Applying Aggregation Operators for Information Access Systems: An Application in Digital Libraries

Enrique Herrera-Viedma,[1,*] Javier López Gijón,[2] Sergio Alonso,[1]
Josefina Vílchez,[2] Concha García,[2] Luis Villén,[2]
Antonio Gabriel López-Herrera[3]
[1]*Department of Computer Science and Artificial Intelligence, University of Granada, 18071, Granada, Spain*
[2]*Department of Library Science, University of Granada, 18071, Granada, Spain*
[3]*Department of Computer Science, University of Jaén, Jaén, Spain*

Nowadays, the information access on the Web is a main problem in the computer science community. Any major advance in the field of information access on the Web requires the collaboration of different methodologies and research areas. In this paper, the concept of aggregation operator playing a role for information access on the Web is analyzed. We present some Web methodologies, as search engines, recommender systems, and Web quality evaluation models and analyze the way aggregation operators help toward the success of their activities. We also show an application of the aggregation operators in digital libraries. In particular, we introduce a Web information system to analyze the quality of digital libraries that implements an important panel of aggregation operators to obtain the quality assessments. © 2008 Wiley Periodicals, Inc.

1. INTRODUCTION

The World Wide Web is a popular and interactive medium to collect, disseminate, and access an increasingly huge amount of information: it constitutes the basic mainstay of the so-called "information and knowledge society." Owing to its spectacular growth, related to both Web resources (pages, sites, and services) and visitors, the Web is nowadays the main information repository.

The explosive growth of the World Wide Web stimulates the development of fast and effective automated systems that support an easy and effective access to the information relevant to specific users' needs.[1] The analysis of some characteristics of the Web could help us to better understand the problem of information access on the Web, and why it offers so many opportunities to research[2]:

*Author to whom all correspondence should be addressed: e-mail: viedma@decsai.ugr.es.

INTERNATIONAL JOURNAL OF INTELLIGENT SYSTEMS, VOL. 23, 1235–1250 (2008)
 Published online in Wiley InterScience (www.interscience.wiley.com). • DOI 10.1002/int.20317

- The Web is an extremely dynamic information source because the information it holds receive constant updates and also because it is constantly growing in a disorganized and uncontrolled way.
- The Web contains information of many types, texts, structured tables, multimedia information (music, images, and movies), and so on.
- The Web presents many redundant and noisy information, and only a small portion of it contains truly relevant or useful information.

Consequently, at present, the definition of systems that help users to automatically access high-quality information relevant to their needs on the Web is a big challenge. It is related with many Web research topics, which include, but are not limited to, the following: Web mining,[3] Web evaluation,[2,4–7] recommender systems,[8] Web intelligent agents,[9] Web search engines,[1] filtering tools,[10] Web personalization,[11] and Semantic Web.[12,13] Furthermore, because of its complexity, it requires the use of interdisciplinary approaches that involve a wide range of tools as preference modeling, knowledge management, aggregation operators, and so on.

Aggregation operators are related to the problem of combining different values (scores, preferences, or ranking results) from various information sources. In the Web context, we identify three important applications:

- *Search engines:* For example, to evaluate multiterm queries or to build metasearch engines.
- *Recommender systems:* For example, to build recommendations according to different user judgments or to choose products according to multiple criteria.
- *Web quality evaluation:* For example, to evaluate the quality of a Web site or a Web service according to multiple criteria.

On the other hand, the developments on the Web are having a great influence on the developments or others information access instruments as digital libraries.[14–17] Since 1990s, the Internet and the Web have become the primary platform for libraries to build and deliver information resources, services, and instructions. Nowadays, in the digital age, we find two kinds of library user information services[15]:

1. *Traditional library user information services*, which are based on a face-to-face personal communication and are developed on-site, as for example, on-site bibliographic instruction, consultation, user technical support, classroom instruction, and so on; and
2. *Electronic library user information services*, which are based on the Web, can be developed on-site or off-site, and are accessible without any geographic and time limitations, as for example, integrated library systems, distance-learning services, e-databases services, Web library catalogs, open source journal information, Web search engines, instant messaging services, virtual reference, and so on.

Depending on the library framework, both services are necessary and complementary to develop library activities. However, electronic services allow us to improve the efficiency of the libraries, and therefore, we find hybrid libraries[18] that keep some traditional services but with a great tendency to create new digital services using all Web possibilities. In this new framework, we have to deal with new challenges and key issues if we want to offer quality library services to the users, as for example[15]: role of academic libraries, quality information resources, Web instructions and trainings, new assessment and evaluation methodologies, and so on.

In this paper, we analyze the important role that aggregation operators can play in different systems related to the automatic access to information on the Web. So, we study some Web methodologies for the information access on the Web as search engines, recommender systems, and Web quality evaluation tools; and we show the way aggregation operators have a hand in their performance. Furthermore, we show a Web information system that we have developed to contribute in the improvement of the quality of digital libraries. In particular, this is a system that implements a new methodology based on fuzzy tools to evaluate the quality of the Spanish digital libraries in an academic context. It is called ECABUD (http://sci2s.ugr.es/biblioreco/). ECABUD is a tool that we are applying in different universities of the Andalusia region to help in the quality evaluation and decision making for academic libraries. An important characteristic of ECABUD is that it implements a great panel of aggregation operators to obtain the quality assessments and supports different frameworks (as for example, it allows us to deal with quantitative information and qualitative one, and to manage criteria with different importance degrees).

The rest of the paper is organized as follows. Section 2 presents several Web methodologies together with an analysis of the use of aggregation operators on the Web. Section 3 presents ECABUD, and finally, our conclusions are outlined.

2. WEB METHODOLOGIES AND AGGREGATION OPERATORS

In this section, we present some Web methodologies that require the use of aggregation operators to carry out information access processes on the Web. In particular, we analyze the following Web methodologies: search engines, recommender systems, and Web quality evaluation models.

2.1. Search Engines

To help ordinary users to find the desired information on the Web, many search engines have been created. Search engines are a very common tool to discover and access information on the Web and have become a fundamental Web service. We can affirm that the dominant model of access in the Web is through search engines. Altavista, Excite, Google, Hotbot, Infoseek, Lycos, and so on, are examples of currently popular search engines.

Most of above search engines are based on the principles of traditional information retrieval.[19] In this sense, a search engine is a Web information retrieval system, which indexes Web documents or pages mainly by itself, allows users to search the Web via keyword-based queries, calculates relevant scores between indexed Web documents and queries, ranks Web documents against the relevant scores, and selects the top-ranked documents to constitute search results. However, the Web presents a new scenario that is quite different with respect to the traditional framework. The characteristics of the Web determine the design of search engines and their performance:

- The Web is possibly the bigger information resource and, thus, search engines have many difficulties to cover and index them totally in a reasonable time.

- The Web presents a structure of linked pages that should be took into account in the design of Web page ranking algorithms of search engines. Ranking algorithms in classical information retrieval only take into account the page content and do not use the graph structure of the Web.
- The Web is growing and being updated at a very high rate. Then, search engines are not able to track the publication of new Web pages, and the only way they have to build their indexes is to collect the documents by crawling the Web graph.[20] Search engines rely on computer programs called spiders or robots or Web crawlers that visit the documents distributed on the Web sites and index them to serve the search engines with a representation of the information available. Search engines can use different Web crawlers adopting different indexing criteria with their own schedules.
- The Web is very heterogeneous. The information available on the Web is presented in different document formats, various authoring styles, and as aforementioned, it can be of many types. Consequently, the Web is not indexed in any standard manner and finding information may be very difficult.

Consequently, the Web poses some new problems to search engines:

- *Problem of Web crawling:* Search engines can cover only a fraction of the whole Web because of its size, structure, and fast growth.[1,2]
- *Problem of Web spamming:* The Web ranking algorithms of search engines can be easily manipulated to promote a Web page near the top of the result set. For example, the PageRank used in Google, could be manipulated by using "link farm."[22]
- *Problem of Web overload:* Usually, when a user seeks information on the Web through a search engine, he receives hundreds of thousands of documents.

To improve search engines effectively, it is essential to apply techniques that allow us to boost the covering of the Web, push the most relevant pages to the top of the result list, and narrow the results. Below, we identify two important developments for boosting the performance of search engines in which the aggregation operators play an important role:

- *Metasearch engines:* The limitations of the search engines (partial crawling and spamming) have led to the introduction of metasearch engines.[23] A metasearch engine searches the Web by making requests to multiple search engines at once. Metasearch engines do not crawl the Web or maintain a database of Web pages. Instead, they act as a middle agent, passing the query on to the major search engines and then returning the results. In such a way, they overcome the crawling problem. SavvySearch, Metacrawler, Profusion, Inquirus, MetaGer, and so on are examples of metasearch engines. In the activity of a metasearch engine, the ability to combine or aggregate the results of multiple search engines is a critical aspect that could limit its performance. Search engines show the results of a query by means of an ordered list of documents according to their computed relevance. Therefore, the aggregation step in an metasearch engine essentially means to combine multiple ordered lists of relevant documents, offered by underlying search engines, into a uniform single ordered list, which is the final answer to the user query. The definition of robust aggregation tools can provide users a certain degree of robustness of search, removing malicious content and linking, and in such a way to overcome the spamming problem.[24] Some studies on aggregation tools in the metasearch engines can be found in Refs. 24–27.
- *Multiterm queries:* Users tend to make queries that result in poor precision and match many documents. The majority of Web users make single-term queries to express their information needs. As it is known, single-term queries return thousands of documents

and unfortunately, ranking the relevance of these documents is a difficult problem, and often the desired documents may not appear near the top of the list.[21] One possibility to improve the precision of results consists in using more query terms, i.e., to facilitate users the formulation of information needs by means of multiterm queries. Usually, this is implemented in the search engines by allowing the use of Boolean query languages. However, two major difficulties appear: to design appropriate user friendly interfaces that facilitate the formulation of Boolean queries (single terms combined via the logic connectives AND, OR, and NOT) and to design appropriate aggregation operators to model the evaluation of the logic connectives that allow to coherently combine the obtained partial results for each single term (called atom) of a Boolean user query. Some studies on aggregation tools in the evaluation of multiterm queries can be found in Refs. 24,28–30.

2.2. Recommender Systems

To filter the great amount of information available across the Web can improve the information access. Information filtering is a name used to describe a variety of processes involving the delivery of information to people who need it. Operating in textual domains, *filtering systems* or *recommender systems* evaluate and filter the great amount of information available on the Web (usually, stored in HTML or XML documents) to assist people in their search processes.[8] In general, recommender systems are an important Web development toward the goal of providing specific customized information to each user.[31]

Traditionally, these systems have fallen into two main categories[32]:

- *Content-based filtering systems* filter and recommend the information by matching user query terms with the index terms used in the representation of documents, ignoring data from other users. These recommender systems tend to fail when little is known about the user information needs, e.g., as happens when the query language is poor.
- *Collaborative filtering systems* use explicit or implicit preferences from many users to filter and recommend documents to a particular user, ignoring the representation of documents. These recommender systems tend to fail when little is known about a user, or when he/she has uncommon interests. An important application of collaborative recommender systems is the e-commerce activities deployed at e-commerce sites on the Web. By analyzing the links between people and the products they purchase or rate, the system recommends to the users what products they might be interested in, based on their and other users' past behavior.

The majority of existing recommender systems are based on collaborative filtering technologies, but it is clear that future recommender systems will incorporate both of these perspectives.[31]

In a content-based recommender system, the recommending is done by the system itself, i.e., the function of the system is to provide recommendations for its individual users, as, for example, the order of the documents in a given collection. In this sense, a recommender system is almost like an information retrieval system.[33] On the other hand, in a collaborative recommender system, the recommending is done by the users of system, i.e., the function of the system is to synthesize multiple users' recommendations of documents in the form of a single ranking for the individual user.[33]

In both types of recommender systems, the generation process of recommendations involves a procedure composed of two steps[31]:

1. *Computation of similarity degrees*. In the case of content-based recommender systems, the similarity degrees are calculated between a new unexperienced document and other documents that user has experienced and rated previously. In the case of collaborative recommender systems, the similarity degrees are calculated between our user profile and other user profiles, without considering the document representations.
2. *Aggregation of ratings*. In the case of content-based recommender systems, the ratings are supplied by the user who receives the recommendation. In the case of collaborative recommender system, the ratings are supplied by other users.

Obviously, the aggregation operators play a very important role in the generation of recommendations.

2.3. Web Quality Evaluation Tools

As we said at the beginning, the Web is dense of noisy, low-quality, and unreliable content. It would be extremely helpful for the Web search engines to be able to identify the quality of Web documents independently of a given user request. Some proposals use link analysis for estimating the quality of Web documents, but as we showed above these proposals can be easily manipulated. Therefore, to assess the quality of a Web document requires additional sources of information.

Recognizing the basic differences between publishing on the Web and publishing on paper may help to understand the lack of quality typical of the World Wide Web. In contrast to the printed paper world, on the World Wide Web anyone can publish information, either by simply acquiring space on a Web site and creating an electronic document (using any of the available formats HTML, XML, PDF, or PostScript) or by paying someone to create it. The fact is that there are neither rules nor standards governing the type and quality of information which a writer can put on the Web, nor a central control on where and how documents are published.[20] In the print world, authors can publish their own studies based on their own expenses, but self-published materials generally reach a limited audience. The Web, on the other hand, facilitates the distribution of self-published works, while significantly reducing the cost of production.[34] Therefore, the amount of Web documents and the number of content-based Web sites on the Internet is continuously and rapidly increasing, although in many cases this happens without an efficient information quality control.

In our opinion, there is not yet a clear and unambiguous definition of the concept of information quality on the Web, and unfortunately, well-founded and theoretical Web quality frameworks are still missing.[7] One can probably find as many definitions for information quality on the Web as there are papers on information quality. The quality evaluation on the World Wide Web is neither simple nor straightforward. Web quality is a complex concept, and its measurement or evaluation is multidimensional in nature.[4] Assuming such a nature, we can agree about the definition of Web quality given in Ref. 35 "we conceive quality as an aggregated value of multiple information quality criteria." Assuming this latter definition of Web quality, we can deduce that any Web quality evaluation methodology to rank Web resources should be composed of two fundamental elements[2,5,35]:

1. *An evaluation scheme:* It establishes the different evaluation criteria or indicators to be considered in the evaluation of Web resources and their importance degrees. Usually, it is appropriate to take into account both subjective and objective criteria and the users participation.[2]
2. *An measurement method:* It establishes how to obtain the ratings associated with each evaluation criteria (e.g., we would have to define a questionnaire to gather users' perceptions) and an aggregation or synthesis mechanism to obtain the quality rating associated with the particular Web resource.

Again, we can observe that the aggregation operators play a very important role in the generation of quality ratings. Furthermore, in this case and in recommender systems the design problem of aggregation operators could be more difficult if we need, for example, to combine quantitative with qualitative information or information assessed on different expression domains.

3. ECABUD: A WEB INFORMATION SYSTEM TO EVALUATE THE QUALITY OF DIGITAL LIBRARIES BASED ON AGGREGATION OPERATORS

In this section, we present the Web information system called ECABUD. As aforementioned, ECABUD is a tool to evaluate the quality of digital libraries, which combines traditional evaluation criteria together with electronic or digital criteria in order to adequately evaluate the new situation of libraries. ECABUD is oriented to academic libraries, which are the first kind of libraries that are getting most benefit from the Web possibilities, to help in teaching, learning, and researching activities.[14,36,37] An important aspect of ECABUD is that it presents a complete panel of aggregation operators based on fuzzy logic which allow (i) to aggregate qualitative information (represented by fuzzy linguistic variables[38]) and quantitative information (represented by numerical scores); (ii) to deal with criteria with different importance degrees; and (iii) to compute the quality assessments adopting different approaches that we could find between an optimist approach and a pessimistic one.

In the following subsections, we will describe its main components, explain how it works and, finally, we will present some results of ECABUD when it is applied to the evaluation of some Spanish university digital libraries.

3.1. Main Components of ECABUD

First, we want to comment some software aspects of ECABUD. ECABUD information system is based on a LAMP stack[39,40] (GNU/Linux, Apache Web server, MySQL database, and PHP programming language), and it is fully Web-based, that is, all its components and options can be accessed through a Web interface. In Figure 1 we show a snapshot of the system.

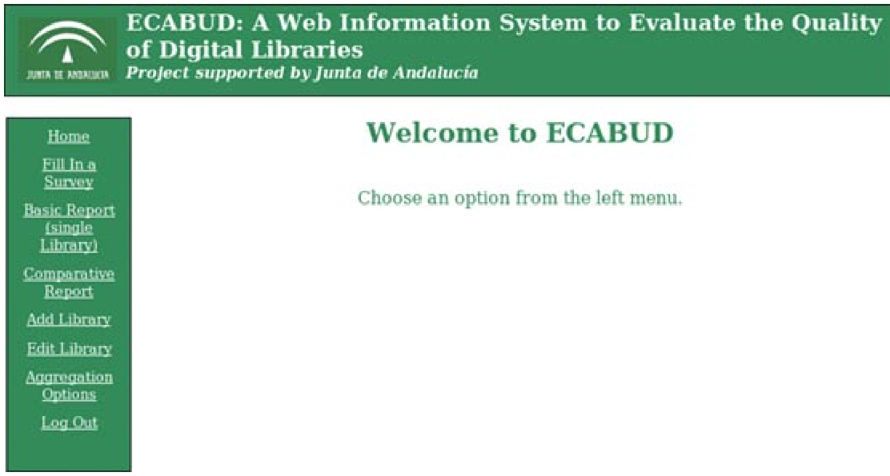

Figure 1. Snapshot of ECABUD information system.

On the other hand, the main components that support the activity of ECABUD are the following:

1. *A conceptual evaluation model.* The conceptual evaluation model is the support of the quality evaluation of the digital libraries developed by ECABUD. Our conceptual evaluation model of quality presents the following characteristics:

   - We introduce new electronic quality criteria to represent the new digital dimension of the libraries.
   - It is oriented to users because, as others existing library evaluation models,[41,42] the user participation in the quality evaluation processes of services is fundamental to correctly draw the situation of the service.
   - Therefore, we use both objective criteria (related with the quantitative data of the library) and subjective criteria (related with the user judgments) to evaluate the quality of the libraries.

   In particular, we evaluate the quality of the academic digital libraries according to the following seven objective criteria:

   (a) per capita accesses to the digital library,
   (b) external hit rate,
   (c) per capita access points to the digital library,
   (d) ratio of digital journals,
   (e) per capita queries,
   (f) per capita megabytes, and
   (g) web impact factor[43];

   and the following eleven subjective criteria:

   (a) you find what you are looking for,
   (b) coverage of the library about search topics,

(c) understandability of the digital library Web site,
(d) information electronic services about new inputs,
(e) added value information profits,
(f) training received,
(g) variety of search tools,
(h) navigability of the digital library Web site,
(i) satisfaction degree with the computing infrastructure,
(j) satisfaction degree with the response time, and
(k) global satisfaction degree.

The objective criteria are computed from the data provided by the different academic libraries, whereas the subjective criteria are obtained from judgments supplied by the users through of a questionnaire (q_1, q_2, ... ,q_{11}) (see Figure 2). The questionnaire is prepared to accept linguistic inputs that are more easily understood by the users than numerical ones. We use fuzzy linguistic variables[38] to represent users' opinions by mean of linguistic labels. Our conceptual evaluation model also allows us to establish different importance degrees among criteria, which can be assigned by the administrator or obtained from a panel of experts.

2. *A survey control module.* Through this module, we define the user questionnaires and establish the linguistic term sets that we use to represent the users' judgments.

A user questionnaire is composed of the eleven queries, one for each subjective criterion, and of some user characteristics, as user's sex, formation level (student, postgraduate student, teacher), his/her teaching topic (see Figure 2).

We use an ordinal fuzzy linguistic approach[2,44–46] to model the linguistic information. It is defined by considering a finite and totally ordered label set $S = \{s_i\}, i \in \{0, \ldots, \mathcal{T}\}$ in the usual sense, i.e., $s_i \geq s_j$ if $i \geq j$, and with odd cardinality (7 or 9 labels). The midterm represents an assessment of "approximately 0.5," and the rest of the terms being placed symmetrically around it. The semantics of the label set is established from the ordered structure of the label set by considering that each label for the pair $(s_i, s_{\mathcal{T}-i})$ is equally informative. For example, we can use the following set of nine labels to provide the user evaluations: {*T* = *Total*, *EH* = *Extremely_High*, *VH* = *Very_High*, *H* = *High*, *M* = *Medium*, *L* = *Low*, *VL* = *Very_Low*, *EL* = *Extremely_Low*, *N* = *None*}.

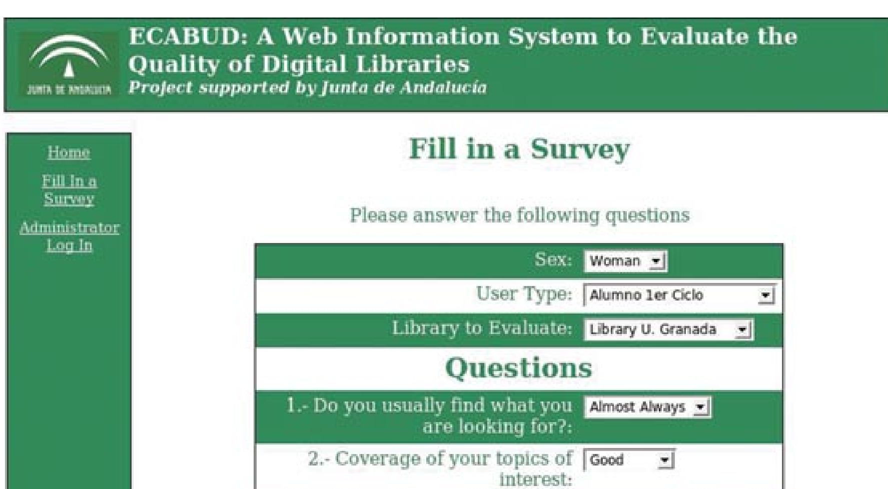

Figure 2. Filling a survey.

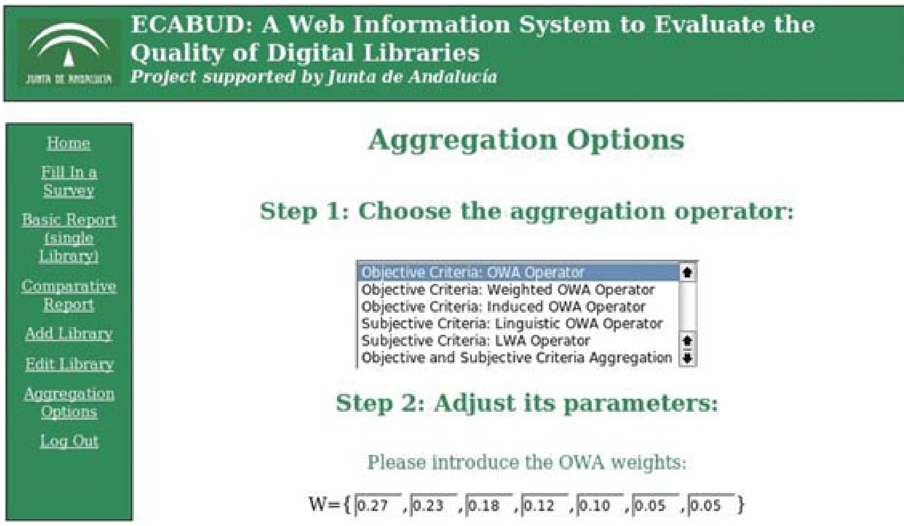

Figure 3. Aggregation operators.

3. *A computation module of quality assessments.* To obtain the partial/global quality assessments for each digital library, we have implemented a complete panel of fuzzy aggregation operators that allow us to develop a great variety of aggregation policies. In particular,

    - To aggregate scores associated with objective criteria, we propose to use OWA operators[47] when they are equally important and weighted OWA operators[48] and induced OWA operators,[49,50] in the other case.
    - To aggregate linguistic opinions associated with subjective criteria, we propose to use linguistic OWA operators[45] when they are equally important and LWA operators[46] in the other case.
    - To combine objective quality assessments together with subjective quality assessments, we use a tool to mix words with numbers that we defined in Ref. 51.

    These operators are shown in Figure 3.

    As it is known, it is possible to control the aggregation behavior of the family of OWA operators through a weighting vector W.[48] Furthermore, in Ref. 47 it was presented a way to compute W from a fuzzy linguistic quantifier.[52] Then system administrator will choose a fuzzy linguistic quantifier to guide the aggregation operators in each case.
4. *A report generation module.* This module is the generator of all data that allow us to better know the quality and performance of each digital library. We generate two kinds of reports to analyze the quality of libraries:

    - *Individual report:* This kind of report shows the situation of each library individually. Then, an individual report for a given library contains the list of quality assessments for each criterion, a global subjective quality assessment, a global objective quality assessment, and a total quality index of that library. To better understand these different quality assessments that draw the quality situation of a library we use different kinds of graphic outputs, as radar plots and ball graphs. We also use some radar plots to compare a particular library against an average of the libraries and, in such a way, to easily detect which particular criteria have better or worst evaluations than the average.

- *Benchmarking report:* This kind of report shows results of different libraries allowing establish comparisons among them. In such a way, it is possible to develop benchmarking procedures among different digital libraries. The purpose of benchmarking is to make improvements to the users and to be that catalyst for change. Benchmarking provides a structured framework for making comparisons between organizations. Clearly, using benchmarking reports a library can detect its drawbacks and those libraries that have overcome those same drawbacks, and therefore, libraries can learn a lot from each other.[53]

5. *An administrator module.* It allows the system administrator to configure the system to develop the evaluation processes to the libraries. For example, the system administrator can carry out some of the following activities:

- *Adding libraries to ECABUD:* The system administrator can add a new library to evaluate in the system. To do so, he is required to introduce the values necessary to compute the objective criteria evaluations for a particular library.
- *Editing libraries:* The administrator can edit and alter any objective criteria evaluation of any library (e.g., when more accurate information about a library has been obtained).
- *Generating reports:* The administrator can generate all kinds of reports for a given library from a report generation module.
- *Aggregation options:* The administrator manages aggregation operators and their parameters to compute the quality assessments.
- *Linguistic representation options:* The administrator manages the linguistic term sets used to represent users' judgments.

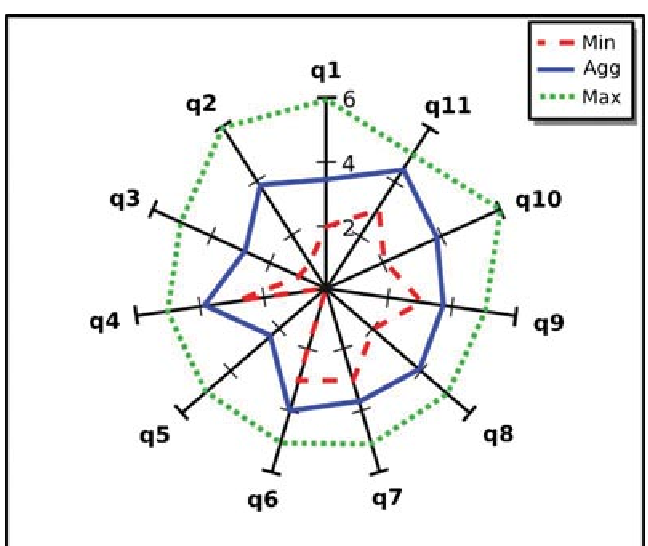

Figure 4. Radar plot of subjective criteria evaluations for the Library of Jaén University.

3.2. Application of ECABUD

In this section, we present an example of application of ECABUD to evaluate the quality of four Spanish academic digital libraries:

1. Library of Granada University
2. Library of Jaén University
3. Library of Córdoba University
4. Library of Málaga University

Actually, we have all data necessary to compute all objective criteria relative to year 2005 and enough surveys processed for each library. Furthermore, the behavior of the OWA operator is configurated as an arithmetic mean. Then, if we generate an individual report for the library of Córdoba University at this moment, we obtain the radar plot of every subjective criteria shown in Figure 4). In this particular example plot, we can see that the subjective criteria *q5* (added value information profits) has a lower evaluation, that is, users of the library think that this particular criterion should be improved. On the other hand, subjective criteria *q11* (global satisfaction degree) has a much better evaluation for this library (in fact, its evaluation is near to the maximum one, which means that almost every user agrees on the good evaluation of *q11*).

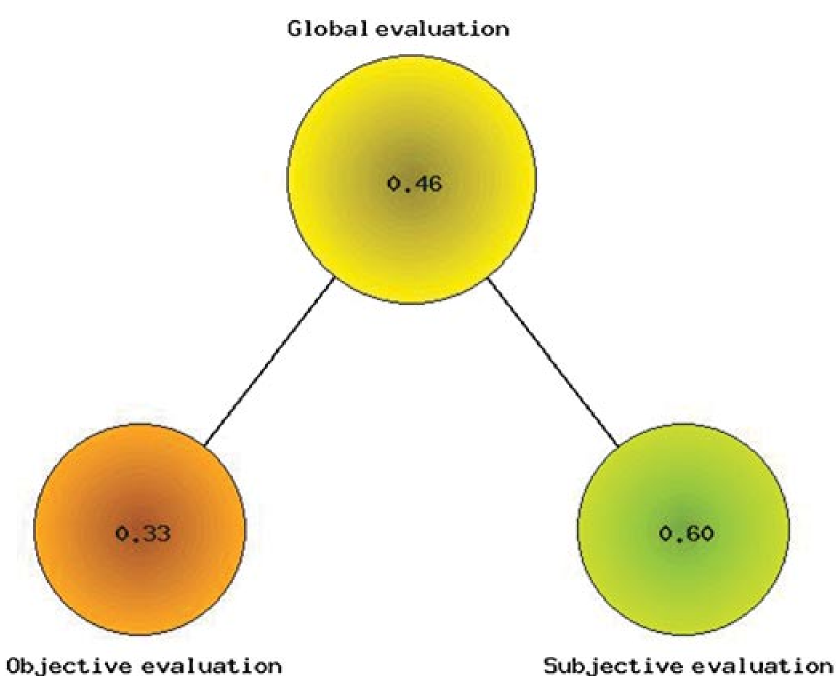

Figure 5. Objective, subjective, and global evaluations graph for library of Córdoba University.

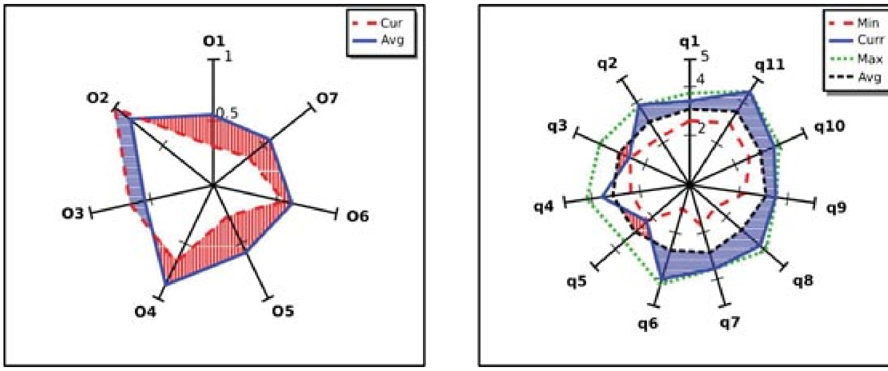

Figure 6. Objective (left) and subjective (right) criteria of the library of Granada University compared with the average of all libraries.

Table I. Rankings of libraries.

Rank	Objective Criteria	Subjective Criteria	Global
1st	Library U. Granada (0.98)	Library U. Granada (0.62)	Library U. Granada (0.8)
2nd	Library U. Córdoba (0.33)	Library U. Córdoba (0.60)	Library U. Córdoba (0.46)
3rd	Library U. Jaén (0.31)	Library U. Málaga (0.51)	Library U. Málaga (0.40)
4th	Library U. Málaga (0.29)	Library U. Jaén (0.39)	Library U. Jaén (0.35)

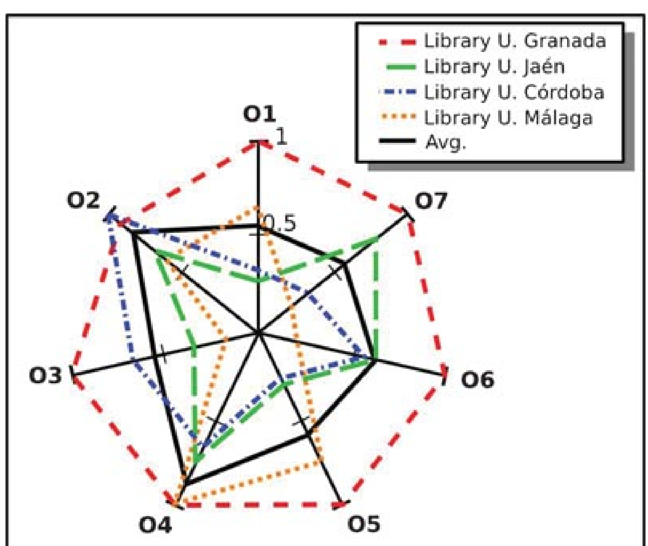

Figure 7. Comparison of objective criteria for all libraries.

We can generate another plot that summarizes both objective, subjective, and global evaluations for the library of Córdoba University (see Figure 5).

Finally, we generate two plots (one for the objective criteria and another for the subjective criteria) that show a comparison of the Library of Granada University with the average evaluations obtained by the rest of libraries, and thus, in such a way, we identify easily the criteria in which the library is better or worse than the average (see Figure 6). Horizontal lined areas represent the criteria in which the library is better than the average, while vertical lined areas represent the criteria in which it is worst than the average.

We can also generate a benchmarking report that allows us to compare between all libraries. If we choose to compare all four libraries, we obtain the rankings shown in Table I.

On the other hand, we can also generate a benchmarking report based on a plot radar to more easily compare all libraries (see, e.g., Figure 7). In the figure we can see that the *Library of the University of Granada* is the best one on almost every objective criteria.

4. CONCLUSIONS

We have analyzed that the aggregation operators play a very important role in the development of new technologies to access information on the Web. In particular, they are important in some Web developments to mitigate the wicked effects of Web spamming, crawling, and overload.

On the other hand, we have shown an application of soft aggregation operators in digital libraries. We have implemented ECABUD, which is a Web information system to evaluate the quality of the academic digital libraries that allows us to carry out benchmarking processes by means of its reports based on graphic outputs. ECABUD is based on objective and subjective criteria and uses a large number of soft aggregation operators based on the OWA operator, which can be used to aggregate numerical information and to aggregate linguistic information.

In the future, we shall research the impact of soft aggregation operators on the development of Web technologies. These operators allow us to perform soft information fusion attending to different aggregation behaviors or criteria that can be fixed by a user or system designer. As it was shown in Ref. 25, they can be very useful in the development of metasearch engines. We think that they could provide positive synergies in other Web technologies as recommender systems.

Acknowledgments

This work has been supported by the Spanish Research Project TIN2007-61079.

References

1. Lawrence S, Giles CL. Searching the World Wide Web. Science 1998;280(5360):98–100.
2. Herrera-Viedma E, Martínez L, Mata F, Chiclana F. A consensus support system model for group decision-making problems with multi-granular linguistic preference relations. IEEE Trans Fuzzy Syst 2005;13:5 644–658.

3. Kosala R, Blockeel H. Web mining research: a survey. ACM SIGKDD Explor 2000;2(1): 1–15.
4. Aladwani AM, Palvia PC. Developing and validating an instrument for measuring user-perceived web quality. Inf Manage 2002;39:467–476.
5. Herrera-Viedma E, Peis E. Evaluating the informative quality of documents in SGML-format using fuzzy linguistic techniques based on computing with words. Inf Process Manage 2003;39(2):195–213.
6. Herrera-Viedma E, Peis E, Morales del Castillo JM, Alonso S, Anaya K. A fuzzy linguistic model to evaluate the quality of Web sites that store XML documents. Int J Approx Reason 2007;46:226–253.
7. Gertz M, Ozsu MT, Saake G, Sattler KU. Report on the Dagstuhl Seminar "Data quality on the Web." Sigmod Rec 2004;33(1):127–132.
8. Reisnick P, Varian HR (Eds.). Special issue on recommender systems. Commun ACM 1997;40(3).
9. Maes P. Agents that reduce work and information overload. Communi ACM 1994;37:31–40.
10. Hanani U, Shapira B, Shoval P. Information filtering: overview of issues, research and systems. User Model User-Adapted Interact 2001;11:203–259.
11. Eirinaki M, Vazirginannis M. Web mining for web personalization. ACM Trans Internet Technol 2003;3(1):1–27.
12. Berners-Lee T, Hendler J, Lassila O. The semantic Web. Sci Am 2001;284(5):34–43.
13. Euzenat J. Research challenges and perspectives of the semantic Web. IEEE Intell Syst 2002;17(5):86–88.
14. Arms WY. Digital libraries. Cambridge, MA: MIT Press; 2001.
15. Bawden D, Vilar P. Digital libraries: to meet or manage user expectations. Aslib Proce: New Inf Perspect 2006;58(4):346–354.
16. Greenstein D. Digital libraries and their challenges. Libr Trends 2000;49(2):290–303.
17. Li L. Leveraging quality Web-based library user services in the digital age. Libr Manage 2006;27(6/7):390–400.
18. Rushbridge C. Towards the hybrid library. D-Lib Mag July/August 1998.
19. Salton G, McGill MH. Introduction to modern information retrieval. New York: McGraw-Hill; 1983.
20. Diligenti M, Gori M, Maggine M. A unified probabilistic framework for Web page scoring systems. IEEE Trans Knowl Data Eng 2004;16(1):4–16.
21. Lawrence S, Giles CL. Searching the Web: general and scientific information access. IEEE Commun Mag 1998;January:116–122.
22. Gori M, Witten I. The bubble of Web visibility. Commun ACM 2005;48(3):115–117.
23. Selberg E, Etzioni O. Multi-servicesearch and comparison using the MetaCrawler. In: Proc of 1995 WWW Conf; 1995.
24. Dwork C, Kumar R, Naor M, Sivakumar D. Rank aggregation methods for the Web. In The 10th WWW Int Conf; 2001. pp 613–622.
25. Bordogna G, Pasi G. A model for a soft fusion of information accesses on the Web. Fuzzy Sets Syst 2004;148:105–118.
26. Gomez M, Gauch S, Wang G. Profusion: intelligent fusion from multiple distributed search engines. J Univ Comput Sci 1996;289:637–649.
27. Yang X, Zhang M. Rational constraints for fusion methods in metasearch engine systems. Int J Intell Syst 2004;19:177–190.
28. Bordogna G, Pasi G. Linguistic aggreation operators of selection criteria in fuzzy information retrieval. J Intell Inf Syst 1995;10:233–248.
29. Lucas W, Topi H. Form and function: the impact of query term and operator usage on Web search results. J Amer Soc Inf Sci Technol 2002;53(2):95–108.
30. Yager RR. A note on weighted queries in information retrieval systems. J Am Soc Inf Sci 1987;38:23–24.
31. Yager RR. Fuzzy logic methods in recommender systems. Fuzzy Sets Syst 2003;136:133–149.

32. Popescul A, Ungar LH, Pennock DM, Lawrence S. Probabilistic models for unified collaborative and content-based recommendation in sparse data environments. In: Proc of the 17th Conf on Uncertainty in Artificial Intelligence (UAI), San Francisco, CA; 2001. pp 437–444.
33. Furner J. On recommending. J Am Soc Inf Sci Technol 2002;53(9):747–763.
34. Tyburski G. Get smart about Web site I.Q. [Online]. Available at http://searchenginewatch.com/searchday/article.php/2159621. Accessed on December 1, 2004.
35. Naumann F. Quality-driven query answering for integrated information systems. Lectures Notes in Computer Science Vol 2261; 2002.
36. Marchionini G, Maurer H. The roles of digital libraries in teaching and learning, Commun ACM 1995;38(4):67–75.
37. Sharifabadi SR. How digital libraries can support e-learning. The Electron Libr 2006;24–3:389–401.
38. Zadeh L A. The concept of a linguistic variable and its applications to approximate reasoning. Part I. Inf Sci 1975;8:199–249 ; Part II. Inf Sci 1975;8:301–357; Part III. Inf Sci 1975;9:43–80.
39. Lawton G. LAMP lights enterprise development efforts. Computer 2005;38(9):18–20.
40. Ramana UV, Prabhakar TV. Some experiments with the performance of LAMP architecture. In The 5th Int Conf on Computer and Information Technology; 2005. pp 916–920.
41. Cook C, et al. Developping a National Science Digital Library (NSDL) LibQUAL+ protocol: An E-service for assessing the library of the 21st century. NSDL Evaluation Workshop; 2003.
42. Marchionini G. Evaluating digital libraries: a longitudinal and multifaceted view. Libr Trends 2000;49(2):304–333.
43. Ingwersen P. The calculation of Web impacts factors. J Doc 1998;54(2):236–243.
44. Herrera-Viedma E. Modeling the retrieval process for an information retrieval system using an ordinal fuzzy linguistic approach. J Am Soc Inf Sci Technol 2001;52(6):460–475.
45. Herrera F, Herrera-Viedma E. Aggregation operators for linguistic weighted information. IEEE Trans Syst, Man Cybern A. 1997;27:646–656.
46. Herrera F, Herrera-Viedma E, Verdegay, JL: Direct approach processes in group decision making using linguistic OWA operators. Fuzzy Sets Syst 1996;79:175–190.
47. Yager RR. On ordered weighted averaging aggregation operators in multicriteria decision making. IEEE Trans Syst, Man Cybern 1998;18:183–190.
48. Yager RR: Families of OWA operators. Fuzzy Sets Syst 1993;59:125–148.
49. Yager RR, FilevDP. Induced ordered weighted averaging operators. IEEE Trans Syst, Man Cybern 1999;29:141–150.
50. Chiclana F, Herrera-Viedma E, Herrera F, Alonso S. Some induced ordered weighted averaging operators and their use for solving group decision-making problems based on fuzzy preference relations. Eur J Oper Res 2007;182(1):383–399.
51. Delgado M, Herrera F, Herrera-Viedma E, Martínez L. Combining numerical and linguistic information in group decision making. Inf Sci 1998;107:177–194.
52. Zadeh LA. A computational approach to fuzzy quantifiers in natural languages. Comput Math Appl 1983;9:149–184.
53. Garrod P, Kinnell M. Benchmarking development needs in the LIS sector. J Inf Sci 1997;23(2):111–118.
54. Herrera-Viedma E, Pasi G, Lopez-Herrera AG, Porcel C. Evaluating the information quality of Web sites: a methodology based on fuzzy computing with words. J Am Soc Inf Sci Technol, 2006;57(4):538–549.