Anne Banks Pidduck
John Mylopoulos
Carson C. Woo
M. Tamer Ozsu (Eds.)

LNCS 2348

Advanced Information Systems Engineering

14th International Conference, CAiSE 2002
Toronto, Canada, May 2002
Proceedings

Springer

Anne Banks Pidduck John Mylopoulos
Carson C. Woo M. Tamer Ozsu (Eds.)

Advanced Information Systems Engineering

14th International Conference, CAiSE 2002
Toronto, Canada, May 27-31, 2002
Proceedings

Springer

Series Editors

Gerhard Goos, Karlsruhe University, Germany
Juris Hartmanis, Cornell University, NY, USA
Jan van Leeuwen, Utrecht University, The Netherlands

Volume Editors

Anne Banks Pidduck
M. Tamer Ozsu
University of Waterloo, School of Computer Science
200 University Avenue West, Waterloo, Ontario, Canada, N2L 3G1
E-mail: {apidduck/tozsu}@uwaterloo.ca

John Mylopoulos
University of Toronto, Pratt Building
6 King's College Road, Toronto, Ontario, Canada, M5S 3H5
E-mail: jm@cs.toronto.edu

Carson C. Woo
University of British Columbia, Faculty of Commerce and Business Administration
2053 Main Mall, Vancouver, B.C. Canada V6T 1Z2
E-mail: Carson.Woo@ubc.ca

Cataloging-in-Publication Data applied for

Die Deutsche Bibliothek - CIP-Einheitsaufnahme

Advanced information systems engineering : 14th international conference ; proceedings / CAiSE 2002, Toronto, Canada, May 27 - 31, 2002. Anne Banks Pidduck ... (ed.). - Berlin ; Heidelberg ; New York ; Barcelona ; Hong Kong ; London ; Milan ; Paris ; Tokyo : Springer, 2002
(Lecture notes in computer science ; Vol. 2348)
ISBN 3-540-43738-X

CR Subject Classification (1998): H.2, H.4-5, H.3, J.1, K.4.3-4, K.6, D.2, I.2.11

ISSN 0302-9743
ISBN 3-540-43738-X Springer-Verlag Berlin Heidelberg New York

Springer-Verlag Berlin Heidelberg New York
a member of BertelsmannSpringer Science+Business Media GmbH

http://www.springer.de

Printed in Germany

Typesetting: Camera-ready by author, data conversion by DA-TeX Gerd Blumenstein
Printed on acid-free paper SPIN 10870009 06/3142 5 4 3 2 1 0

Preface

The explosive growth of the Internet and the Web have created an ever-growing demand for information systems, and ever-growing challenges for Information Systems Engineering. The series of Conferences on Advanced Information Systems Engineering (CAiSE) was launched in Scandinavia by Janis Bubenko and Arne Solvberg in 1989, became an important European conference, and was held annually in major European sites throughout the 1990s. Now, in its 14th year, CAiSE was held for the first time outside Europe, showcasing international research on information systems and their engineering.

Not surprisingly, this year the conference enjoyed unprecedented attention. In total, the conference received 173 paper submissions, the highest number ever for a CAiSE conference. Of those, 42 were accepted as regular papers and 26 as short (poster) papers. In addition, the conference received 12 proposals for workshops of which 8 were approved, while 4 tutorials were selected from 15 submissions.

The technical program was put together by an international committee of 81 experts. In total, 505 reviews were submitted, with every member of the committee contributing. Decisions on all submissions were reached at a program committee meeting in Toronto on January 26-27,2002. Workshop and tutorial proposals were handled separately by committees chaired by Patrick Martin (workshops), and Jarek Gryz and Richard Paige (tutorials).

We wish to extend a great "THANK YOU!" to all members of the program and organizing committees for their volunteer contributions of time and expertise. The fact that so many busy (and famous!) people took the trouble to help us with the organization of this conference and the formation of its technical program speaks well for the future of CAiSE and the field of Information Systems Engineering.

May 2002

Anne Banks Pidduck
John Mylopoulos
Carson Woo
Tamer Özsu

Organization

CAiSE 2002 was organized by the Department of Computer Science, University of Waterloo, and the Department of Computer Science, University of Toronto.

Organizing Committee

General Chair:	Tamer Özsu (University of Waterloo, Canada)
Program Committee Chairs:	John Mylopoulos (University of Toronto)
	Carson Woo (University of British Columbia)
Panel and Tutorial Chairs:	Jarek Gryz (York University, Canada)
	Rich Paige (York University, Canada)
Local and Publicity Chair:	Manuel Kolp (University of Louvain, Belgium)
Workshop and Poster Chair:	Patrick Martin (Queen's University, Kingston, Canada)
Proceedings Chair:	Anne Banks Pidduck (University of Waterloo)

Local Arrangements Committee

Luiz-Marcio Cysneiros
Linda Liu
Patricia Rodriguez-Gianolli
Manuel Kolp (Chair)
Jianguo Lu

Donors

IBM Center for Advanced Studies
Institute for Computer Research
Bell University Laboratories

Academic Supporters

University of Waterloo, Department of Computer Science
University of Toronto, Department of Computer Science
University of Louvain, IAG School of Management
University of British Columbia, Commerce and Business Administration
York University, Department of Computer Science
Queen's University, Kingston, Department of Computing and Information Science

Program Committee

Ken Barker (Canada)
Daniel Berry (Canada)
Alex Borgida (USA)
Paul Bowen (USA)
Sjaak Brinkkemper (The Netherlands)
Michael Brodie (USA)
Christoph Bussler (USA)
Silvana Castano (Italy)
Jan Dietz (The Netherlands)
Johan Eder (Austria)
Marcel Franckson (France)
Mariagrazia Fugini (Italy)
Antonio Furtado (Brazil)
Andreas Geppert (Switzerland)
Paolo Giorgini (Italy)
Martin Glinz (Switzerland)
Sol Greenspan (USA)
Alan Hevner (USA)
Juhani Iivari (Finland)
Sushil Jajodia (USA)
Keith Jeffery (UK)
Manfred Jeusfeld (The Netherlands)
Hermann Kaindl (Austria)
Hannu Kangassalo (Finland)
Gerti Kappel (Austria)
Steven Kimbrough (USA)
Kostas Kontogiannis (Canada)
Manolis Koubarakis (Greece)
Alberto Laender (Brazil)
Julio Leite (Brazil)
Maurizio Lenzerini (Italy)
Michel Leonard (Switzerland)
Chin Li (China)
Marin Litoiu (Canada)
Peri Loucopoulos (UK)
Kalle Lyytinen (Finland)
Neil Maiden (UK)
Sal March (USA)
Pat Martin (Canada)
Nancy Mead (USA)
Misha Missikoff (Italy)
David Monarchi (USA)
Hausi Müller (Canada)
Oscar Nierstrasz (Switzerland)
Andreas Oberweis (Germany)
Antoni Olivé (Spain)
Andreas Opdahl (Norway)
Maria Orlowska (Australia)
Maurizio Panti (Italy)
Mike Papazoglou (The Netherlands)
Jeff Parsons (Canada)
Oscar Pastor (Spain)
Barbara Pernici (Italy)
Alain Pirotte (Belgium)
Dimitris Plexousakis (Greece)
Klaus Pohl (Germany)
Radhakrishnan.T (Canada)
Sudha Ram (USA)
Bill Robinson (USA)
Colette Rolland (France)
Michel Scholl (France)
Michael Schrefl (Austria)
Timos Sellis (Greece)
Mike Shaw (USA)
Olivia Sheng (USA)
Keng Siau (USA)
Jacob Slonim (Canada)
Arne Solvberg (Norway)
Veda Storey (USA)
Eleni Stroulia (Canada)
Alistair Sutcliffe (UK)
Mohan Tanniru (USA)
Bernhard Thalheim (Germany)
Costantino Thanos (Italy)
Aphrodite Tsalgatidou (Greece)
Yannis Vassiliou (Greece)
Yair Wand (Canada)
Benkt Wangler (Sweden)
Tony Wasserman (USA)
Christopher Welty (USA)
Roel Wieringa (The Netherlands)
Eric Yu (Canada)

Tutorials

Fact-Oriented Modeling
Presenter: Terry Halpin, Microsoft Corporation

The Application of Workflow Management Technology in Process-Based B2B Integration
Presenter: Christoph Bussler, Oracle Corporation

The Web-Services Phenomenon: Concepts, Technologies, Current Trends, and Research Directions
Presenter: Michael P. Papazoglou, Tilburg University, The Netherlands

Publish/Subscribe: Applications, Concepts, and Algorithms
Presenter: H-Arno Jacobsen, University of Toronto, Canada

Workshops

Agent-Oriented Information Systems (AOIS 2002)
Organizers: P. Giorgini, Y. Lesperance, G. Wagner, E. Yu

Goal-Oriented Business Process Modeling (GBPM 2002)
Organizers: I. Bider, P. Johannesson

2nd Workshop on Data Integration over the Web (DIWeb 2002)
Organizers: Z. Bellahsene, Z. Lacroix

Workshop on E-Services and E-Business Technology
Organizers: M. Orlowska, C. Bussler, J. Yang, B. Pernici

Design and Management of Data Warehouses (DMDW 2002)
Organizers: J. Hammer, M. Jeusfeld, I. Song, M. Staudt, D. Theodoratos

Evaluation and Modeling Methods in Systems Analysis and Design (EMMSAD 2002)
Organizers: T. Halpin, K, Siau, J. Krogstie

E-Services and the Semantic Web (ESSW 2002)
Organizers: R. Hull, S. McIlraith

Doctoral Consortium
Organizers: A. Hinze, J. Kotlarsky

Table of Contents

Invited Presentations

Developing Web Applications

Knowledge Management

Deployment Issues

Semantics of Information

System Qualities

Integration Issues

Analysis and Adaptation

Retrieval and Performance

Requirements Issues

Schema Matching and Evolution

Workflows

Semantics and Logical Representations

Understanding and Using Methods

Modeling Objects and Relationships

Short Papers

Defining and Validating Measures for Conceptual Data Model Quality

Marcela Genero[1], Geert Poels[2], and Mario Piattini[1]

[1]ALARCOS Research Group, Department of Computer Science
University of Castilla-La Mancha
Paseo de la Universidad, 4 - 13071 - Ciudad Real (Spain)
{mgenero,mpiattin}@inf-cr.uclm.es
[2]Dept. Business Administration, University of Sciences and Arts
Koningsstraat, 336, 1030 Brussels (Belgium)
gpoels@vlekho.wenk.be

Abstract. For assessing conceptual data model quality it is useful to have quantitative and objective measurement instruments. The scarcity of such measurement instruments leaded us to define a set of measures for structural complexity, an internal quality attribute of conceptual data models, with the idea that it is related to the external quality of such models. In order to gather empirical evidence that the proposed measures could be early quality indicators of conceptual data models, we carried out a controlled experiment. The aim of the experiment was to investigate the relation between the structural complexity of conceptual data models and maintainability sub-characteristics such as understandability and modifiability.

1 Introduction

Given that conceptual data models lay the foundation of all later design work their quality has a significant impact on the quality of the database which is ultimately implemented. However, before evaluating and if necessary improving the quality of a conceptual data model, it is necessary to assess it in an objective way. It is in this context that measurement can help database designers to make better decisions during design activities. Even though several quality frameworks for conceptual data models have been proposed, most of them lack valid quantitative measures to evaluate the quality of conceptual data models in an objective way. Papers referring to measures for conceptual data models are scarce. Kesh [4] and Moody [5] have proposed some measures to measure different quality characteristics of ERDs, but their utility in practice has not been demonstrated.

The scarcity of measures that are well-defined and also theoretically and empirically validated leaded us to define a set of measures to quantify various aspects related to one particular, but highly important internal quality attribute of conceptual data models, i.e. their structural complexity. As Whitmire remarked [7] complexity is

A. Banks Pidduck et al. (Eds.): CAiSE 2002, LNCS 2348, pp. 724-727, 2002.

believed to be an indicator of external quality attributes such as understandability, modifiability, etc., but the empirical evidence supporting these relationships is scarce and suspect.

In [3] we have defined and theoretically validated a set of measures for the structural complexity of Entity Relationship Diagrams (ERDs) (see table 1), following the DISTANCE framework [6].

Table 1. Measures for the structural complexity of ERDs

MEASURE	DEFINITION
NE	The total number of entities within an ERD.
NA	The total number of attributes within an ERD.
NDA	The total number of derived attributes within an ERD.
NCA	The total number of composite attributes within an ERD.
NMVA	The total number of multivalued attributes within an ERD.
NR	The total number of relationships within an ERD.
NM:NR	The total number of M:N relationships within an ERD.
N1:NR	The total number of 1:N relationships (including also 1:1 relationships) within an ERD.
NBinaryR	The total number of binary relationships within an ERD.
NIS_AR	The total number of IS_A relationships (generalisation/ specialisation) within an ERD. In this case, we consider one relationship for each child-parent pair within the IS_A relationship.
NRefR	defined as the total number of reflexive relationships within an ERD.
NRR	defined as the number of relationships that are redundant in an ERD.

The aim of this paper is to present a controlled experiment we carried out to gather empirical evidence that the proposed measures could be early quality indicators of ERDs.

2 A Controlled Experiment

The aim of the experiment is to investigate the relationship between the structural complexity of ERDs and two important components of maintainability: understandability and modifiability. The subjects were forty students enrolled in the third year of Computer Science in the Department of Computer Science at the University of Castilla-La Mancha in Spain. The experimental material consisted of a guide explaining the ER notation, and four ERDs (all the experimental material is available at http://alarcos.inf-cr.uclm.es). These diagrams are related to different universes of discourse that are general enough to be easily understood by each of the subjects. The structural complexity of each diagram is different, because the values of the measures are different for each diagram.

Each diagram had a test enclosed which includes two parts:

- Part 1. A questionnaire in order to evaluate if the subjects really understand the content of the ERD. Each questionnaire contained exactly the same number of questions (five) and the questions were conceptually similar and in identical order. Each subject had to write down the time spent answering the questionnaire, by recording the initial time and final time. The difference between the two is what we call the understandability time (expressed in minutes).

- Part 2. Two new requirements for the ERD. The subjects had to modify the ERD according to these new requirements, again writing down the initial time and the final time. The difference between these two times is what we called modifiability time, which includes both the time spent analysing what modifications had to be done and the time needed to perform them.

We selected a within-subject design experiment, i.e. all the tests (i.e. experimental tasks) had to be solved by each of the subjects. The subjects were given the tests in different order. We allowed one hour to do all the tests. Each subject had to work alone. In case of doubt, they could only consult the supervisor who organised the experiment. We collected all the tests controlling if they were complete and the responses were correct. We discarded the tests of 9 subjects, because they included an incorrect answer or a required modification that was done incorrectly. Therefore, we take into account the responses of 31 subjects.

To analyse the data, we first applied the Kolmogorov-Smirnov test to ascertain if the distribution of the data collected was normal or not. As the data was normal we decided to use a parametric test like Pearson's correlation coefficient, with a level of significance $\alpha = 0.05$, which means the level of confidence is 95% (i.e. the probability that we reject H_0 when H_0 is false is at least 95%, which is statistically acceptable). Using Pearson's correlation coefficient, each of the measures was correlated separately to the understandability and the modifiability time (see table 2).

Table 2. Pearson's correlation coefficients between the measures and the understandability and modifiability time (all values are significant)

	NE	NA	NR	NBinaryR	N1:RN	NM:NR
Understandability time	0.7168	0.5588	0.7168	0.7168	0.7168	0.7168
Modifiability time	0.7246	0.5508	0.7246	0.7246	0.7246	0.7246

3 Conclusions

Analysing the Pearson's correlation coefficients shown in table 2, we can conclude that there is a high correlation between the understandability time and the modifiability time and the measures NE, NA, NR, N1:NR, NM:NR, NBinaryR because the correlation coefficient is greater than 0.5, which is a common threshold to evaluate correlation values. Only the NA measure seems to be less correlated to the understandability and modifiability time than the other measures (though the correlation value is still greater than 0.5).

The results obtained in this experiment corroborate, at some extent, the results obtained in a previous similar experiment [2] and the results obtained in a case study using data extracted from 5 real projects [1].

In spite of this we are aware that it is necessary to replicate the experiment and to carry out new ones in order to confirm our results. Also it is necessary to apply these measures to data obtained from "real projects".

Acknowledgements

This research is part of the DOLMEN project supported by CICYT (TIC 2000-1673-C06-06).

References

1. Genero, M., Piattini, M., Calero, C.: An Approach To Evaluate The Complexity Of Conceptual Database Models. 3rd European Software Measurement Conference - FESMA 2000, Madrid (2000).
2. Genero, M., Olivas, J., Piattini M., Romero, F.: Knowledge Discovery For Predicting Entity Relationship Diagram Maintainability, SEKE 2001, Argentina, Proceedings, Knowledge Systems Institute, (2001) 203-211.
3. Genero, M.: Defining and Validating Metrics for Conceptual Models, Ph.D. thesis, University of Castilla-La Mancha. (2002).
4. Kesh, S.: Evaluating the Quality of Entity Relationship Models, Information and Software Technology, Vol. 37 Nº 12, (1995) 681-689.
5. Moody, L.: Metrics for Evaluating the Quality of Entity Relationship Models, Proceedings of the Seventeenth International Conference on Conceptual Modelling (E/R '98), Singapore, November 16-19, (1998) 213-225.
6. Poels, G., Dedene, G.: Distance-based software measurement: necessary and sufficient properties for software measures, Information and Software Technology, Vol. 42 Nº 1, (2000) 35-46.
7. Whitmire, S.: Object Oriented Design Measurement, John Wiley & Sons, Inc. (1997).