V Jornadas de Trabajo DYNAMICA

(DYNamic and Aspect-Oriented Modeling for Integrated Component-based Architectures)

En Valencia, 23 y 24 de noviembre de 2006

Organización:

ISSI Ingeniería del Software y Sistemas de Información

Colaboradores:

UNIVERSITAT POLITÈCNICA DE VALÈNCIA

DSIC DEPARTAMENTO DE SISTEMAS INFORMÁTICOS Y COMPUTACIÓN

ACTAS

V Jornadas de Trabajo DYNAMICA

(DYNamic and Aspect-Oriented Modeling for Integrated Component-based Architectures)

En Valencia, 23 y 24 de noviembre de 2006

Editores:

Jennifer Pérez
Manuel Llavador
Cristóbal Costa
Nour Ali

Financiado por el CICYT (Centro de Investigación Científica Y Tecnológica), Proyecto DYNAMICA (DYNamic and Aspect-Oriented Modeling for Integrated Component-based Architectures)

Editores

Jennifer Pérez
Departamento de Sistemas Informáticos y Computación
Universidad Politécnica de Valencia
Camino de Vera s/n
E-46022 Valencia, Spain
E-mail: jeperez@dsic.upv.es

Manuel Llavador
Departamento de Sistemas Informáticos y Computación
Universidad Politécnica de Valencia
Camino de Vera s/n
E-46022 Valencia, Spain
E-mail: mllavador@dsic.upv.es

Cristóbal Costa
Departamento de Sistemas Informáticos y Computación
Universidad Politécnica de Valencia
Camino de Vera s/n
E-46022 Valencia, Spain
E-mail: ccosta@dsic.upv.es

Nour Ali
Departamento de Sistemas Informáticos y Computación
Universidad Politécnica de Valencia
Camino de Vera s/n
E-46022 Valencia, Spain
E-mail: nourali@dsic.upv.es

Distribuido en España, Grupo de Ingeniería del Software y Sistemas de Información, Departamento de Sistemas Informáticos y Computación, Universidad Politécnica de Valencia
Noviembre, 2006

Presidente de las jornadas

Isidro Ramos Salavert

Comité Organizador

Presidente: Jennifer Pérez Benedí

Vocales:
Nour Ali Irshaid
Silvia Mara Abrahao
Mª Eugenia Cabello Espinosa
Jose Hilario Canós Cerdá
Jose Ángel Carsí Cubel
Cristóbal Costa Soria
Abel Gómez Llana
Emilio Insfrán Pelozo
Patricio Letelier Torres
Rogelio Limón Cordero
Manuel Llavador Campos
Elena Mª Navarro Martínez
Mª Carmen Penadés Gramaje
Carlos Solís Pineda

Comité Técnico

Isidro Ramos Salavert, *Universidad Politécnica de Valencia*

Bárbara Álvarez Torres, *Universidad Politécnica de Cartagena*

Coral Calero Muñoz, *Universidad de Castilla-La Mancha*

Francisco José Rodríguez Urbano, *Universidad Carlos III*

Ambrosio Toval Álvarez, *Universidad de Murcia*

Subproyectos y Grupos Participantes

SUBPROYECTO: PRISMA – Plataforma OASIS para Modelos Arquitectónicos
TIC2003-07804-C05-01
Grupo de Investigación "Ingeniería de Software y Sistemas de Información" (ISSI)
Departamento de Sistemas Informáticos y Computación (DSIC)
Universidad Politécnica de Valencia (UPV)

SUBPROYECTO: ANCLA – Arquitecturas DiNámiCas para Sistemas de TeLeoperAción
TIC2003-07804-C05-02
Grupo DSIE (División de Sistemas e Ingeniería Electrónica
Departamento de Tecnologías de la Información y Comunicaciones
Universidad Politécnica de Cartagena (UPCT)

SUBPROYECTO: CALIPO – CALIdad en POrtales
TIC2003-07804-C05-03
Grupo ALARCOS
Departamento de Informática
Universidad de Castilla-La Mancha (UCLM)

SUBPROYECTO: CARATE – Controlador de Arquitectura Reconfigurable para aplicaciones de Teleoperación
TIC 2003-07804-C05-04
Grupo Laboratorio de Sistemas Inteligentes (LSI)
Departamento de Ingeniería de Sistemas y Automática (ISA)
Universidad Carlos III (UC3M)

SUBPROYECTO: PRESSURE – PREcise Software modelS and reqUirements REuse
TIC 2003-07804-C05-05
Grupo de Investigación en Ingeniería del Software (GIS)
Departamento de Informática y Sistemas
Universidad de Murcia (UMU)

Presentación

El Proyecto DYNAMICA (**DYN**amic and **A**spect-Oriented **M**odeling for **I**ntegrated **C**omponent-based **A**rchitectures) ha sido un proyecto coordinado de tres años de duración financiado por el Ministerio de Ciencia y Tecnología. DYNAMICA surgió por los intereses comunes de cinco grupos de investigación españoles: el grupo de Ingeniería del Software y Sistemas de Información (ISSI) de la Universidad Politécnica de Valencia (UPV), el grupo de División de Sistemas e Ingeniería Electrónica (DSIE) de la Universidad Politécnica de Cartagena (UPCT), el grupo ALARCOS de la Universidad de Castilla-La Mancha (UCLM), el grupo del Departamento de Ingeniería de Sistemas y Automática (ISA) de la Universidad Carlos III (UC3M) y el Grupo de Investigación en Ingeniería del Software (GIS) de la Universidad de Murcia (UMU).

Las últimas jornadas DYNAMICA nos proveen un espacio para compartir los resultados obtenidos a lo largo del proyecto. Así, el objetivo de estas jornadas es presentar los resultados en forma de comunicaciones y "demos" orientadas a enfrentar la dinámica del software desde una visión arquitectónica en dominios con problemas reales como los sistemas de teleoperación, los sistemas hidráulicos y los portales Web.

Se han recibido 16 contribuciones en forma de 11 comunicaciones y 5 "demos" desde los diferentes nodos, reflejando el espíritu e ilusión que hemos invertido en este proyecto. Además, las relaciones inter-nodos han sido intensas, mostrando que somos conscientes que es el único modo de conseguir un avance hacia un objetivo común es mediante el trabajo colectivo. La amistad y sinergia entre todos nosotros ha permitido un excelente trabajo y un gran resultado de este proyecto.

Esto nos ha situado en primera línea de investigación en Ingeniera del Software tanto nacional como internacionalmente y ha propiciado el nacimiento de META, el nuevo proyecto en el que compartiremos los esfuerzos gran parte de los miembros del consorcio.

La voluntad de estar juntos, de aceptarnos mutuamente, con las discusiones necesarias, han sido el acicate dialéctico de avance de DYNAMICA que animamos a continuar en META.

Isidro Ramos.

Presidente de las Jornadas

Índice

Artículos

Demostraciones

Towards a Data Quality Framework for Web Portals

Angélica Caro[1], Coral Calero[2]

[1] Department of Auditoria e Informática, University of Bio Bio
La Castilla s/n, Chillán, Chile
ancaro@inf-cr.uclm.es
[2] Head research of CALIPO project.
Alarcos Research Group. Information Systems and Technologies Department
UCLM-SOLUZIONA Research and Development Institute.
University of Castilla-La Mancha
Coral.Calero@uclm.es

Abstract. Advances in technology and the use of the Internet have favoured the appearance of a great variety of Web applications, among them Web Portals. These applications are important information sources and/or means of accessing information. Many people need to obtain information by means of these applications and they need to ensure that this information is suitable for the use they want to give it. In other words, they need to assess the quality of the data.

In recent years, several research projects were conducted on topic of Web Data Quality. However, there is still a lack of specific proposals for the data quality in Web portals. In this paper we introduce a data quality framework for Web portals. This framework is centred in the point of view of data consumers and uses a probabilistic approach for the data quality evaluation.

Keywords: Data Quality, Information Quality, Web Portal, Data Quality Framework, Bayesian Network.

1 Introduction

A Web portal is a site that aggregates information from multiple sources on the Web and organizes this material in an easy user-friendly manner [25]. Over the past decade the number of organizations which owns Web portals grows dramatically. Their have established portals to complement, substitute or widen existing services to their clients [26]. Many people use data obtained from portals to develop their work and to make decisions. These users or data consumers need to ensure that the data obtained are appropriate for the use they need. Likewise, the organizations owns of Web portals need to deliver data that meet user requirements to achieve the user's preference. So, a common interest between data consumers and portal's owners is the data quality.

In the literature, the concept of Data or Information Quality (DQ hereafter) is often defined as "fitness for use", i.e., the ability of a data collection to meet user

requirements [4, 23]. Besides, the terms “data” and “information” are often used as synonyms. In this work also we will use them as synonymous.
Research on data quality began in the context of information systems [15, 23] and it has been extended to contexts such as cooperative systems, data warehouses or e-commerce, amongst others. Due to the particular characteristics of Web applications and their differences from the traditional information systems, the research community started to deal with the subject of DQ on the Web [10].
However there are no works on DQ that address the particular context of Web portals [5], in spite of the fact that some works highlight the DQ as one of the relevant factors in the quality of a portal [18, 26]. Likewise, except for few works in the DQ area, like [3, 4, 24], most of them have looked at quality from the data producers or data custodians perspective and not from the data consumers perspective [3].
Consequently, our research aim is to create a Data Quality Model for Web portals focused on the data consumer perspective. For this, we have divided our work in two parts. The first, consisted in the definition of a theoretical model [6] named Portal Data Quality Model (PDQM). As a result of this we have identified a set of 34 DQ attributes than can be used to assess the DQ in a portal. The second, now in progress, is to convert PDQM in an operational model, i.e., that it can be used to assess DQ of Web portals. To reach this goal, DQ attributes in PDQM have to be specified in an operational way. This means that we need to define a structure where we can organize the DQ attributes and to associate measures and criteria for them.
Considering the subjectivity own of data consumer perspective and the uncertainty inherent to the quality perception, we have decided to use a probabilistic approach (based on Bayesian networks and fuzzy logic) to convert PDQM in an operational model. In this paper we show the first part of our work and the advances in the second part.
The rest of the paper is organized as follows. Section 2 presents the definition of PDQM. In Section 3 we describe briefly the approach used to convert PDQM in an operational model. Section 4 describes our advances to complete our model. Finally, section 5 shows our conclusions.

2 PDQM

PDQM is a data quality model for Web portals focused on the data consumer perspective. Their development was based on three key aspects:

- **Data consumer perspective**. When data management is conceptualized as a production process [23], we can identify three important roles in this process: (1) data producers (who generate data), (2) data custodians (who provide and manage resources for processing and storing data), and (3) data consumers (who access and use data for their tasks). The last perspective differs from the two others in two important aspects [3]: (1) data consumer has no control over the quality of available data and (2) the aim of consumers is to find data that match their personal needs, rather than provide data that meet the needs of others.

To consider the data consumer perspective in our model we have used the quality expectations of the data consumer on the Internet, proposed in [22]. These

expectations are organized into six categories: Privacy, Content, Quality of values, Presentation, Improvement, and Commitment.

- **Web data quality attributes**. Obtained from DQ frameworks proposed in the literature for different domains in the Web context. The idea was to take advantage of work already carried out in the Web context and apply it to Web portals.
- **Web portal functionalities**. Web portals present basic software functionalities to data consumer deploying their tasks. Under our perspective, the data consumer judges the quality of data by using the application functionalities. So, we used the web portal software functions that Collins proposes in [7] considering them as basics in our model. These functions are as follows: Data Points and Integration, Taxonomy, Search Capabilities, Help Features, Content Management, Process and Action, Collaboration and Communication, Personalization, Presentation, Administration, and Security.

Next, to define PDQM we developed a four-phase process showed follow.

2.1 Identification of Web DQ attributes

The first phase consisted in gathering Web DQ attributes from the literature. For this we have made a systematic review of the relevant literature [14]. Then, we selected works proposed for different domains in the Web context (Web sites [8, 13, 19], integration of data [2, 20], e-commerce [12], Web information portals [26], cooperative e-services [9], decision making [11], organizational networks [17] and DQ on the Web [10]). As result and after summarizing the collected initial set of attributes, we obtained 41 DQ attributes (see top of Table 1).

2.2 Definition of a Classification Matrix for Web DQ attributes

In the second phase, we have built a matrix for the classification of the DQ attributes obtained in previous phase. This matrix relates two basic aspects considered in our model: the data consumer perspective by means their DQ expectations on Internet [22] and the basic functionalities in a Web portal. On this matrix we carried out an analysis of what expectations were applicable in each different functionality of a Web portal, represented in Figure 1 with a "√" mark.

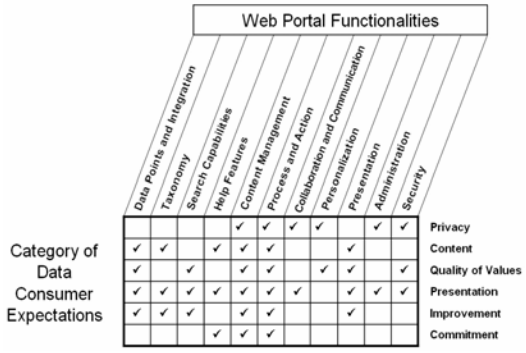

Figure 1. Matrix to classify Web DQ attributes.

2.3 Classification of Web DQ attributes in the Matrix

In the third phase, we used the obtained matrix to classify the Web DQ attributes identified in phase 1. Then for each relationship between functionality and expectation, we assigned the Web DQ attributes that could be used by the data consumer to evaluate the DQ in a portal. We did it by studying the appropriateness of each attribute (based on its definition), in relation to the objective of each portal functionality and the user DQ expectation. On Table 1, we have summarized the attributes assigned for functionality.

Functionalities	Accessibility	Accuracy	Amount of data	Applicability	Attractiveness	Availability	Believability	Completeness	Concise Representation	Consistent Representation	Cost effectiveness	Customer support	Currency	Documentation	Duplicates	Ease of operation	Expiration	Flexibility	Granularity	Interactive	Internal consistency	Interpretability	Latency	Maintainable	Novelty	Objectivity	Ontology	Organization	Price	Relevancy	Reliability	Reputation	Response time	Security	Specialization	Source's information	Timeliness	Traceability	Understandability	Validity	Value-added	Total of Attributes
Data Points and Integration	✓	✓	✓				✓		✓	✓		✓	✓				✓								✓					✓	✓			✓				✓	✓	✓		15
Taxonomy	✓		✓				✓		✓			✓				✓														✓	✓							✓	✓	✓		11
Search Capabilities	✓		✓				✓	✓	✓			✓	✓			✓									✓					✓	✓							✓	✓			13
Help Features	✓		✓						✓			✓				✓														✓								✓	✓		8
Content Management	✓		✓	✓		✓	✓	✓	✓			✓	✓	✓	✓	✓						✓			✓	✓		✓		✓	✓	✓		✓	✓			✓	✓	✓	24
Process and Action	✓	✓	✓	✓		✓	✓	✓	✓			✓				✓	✓					✓				✓				✓	✓	✓		✓	✓			✓	✓		21
Collaboration and Communication						✓						✓										✓								✓				✓				✓			6
Personalization		✓						✓					✓				✓																				✓	✓	✓		7
Presentation			✓		✓		✓		✓			✓	✓			✓	✓	✓				✓								✓	✓				✓			✓	✓	✓		15
Administration			✓						✓	✓						✓																	✓				✓				6
Security	✓	✓	✓					✓	✓							✓						✓												✓				✓	✓		10
Number of References	7	4	9	2	1	3	6	5	9	1	0	8	5	1	1	8	4	1	0	0	0	5	0	0	3	2	0	1	0	7	7	2	0	5	3	1	0	7	11	8	1	

Table 1. Data quality attributes assigned for functionality

As a result of this work we have a set of 34 DQ attributes that can be used for the DQ evaluation in portals, considering the data consumer perspective.

2.4 Validation

The fourth phase consists on the validation of the set of DQ attributes selected. Despite that this phase is not essential still, because later we will develop another validation with the whole model; we have decided to carry out this previous validation by means a survey. In this survey, data consumers of web portals to be asked about the importance for them of the DQ attributes considered in PDQM. The survey is run on http://alarcos.inf-cr.uclm.es/PDQM/.

3 A probabilistic approach to make PDQM an operational model

With the aim to use PDQM in an evaluation process, we need to convert it in an operational model. For this, and considering the subjectivity of data consumer perspective and the uncertainty inherent to the quality perception, we have decided to use a probabilistic approach as the proposed in [16]. This approach involves Bayesian

networks and fuzzy logic. Bayesian networks model the problems that involve uncertainty, and combine the advantages of an intuitive visual representation with a sound mathematical basis in Bayesian probability [21]. Finally, fuzzy logic provides an effective conceptual framework for dealing with the problem of knowledge representation in an environment of uncertainty and imprecision [27].

A Bayesian Network (BN) is a directed acyclic graph, whose nodes are the uncertain variables and edges are the causal or influential links between variables. A conditional probability function models the uncertain relationship between each node and its parents [21]. In our context, BNs offer an interesting framework with which it is possible to:

- Represent the interrelations among DQ attributes in an intuitive and explicit way by connecting influencing factors to influenced ones. Such a representation facilitates the comprehension of the model, its validation, its evolution and its exploitation.
- Circumvent the problems of subjectivity uncertainty.
- Actually use the obtained network to predict/estimate the data quality of a portal.
- Isolate responsible factors in the case of low data quality.

Another interesting property of the Bayesian approach is the fact that it considers the probability as being a dynamic entity that can be updated as more data arrive (self learning mechanism). New data may naturally improve the degree of belief in certain propositions [1]. Consequently, a BN model is particularly adapted to the changing domain of web portals.

4 Process to convert PDQM in an operational model

To convert PDQM in an operational model, we have developed a process with the following phases:

a. To organize hierarchically the DQ attributes of PDQM. This phase consisted on the definition of criteria to organize and classify the DQ attributes of PDQM.
b. To build a graph to represent PDQM. In this phase the DQ attributes organized in the previous phase, will be represented in the form of a BN.
c. To prepare the BN for the evaluation process. This phase is developed in sub-phases due to the size of the obtained BN. In each sub-phase we have selected a sub-network of the BN developing the next activities on it:
    - To define quantifiable variables for the DQ attributes in the last level (entry nodes) of the sub-network.
    - To define the node probability table for each intermediate node of the sub-network.
d. To develop an experiment to compare the judgements of a set of portal-user subjects with the evaluation results produced by the BN.

On the first phase we have selected the conceptual DQ framework developed in [24] as criteria of classification to organize the DQ attributes of PDQM. With the idea of considering some aspects inherent to the Web context, in our work we have renamed and redefined the Accessibility category as Operational category, emphasizing with

this the importance of the role of systems not only with respect to accessibility and security but also to aspects as personalization, collaboration, etc. So, the final DQ categories to classify the DQ attributes of PDQM are: Intrinsic, Operational, Contextual and Representational.

After this, and considering the definition of each DQ category, we have classified all the DQ attributes of PDQM into these categories.

In the second phase, we have generated new levels in the BN based on the relationships of direct influences among the DQ attributes in each category. We used the definitions of the DQ attributes to establish these relationships. Our aim was to establish which DQ attribute in a category had direct influence on other DQ attributes in the same category, and eventually on attributes in other category.

Based on these relationships we have built the graph of the BN which represents PDQM, see Figure 2. In this BN we can distinguish the following levels:

- Level 0, where the PDQ is the node that represents DQ in the whole portal.
- Level 1, where node represents the DQ in each DQ category in a portal. Obviously, the node PDQ is defined in terms of the other four nodes.
- Level 2, where nodes represent the DQ attributes with a direct influence over each one of the DQ categories.
- Level 3, where nodes represent the DQ attributes with a direct influence over each one of the DQ attributes in Level 2.
- Level 4, with only one node that represents a direct influence over each one of the DQ attributes in Level 3.

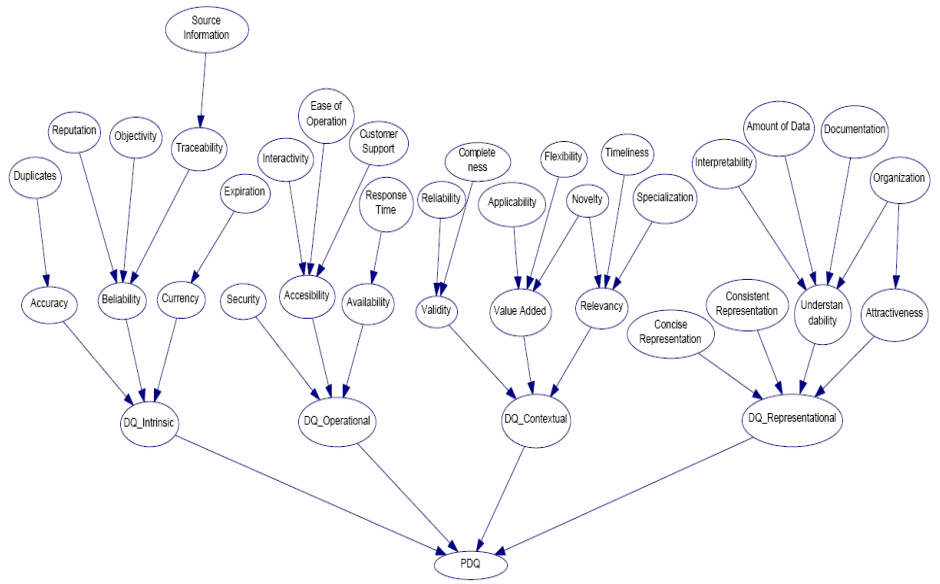

Figure 2. Graph of the BN that represents PDQM.

As it can be seen in the resultant graph, some relationships are more complex than others. For example: the variables Accessibility and Understandability have three and four parents respectively. Then with the aim of facilitating the use of the BN, will be

necessary to create new nodes (intermediate nodes) that permit to reduce the number of parents for each node.

In the third phase, corresponding to the preparation of the BN for the evaluation process, we will work separately with each sub-network of the BN created. For each one, we will develop two sub-phases: first, to define one quantifiable variable (indicator) for each DQ attribute in the last level of the BN, and second, to define the probability tables for each intermediate node.

As a point of start, we have selected the DQ_Representational sub-network. As we can see in Figure 2, the original sub-network had two nodes with four parents (Understandability and DQ_Representational) and was necessary to create two synthetic nodes (Representation and Volume of Data) in order to reduce the combinatory explosion in the next step during the preparation of the probability tables, obtaining the sub-network showed in Figure 3.

For this sub-network we have defined one indicator for each entry node in the BN (see on Figure 3 the nodes in the last level). In general we have selected and defined the measures according to each attribute (entry node) definition. To exemplify this in the Figure 3 we show the definition of indicator LCsR (Level of Consistent Representation). For the calculation and definition of each indicator we have used several base and derived measures. All the indicators take a numerical result between 0 and 1. Consequently, with the fuzzy approach used, we have defined for each one the labels that represent the fuzzy sets associated with it. Finally, a membership function was defined to determine the degree of membership of each indicator respect of the fuzzy labels.

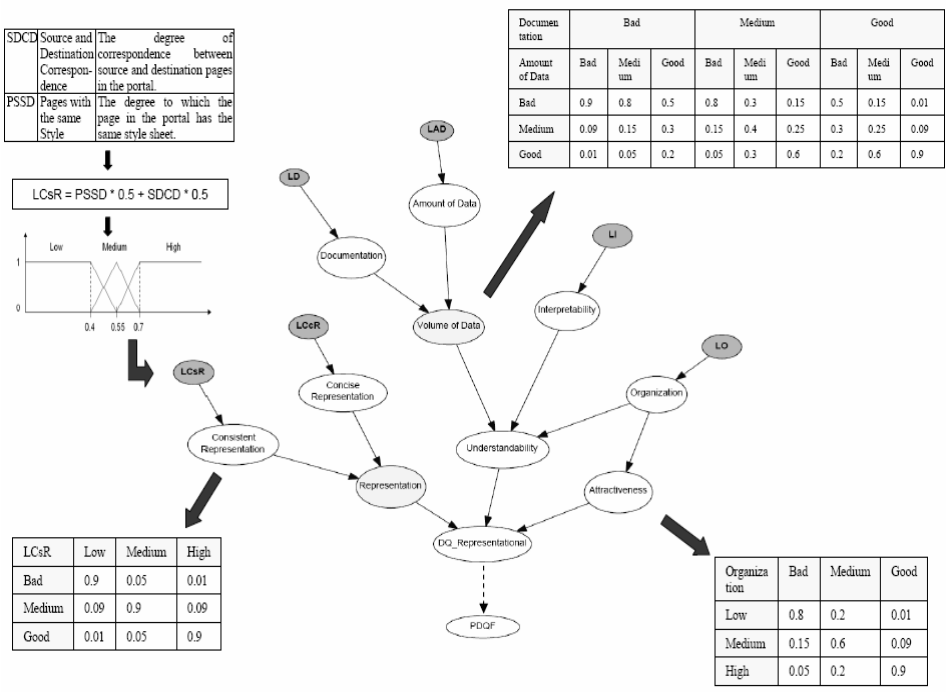

Figure 3. Preparation of sub network DQ_Representational for evaluation process.

The intermediate nodes are nodes defined by their parents and not directly measurable. So, whereas these nodes are not measurable, their probability distribution was given by expert judgments. On the other hand, we considered that this probability distribution can differ depending on the Web portal context.
In this work as a matter of example, we have considered the educational context and we have defined the nodes probability tables considering that this BN will be applied to university portals. Figure 3 shows some nodes probability tables for sub-network.
For the last phase an experiment will be developed. This study will give us the judgments of a group of subjects about the DQ Representational in a set of Web portals in the context selected. The idea is to compare the valuations of the subjects with the results of the BN, and based on this comparison adjust the sub-network.

5 Conclusions and the future work

In this paper, we have shown the development of a DQ model for Web portals (PDQM) centered in the data consumer perspective. This work has been developed in two parts. The first consisted in the definition of the theoretical foundations of the model and in the identification of a set of 34 DQ attributes that can be used for the DQ evaluation in Web portals. The second part, now in progress, consists in the transformation of PDQM into an operational DQ model. For this, we are using a probabilistic approach. As a result of this part, we have generated a BN that represents PDQM and we have implemented a sub-network of it for a specific Web portal context.
As future work we will develop an experiment to validate the sub-network implemented and after we will work with the other sub-networks of PDQM until complete and validate the whole model.
The choice of a probabilistic approach to generate the framework is motivated by the fact that many issues in quality assessment are circumvented: threshold value definition, measure combination, and uncertainty.
One of the advantages of our framework will be its flexibility. Indeed, the idea is to develop a global framework that could be adapted for both the goal and the context of evaluation. From the goal perspective, the user can choose the sub-network that evaluates the characteristics he is interested in. From the context point of view, the parameters (probabilities) can be changed to consider the specific context of the evaluated portal. This operation can be performed using available historical data from the organization.

Acknowledgments. This research is part of the following projects: CALIPO (TIC2003-07804-C05-03), CALIPSO (TIN20005-24055-E) supported by the Ministerio de Educación y Ciencia (Spain) and COMPETISOFT (506PI0287) financed by CYTED.

References

1. Baldi, P., Frasconi, P., and Smyth, P., *Modeling the Internet and the Web; Probabilistic Methods and Algorithms.* 2003: Wiley.
2. Bouzeghoub, M. and Peralta, V. A Framework for Analysis of data Freshness. in International Workshop on Information Quality in Information Systems, (IQIS2004). 2004. Paris, France: ACM.
3. Burgess, M., Fiddian, N., and Gray, W. Quality Measures and The Information Consumer. in Proceeding of the Ninth International Conference on Information Quality. 2004.
4. Cappiello, C., Francalanci, C., and Pernici, B. Data quality assessment from the user´s perspective. in International Workshop on Information Quality in Information Systems, (IQIS2004). 2004. Paris, Francia: ACM.
5. Caro , A., Calero, C., Caballero, I., and Piattini, M. Data quality in web applications: A state of the art. in IADIS International Conference WWW/Internet 2005. 2005. Lisboa-Portugal.
6. Caro, A., Calero, C., Caballero, I., and Piattini, M. Defining a Data Quality Model for Web Portals. in WISE2006, The 7th International Conference on Web Information Systems Engineering. 2006. Wuhan, China: Springer LNCS 4255.
7. Collins, H., *Corporate Portal Definition and Features.* 2001: AMACOM.
8. Eppler, M., Algesheimer, R., and Dimpfel, M. Quality Criteria of Content-Driven Websites and Their Influence on Customer Satisfaction and Loyalty: An Empirical Test of an Information Quality Framework. in Proceeding of the Eighth International Conference on Information Quality. 2003.
9. Fugini, M., Mecella, M., Plebani, P., Pernici, B., and Scannapieco, M. (2002). Data Quality in Cooperative Web Information Systems.Personal Communication. citeseer.ist.psu.edu/fugini02data.html.
10. Gertz, M., Ozsu, T., Saake, G., and Sattler, K.-U., Report on the Dagstuhl Seminar "Data Quality on the Web". SIGMOD Record, 2004. vol. 33, Nº 1: p. 127-132.
11. Graefe, G. Incredible Information on the Internet: Biased Information Provision and a Lack of Credibility as a Cause of Insufficient Information Quality. in Proceeding of the Eighth International Conference on Information Quality. 2003.
12. Katerattanakul, P. and Siau, K., *Information quality in internet commerce desing*, in *Information and Database Quality*, M. Piattini, C. Calero, and M. Genero, Editors. 2001, Kluwer Academic Publishers.
13. Katerattanakul, P. and Siau, K. Measuring Information Quality of Web Sites: Development of an Instrument. in Proceeding of the 20th International Conference on Information System. 1999.
14. Kitchenham, B. (2004). Procedures for Performing Systematic Reviews.RN: 0400011T.1. http://www.idi.ntnu.no/emner/empse/papers/kitchenham_2004.pdf.
15. Lee, Y., AIMQ: a methodology for information quality assessment. Information and Management. Elsevier Science, 2002: p. 133-146.
16. Malak, G., Sahraoui, H., Badri, L., and Badri, M. A Proposal of a Probabilistic Framework for Web-Based Applications Quality. in 10th ECOOP Workshop on Quantitative Approaches in Object-Oriented Software Engineering, (QAOOSE06). 2006.
17. Melkas, H. Analyzing Information Quality in Virtual service Networks with Qualitative Interview Data. in Proceeding of the Ninth International Conference on Information Quality. 2004.
18. Moraga, M.Á., Calero, C., and Piattini, M., Comparing different quality models for portals. Online Information Review., 2006. Vol. 30(5): p. 555-568.
19. Moustakis, V., Litos, C., Dalivigas, A., and Tsironis, L. Website Quality Assesment Criteria. in Proceeding of the Ninth International Conference on Information Quality. 2004.

20. Naumann, F. and Rolker, C. Assesment Methods for Information Quality Criteria. in Proceeding of the Fifth International Conference on Information Quality. 2000.
21. Neil, M., Fenton, N.E., and Nielsen, L., Building large-scale Bayesian Networks. The Knowledge Engineering Review, 2000. 15(3): p. 257-284.
22. Redman, T., *Data Quality: The field guide*. 2000, Boston: Digital Press.
23. Strong, D., Lee, Y., and Wang, R., Data Quality in Context. Communications of the ACM, 1997. Vol. 40, Nº 5: p. 103 -110.
24. Wang, R. and Strong, D., Beyond accuracy: What data quality means to data consumers. Journal of Management Information Systems; Armonk; Spring 1996, 1996. 12(4): p. 5-33.
25. Xiao, L. and Dasgupta, S., *User Satisfaction with Web Portals: An empirical Study*, in *In Web Systems Design and Online Consumer Behavior*, Y. Gao, Editor. 2005, Idea Group Publishing, Hershey, PA. p. 193-205.
26. Yang, Z., Cai, S., Zhou, Z., and Zhou, N., Development and validation of an instrument to measure user perceived service quality of information presenting Web portals. Information and Management. Elsevier Science, 2004. 42: p. 575-589.
27. Zadeh, L.A., Knowledge Representation in Fuzzy Logic. IEEE Transactions on Knowledge and Data Engineering, 1989. 1(1): p. 89-100.