

Evaluating Cultural Heritage Information Access Systems (Panel)

Tiziana Catarci¹, Maria Guercio^{1,2},
Giuseppe Santucci¹, and Francesca Tomasi³

¹ Sapienza Università di Roma, Italy
{catarci,santucci}@dis.uniroma1.it, maria.guercio@uniroma1.it

² Digilab, Roma, Italy

³ University of Bologna, Italy
francesca.tomasi@unibo.it

1 Introduction

IRCDL¹ is a yearly deadline for Italian researchers on Digital Libraries related topics. This year the focus of IRCDL 2013 was on emphasizing the multidisciplinary nature of the research on digital libraries which not only goes from humanities to computer science but also crosses among areas in the same field ranging, for example, from archival to librarian sciences or from information systems to human-computer interaction.

The panel on “Evaluating Cultural Heritage Information Access Systems” was also characterized by this interdisciplinary flavour. Indeed, the panelists reflected the perspective of different fields, ranging from computer science (Prof. Giuseppe Santucci), to archival and librarian sciences (Prof. Mariella Guercio and Dott. Maurizio Messina) and digital humanities (Dott.ssa Francesca Tomasi).

It is evident that in these last years there is a growing and persistent demand for more and more digital content in many different areas and for diverse purposes, with a particular emphasis in the cultural heritage sector. Typically, raw digital content is assembled in digital collections, but only the curation and enrichment of the raw material make it usefully available for working with it and exploiting it, and this subsequent embellishment phase is carried out in digital libraries. Digital libraries basically consist of large digital collections plus a set of tools that make content alive, that help the users to find it, make sense out of it, annotate it, comment it, share it in a community, collaborate on it, and so on. Thus, the evaluation of digital libraries and information access systems has to consider not only the quality of the digital collections and the preservation approach, but also the effectiveness and efficacy of the user-oriented tools they provide, in an overall user experience, having the ultimate goal of creating new knowledge.

Following the above idea, the panelists were presented with four specific topics:

- Cultural Heritage Information Access System vs Digital Library: similarities and differences

¹ <http://ims.dei.unipd.it/websites/ircdl/home.html>

- Accessing Raw Material vs interpretation and curated presentation of such material
- Dimensions of the User Experience
- Evaluation coordinates: quality and completeness of the information, tools to make sense out of it, usability, accessibility, open access,

And asked to give their view with respect to one or more of them. Summaries of their presentations are presented in the following sections.

Finally, the panel - and IRCDL 2013 in general - would not have been possible without the generosity of our sponsors, namely Sapienza Università di Roma, PROMISE (Participative Research labOratory for Multimedia and Multilingual Information Systems Evaluation - Network of Excellence co-funded by the 7th Framework Programme of the European Commission, grant agreement no. 258191) and the CINECA Consortium.

2 A Contribution from the Archival Domain

The presentation has focussed the attention to only one specific aspect of the complex issues under discussion in the panel: the capacity of representing the archival cultural heritage when the representation has to be included as part of a large and multidisciplinary digital library. In particular the presentation has analysed the challenges which always emerge when a cross-domain heritage is involved and variegated and rich contexts and provenance have to be described and, even more, rendered in large web environments. This aspect have been considered from the perspective of the archival heritage and its main (and not avoidable) complexity but, as mentioned, the presentation has not underestimated the general need for cultural heritage collections (of any kind) to be represented and made available within their original structure and relations.

From this point of view, the archival attention for the provenance and the contextual information can provide a fruitful methodology for future qualified solutions, for a better structured information representation and for a higher degree of intelligibility of the digital resources made available on the web. As discussed in the panel, the limits of the present evolution is partially due to the lack of participation of digital curators other than librarians in the processes related to the communication processes of digital heritage: the archivists but also the professionals active in the cultural heritage sector have not realized the strategic relevance of this cooperation and have not sufficiently contributed to the definition of a more comprehensive integrated approach to the digitization process and to the interoperability of descriptive metadata. The consequences of this absence have been: too granular and flat solutions, no capacity of collecting relations and supporting contexts and provenance information, a very limited number of innovative proposals in the field. Among others, the presentation has discussed:

- the *key questions* still unsolved for defining the quality of a digital library according to the DELOS requirements like a not yet well defined assessment framework, the lack of attention for an efficient capacity of retrieving and

managing digital library contents when peculiar representations and information structures have to be supported,

- the limits of the available solutions to support the *complex translation into the web of archival finding aids systems* and resources with reference to their uniqueness and to the complexity of their reciprocal relations (strategic for making them understandable), but also in consideration of the massive volumes (approximately 8000 km in Italy of unique resources) which limit the role of the digitization process in this area (the finding aids system will necessarily focus, at least for the next decade, on analogue records to be accessed in the traditional reference services and only few resources will have the “privilege” to be digitized), the risk of fragmentation and arbitrary criteria for selection: the thematic approach generally supported by the digital library environment can be useful for attracting users but not for providing qualified and inclusive services to the scientific communities and to the citizens needs,
- the *ambiguity of the concepts involved* which has prevented the archival sector by using even the terms involved in this new environments: archival websites or archival information systems are the preferred terms and the preferred scenarios for developing services and functionality, while the term digital library is at the moment rarely used for proposing online publication of digital archival resources in dedicated environments.

For providing concrete and technical elements for a more detailed analysis, the presentation had considered best research projects and, if possible, good practices already available at national level to sustain future cooperation in this field (just as examples of promising services for future development and a more comprehensive analysis). They have included the following cases:

- SIAR (Sistema informativo archivistico regionale del Veneto), created with the aim of integrating distributed archives by using both standards, methodology and tools developed for digital libraries and for specific domain, but also for exchanging or sharing metadata, even if, at the moment, the integration for research services is not explicit and the catalogues are maintained separate,
- Biblioteca digitale della Lombardia, whose aim is the digitization and online publication of cross-domain cultural resources able to testify the regional cultural heritage and whose basic requirements are: the acquisition and online availability (through a regional web portal) of significant resources of regional culture, the integration of resources preserved and created in different environment, including archival documents, the capacity of ensuring the long-term preservation of digital resources, in compliance with the main standards and in cooperation with the main national and international projects (Europeana, CulturaItalia and Sistema archivistico nazionale),
- Sapienza Digital Library, whose ambitious goals are: 1) the aggregation and accessibility in a digital form of cross-domain information contents created by Sapienza University research communities or made available by corporate bodies or individuals in relation to the academic environment; 2) the

harmonization of the descriptive practices for new resources (not easily ascribable to a specific domain) by adopting with some degree of “creativity and imagination” national standards and recommendations with reference to the use of controlled vocabularies, ke PICO (Portale della Cultura Italiana) 4.3 MibAc and *Nuovo soggettario di Firenze* for subjects, TGN (Thesaurus of geographic names) Getty and Geonames, but also VIAF Virtual International Authority file and other internationally recognized vocabularies based on specific disciplines.

The critical state of art of the Italian projects has been also briefly examined and some conclusions made, as here summarized:

- provenance and context are not always identified as crucial components;
- archival standards are recognized for their general value, but not yet completely implemented outside the archival information systems,
- compliance with European standards is generally stated and partially ensured but mainly as a static and flat model for representation,
- the main difficulties concern: the differentiation of digital resources (many projects are limited to the identification of single resources and have developed simple research interfaces), the low level of integration and cooperation among institutions both at regional and at national level and of course and the lack of financial resources
- the integration (and not convergence) among heterogeneous cultural information access systems is here the key term, but a balance is required between specificity, details and general perspective: the functionality for retrieval must be easy to use but not trivial and new forms for intermediation are required (particularly when digital resources are complex and articulated).

3 Evaluating Cultural Heritage Information Access Systems through Visual Analytics

The evaluation of an Information Access System is a non trivial task and several research activities focus on supporting it (see, e.g., *Participative Research labOratory for Multimedia and Multilingual Information Systems Evaluation (PROMISE)*). Different systematic methodologies and techniques are used in such a context and all of them share a common outcome: they produce a lot of complex data, whose interpretation is a challenging activity by itself. Lessons learned point out some key areas that can improve the overall process:

- Building a community focused on common objectives: that allows for better shaping methodologies and techniques, sharing ideas, experiences, and test cases. Moreover that allows for a sound and productive comparison of evaluation results;
- Using a robust infrastructure: computer support is mandatory for managing and providing access to the scientific data produced during evaluation activities, with the final goal of supporting the organization and the execution of evaluation activities, increasing the automation in the evaluation process;

- Visualizations: suitable visual analysis can foster and improve the usage and the interpretation of the managed evaluation data. Visualizations must be integrated in the system and an overall methodology like *Visual Analytics* (VA) is needed.

VA [1] is an emerging multi-disciplinary area that takes into account both ad-hoc and classical *Data Mining* (DM) algorithms and *Information Visualization IV* (IV) techniques, combining the strengths of human and electronic data processing. Visualisation becomes the medium of a semi-automated analytical process, where human beings and machines cooperate using their respective distinct capabilities for the most effective results. Decisions on which direction analysis should take in order to accomplish a certain task are left to the user. Although IV techniques have been extensively explored [2], combining them with automated data analysis for specific application domains is still a challenging activity [3].

In order to apply VA techniques to *Cultural Heritage Information Access Systems* (CHIAS) evaluation, improving the visualizations, analysis, and interpretation of experimental data, it is mandatory to understand the structure of the data that are actually used for evaluation purposes. A full discussion of this aspect is out of the scope of this panel. Here we report some results coming from the PROMISE NoE project, whose experimental datasets rely on the idea of evaluation campaigns and have a quite broad validity.

An evaluation campaign provides large test collections (like multimedia, multilingual, text, images), specifying a set of topics and a relevance assessment.

Participants evaluates their searching algorithm(s) (i.e., experiments) against a specific collection, producing a (ranked) result set on which different metrics are computed and stored. These data can be represented by the TME (Topics-Metrics-Experiment) cube shown on Figure 1.

Starting from this cube, it is possible to aggregate or manipulate data in different ways, according to different evaluation needs that, roughly speaking,

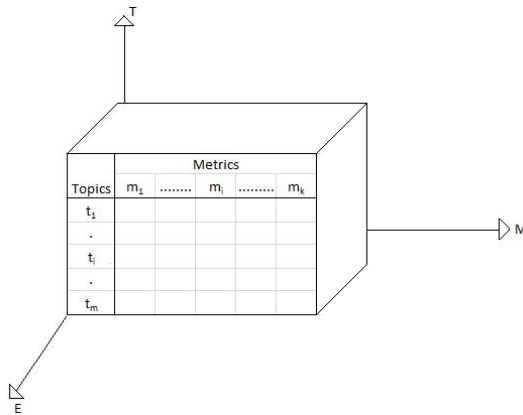


Fig. 1. The TME Data cube

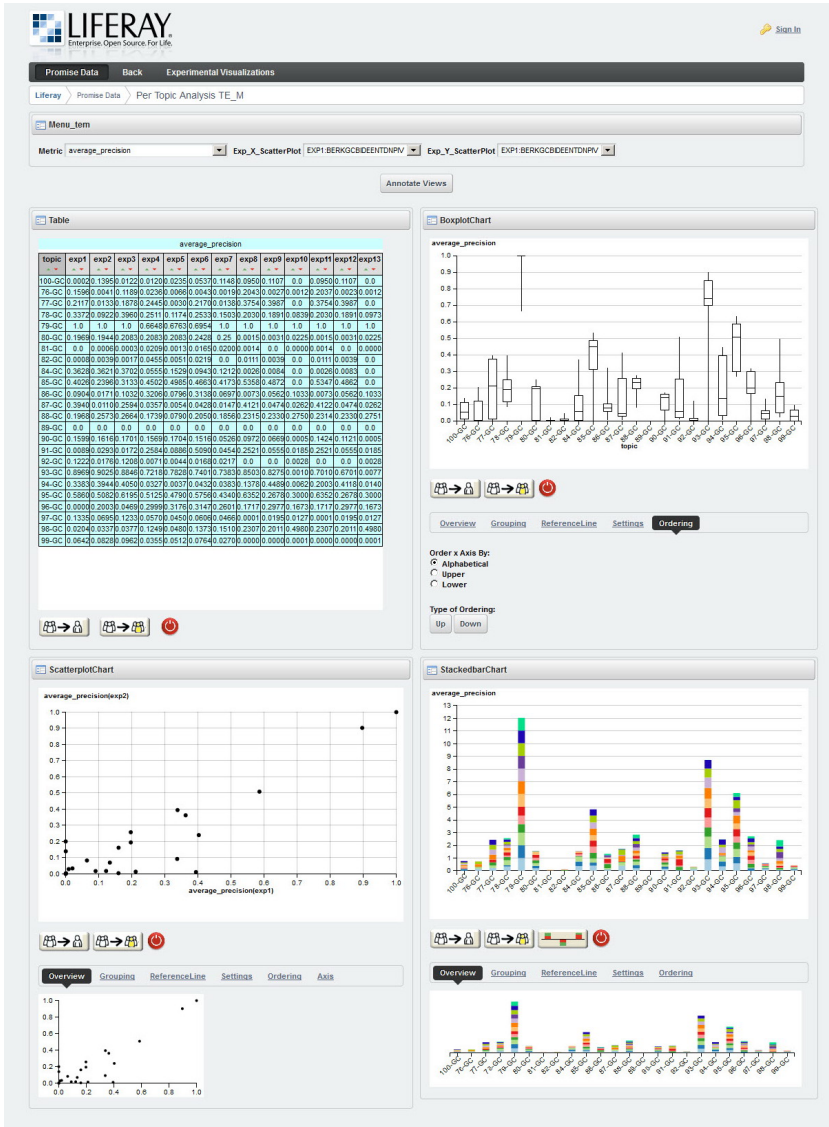


Fig. 2. Per topic analysis

correspond to a) evaluating a single search engine, either topic by topic or as a whole or b) to compare two or more search engines.

According to these two evaluation analysis patterns a VA prototype has been developed within PROMISE, implementing two visual analysis patterns, namely *Per topic analysis* and *Per experiment analysis*. In the following we describe such a prototype to provide a practical example of how a suitable visual analysis can foster and improve the usage and the interpretation of the managed evaluation data.

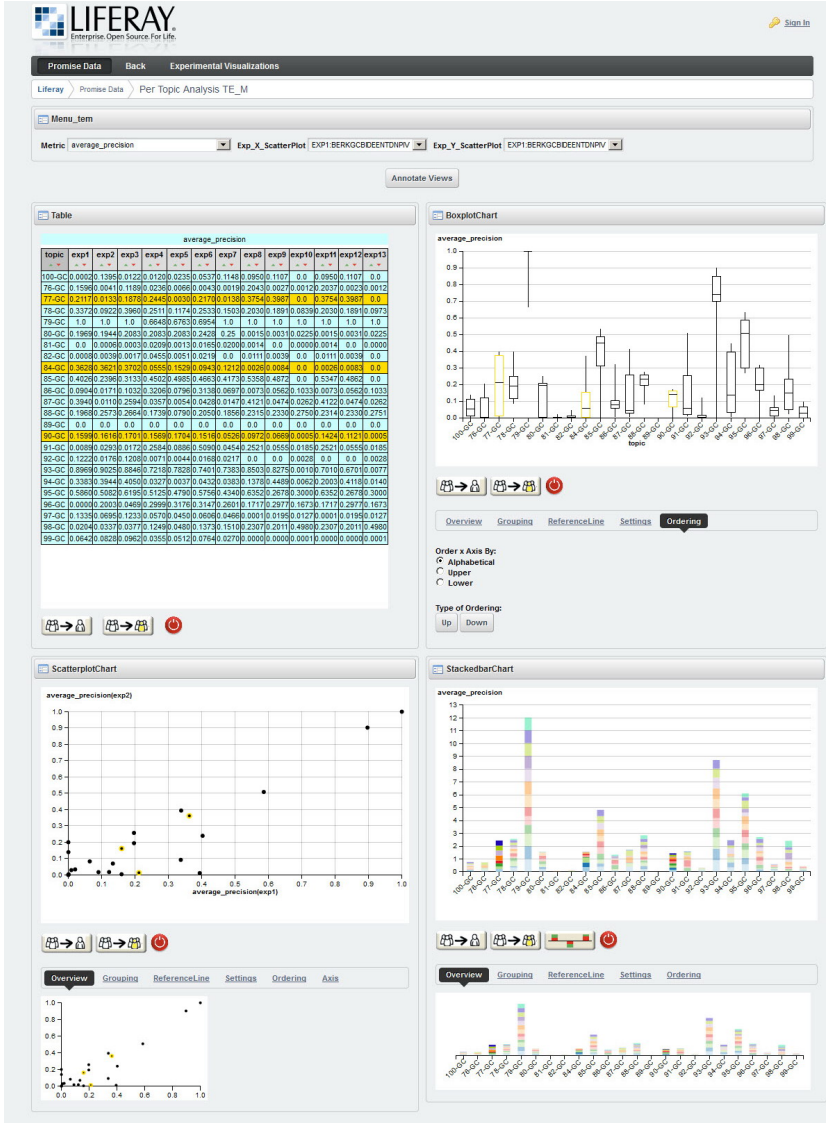


Fig. 3. Per topic analysis: a highlight operation

Per topic analysis Per topic analysis allows for comparing a set of search engines on each topic with respect to a chosen metric. Therefore the first step for an evaluator is to choose a metric m and, because the analysis implies a comparison on each topic, we represent topics on the x-axis in each available visualization. We foresee four coordinated visualizations: a table, a boxplot chart, a bi-dimensional scatter plot, and a stacked bar chart.

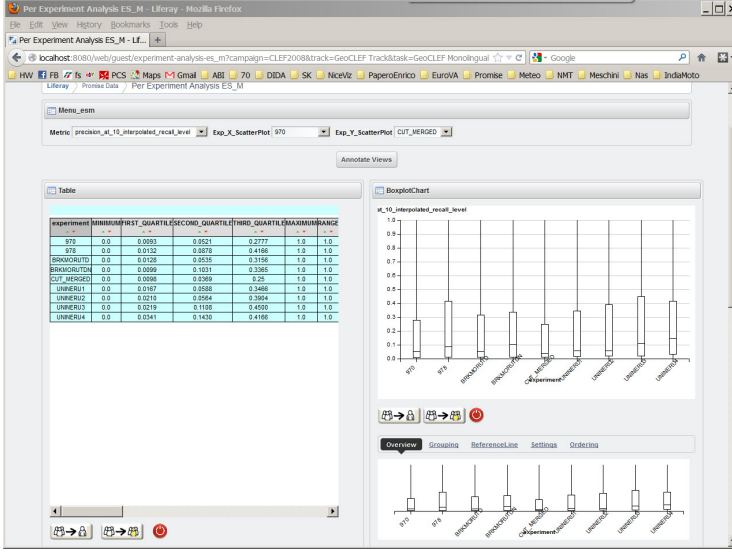


Fig. 4. Per experiment analysis: table and box plot

- The table represents a projection of the the $TME(m)$ cube on the chosen metric, allowing for looking up details (see Figure 2, upper left);
- The box plot chart (see Figure 2, upper right) is used to evaluate the trend of topics among experiments with respect to the chosen metric m . It contains a box plot for each topic (x-axis) and the chosen metric (y-axis). Each box plot is built calculating statistical indicators on the set of data represented by a single $TE(m)$ row;
- The bi-dimensional scatter plot (see Figure 2, lower left) allows for comparing topic behavior with respect to experiments. Each topic is represented by a point, according to the values it shows on the two experiments.
- The stacked bar chart (see Figure 2, lower right) has the same purpose as the box plot chart: to evaluate the trend of a topic among experiments with respect to a chosen metric. In such a visualization each topic is associated with all the values the explored metrics m exhibits in all experiments and the height of the bar represents the sum of all these values.

The evaluator can change the metric under analysis and restrict his or her focus on data subsets. As an example, Figure 3 shows three topics highlighted in all the four visualizations.

Per experiment analysis Per experiment analysis allows for analyzing a search engine as a whole and/or comparing the performances of a set of search engines with respect to a chosen descriptive statistics. As an example, on Figure 4, left side, the table chart represents a search engine in each row, showing the descriptive statistics of the metric average_precision (min, max, median, etc.). The box plot chart on Figure 4, right side, shows the percentile values of the observed metric for each experiment represented through boxplots.

Summarizing, we have introduced VA, a new, challenging methodology for analyzing large and complex dataset showing how apply it in the context of the evaluation of CHIAS, dealing with data structure and visualization requirements. We have used the European NoE PROMISE and the *Cross-Language Evaluation Forum (CLEF)* conference series as bed tests, but the ideas and the results presented in this panel have a quite broader scope.

4 Modeling Heterogeneous Humanistic Data in a User-Oriented Perspective

The primary step in the evaluation of a cultural heritage information access system from a digital humanist researchers point of view, consists in state clear what Digital Humanities (DH) is. DH is the name given to the alliance between associations involved in the study of relations between humanities and computer science (ADHO (Alliance of Digital Humanities Organizations); DH is a word referring to the main conference² in the field; DH is a open-access peer-reviewed electronic journal (Digital Humanities Quarterly³); DH is the title of publications, centers, blogs, projects and tools (see the “guide to digital humanities & arts”⁴). DH is a label that only recently started to identify a field historically known as “humanities computing” or “humanities computer science”; that is “computing” was replaced with digital. This terminological shift [4] corresponds to a new method of thinking about cultural objects as resulting from media integration. We assisted to the migration from a text-based computability process (i.e. markup languages, vocabularies and schemas like TEI, text analysis, text mining, text processing) to a process of representation and description of each informational resource in a social dimension (images, audio and video and their integration in computational systems for information retrieval/extraction/mining).

The purpose of DH is conceptualization, modelization, formalization of humanistic data/content (that is the domain); the goal of DH is to define methods and develop strategies of domain representation in order to computability. In this perspective the Cultural Heritage (CH) is a multimedia humanistic domain that requires to deal with formats and data types, to manage metadata and controlled vocabularies, to define procedures regarding these data selection and dissemination. The information access system, instead, is connected to Digital Library (DL) and regards the choice of the infrastructure, the definition of the services offered, the digital objects (in this case deriving from the CH domain) and the final users [5].

In this scenario, DH is procedure of modeling heterogeneous data, that is of managing information organization methods [6], in order to produce complex digital objects with the aim to acquire knowledge. This means that DH works by abstraction, enucleating the hermeneutical pertinent elements of a collection, for what concerns the computational objectives; this abstraction requires then to

² <http://adho.org/conference>

³ <http://www.digitalhumanities.org/dhq/>

⁴ <http://www.arts-humanities.net>

reflect on the data model and finally on the information architecture, i.e. user access/interface and user needs.

In this context the user has a crucial role. A good data model potentially allows a good user access. Reading, browsing, tagging, query are examples of services offered by the DL; they depend both on the modelization at information system level and, consequently, the DL access criteria. The browsing method is one of the main evaluation methods of a CH information access system. In particular, browsing by relationships is an essential issue. It depends on both the data model and the user access. The data model, that is the ontological representation of the domain, is fundamental, because semantic relationships constitute the most exhaustive exploration method. Different levels of relationships, semantically declared, could be established: lexical networks (hyponyms; hyperonyms; synonyms; meronyms; related terms), structural connections (intertext, paratext, hypertext, architext, metatext, following the Genette classification [7]), concepts and topics (overlapping meanings, discipline and the object of study, persons and occupations, products and institutions). A good CH information system has to provide methods for acquiring knowledge through browsing concepts explicitly related at a certain level. The user access finds a method of exploration of the modeled collection by a “faceted navigation” [8]. Facets represent classes and the data filtering is expression of the predicates. The data model could find a classification output through facets, that control heterogeneous data expressed by different media.

One last consideration. Both the data model and the facets are expression of a specific point of view. The interpretation of the collection not only determines the user satisfaction but also fulfills the user needs. User log is a gold mine useful to understand effectiveness and efficiency of an information system both at data model level as well as at navigation behavior level.

References

1. Keim, D., Andrienko, G., Fekete, J.-D., Görg, C., Kohlhammer, J., Melançon, G.: Visual analytics: Definition, process, and challenges. In: Kerren, A., Stasko, J.T., Fekete, J.-D., North, C. (eds.) *Information Visualization*. LNCS, vol. 4950, pp. 154–175. Springer, Heidelberg (2008)
2. Shneiderman, B.: The eyes have it: a task by data type taxonomy for information visualizations. In: *Proceedings of the IEEE Symposium on Visual Languages*, pp. 336–343 (1996)
3. Keim, D.A., Kohlhammer, J., Santucci, G., Mansmann, F., Wanner, F., Schaefer, M.: Visual analytics challenges. In: *Proceedings of the eChallenges 2009* (2009)
4. Svensson, P.: Humanities computing as digital humanities. *Digital Humanities Quarterly* (2009)
5. Fuhr, N., Tsakonas, G., Aalberg, T., Agosti, M., Hansen, P., Kapidakis, S., Klas, C.P., Kovács, L., Landoni, M., Micsik, A., et al.: Evaluation of digital libraries. *International Journal on Digital Libraries* 8, 21–38 (2007)
6. Taylor, A.G., Joudrey, D.N.: *The organization of information*. Libraries Unlimited, Westport (2004)
7. Genette, G.: *Palimpsestes*. Seuil, Paris (1982)
8. Broughton, V.: The need for a faceted classification as the basis of all methods of information retrieval. In: *Aslib Proceedings*, vol. 58, pp. 49–72. Emerald Group Publishing Limited (2006)