. Both spelling issues and lexical issues were solved using phonetic and phonology studies, historical linguistics and etymology: what we call today "*camphor*" in English is found in IPSA as "*canfora*", "*camfora*", "*canphora*" and "*camphora*", and there is no explicit link between all these variants, even though they look similar to the human eye. On the other hand, lexical issues are more complicated to solve than spelling issues, partly because they need philological and linguistic research to solve them. For example, we can use the case of the plant that today is called "*alkanet*" in English. In IPSA this plant is called "*buglossa*" in one manuscript and "*lingua bovis*" in another. It is clear that human intervention is needed to establish an explicit link between these two names, thus allowing users to find both plants in a single query since these two different names indicate the very same plant. Thanks to the normalization of plant names, many links can now be established between plants that did not seem to have anything in common before.

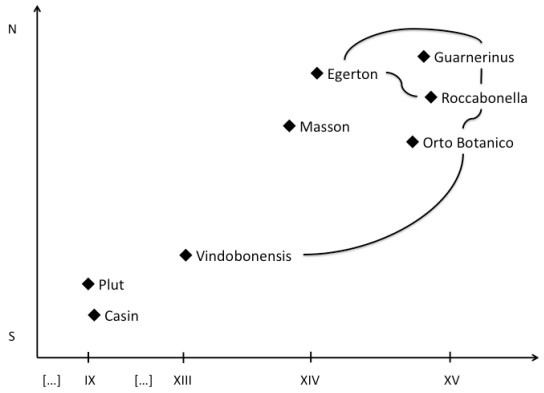

Fig. 3. Graphical representation of possible paths across the collection based on the co-occurrence of illustrations in manuscripts

Additional recommendations can be made, including computing the similarity between manuscripts based on the overlap in their visual content. In fact, it is likely that when one manuscript inspired the development of another, the choice of the plants to be represented was influenced as well. This situation is exemplified by the relation between Egerton and Roccabonella, because the former largely inspired the development of the latter, and in fact two thirds (66%) of the subjects represented in Egerton are also represented in Roccabonella. Another piece of evidence is the relation between Orto Botanico and Vindobonensis. As described in the narrative reported in Section 3, historians of art believe that the former was influenced by previous models developed at the court of Federico II. This result is also reflected by the overlap between the two

manuscripts, since all the illustrations of Vindobonensis that are common to at least one other manuscript (one third of the total, about 35%) are also depicted in Orto Botanico.

The analysis of overlapping images also highlights a group of manuscripts that have in common a large number of illustrations: Egerton, Roccabonella, Guarnierinus and Orto Botanico. As can be seen in Figure 1, they share the same provenance and they were created in the same historical period, meaning that automatic analysis can provide relevant hints about the relations between these manuscripts.

Figure 3 shows the presence of possible thematic paths which were computed by starting from metadata content, as an automatic alternative to the narratives provided by scholars.

5 Visual Similarity as Evidence for Alternative Paths

Another kind of analysis can be carried out on image content. Visual descriptors have become a popular tool for content-based information retrieval, in particular in the context of MPEG-7[5]. Although these descriptors were not developed for historical illustrations, they capture some characteristics that can be relevant also from the point of view of an art historian. In particular, according to a joint analysis with an expert in illuminated manuscripts, we selected three of the standard MPEG-7 descriptors:

- *Color Structure Descriptor* (CSD): providing that some normalization is carried out to compensate for the background color of the parchment, this takes into account the color palette used by the illustrator. In particular, the artistic quality of the manuscript and the choice of a realistic representation versus a simplified one are likely to be correlated with the color structure of the illustration.
- *Edge Histogram Descriptor* (EHD): this is correlated with the process of creating illuminations. In fact, depending on stylistic choices, some illustrators decided to highlight the contours with black ink, while others created a pencil sketch and did not delete the pencil drawing after painting. Sometimes, they covered the whole pencil sketch with colors, so we do not have any evidence of the existence of underdrawing. Hence, edges can be of many different types, colors and dimensions.
- *Region Shape Descriptor* (RSD): this is correlated with the choice on how the illustration is integrated with the text of the manuscript. Typically, the illustrations were made after the calligrapher wrote the manuscript, using the available space that was left on the page.

We carried out an initial evaluation of how visual similarity can be correlated with the relations highlighted in the narratives, using the plants that were common to almost all the manuscripts in the collection. To this end we manually cropped the images from the full manuscript pages contained in IPSA and we computed the similarity using the cosine distance between images of the same subject. We then qualitatively analyzed the individual effect of each descriptor.

[5] `http://mpeg.chiariglione.org/standards/mpeg-7`

The relation between Egerton and Roccabonella, which present a high level of stylistic resemblance, is also highlighted by visual descriptors. When an illustration from Roccabonella is used as a query, the corresponding illustration from Egerton had the highest rank for both CSD and EHD, while it was among the first two for RSD. Similarly, also the relation between Vindobonensis and Orto Botanico (described in the previous section) is correlated with their visual similarity: using illustrations from Orto Botanico as queries gave the corresponding illustrations of Vindobonensis at top rank. Using both CSD and EHD the two manuscripts created in Montecassino showed the highest similarity, although this result is not surprising considering that the illustrations in both case are monochromatic. Similarly, the more recent manuscripts in the IPSA collection tended to cluster together, with Guarnierinus very close to Roccabonella and Orto Botanico. This characteristic is particularly evident using RSD.

We think that these initial results can be an interesting starting point for the development of computer-aided narratives. Clearly, the experience of art historians cannot be simulated by the use of these general-purpose descriptors, yet we believe it is interesting that some correlation exists between different sources of evidence. An example of these paths is shown in Figure 4.

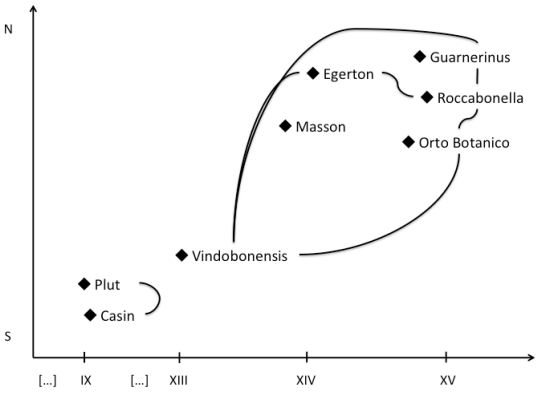

Fig. 4. Graphical representation of possible paths across the collection based on visual similarity

6 Conclusions and Future Work

In this paper we presented different possible approaches for engaging users with a digital collection. On the one hand, the experience of scholars can be used to create narratives through a collection of manuscripts in order to present relevant content to the general public. On the other hand, automatic analysis of metadata, and visual descriptors can be used as an alternative way to recommend new items to end users. In this way, users can access digital cultural heritage both following the paths prepared by experts in the domain and exploring content preselected by automatic methods. As an additional feature, users could be taught not only with the content of the collection

but also they would get to know art historian research methodologies, expecially the way they analyze and compare images.

Further experiments are underway to see if some support to narratives can be automatically prepared by using the results of the visual similarity computed on the enriched metadata available in the cultural heritage environment in use in the project.

Acknowledgements. The work reported has been partially supported by the CULTURA project (reference: 269973) within the Seventh Framework Programme of the European Commission.

References

1. Elgammal, E.: Human-centered multimedia: representations and challenges. In: Proc. of the 1st ACM Int. Workshop on Human-Centered Multimedia (HCM 2006), pp. 11–18. ACM, New York (2006)
2. Hampson, C., Lawless, S., Bailey, E., Yogev, S., Zwerdling, N., Carmel, D., Conlan, O., O'Connor, A., Wade, V.: CULTURA: A Metadata-Rich Environment to Support the Enhanced Interrogation of Cultural Collections. In: Dodero, J.M., Palomo-Duarte, M., Karampiperis, P. (eds.) MTSR 2012. CCIS, vol. 343, pp. 227–238. Springer, Heidelberg (2012)
3. Adacher, S.: La trasmissione della cultura medica a Montecassino tra la fine del IX secolo e l'inizio del X secolo. In: Atti del II convegno di studi sul medioevo meridionale. Montecassino. Dalla prima alla seconda distruzione. Momenti e aspetti di storia cassinese (secc. VI-IX), pp. 385–400. Pubblicazioni cassinesi, Montecassino (1987)
4. Mariani Canova, G.: Per Cultura: le immagini dei manoscritti della scienza a Padova dal Medioevo al Rinascimento. Atti e Memorie dell'Accademia Galile-iana di Scienze, Lettere ed Arti, vol. CXXIV, pp. 81–90 (2011-2012)