Lecture Notes in Computer Science

The LNCS series reports state-of-the-art results in computer science research, development, and education, at a high level and in both printed and electronic form. Enjoying tight cooperation with the R&D community, with numerous individuals, as well as with prestigious organizations and societies, LNCS has grown into the most comprehensive computer science research forum available.

The scope of LNCS, including its subseries LNAI and LNBI, spans the whole range of computer science and information technology including interdisciplinary topics in a variety of application fields. The type of material published traditionally includes

- proceedings (published in time for the respective conference)
- post-proceedings (consisting of thoroughly revised final full papers)
- research monographs (which may be based on outstanding PhD work, research projects, technical reports, etc.)

More recently, several color-cover sublines have been added. Beyond a collection of papers, various added-value components; sublines include

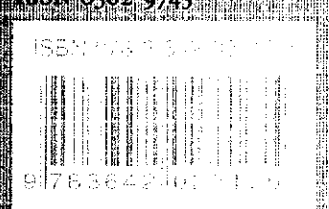
- tutorials (textbook-like monographs or collections of lectures given at advanced courses)
- state-of-the-art surveys (offering complete and mediated coverage of a topic)
- hot topics (introducing emergent topics to the broader community)

In parallel to the printed book, each new volume is published electronically in LNCS Online.

Detailed information on LNCS can be found at www.springer.com/lncs

Proposals for publication should be sent to LNCS Editorial, Tiergartenstr. 17, 69121 Heidelberg, Germany. E-mail: lncs@springer.com

ISSN 0302-9743



Lecture Notes in Computer Science

LNCS LNAI LNBI

springer.com

Gaedke • Grossniklaus
Díaz (Eds.)

LNCS 5648

Web Eng

9th International Conference, ICWE 2009
San Sebastián, Spain, June 2009
Proceedings

Springer

Lecture Notes in Computer Science 5648

Commenced Publication in 1973
Founding and Former Series Editors:
Gerhard Goos, Juris Hartmanis, and Jan van Leeuwen

Martin Gaedke Michael Grossniklaus
Oscar Díaz (Eds.)

Web Engineering

9th International Conference, ICWE 2009
San Sebastián, Spain, June 24-26, 2009
Proceedings

Springer

Volume Editors

Martin Gaedke
Chemnitz University of Technology
Faculty of Computer Science
Straße der Nationen 62, 09111 Chemnitz, Germany
E-mail: martin.gaedke@cs.tu-chemnitz.de

Michael Grossniklaus
Politecnico di Milano
Dipartimento di Elettronica e Informazione
Piazza Leonardo da Vinci 32, 20133 Milano, Italy
E-mail: grossniklaus@elet.polimi.it

Oscar Díaz
University of the Basque Country
Department of Computer Languages and Systems
Paseo de Manuel Lardizabal 1, 20018 San Sebastián, Spain
E-mail: oscar.diaz@ehu.es

Library of Congress Control Number: Applied for

CR Subject Classification (1998): H.5, H.4, K.6, D.2, C.2, H.3.5, H.5.3

LNCS Sublibrary: SL 3 – Information Systems and Application, incl. Internet/Web and HCI

ISSN 0302-9743
ISBN-10 3-642-02817-9 Springer Berlin Heidelberg New York
ISBN-13 978-3-642-02817-5 Springer Berlin Heidelberg New York

Printed in Germany

Typesetting: Camera-ready by author, data conversion by Scientific Publishing Services, Chennai, India
Printed on acid-free paper SPIN: 12711277 06/3180 5 4 3 2 1 0

Preface

As of 2009, the discipline of Web engineering is a well-established and mature field of research within the software engineering, database, information technology, and other related communities. By its very nature, Web engineering is, therefore, a multidisciplinary field that is beginning to establish ties even outside the domain of computer science. As a discipline, Web engineering systematically applies the knowledge of Web science to the development and evolution of Web-based applications and systems. This volume contains the proceedings of the 9th International Conference on Web Engineering (ICWE 2009), which was held in San Sebastián, Spain in June 2009. The ICWE conferences are among the most essential events of the Web engineering community. This fact is manifested both by the number of accomplished researchers that support the conference series with their work and contributions as well as by the continuing patronage of several international organizations dedicated to promoting research and scientific progress in the field of Web engineering.

ICWE 2009 followed conferences in Yorktown Heights, NY, USA; Como, Italy; Palo Alto, CA, USA; Sydney, Australia; Munich, Germany; Oviedo, Spain; Santa Fe, Argentina; and Cáceres, Spain. With San Sebastián as this year's venue, the conference series visits the country where it was originally launched in 2001 for the third time.

This year's call for papers attracted a total of 90 submissions from 33 countries spanning all continents of the world with a good coverage of all the different aspects of Web engineering. Topics addressed by the contributions include areas ranging from more traditional topics such as component-based Web engineering, model-driven Web engineering, navigation, search, Semantic Web, quality, and testing to novel domains such as the Web 2.0, rich internet applications, and mashups. All submitted papers were reviewed in detail by at least three members of the Program Committee which was composed of experts in the field of Web engineering from 23 countries. Based on their reviews, 22 submissions were accepted as full papers (24%) and 15 as short papers (22%). The program was completed by 8 posters and 10 demonstrations that were presented in dedicated sessions at the conference. Finally, the conference was also host to keynotes by James A. Hendler (Rensselaer Polytechnic Institute, USA), Jaideep Srivastava (University of Minnesota, USA), and Juan Jose Hierro (Telefonica, Spain) as well as an outstanding collection of four tutorials and four workshops.

We would like to express our gratitude to all the sponsors that supported ICWE 2009 financially, namely, the Regional Council of Gipuzkoa, the Association of Industries for Electronic and Information Technologies in the Basque Country (GAIA), LKS Co., and the University of the Basque Country (Summer Course Board). The conference would not have been possible without the endorsement of the International World Wide Web Conference Committee (IW^{3}C^2)

and the International Society for Web Engineering (ISWE). In this context, we would especially like to thank Bebo White and Geert-Jan Houben for their work as our liaisons to these two organizations. We are also indebted to the various Chairs (Josu Aramberri, Francisco Curbera, Florian Daniel, Peter Dolog, Jon Iturrioz, Oscar Pastor, Mario Piattini, Gustavo Rossi, Takehiro Tokuda, and Antonio Vallecillo) and to the local organizers who helped with their enthusiastic work to make ICWE 2009 a reality. Finally, a special thanks to all the researchers who contributed their work and participated in the conference. After all, as with any other conference, exchanging ideas and forging connections is what it is all about.

May 2009

Martin Gaedke
Michael Grossniklaus
Oscar Díaz

Organization

General Chair

Oscar Díaz	University of the Basque Country, Spain

Program Chairs

Martin Gaedke	Chemnitz University of Technology, Germany
Michael Grossniklaus	Politecnico di Milano, Italy

Program Committee

Silvia Abrahão	Polytechnic University of Valencia, Spain
Virgilio Almedia	Federal University of Minas Gerais, Brazil
Boualem Benatallah	University of New South Wales, Australia
Maria Bielikova	Slovak University of Technology in Bratislava, Slovakia
Judith Bishop	University of Pretoria, South Africa
Marco Brambilla	Politecnico di Milano, Italy
Chris Brooks	University of San Francisco, USA
Jordi Cabot	University of Toronto, Canada
Fabio Casati	University of Trento, Italy
Sven Casteleyn	Vrije Universiteit Brussel, Belgium
Dan Chiorean	University Babes-Bolyai, Romania
Maria da Graça Pimentel	University of São Paulo, Brazil
Paul Dantzig	IBM T.J. Watson Research Center, USA
Peter Dolog	Aalborg University, Denmark
Schahram Dustdar	Vienna University of Techonology, Austria
Flavius Frasincar	Erasmus University of Rotterdam, The Netherlands
Dragan Gasevic	Simon Fraser University, Canada
Athula Ginige	University of Western Sydney, Australia
Angela Eck Soong Goh	Nanyang Technological University, Singapore
Jaime Gomez	University of Alicante, Spain
Mei Hong	Peking University, China
Geert-Jan Houben	Delft University of Technology, The Netherlands
Arun Iyengar	IBM, USA
Stanislaw Jarzabek	National University of Singapore, Singapore
Gerti Kappel	Vienna University of Technology, Austria

Nora Koch	Ludwig-Maximilians-Universität München, Germany
David Lowe	University of Technology Sydney, Australia
Ioana Manolescu	Inria Futurs, France
Maristella Matera	Politecnico di Milano, Italy
Emilia Mendes	University of Auckland, New Zealand
San Murugesan	University of Southern Cross, Australia
Moira C. Norrie	ETH Zurich, Switzerland
Luis Olsina	National University of la Pampa, Argentina
Satoshi Oyama	Kyoto Univerisity, Japan
Oscar Pastor	Polytechnic University of Valencia, Spain
Vicente Pelechano	Polytechnic University of Valencia, Spain
Claudia Pons	National University of la Plata, Argentina
Birgit Pröll	University of Linz, Austria
I.V. Ramakrishnan	Stony Brook University, USA
Simos Retalis	University of Crete, Greece
Gustavo Rossi	National University of la Plata, Argentina
Klaus-Dieter Schewe	Massey University, New Zealand
Daniel Schwabe	Pontifical Catholic University of Rio de Janeiro, Brazil
Weisong Shi	Wayne University, USA
Katsumi Tanaka	Kyoto University, Japan
Bernhard Thalheim	Christian Albrechts University Kiel, Germany
Giovanni Toffetti Carughi	Università della Svizzera Italiana, Switzerland
Takehiro Tokuda	Tokyo Institute of Technology, Japan
Riccardo Torlone	Univesità di Roma, Italy
Jean Vanderdonckt	Université Catholique de Louvain, Belgium
Petri Vuorimaa	Helsinki University of Technology, Finland
Vincent Wade	Trinity College Dublin, Ireland
Bebo White	Stanford Linear Accelerator Center, USA
Marco Winckler	Université Paul Sabatier, France
Bin Xu	Tsinghua University, China

Posters Chair

Florian Daniel	University of Trento, Italy

Posters Committee

Maria Bielikova	Slovak University of Technology in Bratislava, Slovakia
Alexander Birukou	University of Trento, Italy
Alessandro Bozzon	Politecnico di Milano, Italy
Sara Comai	Politecnico di Milano, Italy
Vincenzo D'Andrea	University of Trento, Italy

Federico Michele Facca	STI Innsbruck, Austria
Michael Grossniklaus	Politecnico di Milano, Italy
Nora Koch	Ludwig-Maximilians-Universität München, Germany
David Lowe	University of Technology Sydney, Australia
Maristella Matera	Politecnico di Milano, Italy
Moira C. Norrie	ETH Zurich, Switzerland
Giovanni Toffetti Carughi	Universitá della Svizzera Italiana, Switzerland
Jin Yu	Tiburon, USA

Demonstrations Chair

Antonio Vallecillo	University of Malaga, Spain

Demonstrations Committee

Silvia Abrahão	Polytechnic University of Valencia, Spain
Michel Chaudron	Eindhoven University of Technology, The Netherlands
Piero Fraternali	Politecnico di Milano, Italy
Nora Koch	Ludwig-Maximilians-Universität München, Germany
David Lowe	University of Technology Sydney, Australia
Alfonso Pierantonio	Università de L'Aquila, Italy
Vicente Pelechano	Polytechnic University of Valencia, Spain
Gustavo Rossi	National University of la Plata, Argentina
Fernando Sánchez-Figueroa	University of Extremadura, Spain
Manuel Wimmer	Vienna University of Techonology, Austria

Doctoral Consortium Chair

Gustavo Rossi	National University of la Plata, Argentina

Doctoral Consortium Committee

Sven Casteleyn	Vrije Universiteit Brussel, Belgium
Florian Daniel	University of Trento, Italy
Damiano Distante	Università Telematica Telma, Italy
Marí Escalona	University of Seville, Spain
Fernando Sánchez-Figueroa	University of Extremadura, Spain
Joan Fons	Polytechnic University of Valencia, Spain
Daniela Godoy	ISISTAN Research Institute, Argentina

Nora Koch	Ludwig-Maximilians-Universität München, Germany
Luis Olsina	National University of la Pampa, Argentina
Gustavo Rossi	National University of la Plata, Argentina
Wieland Schwinger	Johannes Kepler University of Linz, Austria

Workshop Chairs

Peter Dolog	Aalborg University, Denmark
Takehiro Tokuda	Tokyo Institute of Technology, Japan

Tutorial Chair

Oscar Pastor	Polytechnic University of Valencia, Spain

Publicity Chair

Mario Piattini	University of Castilla-La Mancha, Spain

Industrial Chair

Francisco Curbera	IBM Research, USA

ISWE Conference Steering Committee Liaison

Geert-Jan Houben	Delft University of Technology, The Netherlands

IW3C2 Liaison

Bebo White	Stanford Linear Accelerator Center, USA

Local Organization Chairs

Jon Iturrioz	University of the Basque Country, Spain
Josu Aramberri	University of the Basque Country, Spain

Additional Referees

Ahmed, Faisal	Bartalos, Peter	Cachero, Cristina
Al-Naymat, Ghazi	Beheshti, Seyed M. R.	Castillo, Claris
Baez, Marcos	Borodin, Yevgen	Covella, Guillermo
Barla, Michal	Brosch, Petra	Ferreira, Renato

Garg, Supriya	Luo, Sen	Simko, Marian
Garrigós, Irene	Mahmud, Jalal	van der Sluijs, Kees
Gu, Zhifeng	Meinecke, Johannes	de Spindler, Alexandre
Guabtni, Adnene	Melia, Santiago	Subramanian, Revathi
Gómez, Cristina	Molina, Hernan	Syukur, Evi
Hidders, Jan	Moro, Mirella	Tan, Puay Siew
Huang, Gang	Noack, Rene	Torres, Victoria
Insfran, Emilio	Oyama, Satoshi	Valverde, Francisco
Islam, Asiful	Paraschiv, Sergiu	Van Woensel, William
Jatowt, Adam	Parra, Cristhian	Veloso, Adriano
Jiao, Wenpin	Pereira, Adriano	Vu, Hung
Kawai, Hideki	Prates, Raquel	Wang, Yasha
Kotsis, Gabriele	Puzis, Yury	Wimmer, Manuel
Kroiss, Christian	Ravi, Jayashree	Wu, Qinyi
Le, Duy Ngan	Retschitzegger, Werner	Yan, Yixin
Leonardi, Erwin	Sakr, Sherif	Zapletal, Marco
Leone, Stefania	Schall, Daniel	Zhang, Lu
Li, Fei	Seidl, Martina	Zhang, Yan
Li, Ge	Silva-Lepe, Ignacio	Zhao, Junfeng

Table of Contents

Navigation

Process, Planning and Phases

Quality

Rich Internet Applications

Search

Testing

Web Services, SOA and REST

Web 2.0

Posters

Demonstrations

8. Ko, A.J., Myers, B.A., Aung, H.H.: Six learning barriers in end-user programming systems. In: VL/HCC, pp. 199–206. IEEE Computer Society, Los Alamitos (2004)
9. Ellis, B., Stylos, J., Myers, B.A.: The Factory Pattern in API Design: A Usability Evaluation. In: ICSE, pp. 302–312. IEEE Computer Society, Los Alamitos (2007)
10. Jeong, S.Y., Xie, Y., Beaton, J., Myers, B., Stylos, J., Ehret, R., Karstens, J., Efeoglu, A., Busse, D.K.: Improving Documentation for eSOA APIs through User Studies. In: Proc. of the Second International Symposium on End User Development (IS-EUD 2009), Siegen, Germany, March 2–4 (2009)
11. Cappiello, C.: Analyzing the Success of Mashup Components. Technical report, Politecnico di Milano (2009)
12. Redman, T.: Data Quality for the Information Age. Artech House (1996)
13. Wang, R., Strong, D.: Beyond Accuracy: What Data Quality Means to Data Consumers. Journal of Management Information Systems 12 (1996)
14. Ballou, D., Wang, R., Pazer, H., Tayi, G.: Modeling Information Manufacturing Systems to Determine Information Product Quality. Management Science 44 (1998)
15. Nielsen, J.: Web Usability. New Riders, Indianapolis (2000)
16. Matera, M., Rizzo, F., Carughi, G.T.: Web Usability: Principles and Evaluation Methods. In: Web Engineering, pp. 109–142. Springer, Heidelberg (2005)
17. Consortium, W.: Wai guidelines and techniques. Technical report (2007), http://www.w3.org/WAI/guid-tech.html

Towards the Discovery of Data Quality Attributes for Web Portals*

Carmen Moraga[1], Mª Ángeles Moraga[1], Coral Calero[1], and Ángélica Caro[2]

[1] Alarcos Research Group – Institite of Information Technologies & Systems, Dept. Information Technologies & Systems – Escuela Superior de Informática, University of Castilla-La Mancha, Spain
Carmen.Moraga@alu.uclm.es, {MariaAngeles.Moraga,Coral.Calero}@uclm.es
[2] Department of Computer Science and Information Technologies, University of Bio Bio, Chillán, Chile
mcaro@ubiobio.cl

Abstract. The Internet has become in a place for the exchange and publication of data. Nowadays, Web portals serve as an important means to access information. In this context, the concepts of quality in general and of data quality in particular are highly relevant. The objective of this paper is to carry out a systematic literature review (SLR) in order to discover the state-of-the art in data quality for Web portals, and to evaluate the evolution of data quality since 2006, when another SLR was carried out, and in which a PDQM (Portal Data Quality Model) was defined. As a result, 39 attributes have been considered relevant for the assessment of data quality in Web portals.

Keywords: Data/Information quality, Web portals, data quality attributes.

1 Introduction

One of the aims of many web portals is to select, organize and distribute content (information or other services and products) in order to satisfy their users/customers [1]. However, unnecessary, out of date or erroneous data are also included. Data quality is an actual factor in competitiveness.

Bearing in mind the importance of data quality, the main goal of this paper is to discover the state-of-the-art in Web portal data quality through a systematic literature review (SLR). This SLR is based on a previous SLR [2], which covered the years 1996 to 2005, and in which 33 attributes considered to be relevant for Web portal data quality were chosen. These attributes were then used to define a quality model for the assessment of Web portal data quality, namely PDQM (Portal Data Quality Model), in [3]. The SLR which is presented here covers 2006 to the end of 2008. The objective of this SLR is to establish the evolution of Web portal data quality

* This work is part of the projects: INCOME (PET2006-0682-01) from Ministerio de Educación and IVISCUS (PAC08-0024-5991) from the Consejería de Educación y Ciencia (JCCM) and DQNet (TIN2008-04951-E) supported by the Spanish Ministerio of Educación y Ciencia.

M. Gaedke, M. Grossniklaus, and O. Díaz (Eds.): ICWE 2009, LNCS 5648, pp. 251 – 259, 2009.

attributes. As a result, it will be possible to evaluate whether the PDQM attributes are still valid and to identify new relevant attributes.

This paper is organized as follows. In Section 2, the SLR process, including the planning and conduction phases, is presented. The main results obtained from the SLR are reported in Section 3. Finally, our conclusions and future works are outlined in Section 4.

2 Review Process

This section details the activities performed in each of the two main phases of the procedure for performing an SLR, as proposed by [4]: "Planning the review" and "Conducting the review".

a) Planning the review: The most important pre-review activities are identified by the research questions(s) that the systematic review will address, and by producing a review protocol (i.e. plan) which defines the basic review procedures. In this phase, the following steps have been carried out:

1. Identification of the need for a review: The SLR has been planned in an attempt to identify the most important attributes related to Web data quality. Therefore, the main goal of our SLR is to discover the state-of-the-art in data quality for the Web since 2006.

2. Specifying the research questions. The following research questions (see Table 1) guided the design of the review process.

Table 1. Research Questions

Research Questions	Main Motivation
RQ1: "Which Web data quality attributes are addressed by researchers?"	To identify the Web data quality attributes which have been researched in the Web context
RQ2: "From which point of view is the Web data quality analyzed?"	To discover whether the Web data quality is from the manager's, programmer's or consumer's/user's perspective
RQ3: "In what context is the Web data quality evaluated?"	To identify whether the work is focused on the Web in general, a Web site or a Web portal
RQ4: "Is a quality model defined?"	To evaluate whether a set of attributes and the relationships between them have been defined
RQ5:"Do any measures for Web data quality exist?"	To discover whether measures are defined
RQ6: "Does a tool which supports the proposed approach exist?"	To determine whether a tool that assesses Web data quality exists.

3. Developing a review protocol. The development of the review protocol is the most relevant activity of the review process, since it establishes the basis of the search.

Source selection. The planned list of sources with which to carry out the SLR review was: 1.- Digital libraries; 2.- Grey literature comprises some papers considered to be relevant by experts which were not included in the aforementioned digital sources, and 3.- The manual revision of the Conference Proceeding of WISE (Web Information Systems Engineering) and ICWE (International Conference on Web Engineering) of 2006 and 2007.

Search string. The following strings were defined: 1) "data quality" AND web; 2) "information quality" AND web; 3) ("data quality" and web) and ("information quality"); 4) web and ("information quality" or "data quality"); 5) ("data quality" and web) AND (aspect OR dimension OR characteristic OR factor OR criterion OR criteria OR attribute OR model); 6) ("information quality" and web) AND (aspect OR dimension OR characteristic OR factor OR criterion OR criteria OR attribute OR model). The results obtained by using search strings 3, 4, 5 and 6, are very similar in the majority of cases. This signifies that the terms "Data Quality" and "Information Quality" are used interchangeably in literature.

Inclusion and Exclusion criteria. The inclusion criteria defined for this review were papers that present approaches or proposals that: a)were written in English, b)were published after 2005, c)identified a set of data quality characteristics, attributes or measures. And the following were defined as exclusion criteria: a) the work is previous to 2006, b) the paper is not within the scope of data quality in the Web, c) the paper does not propose a data quality attribute, or is not relevant, d) it does not contain the terms 'data quality' or 'information quality' either in the title or in the abstract e) studies are only available in the form of abstracts or Powerpoint presentations, f) duplicate studies, g) quality is not a part of the contributions of the paper.

b) Conducting the review: Once the protocol has been agreed, the review can begin. In this phase, the following steps have been made:

1. Selection of primary studies. The search process was completed on 31/12/2008 in the digital libraries previously mentioned, and 4105 papers were found. Many of the papers were eliminated owing to the fact that the use of different search strings in the digital libraries had caused them to be duplicated. Once these papers had been discarded, 1332 papers remained. These papers were then analyzed. This was done by first analysing the title and the abstract, and a total of 173 papers were selected. The full texts were then read, and once the inclusion and exclusion criteria had been applied, 69 papers were obtained.

2. Data extraction and monitoring. Once the primary studies had been chosen, the extraction of the relevant information for this SLR was stored in a data extraction form which was structured as follows: a) Data of the paper, including the search engine, title, year, type of publication and authors; b) Data of the classification, considering the following dimensions: quality attributes, point of view, context, application domain, quality model, measure and tool.

3 Results

This section provides an in depth presentation of the "Reporting the review" phase. For the purpose of our analysis, the papers were classified in order to answer the research questions listed in Table 1.

Our classification will hereafter be used in this section to present the answers to each research question.

RQ1: "Which Web data quality attributes are addressed by researchers?"
A total of 130 attributes were initially obtained. Bearing in mind that our objective is to select the set of most relevant attributes, the attributes which did not contain descriptions were first discarded. 20 attributes were thus eliminated. Next, we

analyzed the attributes specific to Web Site or Web that could be applicable for Web portals. 63 attributes were defined for data quality in the Web or in a Web site. However, after studying their applicability to the Web portal context only 22 attributes were selected. We then analyzed the definition of the attributes and detected that there were different names which were related to the same concept.

A total of 39 attributes related to data quality attributes were eventually selected (see Table 2).

Table 2. Attributes by origin

Attributes not obtained in our SLR (for authors other than those of [2])	Attributes obtained in our SLR also included in [2]			New attributes obtained in our SLR and not in [2]
Customer Support	Accessibility	Consistent Representation	Reliability	Effectiveness
Documentation	Accuracy	Currency	Reputation	Efficiency
Duplicates	Amount of data	Ease of operation	Security	Readability
Expiration	Applicability	Interactivity	Specialization	Usability
Flexibility	Attractiveness	Interpretability	Timeliness	Usefulness
Response Time	Availability	Novelty	Understandability	Verifiability
Traceability	Believability	Objectivity	Validity	
	Completeness	Organization	Value-added	
	Concise Representation	Relevancy		

As was previously stated, in [2] the best attributes for Web portals were selected, but this selection was made in 2005. Therefore, we wish to make a comparison between the PDQM attributes and the attributes detected in our SLR. The main findings of this comparison are shown in Table 2. Note that there is a set of attributes which was only selected in [2]. These attributes are presented in the first column. The second column shows the attributes which were chosen both in this paper and in PDQM. Finally, the last column presents 6 attributes which have been added in this paper as a result of the SLR.

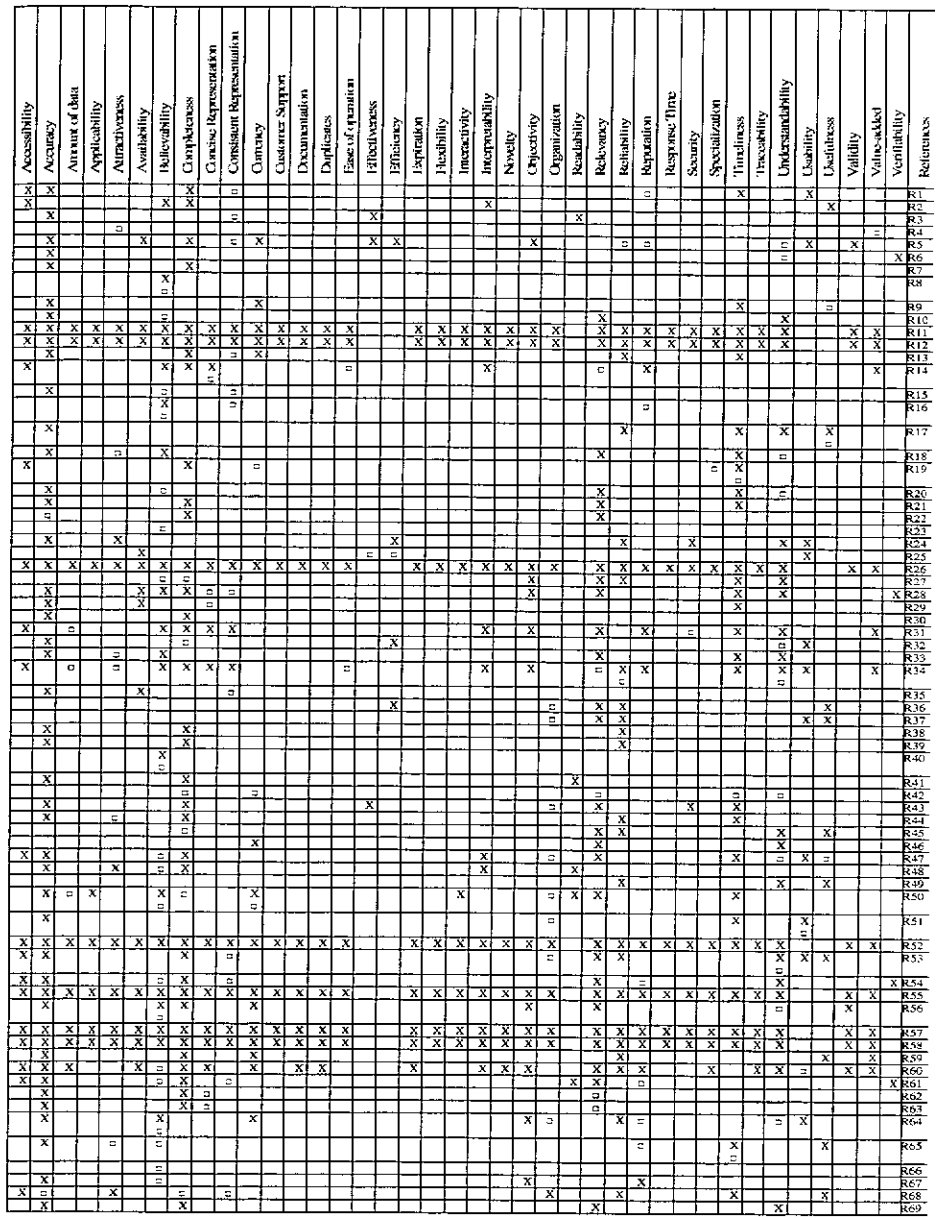
Table 3 presents the data quality attributes. The table shows the selected papers. The attributes which appear in the papers under the same name have been marked with an "X", and those which have a different name but explain the same concept have been marked with a "□". References are detailed in Appendix Table I.

RQ2: "From what point of view is the Web data quality analyzed?"
In this paper, we consider that data quality can be analyzed from three points of view: manager, programmer and consumer. Fig 1 summarizes the number of papers for each point of view.

We can deduce that the majority of the papers are related to the consumer or user. This signifies that researchers are more concerned about the quality of the data with which the consumer is provided. We can therefore affirm that it is necessary to make an effort to study the data quality from the other perspectives, since all the perspectives are obviously related.

RQ3: "In which context the Web data quality is evaluated?"
As was previously stated, in order to attain a wide knowledge of data quality attributes the following contexts were included: Web portal, Web site and the Web in general. However, each attribute has been reviewed, and only those which are of interest in the context of Web portal data quality have been selected. Fig 2 illustrates the classification of papers according to their context. It must be stressed that Web sites are studied more frequently than Web portals.

Table 3. Quality attributes investigated in the context of SPL



RQ4: "Is a quality model defined?"
With regard to whether a quality model is proposed, as Fig 3 shows, 60% of the works defined a quality model. This means that there is a tendency towards covering

all the aspects related to quality, and that the works are not limited to the simple definition of attributes.

RQ5: "Do measures for Web data quality exist?"
Only 22% of the papers do not include measures, as is shown in Fig 4. Therefore, unlike the situation of some years ago, researchers have now realized the importance of measurement and almost all the proposals define measures with which to assess the data quality level. We believe that this is a very positive aspect, since without measures it is obviously not possible to evaluate quality.

RQ6: "Does a tool with which to support the proposed approach exist?"
The proposal was considered to contain a tool when the authors affirmed that a new tool had been created or when one or several existing tools could support their proposal. Only 8 of the 69 selected papers provided a support tool, which represents 12% of the total.

In conclusion, we can state that although the majority of the proposals define measures (as was mentioned in the previous section), their assessment it not automated. This reveals the difficulty of automating the proposed measures. Hence, as a future work it will be necessary to work on the automation of measure assessment.

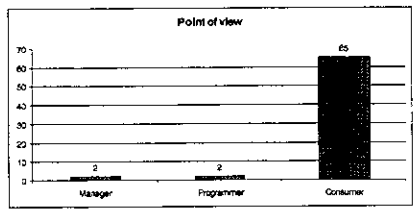


Fig. 1. Papers according to whose point of view they are directed towards

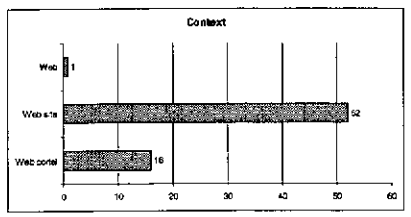


Fig. 2. Papers according to their context

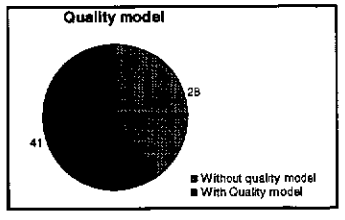


Fig. 3. Proposals with a quality model

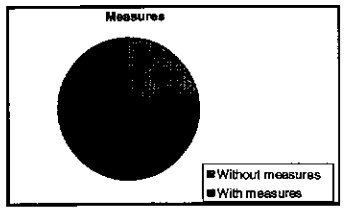


Fig. 4. Proposals which define measures

4 Conclusions and Future Works

In this paper, a systematic literature review has been carried out in order to obtain the portal data quality attributes that have been proposed in literature.

Moreover, the realization of this SLR has led us to certain conclusions. Firstly, the majority of the papers study Web data quality from the consumer's perspective, and more effort should therefore be made to study this from the other perspectives. Secondly, it should be noted that Web sites are studied more frequently than Web portals. Thirdly, more than half the proposals define a quality model. This means that there is a tendency towards covering all the aspects related to quality and that the works are not limited to the simple definition of attributes. Fourthly, it is also of interest to stress that researchers have realized the importance of measurement, and that almost all the proposals define measures with which to assess the data quality level.

Without measures it is obviously not possible to evaluate quality. However, these measures are not easy to calculate automatically since only 12% of the proposals have developed a tool for their assessment.

Finally, in comparison with the previous work [2], we conclude that a greater number of papers related to data quality attributes were selected, and a greater number of attributes were identified. We have therefore included all the attributes of [2] and have added other attributes detected in this SLR which we consider to be relevant to our study.

In the future we shall compare both the data quality attributes obtained and SQUARE [5]. Since some PDQM attributes are now obsolete and are included in this work, it is necessary to review and analyze the possibility of discarding them. In order to do this, we shall first study the data quality attributes of SQUARE [5], and shall then analyze both the attributes obtained in this SLR and the PDQM attributes, which will eventually be compared in order to select the most relevant attributes.

References

1. Domingues, M.A., Soares, C., Jorge, A.M.: A Web-Based System to Monitor the Quality of Meta-Data in Web Portals. In: IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology (WI-IATW 2006), pp. 188–191 (2006)
2. Caro, A., Calero, C., Caballero, I., Piattini, M.: A proposal for a set of attributes relevant for Web portal data quality. Software Quality Journal 16, 513–542 (2008)
3. Caro, A., Calero, C., Caballero, I., Piattini, M.: Defining a data quality model for web portals. In: Aberer, K., Peng, Z., Rundensteiner, E.A., Zhang, Y., Li, X. (eds.) WISE 2006. LNCS, vol. 4255, pp. 363–374. Springer, Heidelberg (2006)
4. Kitchenham, B., Charters, S.: Guidelines for performing systematic literature reviews in software engineering. Technical Report EBSE-2007-01, School of Computer Science and Mathematics, Keely University (2007)
5. [ISO/IEC-FDIS-25012]: Software engineering - Software product Quality Requirements and Evaluation (SQuaRE) - Data quality model (2008)

Appendix

Table I. References of the SLR

ACM
R1: Nichols, D.M., Chan, C., Bainbridge, D., McKay, D., Twidale, M.B.: A lightweight metadata quality tool. In: JCDL '08:Proceedings of the 8th ACM/IEEE-CS joint conference on Digital libraries. (2008)
R2: Gelman, I., A.Barletta, A.L.: A "quick and dirty" website data quality indicator. (2008)
R3: Scaffidi, C., Myers, B., Shaw, M.: Topes: reusable abstractions for validating data. In: ICSE '08: Proceedings of the 30th international conference on Software engineering. (2008)
R4: Caro, A., Calero, C., Caballero, I., Piattini, M.: Defining a quality model for portal data. In: ICWE '06: Proceedings of the 6th international conference on Web engineering. (2006)
R5: Costa, C.J., Nhampossa, J.L., Aparicio, M.: Wiki content evaluation framework. In: SIGDOC '08: Proceedings of the 26th annual ACM international conference on Design of communication. (2008)
R6: Kittur, A.R.E. K.: Harnessing the wisdom of crowds in wikipedia: quality through coordination. In: CSCW '08: Proceedings of the ACM 2008 conference on Computer supported cooperative work. (2008)
R7: Wilkinson, D.M., Huberman, B.A.: Cooperation and quality in wikipedia. In: WikiSym '07: Proceedings of the 2007 international symposium on Wikis. (2007)
R8: McKnight, D.H.Kacmar, C.J.: Factors and effects of information credibility. In: ICEC '07: Proceedings of the ninth international conference on Electronic commerce. (2007)
R9: Rodgers, W., Negash, S.: The effects of web-based technologies on knowledge transfer. Communications of the ACM. 50, 117-122 (2007)
R10: Belanger, F., Fan, W., Schaupp, L.C., Krishen, A., Everhart, J., Poteet, D.Nakamoto, K.: Web site success metrics: addressing the duality of goals. Communications of the ACM. 49 (2006)
IEEE
R11: Caro, A., Calero, C., Enriquez de Salamanca, J., Piattini, M.: Refinement of a Tool to Assess the Data Quality in Web Portals. In: Seventh International Conference on Quality Software (QSIC 2007). (2007)
R12: Caro, A., Calero, C., Mendes, E., Piattini, M.: A Probabilistic Approach to Web Portal's Data Quality Evaluation. In: 6th International Conference on the Quality of Information and Communications Technology (QUATIC 2007). (2007)
R13: Vorochek, O., Biletskiy, Y.: Toward Assessing Data Quality of Ontology Matching on the Web. In: Fifth Annual Conference on Communication Networks and Services Research (CNSR '07). (2007)
R14: Domingues, M.A., Soares, C., Jorge, A.M.: A Web-Based System to Monitor the Quality of Meta-Data in Web Portals. In: 2006 IEEE/WIC/ACM International Conferences on Web Intelligence and Intelligent Agent Technology - Workshops. (2006)
R15: Yin, X., Han, J.Yu, P.S.: Truth Discovery with Multiple Conflicting Information Providers on the Web. IEEE Transactions on Knowledge and Data Engineering. 20, 796-808 (2008)
R16: Prat, N., Madnick, S.: Measuring Data Believability: A Provenance Approach. In: Proceedings of the 41st Annual Hawaii International Conference on System Sciences (HICSS 2008). (2008)
R17: Hadaya, P., Ethier, J.: Online Purchasing of Simple Retail Goods: The Impact of e-Service Quality as Provided by Electronic Commerce Functionalities. In: Proceedings of the 41st Annual Hawaii International Conference on System Sciences (HICSS 2008). (2008)
R18: Tate, M., Evermann, J., Hope, B., Barnes, S.: Perceived Service Quality in a University Web Portal: Revising the E-Qual Instrument. In: 40th Annual Hawaii International Conference on System Sciences (HICSS'07). (2007)
R19: Prestipino, M., Aschoff, F.-R., Schwabe, G.: How up-to-date are Online Tourism Communities? An Empirical Evaluation of Commercial and Non-commercial Information Quality. In: 40th Annual Hawaii International Conference on System Sciences (HICSS'07). (2007)
R20: Rabjohn, N., Cheung, C.M.K.Lee, M.K.O.: Examining the Perceived Credibility of Online Opinions: Information Adoption in the Online Environment. In: Proceedings of the 41st Annual Hawaii International Conference on System Sciences (HICSS 2008). (2008)
R21: Lin, H., Fan, W., Wallace, L., Zhang, Z.: An Empirical Study of Web-Based Knowledge Community Success. In: 40th Annual Hawaii International Conference on System Sciences (HICSS'07). (2007)
R22: Schaupp, L.C., Fan, W., Belanger, F.: Determining Success for Different Website Goals. In: Proceedings of the 39th Annual Hawaii International Conference on System Sciences (HICSS'06) Track 6. (2006)
R23: McKnight, H., Kacmar, C.: Factors of Information Credibility for an Internet Advice Site. In: Proceedings of the 39th Annual Hawaii International Conference on System Sciences (HICSS'06) Track 6. (2006)
R24: Franch, X., Quer, C., Canton, J.A., Salietti, R.: Experience Report on the Construction of Quality Models for Some Content Management Software Domains. In: Seventh International Conference on Composition-Based Software Systems (ICCBSS 2008). (2008)
OTHER
R25: Mich, L.: Subjectivity in Web site quality evaluation: the contribution of Soft Computing. In: Workshops of 7th International Conference on Web Engineering. (2007)
R26: Caro, A., Calero, C., Sahraoui, H.A., Piattini, M.: A Bayesian Network to represent a data quality model. International Journal Information Quality. 1, 272-294 (2007)
SCIENCE DIRECT
R27: Lee, J., Park, D.-H., Han, I.: The effect of negative online consumer reviews on product attitude: An information processing view. Electronic Commerce Research and Applications. 7, 341-352 (2008)
R28: Bizer, C., Cyganiak, R.: Quality-driven information filtering using the WIQA policy framework. Web Semantics: Science, Services and Agents on the World Wide Web. 7, 1-10 (2009)
R29: Prybutok, V.R., Zhang, X., Ryan, S.D.: Evaluating leadership, IT quality, and net benefits in an e-government environment. Information & Management. 45, 143-152 (2008)
R30: Gonçalves, M.A., Moreira, B.L., Fox, E.A., Watson, L.T.: "What is a good digital library?" – A quality model for digital libraries. Information Processing & Management. 43, 1416-1437 (2007)
R31: Herrera-Viedma, E., Peis, E., Morales-del-Castillo, J.M., Alonso, S., Anaya, K.: A fuzzy linguistic model to evaluate the quality of Web sites that store XML documents. International Journal of Approximate Reasoning. 46, 226-253 (2007)
R32: Chung, W., Bonillas, A., Lai, G., Xi, W., Chen, H.: Supporting non-English Web searching: An experiment on the Spanish business and the Arabic medical intelligence portals. Decision Support Systems. 42, 1697-1714 (2006)
R33: Barnes, S.J., Vidgen, R.T.: Data triangulation and web quality metrics: A case study in e-government. Information & Management. 43, 767-777 (2006)
R34: Chung, W.: Studying information seeking on the non-English Web: An experiment on a Spanish business Web portal. International Journal of Human-Computer Studies. 64, 811-829 (2006)
R35: Chen, C.-C., Wu, C.-S., Wu, R.C.-F.: e-Service enhancement priority matrix: The case of an IC foundry company. Information & Management. 43, 572-586 (2006)
R36: Grigoroudis, E., Litos, C., Moustakis, V.A., Politis, Y., Tsironis, L.: The assessment of user-perceived web quality: Application of a satisfaction benchmarking approach. European Journal of Operational Research. 187, 1346-1357 (2008)
R37: Tsakonas, G., Papatheodorou, C.: Exploring usefulness and usability in the evaluation of open access digital libraries. Information Processing & Management. 44, 1234-1250 (2008)
R38: Kim, D.J., Steinfield, C., Lai, Y.-J.: Revisiting the role of web assurance seals in business-to-consumer electronic commerce. Decision Support Systems. 44, 1000-1015 (2008)
R39: Kim, D.J., Ferrin, D.L., Rao, H.R.: A trust-based consumer decision-making model in electronic commerce: The role of trust, perceived risk, and their antecedents. Decision Support Systems. 44, 544-564 (2008)
R40: Robins, D., Holmes, J.: Aesthetics and credibility in web site design. Information Processing & Management. 44, 386-399 (2008)
R41: Wang, Y., Liu, Z.: Automatic detecting indicators for quality of health information on the Web. International Journal of Medical Informatics. 76, 575-582 (2007)
R42: Wang, Y.-S., Wang, H.-Y., Shee, D.Y.: Measuring e-learning systems success in an organizational context: Scale development and validation. Computers in Human Behavior. 23, 1792-1808 (2007)
R43: Yen, B., P.J.-H., H., Wang, M.: Toward an analytical approach for effective Web site design: A framework for modeling, evaluation and enhancement. Electronic Commerce Research and Applications. 6, 159-170 (2007)
R44: Ahn, T., Ryu, S.J., H.: The impact of Web quality and playfulness on user acceptance of online retailing. Information & Management. 44, 263-275 (2007)
R45: Song, J., Jones, D., Gudigantala, N.: The effects of incorporating compensatory choice strategies in Web-based consumer decision support systems. Decision Support Systems. 43, 359-374 (2007)
R46: Lee, Y., .Kozar, K.A.: Investigating the effect of website quality on e-business success: An analytic hierarchy process (AHP) approach. Decision Support Systems. 42, 1383-1401 (2006)
R47: Kang, Y.-S., Kim, Y.J.: Do visitors' interest level and perceived quantity of web page content matter in shaping the attitude toward a web site? Decision Support Systems. 42, 1187-1202 (2006)
R48: Sillence, E., Briggs, P., Harris, P., Fishwick, L.: A framework for understanding trust factors in web-based health advice. International Journal of Human-Computer Studies. 64, 697-713 (2006)

Table I. (*continued*)

R49: Ethier, J., Hadaya, P., Talbot, J., Cadieux, J.: B2C web site quality and emotions during online shopping episodes: An empirical study. Information & Management. 43, 627-639 (2006)
R50: De Wulf, K., Schillewaert, N., Muylle, S., Rangarajan, D.: The role of pleasure in web site success. Information & Management. 43, 434-446 (2006)
R51: Zviran, M., Glezer, C., Avni, I.: User satisfaction from commercial web sites: The effect of design and use. Information & Management. 43, 157-178 (2006)
SCOPUS
R52: Calero, C., Caro, A., Piattini, M.: An applicable data quality model for web portal data consumers. World Wide Web. 11, 465-484 (2008)
R53: Cheung, C.M.K., Lee, M.K.O.: The structure of web-based information systems satisfaction: Testing of competing models. Journal of the American Society for Information Science and Technology. 59, 1617-1630 (2008)
R54: Sivilia, B., Twidale, M.B., Smith, L.C., Gasser, L.: Information quality work organization in Wikipedia. Journal of the American Society for Information Science and Technology. 59, 983-1001 (2008)
R55: Caro, A., Calero, C., Caballero, I., Piattini, M.: A proposal for a set of attributes relevant for Web portal data quality. Software Quality Journal. 16, 513-542 (2008)
R56: Kelton, K., Fleischmann, K.R., Wallace, W.A.: Trust in digital information. Journal of the American Society for Information Science and Technology. 59, 363-374 (2008)
R57: Caro, A., Calero, C., Piattini, M.: Development process of the operational version of PDQM. Lecture Notes in Computer Science. LNCS 4831, 436-448 (2007)
R58: Caro, A., Calero, C., Caballero, I., Piattini, M.: Defining a data quality model for web portals. 7th International Conference on Web Information System Engineering (WISE 2006) Lecture Notes in Computer Science. Vol. 4255, 363-374 (2006)
R59: Moraga, A., Calero, C., Piattini, M.: Comparing different quality models for portals. Online Information Review. 30, 555-568 (2006)
R60: Caro, A., Calero, C., Caballero, I., Piattini, M.: A first approach to a data quality model for web portals. Lecture Notes in Computer Science. LNCS 3982, 984-993 (2006)
R61: Ehmann, K., Large, A., Beheshti, J.: Collaboration in context: Comparing article evolution among subject disciplines in Wikipedia. First Monday. 13 (2008)
R62: Verbert, K., Ochoa, X., Duval, E.: The ALOCOM framework: Towards scalable content reuse. Journal of Digital Information. 9 (2008)
R63: Verbert, K., Duval, E.: Evaluating the ALOCOM approach for scalable content repurposing. Lecture Notes in Computer Science. LNCS 4443, 325-336 (2007)
R64: Metzger, M.J.: Making sense of credibility on the web: Models for evaluating online information and recommendations for future research. Journal of the American Society for Information Science and Technology. 58, 2078-2091 (2007)
R65: Dondio, P., Barrett, S.: Computational trust in web content quality: A comparative evaluation on the Wikipedia project. Informatica. 31, 151-160 (2007)
R66: Hong, T.: The influence of structural and message features on web site credibility. Journal of the American Society for Information Science and Technology. 57, 114-127 (2006)
WILEY
R67: Herrera-Viedma, E., Pasi, G., Lopez-Herrera, A.G., Porcel, C.: Evaluating the information quality of Web sites: A methodology based on fuzzy computing with words. Journal of the American Society for Information Science and Technology. 57, 538-549 (2006)
R68: Katerattanakul, P., Siau, K.: Factors affecting the information quality of personal Web portfolios. Journal of the American Society for Information Science and Technology. 59, 63-76 (2008)
R69: Chiu, C.-M., Chiu, C.-S., Chang, H.-C.: Examining the integrated influence of fairness and quality on learners' satisfaction and Web-based learning continuance intention. Information Systems Journal. 17, 271-287 (2007)