World Scientific Proceedings Series on
Computer Engineering and Information Science 2

Intelligent Decision Making Systems

Proceedings of the
4th International
ISKE Conference

Koen Vanhoof
Da Ruan
Tianrui Li
Geert Wets
editors

World Scientific

Published by

World Scientific Publishing Co. Pte. Ltd.

5 Toh Tuck Link, Singapore 596224

USA office: 27 Warren Street, Suite 401-402, Hackensack, NJ 07601

UK office: 57 Shelton Street, Covent Garden, London WC2H 9HE

British Library Cataloguing-in-Publication Data

A catalogue record for this book is available from the British Library.

World Scientific Proceedings Series on Computer Engineering and Information Science – Vol. 2

INTELLIGENT DECISION MAKING SYSTEMS

Proceedings of the 4th International ISKE Conference on Intelligent Systems and Knowledge Engineering

Copyright © 2010 by World Scientific Publishing Co. Pte. Ltd.

All rights reserved. This book, or parts thereof, may not be reproduced in any form or by any means, electronic or mechanical, including photocopying, recording or any information storage and retrieval system now known or to be invented, without written permission from the Publisher.

For photocopying of material in this volume, please pay a copying fee through the Copyright Clearance Center, Inc., 222 Rosewood Drive, Danvers, MA 01923, USA. In this case permission to photocopy is not required from the publisher.

ISBN-13 978-981-4295-05-5

ISBN-10 981-4295-05-1

Printed in Singapore by World Scientific Printers

QUALITY EVALUATION OF DIGITAL LIBRARIES BASED ON LINGUISTIC INFORMATION

F. J. CABRERIZO

Department of Software Engineering and Computer Systems,
Distance Learning University of Spain,
Madrid, 28040, Spain
E-mail: cabrerizo@issi.uned.es

J. LÓPEZ-GIJÓN

Department of Library Science, University of Granada,
Granada, 18071, Spain
E-mail: jgijon@ugr.es

I. J. PÉREZ and E. HERRERA-VIEDMA

Department of Computer Science and A.I., University of Granada,
Granada, 18071, Spain
E-mail: {ijperez,viedma}@decsai.ugr.es

In this work, we present a model based on linguistic information to evaluate the quality of digital libraries. The quality evaluation of digital libraries is defined using users' perceptions on the quality of digital services provided through its Web site. The concept of quality of digital libraries is characterized by means of subjective quality indicators measured on their Web sites. We assume a fuzzy linguistic modelling to represent the users' perception and apply automatic tools of fuzzy computing with words based on the LOWA and LWA operators to compute global quality evaluations of digital libraries.

Keywords: Quality; Digital Libraries; Web Site; Fuzzy Linguistic Modelling.

1. Introduction

The explosive growth of the Web stimulates the creation of fast and effective automated systems that support an easy and effective access to the information relevant to specific users' needs.[1] It is changing the information access processes and is one of the most important information media. Thus, the developments on the Web are having a great influence over the developments on others information access instruments as digital libraries.[2,3]

Digital libraries are information collections that have associated services delivered to user communities using a variety of technologies. The information collections can be scientific, business or personal data, and can be represented as digital text, image, audio, video, or other media. This information can be digitalized paper or born digital material and the services offered on such information can be varied and offered to individuals or user communities. Internet access has resulted in digital libraries that are increasingly used by diverse communities for diverse purposes, and which sharing and collaboration have become important social elements. As digital libraries become commonplace, as their contents and services become more varied, people expect more sophisticated services from them.[4] Thus, the quality evaluation of digital libraries is an important task.

The main of this paper is to present a model to evaluate the quality of digital libraries. This evaluation model presents a set of subjective criteria related to the Web sites of digital libraries and a computation instrument of quality assessments. We assume that the quality of a digital library is measured through users' perceptions on the digital services offered through its Web site. For this reason, we use an ordinal fuzzy linguistic modelling[5] to represent the users' perceptions and tools of computing with words based on the linguistic aggregation operators LOWA[5] and LWA[6] to compute the quality assessments. To do so, users are invited to fill in a survey built on the set of subjective criteria.

The rest of the paper is set out as follows. Section 2 presents the preliminaries, i.e., the ordinal fuzzy linguistic modelling and fuzzy computing with words. Section 3 describes the model to evaluate the quality of digital libraries. Finally, Section 4 draws our conclusions.

2. Preliminaries

The *ordinal fuzzy linguistic approach*[5,6] is a very useful kind of fuzzy linguistic approach used for modeling the computing with words process as well as linguistic aspects of problems. 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 mid term 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: $\{T = Total, EH = Extremely\_High, VH = Very\_High, H = High, M =$

$Medium, L = Low, VL = Very\_Low, EL = Extremely\_Low, N = None\}$.

In any linguistic approach we need management operators of linguistic information.[5,6] Usually, the ordinal fuzzy linguistic model for computing with words is defined by establishing i) a negation operator, $Neg(s_i) = s_j \mid j = \mathcal{T} - i$, ii) comparison operators based on the ordered structure of linguistic terms: *Maximization operator*, $MAX(s_i, s_j) = s_i$ if $s_i \geq s_j$; and *Minimization operator*, $MIN(s_i, s_j) = s_i$ if $s_i \leq s_j$, and iii) adequate aggregation operators of ordinal fuzzy linguistic information such as the LOWA and LWA operators.

The *Linguistic Ordered Weighted Averaging* (LOWA) is an operator used to aggregate non-weighted ordinal linguistic information, i.e., linguistic information values with equal importance.[5]

Definition 2.1. Let $A = \{a_1, \ldots, a_m\}$ be a set of labels to be aggregated, then the LOWA operator, ϕ, is defined as $\phi(a_1, \ldots, a_m) = W \cdot B^T = \mathcal{C}^m\{w_k, b_k, k = 1, \ldots, m\} = w_1 \odot b_1 \oplus (1 - w_1) \odot \mathcal{C}^{m-1}\{\beta_h, b_h, h = 2, \ldots, m\}$, where $W = [w_1, \ldots, w_m]$, is a weighting vector, such that, $w_i \in [0, 1]$ and $\Sigma_i w_i = 1$. $\beta_h = w_h / \Sigma_2^m w_k, h = 2, \ldots, m$, and $B = \{b_1, \ldots, b_m\}$ is a vector associated to A, such that, $B = \sigma(A) = \{a_{\sigma(1)}, \ldots, a_{\sigma(m)}\}$, where, $a_{\sigma(j)} \leq a_{\sigma(i)}\ \forall\ i \leq j$, with σ being a permutation over the set of labels A. $\mathcal{C}^m$ is the convex combination operator of m labels and if m=2, then it is defined as $\mathcal{C}^2\{w_i, b_i, i = 1, 2\} = w_1 \odot s_j \oplus (1 - w_1) \odot s_i = s_k$, such that, $k = \min\{\mathcal{T}, i + round(w_1 \cdot (j - i))\}$ s_j, $s_i \in S$, $(j \geq i)$, where *"round"* is the usual round operation, and $b_1 = s_j$, $b_2 = s_i$. If $w_j = 1$ and $w_i = 0$ with $i \neq j\ \forall i$, then the convex combination is defined as: $\mathcal{C}^m\{w_i, b_i, i = 1, \ldots, m\} = b_j$.

The LOWA operator is an "or-and" operator[5] and its behavior can be controlled by means of W. In order to classify OWA operators with regards to their localization between "or" and "and", Yager[7] introduced a measure of *orness*, associated with any vector W : $orness(W) = \frac{1}{m-1} \sum_{i=1}^{m} (m-i) w_i$. This measure characterizes the degree to which the aggregation is like an "or" (MAX) operation. Note that an OWA operator with $orness(W) \geq 0.5$ will be an *orlike*, and with $orness(W) < 0.5$ will be an *andlike* operator.

The *Linguistic Weighted Averaging* (LWA) operator is another important operator which is defined to aggregate weighted ordinal linguistic information, i.e., linguistic information values with non equal importance.[6]

Definition 2.2. The aggregation of a set of weighted linguistic opinions, $\{(c_1, a_1), \ldots, (c_m, a_m,)\}$, $c_i, a_i \in S$, according to the LWA operator Φ is defined as $\Phi[(c_1, a_1), \ldots, (c_m, a_m)] = \phi(h(c_1, a_1), \ldots, h(c_m, a_m))$, where a_i

represents the weighted opinion, c_i the importance degree of a_i, and h is the transformation function defined depending on the weighting vector W used for the LOWA operator ϕ, such that, $h = MIN(c_i, a_i)$ if $orness(W) \geq 0.5$ and $h = MAX(Neg(c_i), a_i)$ if $orness(W) < 0.5$.

3. A Model Based on Linguistic Information to Evaluate the Quality of Digital Libraries

We use the information quality framework[8] defined in the context of management information systems as basis of our model to evaluate the quality of digital libraries. It has been satisfactorily applied to previous quality models for personal Web sites,[9] mobile Internet services[10] and Web sites that store Web documents.[11] In this information quality framework is established that the quality of the information systems cannot be assessed independently of the information users' opinions. This framework defines four major quality dimensions:[8]

(1) *Intrinsic quality.* The main criterion of this dimension is the accuracy of the information. If a reputation for inaccurate information becomes common knowledge for a particular information system, this system is viewed as having little added value and will result in a reduction of use.
(2) *Contextual quality.* It highlights the requirement that information quality must be considered within the context of the task in hand; it must be relevant, timely, complete, and appropriate in terms of amount, so as to add value to the tasks for which the information is provided.
(3) *Representational quality.* It requires information systems to present their information in such a way that it is interpretable, easy to understand, easy to manipulate, and is represented concisely and consistently.
(4) *Accessibility quality.* It requires the information system to be accessible but secure.

We adapt this information quality framework to develop our model to evaluate the quality of digital libraries. However, as it is oriented to users because the user participation in the quality evaluation processes of services is fundamental to correctly draw the situation of the service, we are going to define a low number of subjective criteria being easily understandable by the users in order to they do not cause the rejection of the users.

Taking into account these considerations, we define a model to evaluate the quality of digital libraries focused on their Web sites that present two

elements: *evaluation scheme* that contains the subjective criteria and a *computation method* to generate quality assessments of digital libraries.

3.1. *Evaluation scheme*

According to the quality framework,[8–11] the evaluation scheme contains the following four quality dimensions together with their digital quality criteria:

(1) *Intrinsic quality of digital libraries:* To evaluate the intrinsic quality or accuracy of digital libraries, we define the following subjective criterion: you find what you are looking for.
(2) *Contextual quality of digital libraries:* To evaluate the information quality of the digital libraries within the context, the following subjective criteria are defined: coverage of the library about search topics, information electronic services about new inputs, added value information profitsand also global satisfaction degree.
(3) *Representational quality of digital libraries:* It is evaluated taking into account the following subjective criteria: understandability of the digital library Web site and training received.
(4) *Accessibility and interaction quality of digital libraries:* It is measured considering the following subjective criteria: variety of search tools, navigability of the digital library Web site, satisfaction degree with the computing infrastructure and satisfaction degree with the response time.

3.2. *Computation method*

Firstly, we define a quality evaluation questionnaire providing questions for each one of the subjective criteria proposed in the evaluation scheme, i.e., there are eleven questions: $\{q_1, \ldots, q_{11}\}$. For example, for the subjective criterion you find what you are looking for, the question q_1 can be: "Do you usually find what you are looking for?". The concept behind each question is rated on a linguistic term set S. To do so, we can use the set of linguistic terms proposed in Section 2 to rate all the questions. In addition, we assume that each subjective criteria does not have the same importance in the evaluation scheme, i.e., it is assigned a relative linguistic importance degree for each subjective criterion: $\{I(q_1), \ldots, I(q_{11})\}$, $I(q_i) \in S$. These importance degree could be obtained from a set of experts or users' judgements.[12]

Then, assuming that we have a group of users $\{e_1, \ldots, e_L\}$ that have filled in the questionnaire and given a digital library $\mathcal{A}_m$, the computation method generates its quality assessment $r^m \in S$ using the linguistic aggregation operators LOWA and LWA in the following steps:

(1) Calculate for each subjective criterion q_i the global quality assessment $r_i^m \in S$ by means of LOWA operator ϕ, $r_i^m = \phi(e_1(q_i), \ldots, e_L(q_i))$, where $e_l(q_i) \in S$ is the linguistic preference provided by the e_l on subjective criteria represented by the question q_i.
(2) Calculate the quality assessment $r^m \in S$ by means of LWA operator Φ, $r^m = \Phi((I(q_1), r_1^m), \ldots, (I(q_{11}), r_{11}^m))$.

4. Conclusions

We have presented a linguistic model based on users' perceptions to evaluate the quality of the digital libraries. In the future, we extend the concept of quality to other new Web technologies, such as mobile Internet, which are being incorporated in the digital libraries.

5. Acknowledgment

This paper has been developed with the financing of FEDER funds with the PETRI project (PET2007-0460) and FUZZYLING project (TIN2007-61079).

References

1. S. Lawrence and C. Giles, *Science* **280**, 98 (1998).
2. W. Arms, *Digital Libraries* (MIT Press, 2001).
3. L. Li, *Library Management* **27**, 390 (2006).
4. M. Renda and U. Straccia, *Information Processing and Management* **41**, 5 (2005).
5. F. Herrera, E. Herrera-Viedma and J. Verdegay, *Fuzzy Sets and Systems* **79**, 175 (1996).
6. F. Herrera and E. Herrera-Viedma, *IEEE Trans. on Sys. Man and Cyb. Part. A: Syst. and Humans* **27**, 646 (1997).
7. R. Yager, *IEEE Trans. on Sys. Man and Cyb.* **18**, 183 (1988).
8. K. Huang, Y. Lee and R. Wang, *Quality Information and Knowledge* (Upper Saddle River, NJ: Prentice Hall, 1999).
9. P. Katerattanakul and K. Siau, Measuring Information Quality of Web sites: Development of an Instrument, in *Proc. of 20th Int. Conf. on Inf. Sys.*, (Charlotte, NC, 1999).
10. M. Chae and J. Kim, *Electronic Markets* **12**, 38 (2002).
11. E. Herrera-Viedma, G. Pasi, A. López-Herrera and C. Porcel, *Journal of American Society for Information Science and Technology* **57**, 538 (2006).
12. P. Zhang and G. V. Dran, Expectations and Rankings of Websites Quality Features: Results of Two Studies on User Perception, in *Proc. of the 34th Hawaii Inter. Conference on System Sciences*, (Hawaii, 2001).