ASSE' 2000

Proceedings

Argentine Symposium on Software Engineering

29 JAIIO

29 Jornadas Argentinas de Informática e Investigación Operativa

SADIO
40 AÑOS
1960 - 2000

Sociedad Argentina de Informática e Investigación Operativa.

ASSE' 2000
Proceedings
Argentine Symposium on Software Engineering

Organizan:

SADIO

Sociedad Argentina de Informática e Investigación Operativa
SADIO - Uruguay 252, 2° "D"
C1015ABF - Buenos Aires - Argentina
Tel./fax: 54 (11) 4371-5755 / 4372-3950
E-mail: jaiio@sadio.org.ar
sadio@mbox.servicenet.com.ar
http://www.sadio.org.ar

ISISTAN
Instituto de Sistemas de Tandil
Facultad de Ciencias Exactas
Universidad Nacional del Centro

Tandil - Argentina
4 al 9 de Septiembre de 2000

29 JAIIO

29 Jornadas Argentinas de Informática e Investigación Operativa

Argentine Symposium on Software Engineering

Proceedings

Edited by

Claudia Marcos and Alvaro Ortigosa

ISISTAN, UNICEN, Tandil

Tandil, September 4-9, 2000

Organized by

SADIO, Sociedad Argentina de Informática e Investigación Operativa and

ISISTAN, Instituto de Sistemas Tandil

ASSE 2000

Chairs

Horacio Leone
Universidad Tecnológica Nacional
Fac. Reg. Santa Fe, Argentina

Alvaro Ortigosa
Universidad Nacional del Centro de
la Pcia. de Bs. As.
Instituto de Sistemas Tandil,
Argentina

Program Committee

Jorge Boria
TeraQuest Metrics Inc., USA

Paul Clements
Software Engineering Institute, CMU, USA

Mohamed Fayad
University of Nebraska, Lincoln, USA

Miguel Felder
Universidad de Buenos Aires, Argentina

Silvia Gordillo
Lifia, Universidad Nac. de La Plata, Argentina

Carlos Heuser
Universidade Federal do Rio Grande do Sul, Porto Alegre, Brasil

Ernesto Pimentel Sánchez
Universidad de Málaga, España

Alain Pirotte
Université Catholique de Louvain, Belgique

Roberto Tom Price
Universidade Federal do Rio Grande do Sul, Porto Alegre, Brasil

Will Tracz
Lockheed Martin Federal Systems, USA

Marcello Visconti Zamora
Universidad de Santa María, Valparaíso, Chile

Esteban Zimanyi
Université Libre de Bruxelle, Belgique

Organizing Committee

Claudia Marcos
Instituto de Sistemas Tandil, UNICEN, Argentina

Additional Reviewers

Edgardo Belloni
Instituto de Sistemas Tandil, UNICEN, Argentina

Marcelo Campo
Instituto de Sistemas Tandil, UNICEN, Argentina

Alejandro Clausse
Instituto de Sistemas Tandil, UNICEN, Argentina

George Fernandez
Dept of Computer Science, RMIT University, Australia

Jacinto Marchetti
INTEC, Rosario, Argentina

Claudia Marcos
Instituto de Sistemas Tandil, UNICEN, Argentina

Roberto Moriyon
Universidad Autónoma de Madrid, Madrid, España

Jane Pryor
Instituto de Sistemas Tandil, UNICEN, Argentina

Gustavo Rossi
Lifia, Universidad Nac. de La Plata, Argentina

Auspices of 29 JAIIO

Departamento de Ingeniería Eléctrica - Universidad Nacional del Sur.

Facultad de Ciencias Económicas y Estadística, Universidad Nacional de Rosario.

Facultad de Ciencias Económicas, Universidad Nacional de Córdoba.

Facultad de Ciencias Económicas, Universidad Nacional de Lomas de Zamora.

Facultad de Ciencias Exactas, Universidad Nacional del Centro de la Provincia de Buenos Aires.

Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires.

Facultad de Ciencias Exactas y Tecnológicas, Universidad Nacional de Santiago del Estero.

Facultad de Ciencias Exactas, Físico - Químicas y Naturales, Universidad Nacional de Río Cuarto.

Facultad de Ciencias Fisicomatemáticas e Ingeniería, Universidad Católica Argentina.

Facultad de Informática, Ciencias de la Comunicación y Técnicas Especiales, Universidad de Morón.

Facultad de Ingeniería, Universidad Católica de Córdoba.

Facultad de Ingeniería, Universidad Nacional de San Juan.

Facultad de Matemática Aplicada, Universidad Católica de Santiago del Estero.

Universidad del Salvador.

Universidad Nacional de Entre Ríos.

Universidad Nacional del Centro de la Provincia de Buenos Aires.

Universidad Nacional de Luján.

Universidad Nacional de Río Cuarto.

Universidad Tecnológica Nacional - Facultad Regional Santa Fé.

SADIO

President

Jorge Clot

Vice President

Irene Loiseau

Treasurer

Mario Weber

Co-Treasurer

Alejandro Bras Harriott

Secretary

Juan Carlos Frankel

Committee Members

Arnoldo Palma
Conrado Estol
Federico Shortrede
Basilio Jezieniecki

Substitute Committee Member

Jane Pryor

Auditors

Clara Fuks
Guillermo Arechaga

Foreword

The vertiginous development undergone by information technology and the wide diversity of its applications over the last decade, have led to the construction of highly complex systems, which support the organizations' operations all their aspects, even to the extent of conditioning and modifying their own structures and ways of functioning. These circumstances impose methodological and formal requirements on the software development process that are comparable to ones found in any other engineering discipline. For this purpose, methodologies, tools, architectures and project-management techniques are integrated.

The First Argentine Symposium on Software Engineering (ASSE 2000) is an attempt to allow researchers and professionals involved in the systems area to put into practice this need for integration between the multiple facets and disciplines that converge at the software development process. ASSE 2000 results from the melting of the Argentine Symposium on Object-Orientation and the Argentine Symposium on Software Technology, which have taken place within the framework of JAIIO.

ASSE 2000 has received valuable contributions by researchers and professionals from Argentina, Brazil, Ecuador, Australia, Sweden and Spain. The Program Committee has chosen 14 papers to be presented and published in the Symposium proceedings. The contributions address main issues relating to Software Development Process, Software Quality as well as Software and Business Process Supporting Tools.

ASSE 2000 will also receive the contribution of important researchers from Argentina and other countries. They will address both, the state of the art as well as future perspectives of tools, techniques and methodologies in software engineering disciplines, like software quality, architectures, software process development, etc. The main contributors are going to be: Alvaro Ortigosa (ISISTAN, UNCPBA), George Fernández (Royal Melbourn Institute of Technology, Australia), Jorge Boria (TeraQuest Metric Inc., USA), Mohamed Fayad (University of Nebraska, USA) and Marcello Visconti Zamora (Universidad de Santa María, Chile).

Finally, we are grateful to all the reviewers that have collaborated in the review process. Also, we would like to thank Ing. Claudia Marcos who gave us an invaluable help in the organization of the review process and the edition of ASSE 2000 proceedings.

Alvaro Ortigosa
Symposium Chair

Horacio Leone
Symposium Chair

Contents

Selected Papers

Invited Talks and Tutorials

Metrics for Entity Relationship Diagrams

Marcela Genero, Mario Piattini, Coral Calero, Joaquín Rincón-Cinca

{mgenero, mpiattin, ccalero}@inf-cr.uclm.es
Grupo ALARCOS
Departamento de Informática
Escuela Superior de Informática
Universidad de Castilla-La Mancha
Ronda de Calatrava, 5 - 13071 - Ciudad Real - ESPAÑA
Tel.: + 34 926 29 53 00 ext. 3715
fax: + 34 926 29 53 54

Abstract. It is well known that the quality of information systems depends greatly on the accuracy of the requirements specification, and the greatest should focus on improving the early stages of developments. Conceptual data models lay the foundation for all later design work and determine what information can be represented by an information system. So, their quality can have a significant impact on the quality of the information system which is ultimately implemented. Improving the quality of conceptual data models will therefore be a major step towards the quality improvement of information system development.

Unfortunately, most of the work related to conceptual data models quality merely lists properties, without giving quantitative measures that assess the quality of such models in an objective way. In this work, we propose a set of metrics for measuring the complexity of the well known Entity Relationship diagrams. The early availability of metrics allows designers to measure the complexity of conceptual data models in order to improve the quality of the information systems from the early stages of their life cycle. We put the proposed metrics under theoretical validation following Zuse´s framework, which is based on the measurement theory. And we also put them under empirical validation in order to ascertain if they may be used as early quality indicators in the information system life cycle.

Keywords: Conceptual Data Models, Entity Relationship Diagrams, Software Metrics, Quality in Conceptual Modelling, Theoretical Validation, Empirical Validation

1. Introduction

Nowadays, in a global and increasingly competitive market, quality is gaining importance in all economical and organisational aspects, and especially in Information Systems. It is well known that the quality of information systems depends greatly on the accuracy of the requirements specification, and the greatest effort should focus on improving the early stages of developments. Conceptual data models form the basis of requirements specification, lay the foundation for all later design work and determine what information can be represented by an information system (Feng, 1999). So, their quality can have a significant impact on the quality of the information system which is ultimately implemented (Shanks and Darke, 1997), which becomes even bigger if we take into account the size and complexity of current information systems. Improving the quality of conceptual data models will therefore be a major step towards the quality improvement of information system development. Unfortunately, most of the work related to conceptual data models quality merely lists properties, without giving quantitative measures that assess the quality of such models in an objective way (Moody and Shanks, 1994; Krogstie et al., 1995; Shanks and Darke, 1997; Moody et al., 1998).

Our objective should be to replace intuitive notions of quality in conceptual data models with formal, quantitative measures in order to reduce subjectivity and bias in the evaluation process.

Within the field of software engineering a plethora of metrics has been proposed for measuring software products, processes and resources (Melton, 1996; Fenton and Pflegger, 1997; Henderson-Sellers, 1996). The only works that propose metrics for conceptual data models are Eick (1991) Gray et al. (1991), Moody (1998) and Kesh (1995).

Although all of these metric proposals are a good starting point to think about quality in conceptual modelling in a numeric scale, most of them are subjective and lack empirical and theoretical validation. Thus, there is a need for metrics and quality models that can be applied at the early stages of information system design, particularly in what we concern applied to Entity Relationship (ER) diagrams (Chen , 1976) , to ensure that that designs have favourable internal properties that will lead to the development of quality information systems.

As in other aspects of Software Engineering, proposing techniques and metrics is not enough, it is also necessary to put them under theoretical and empirical validation, in order to assure their utility in practice. Validation is critical to the success of software measurement (Kitchenham, 1995; Fenton and Pflegeer, 1997; Kitchenham, 1995; Schnneidewind, 1992; Basili et al., 1999). Empirical validation is also necessary to give some limits which can be useful for designers. However, as DeChampeaux (1997) remarks, we must be conscious that "associating the qualifications *good* and *bad* with numeric ranges is the hard part".

Related to theoretical validation, for every measurement we have to be aware of its scale type (Zuse, 1998). Knowledge of scale type tell us about limitations on the kind of mathematical manipulations that can be performed. The scale type of a measure affects the types of operations and statistical analyses that can be sensibly applied to the data values (Fenton and Pflegger, 1997).

Equally important is the empirical validation, in order to demonstrate that the proposed metrics really function in practice. In this work we want to validate if there exist correlation between our metrics and maintenance time.

In section 2 we propose a set of metrics for ER diagrams, thus allowing designers to measure the complexity of their designs since the early stages of information systems life cycle. Next, in section 3 we validate them following the framework of software measurement proposed by Zuse (1998) with the goal of determining some properties of the proposed metrics, as well as each scale type. We empirically validate the proposed metrics in section 4, with the goal of determining if correlation exists between each of the ER complexity metrics and the maintenance time. Lastly, section 5 summarises the paper, draws our conclusions, and presents our future research directions.

2. A Proposal of Metrics for ER Diagrams

In this section we propose a set of metrics to assess the complexity of ER diagrams. According to the taxonomy of quality measurements proposed in (Tian, 1999), our metrics are categorised as product internal measurements. Since the aim is that of simplifying the ER diagram, the objective will be to minimise the value of these metrics. It is common understanding that the greatest complexity is strongly correlated with the development and maintenance efforts and the overall quality of information systems. We classify these metrics into the following categories:

2.1 Entity Metrics

NE metric. We define the Number of Entities metric (NE) as the number of entities within the ER diagram.

2.2 Attribute Metrics

NA Metric. We define the Number of Attribute metric (NA) as the number of attributes that exist within the ER diagram. In this number we include simple attributes, composite attributes and also multivalued attributes, each of which take the value 1.

DA Metric. An ER diagram is minimal when every aspect of the requirements appears once in the diagram, i.e. an ER diagram is minimal if it does not have any redundancies. One of the sources of redundancies in the ER diagrams is the existence of derived attributes. An attribute is derived when its value can be calculated or deduced from the values of other attributes.
We define.the Derived Attributes metric (DA) as the number of derived attributes existing in the ER diagram.

CA Metric. We define the Composite Attributes metric (CA) as the number of composite attributes within an ER diagram. A composite attribute is an attribute composed of a set of simple attributes.

MVA Metric. The Multivalued Attributes metric (MVA) is defined as the number of multivalued attributes within the ER diagram. A multivalued attribute is an attribute that can take several values for an individual entity.

2.2 Relationship Metrics

NR Metric. We define the Number of Relationships metric (NA) as the number of relationships within the ER diagram.

M-NR Metric. The M:N Relationships metric (M:NR) is defined as the number of M:N relationships within the ER diagram.

N-AryR Metric. The N-ary Relationships metric (N-AryR) is defined as the number of N-ary relationships (not binary) within the ER diagram.

BinaryR Metric. The Binary Relationships metric (BinaryR) is defined as the number of binary relationships within the ER diagram.

NIS_AR Metric. We define the Number of IS_A Relationships metric (NIS_AR) as the number of relationships IS_A (generalisation or specialisation) that exist within the ER diagram. In this case, we consider one relationship for each pair child-parent within the IS_A relationship.

RR Metric. Another source of redundancy in an ER diagram is the existence of redundant relationships. We define the Redundant Relationship metric (RR) as the number of relationships that are redundant in the ER diagram.

3. Theoretical Validation of Metrics

Several frameworks for measure characterisation have been proposed (Briand et al., 1996; Morasca and Briand, 1997; Weyuker, 1988; Zuse, 1998). In this paper we will follow Zuse´s framework (Zuse, 1998) with the goal of determining some properties of the proposed metrics, as well as each scale type.

3.1 Zuse´s Formal Framework

This framework is based on an extension of the classical measurement theory. Measurement theory gives clear definitions of terminology, a sound basis of software measures, criteria for experimentation, conditions for validation of software measures, foundations of prediction models, empirical properties of software measures, and criteria for measurement scales.

A discussion of measurement scales (nominal, ordinal, interval, ratio and absolute scale) is mostly ahead practical applications in software engineering. Mostly people do not consider measurement scales in practice. In 1946 measurement scales were introduced by Stevens (1946) in his seminal paper. However, most research in the software measurement area does not address measurement scales. Much of it argues that scales are not so important. This forgets that empirical properties of software measures are hidden behind scales. Units are also closely connected to measurement scales. The discussion of scale types is important for statistical operations. Meaningful statistical operations are essential in experimental sciences.

Table 1 shows the admissible transformations, the statistics and the statistic tests that can be applied to each scale type. Software measurement starts at ordinal scale (Zuse, 1998), because with nominal measures it is no possible to do much.

SCALE	ADM. TRANSF.	STATISTIC	STATISTIC TEST
Nominal	Any one-to-one	Frequency, mode	Nonparametrics order indep.
Ordinal	g: strictly increasing monotonic function	Median, quartiles	Nonparametrics, coef. de Karrell, coef. of Spearman y Kendall
Interval	g(x)=ax+b, a>0	Arithmetic mean, standard deviation, variance	Parametrics and nonparametrics
Ratio	g(x)=ax, a>0	Mean, coef. of	Parametrics and

		variation	nonparametrics
Absolute	Identity	Mean, coef. of variation	Parametrics and nonparametrics

Table 1. Scale types

•

Because many empirical and numerical conditions are not covered by a certain scale type, the consideration of the empirical and numerical conditions is necessary and very important, too.

People are interested in establishing "empirical relations" between objects, such as "higher than" or "equally high or higher than". These empirical relations will be indicated by the symbols "• >" and "• >=" respectively. We called Empirical Relational System a triple: A = (A, • >=, o), where A is a non-empty set of objects, • >= is an empirical relation to A and o is a closed binary (concatenation) operation on A. The concatenation operations allow us to define powerful measurement structures (see table 2) which give us a more precise interpretation of numbers. Concatenation operations are directly connected with a measure. But sometimes, a measure assumes a non-intuitive empirical concatenation operation.

Zuse (1998) defines a set of properties for measures, which characterise different measurement structures. The most important ones are shown in table 2.

MODIFIED EXTENSIVE STRUCTURE	INDEPENDENCE CONDITIONS	MODIFIED RELATION OF BELIEF
Axiom1: (A, • >=) (weak order) **Axiom2**: A1 o A2 • >= A1 (positivity) **Axiom3**: A1 o (A2 o A3) ≈ (A1 o A2) o A3 (weak associativity) **Axiom4**: A1 o A2 ≈ A2 o A1 (weak commutativity) **Axiom5**: A1 • >= A2 ⇒ A1 o A • >= A2 o A (weak monotonicity) **Axiom6**: If A3 • > A4 then for any A1, A2, then there exists a natural number n, such that A1o nA3 • >A2 o nA4 (Archimedean axiom)	**C1**: A1 ≈ A2 ⇒ A1 o A ≈ A2 o A and A1 ≈ A2 ⇒ A o A1 ≈ A o A2 **C2**: A1 ≈ A2 ⇔ A1 o A ≈ A2 o A and A1 ≈ A2 ⇔A o A1 ≈ A o A2 **C3**: A1 • >= A2 ⇒ A1 o A • >= A2 o A, and A1 • >= A2 ⇒ A o A1 • >= A o A2 **C4**: A1 • >= A2 ⇔ A1 o A • >= A2 o A, and A1 • >= A2 ⇔A o A1 • >= A o A2	**MRB1**: ∀ A, B ∈ ℑ: A • >= B or B • >= A (completeness) **MRB2**: ∀ A, B, C ∈ ℑ: A • >= B and B • >= C ⇒ A • >= C (transitivity) **MRB3**: ∀ A ⊇ B ⇒ A • >= B (dominance axiom) **MRB4**: ∀ (A ⊃ B, A ∩ C = ϕ) ⇒ (A • >= B ⇒ A U C • > B U C) (partial monotonicity) **MRB5**: ∀ A ∈ ℑ: A • >= 0 (positivity)
As we know, binary relation • >= is called weak order if it is transitive and complete: A1 • >= A2, and A2 • >= A3 ⇒ A1 • >= A3 A1 • >= A2 or A2 • >= A1	Where A1 ≈ A2 if and only if A1 • >= A2 and A2 • >= A1, and A1 • > A2 if and only if A1 • >= A2 and not (A2 • >= A1).	

Table 2. Zuse's formal framework properties

It is important to note that when a metric accomplishes the weak order of the extensive modified structure axiom, it also accomplishes the completeness and the transitivity axioms of the belief structure.

Measures may be classified in a scale type, depending on whether they assume an extensive structure or not. When a measure accomplishes this structure, it also accomplishes the independence conditions and can be used on the ratio scale levels.

If a measure does not satisfy the modified extensive structure, the combination rule (that describes the properties of the software measure clearly) will exist or not depending on the independence conditions. When a measure assumes the independence conditions but not the modified extensive structure, the scale type is the ordinal scale (the characterisation of measures above the ordinal scale level is very important because we cannot do very much with ordinal numbers).

In the next subsection we present the formal description of the DA metric. First we define the concatenation operation and the combination function, after we prove the modified extensive structure.

3.2 Theoretical Validation of the Proposed Metrics

For our purposes, the Empirical Relational System could be defined as:

E =(E, •>=, o)

Where E is a non-empty set of ER diagrams, •>= is the empirical relation "equal or more complex than" on E and o is a closed binary (concatenation) operation on E. In our case we will consider the concatenation operation ERCon. Two ER diagrams, E1 and E2 are concatenated by the concatenation operation ERCon, adding a new relationship between them, as it is shown in figure 1.

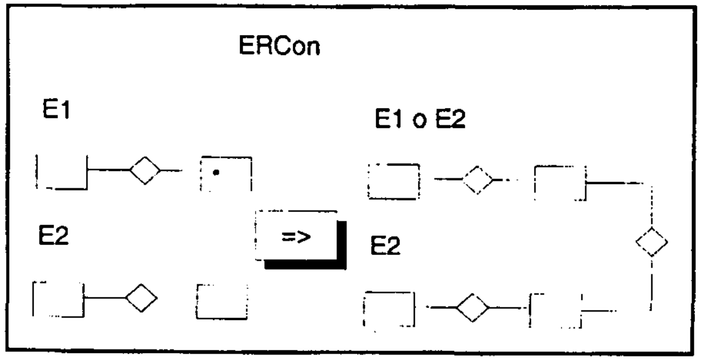

Figure 1. Entity Relationship Concatenation

Theoretical Validation of DA Metric. DA metric is a mapping: DA:E->ℜ such as the following holds for all ER diagrams Ei and Ej ∈ E: Ei •>= Ej ⇔ DA(Ei) >=DA(Ej)

We can define the combination rule for DA in the following way:

DA(Ei o Ej) = DA(Ei) + DA(Ej), ie., the number of derived attributes of E1oE2, is equal to the sum of the number of derived attributes of E1 and E2. We do not show attributes in figure 1, for the sake of brevity.

We will verify if DA metric fulfils all of the axiom of the Modified Extensive Structure.

Axiom 1. DA fulfils the first axiom of weak order, because if we have two ER diagrams E1 and E2, it is obvious that DA(E1) >= DA(E2) or DA(E2) >= DA(E1) (completeness) and let E1, E2 and E3 three ER diagrams, transitivity is always fulfilled: DA(E1) >= DA(E2) or DA(E2) >= DA(E3), then DA(E1) >= DA(E3).

Axiom 2. DA also fulfils positivity, because the number of derived attributes of E1 o E2 will be always greater or equal than the number of derived attributes of E1. In the case that E2 has no derived attributes DA(E1 o E2) = DA(E1), and if E2 has derived attributes DA(E1 o E2) > DA(E1).

Axiom 3. DA also fulfils weak associativity, because the number of derived attributes does not depend on the order which we associate the ER diagrams to apply the concatenation operation ERCon.

Axiom 4. DA also fulfils weak conmutativity. Taking into account the definition of ERCon, the order in which we concatenate the ER diagrams does not affect the number of derived attributes.

Axiom 5. DA also fulfils weak monotonicity, because if the number of derived attributes of E1 is greater than or equal to the number of derived attributes of E2, and after we do E1 o E and E2 o E, DA(E1oE) >=DA(E2oE) will result.

Axiom 6. DA also fulfils the Arquimedean axiom. Let E1, E2, E3 and E4 four ER diagrams, and DA(E3)>D(E4) it is easy to see that one number exits "n" such that DA(E1onE3) > DA(E2onE4), ie. if we concatenate n times E1 with E3, as DA(E3)>DA(E4), for some value of n it will happen that DA(E1 o nE3) > DA(E2 o nE4).

Seeing that DA metric fulfils all of the axiom of the Modified Extensive Structure, we can conclude that this metric is in ratio scale.

Analogously, it is easy to show that all of the metrics proposed in section 2, accomplish the modified extensive structure. So that, all of them are in the ratio scale.

4. Empirical Validation of the Proposed Metrics

Defining metrics is a very hard task, because sometimes we define metrics with the intention of measuring something but when we put them in practice, we realise that they do not work as we have expected. So, it is essential to put metrics under empirical validation.

In order to validate the metrics proposed in section 2 we have chosen five ER diagrams taken from real implemented information systems. All of them have been built using a tool called Data Architect.

First of all, we briefly describe each of them:

ER 1) WORK_CERTIFICATION: Dedicated to the administration of certificates to the builder CABBSA.

ER 2) ACCOUNTING_ANALYSIS: Dedicated to the administration of accounts for the builder CABBSA.

ER 3) BILLING_ZONE: Dedicated to the control of billing of a series of jobs for the builder CABBSA.

ER 4) SUPPLIERS: dedicated to the administration of CABBSA´suppliers..

ER 5) SOFIA: Dedicated to the control of the offers, evaluations, and product catalogue of one of the units of ERICSSON.

Table 3 shows the values of the metrics NE, NA, NR, M:NR, 1:NR and BinaryR and the last column shows maintenance time (expressed in hours) in the initial six months from system delivery. All of the metrics have been collected using a metric tool MANTICA which was developed inside our research group. We only considered these metrics due to the fact that the rest of the metrics were insignificant, as in each case they took a zero value.

	NE	**NA**	**NR**	**M:NR**	**1:NR**	**BinaryR**	**Maintenance time (hours)**
ER 1	9	98	6	0	6	6	12
ER 2	17	72	18	0	18	18	17
ER 3	13	84	13	0	13	13	12

ER 4	9	80	9	0	9	9	12
ER 5	48	178	109	2	101	103	208

Table 3. Values of the proposed metrics and the maintenance time

Pearson´s correlation was used to determine the correlation of the nonparametric data in table 3. The correlation coefficient is a measure of the ability of one variable to predict the value of another variable. Using Pearson´s correlation coefficient, each of the metrics was correlated separately to the maintenance time.

We would like to test the hypothesis that there is a significant correlation between the current metric data set (NE, NA, NR, M:NR, BinaryR) and the maintenance time.

Analysing the Pearson´s correlation coefficients shown in table 4, we can conclude that a high correlation exists between all of the metrics and the maintenance time, as we intuitively think.

Maintenance Time With NE Metric	Maintenance Time with NA Metric	Maintenance Time with NR Metric	Maintenance Time with M:NR Metric	Maintenance Time With 1:NR Metric	Maintenance Time with BinaryR Metric
0.989	0.971	0.997	1	0.996	0.996

Table 4. Correlation between ER complexity metrics and the Maintenace Time

Even though the sample size (five real cases) is not enough in order to use this conclusion as a final conclusion, we think that it is a good starting point in order to think about conceptual data models in a numeric terms. We are aware that it is necessary to replicate this experiment with a bigger sample than that which is used in this work.

4. Conclusions

Due to the growing complexity of information systems, continuous attention to and assessment of the conceptual data models are necessary to produce quality information systems. Following this idea, we have presented a set of objective and automatically computed metrics for evaluating the complexity of ER diagrams.

We put them under theoretical validation following Zuse´s formal framework in order to demonstrate all of the properties that a metric fulfils and the scale type of each metric. All of the proposed metrics are in ratio scale, which, as it was cited above, have an important significance in the scope of software measurement.

We also put them under empirical validation, corroborating that some of the proposed metrics (NE, NA, NR, M:NR, 1:NR, BinaryR) have a high correlation with the maintenance time. We are aware of the fact that we must perform more experimentation in order to validate these metrics as maintenance time predictors.

We want to highlight that our proposal cannot be considered as a final proposal. Instead, it is a starting point and we require feedback to improve it.

Due to the increasing and fast diffusion of the object oriented (OO) paradigm, we are tailoring the proposed metrics (when it is possible) or defining new ones, in order to address the complexity of IS using UML (Booch, 1998). We have already performed some research regarding OO conceptual modelling (Genero et al., 1999; Genero et al., 2000a). Furthermore, we will not only address complexity, we also have to focus our research towards measuring other quality factors like the ones proposed in the ISO 9126 (1999).

We have built a metric tool, called MANTICA, for collecting, and visualising metric values. Now we are working on building a tool to analyse measurement empirical data, using a novel data analysis approach based on regression and classification fuzzy trees (Genero, et al., 2000b).

Acknowledgements

This research is part of the MANTICA project, partially supported by CICYT and the European Union (CICYT-1FD97-0168).

References

Basili, V., Shull, F. and Lanubile, F.: Building knowledge through families of experiments. IEEE Transactions on Software Engineering, Vol. 25 Num. 4. (1999) 435-437.

Booch, G., Rumbaugh, J. and Jacobson, I.: The Unified Modeling Language User Guide. Addison-Wesley (1998).

Briand, L., Morasca, S. and Basili, V.: Property-Based Software Engineering Measurement. IEEE Transactions on Software Engineering, Vol. Num. (1996) 68-86.

Chen, P.: The Entity-Relationship Model: toward a unified view of data. ACM Transactions on Database Systems, Vol. 1. (1976) 9-36.

De Champeaux, D.: Object-oriented development process and metrics. Upper Saddle River, Prentice-Hall. (1997).

Eick, C.: A Methodology for the Design and Transformation of Conceptual Schemas. Proc. of the 17th International Conference on Very Large Data Bases. Barcelona. (1991).

Feng, J.: The "Information Content" problem of a conceptual data diagram and a possible solution. Proceedings of the 4th UKAIS Conference: Information Systems-The Next Generation, University of York. (1999) 257-266.

Fenton, N. and Pfleeger, S.: Software Metrics: A Rigorous Approach. 2nd. edition. London, Chapman & Hall. (1997).

Genero, M., Manso, Mª E., Piattini, M. and García, F.: Assessing the Quality and the Complexity of OMT Models. 2nd European Software Measurement Conference - FESMA 99, Amsterdam, The Netherlands. (1999) 99-109.

Genero, M., Piattini, M. and Calero, C. : Una propuesta para medir la calidad de los diagramas de clases en UML. IDEAS'2000. Cancún, México. (2000a).

Genero, M., Piattini, M. and Calero, C.. Measuring the Quality of Entity Relationship Diagrams. Proceedings of the Entity-Relationship 2000, Salt Lake City, November. (2000b).

Gray, R., Carey, B., McGlynn, N. and Pengelly A.: Design metrics for database systems. BT Technology, Vol. 9 Num. 4. (1991)

Henderson-Sellers, B. (1996). Object-oriented Metrics - Measures of complexity. Prentice-Hall, Upper Saddle River, New Jersey.

ISO/IEC 9126-1: Information technology- Software product quality – Part 1: Quality model. (1999).

Kesh, S.: Evaluating the Quality of Entity Relationship Models. Information and Software Technology, Vol. 37 Num. 12. (1995) 681-689.

Kitchenham, B., Pflegger, S. and Fenton, N.: Towards a Framework for Software Measurement Validation. IEEE Transactions of Software Engineering, Vol. 21 Num. 12. (1995) 929-943.

Krogstie, J., Lindland, O. and Sindre, G.: Towards a Deeper Understanding of Quality in Requirements Engineering, Proceedings of the 7th International Conference on Advanced Information Systems Engineering (CAISE), Jyvaskyla, Finland. (1995) 82-95.

Melton, A.: Software Measurement, London, International Thomson Computer Press. (1996).

Moody, L. and Shanks G.: What Makes A Good Data Model? Evaluating The Quality of Entity Relationships Models. Proceedings of the 13th International Conference on Conceptual Modelling (ER '94), Manchester, England. (1994) 94-111.

Moody, L.: Metrics For Evaluating the Quality of Entity Relationship Models. Proceedings of the Seventeenth International Conference on Conceptual Modelling (ER '98), Singapore, November. (1998) 16-19.

Moody, L., Shanks, G. and Darke, P.: Improving the Quality of Entity Relationship Models – Experience in Research and Practice. Proceedings of the Seventeenth International Conference on Conceptual Modelling (ER '98), Singapore. (1998) 255-276.

Morasca, S. and Briand, L.: Towards a Theoretical Framework for measuring software attributes. Proceeding of the Fourth International, Software Metrics Symposium (1997) 119-126.

Pfleeger, S.: Assessing Software Measurement. IEEE Software, March/April. (1997) 477-482.

Shanks, G. and Darke, P.: Quality in Conceptual Modelling: Linking Theory and Practice. Proceedings of the Pacific Asia Conference on Information Systems ,PACIS'97. (1997) 805-814.

Schneidewind, N.: Methodology For Validating Software Metrics. IEEE Transactions of Software Engineering . Vol. 18 Num. 5. (1992) 410-422.

Stevens, S.: On the Theory of Scales and Measurements. Science 103. (1946) 677-680.

Tian, J.: Taxonomy and Selection of Quality Measurements and Models. Proceedings of SEKE'99, The 11th International Conference on Software Engineering & Knowledge Engineering, June 16-19. (1999) 71-75.

Weyuker, E.: Evaluating software complexity measures. IEEE Transactions on Software Engineering,. Vol. 14 Num. 9, (1988) 1357-1365.

Zuse, H.: A Framework of Software Measurement. Berlin, Walter de Gruyter. (1998).