

User Profiles in Commercial Interaction Web Portals

Carmen Moraga¹, M^a Ángeles Moraga¹, Angélica Caro², Coral Calero¹ and Rodrigo Romo Muñoz³

¹*Alarcos Research Group-Instituto de Tecnología y Sistemas de la Información, University of Castilla-La Mancha, Paseo de la Universidad 4, Ciudad Real, Spain*

²*Department of Computer Science and Information Technologies, University of Bio Bio, Chillán, Chile*

³*Department of Business Management, University of Bio Bio, Chillán, Chile*

Keywords: Data Quality, Web Portal, Statistical Method, Commercial Interaction.

Abstract: The use of Web portals to carry out on-line transactions is continuously increasing. This paper studies this type of Web portals (which we have denominated as Commercial Interaction) from the perspective of data quality, since we consider that this is an essential element if Web portals are to be competitive and their use is to be boosted. The quality of data in a Web portal is determined on the basis of a set of data quality characteristics. However, we have established that some data quality characteristics are more important than others according to various user profiles, based on demographic aspects (gender, age range, level of studies and type of organisation). In order to do that, Commercial Interaction Web portal users were surveyed and, as a result of that, three different user profiles were identified. This paper describes the survey development and the generation of the user profiles.

1 INTRODUCTION

The amount of activities that can be carried out via the Internet is increasing on a daily basis, and the Internet has come to be used in all aspects of our life in recent years (Komathi and Maimunah, 2009). One means to access information on the Internet is through Web portals.

Web portals can be classified into various groups according to the principal type of activity that users wish to carry out. These groups are:

- ‘The Search for and Reading of Information’: defined as those portals that the user principally uses to obtain information (e.g., a newspaper Portal, etc.)
- ‘Commercial Interaction’: this is used to carry out some kind of on-line transaction, such as buying train or airline tickets.
- ‘Interaction with other People’: the important aspect here is the ability to relate to or get in contact with other people, known or otherwise, as is the case of social networks.

We focus on ‘Commercial Interaction’ Web portals, in which all types of on-line transactions are carried out, due to they are increasingly used.

In these Web portals, the data quality (DQ), which is often defined as the ability of a collection of data to meet user requirements (Cappiello et al., 2004);

(Strog et al., 1997); (Wang and Strong, 1996), is increasingly more important if user loyalty is to be maintained and new users are to be attracted. Data quality is a multi-dimensional concept, and is relative to the context in which it is applied (Katerattanakul and Siau 1999); (Shanks and Corbitt 1999). Bearing this in mind, we have therefore studied the DQ in the context of ‘Commercial Interaction’ Web Portals by using the data quality model denominated as SPDQM (SQuaRE-Aligned Portal Data Quality Model) (Moraga et al., 2009).

In this paper, we consider that it is interesting to establish whether different user profiles exist as regards preferences towards the various DQ characteristics in ‘Commercial Interaction’ Web Portals. This is done by using an initial set of DQ characteristics and surveying ‘Commercial Interaction’ Web portal users. Our study allows us to determine, first, whether all the DQ characteristics are important for users, and second, which are most relevant in comparison to the others according to the different user profiles established.

The remainder of this paper is organized as follows: Section 2 shows the works related to our study. Section 3 describes the DQ characteristics in the Contextual category. Section 4 describes the method used to create the survey, and how its results were obtained. Section 5 shows an analysis of the

users according to their various demographic aspects. In Section 6, the DQ characteristics that the various user profiles consider to be most important for 'Commercial Interaction' Web portals are obtained. Section 7 shows guidelines for designers and developers. Finally, Section 8 presents our conclusions and future work.

2 RELATED WORK

Our search of literature has led to the discovery of several studies whose objective is to analyse the differences in the use of the Internet according to various demographic aspects. For example, (Durndell and Haag, 2002); (Hupfer and Detlor 2006) state that the user's preferences may be different owing to, among other things, their demographic aspects, such as their gender.

The level of a user's studies also influences the use of Web portals, as is stated by (Şahin, 2011) who point out that there are different levels of addiction to Internet use among students and other professional groups.

And in general, as is stated by (Lenhart et al., 2010), gender, education and even social status have an influence on Internet use, principally as regards wireless access.

All of the above has led us to believe that it would be interesting to consider Web portal users' demographic aspects when determining the data quality in Web portals.

3 THE SELECTION OF DQ CHARACTERISTICS FOR OUR STUDY

SPDQM is a model which allows DQ to be characterised in a Web portal on the basis of a set of 42 DQ characteristics organized in 4 categories: Intrinsic, Contextual, Representational and Operational.

In this study we focus on the DQ characteristics corresponding to the Contextual category, since it is the only category in which the importance that users place on certain DQ characteristics with regard to others may vary according to the type of Web portal.

The Contextual category represents the data quality needs in the context of the task in hand. The context is associated with the tasks that will be carried out, and these tasks are influenced by the types of Web portals. Of the three types of Web portal identified in the previous section, in this paper

we have decided to begin with Web portals of the 'Commercial Interaction' type for this category. This category is composed of 10 characteristics and 6 sub-characteristics. Further information on SPDQM can be found in (Moraga et al., 2009).

In our study, we shall ask for both the characteristics and the sub-characteristics since they all refer to aspects of data quality whose importance we wish to determine.

4 CARRYING OUT THE SURVEY

We decided to discover users' opinions of the DQ characteristics and to determine which of the 16 DQ characteristics are most important in the context of transactional activities by asking the users themselves, and we therefore prepared a survey. In order to do that, we used "the principles of survey research" proposed in (Kitchenham and Pflieger 2002).

- Survey design: the questionnaire in the survey was made up of 21 questions, of which 4 were related to demographic aspects (on Table 1), 16 corresponded to the DQ characteristics in the Contextual category and 1 was related to the definition of the term 'Contextual' itself. Table 2 shows some of these questions.
- Preparing the data collection instrument: The questionnaire was written in simple language, with questions that were easy to understand without negative questions. We employed closed questions with only one possible answer, using the Likert scale of 1 to 10 in which totally disagree was (0) and totally agree was (10).
- Validating the instrument: The initial questionnaire used in the survey was validated in a control test with 10 participants who had experience in the use of Web portals.
- Selection of participants: The survey was distributed to a heterogeneous group of 200 'Commercial Interaction' Web portal users from Europe and Latin America.
- Administration and recovery of data from the survey: The survey was distributed and collected by e-mail or in printed format (manually). Of the total of 200 surveys that were distributed, 192 were returned.

Once the data had been obtained it was necessary to determine whether the results were reliable, and the Cronbach's alpha was therefore calculated, whose value must be over 0.6 to be considered appropriate. In our case its value was 0.942, and this value therefore indicated that the results were reliable and that there was good internal consistency.

Table 1: Questions Concerning Demographic Aspects.

Gender: Male/Female
Level of studies COMPLETED: Primary Education Qualification / Secondary Education Qualification / Vocational Training/ University/ Post Graduate.
Type of organization with which you are linked (for study or work purposes). If there are various, please place them in the order in which most time is dedicated to them, from greatest to least: Education / Industrial / Commercial / Service Sector / Financial / Other (Please state which).
Age range: Under 25 / Between 25 and 35 / Between 35 and 45 / Between 45 and 55 / Between 55 and 65 / over 65.

Table 2: Questions in the Questionnaire for the Contextual Category.

1.- The data should be sufficiently detailed to facilitate the task at hand.
2.- The data obtained from a Web portal should be true and reliable (believable).

6.- The data provided by a Web portal should contain the appropriate and specific information for the use to which they will be put.
7.- The data should adapt to user needs (e.g., they should be integrated into other applications or presented in different formats).
8.- The data should be useful and specially oriented towards the user community that will utilize them.

14.- It should be possible to obtain the data by using the appropriate quantities and types of resources (by, for example, using the smallest possible number of links to locate the data desired).
15.- The data obtained from Web portals should be exact and concise, thus helping you to find relevant results.
16.- The data in Web portals should provide the information that users are seeking.
17.- The data provided by Web portals should have a level of quality that accords with the specific use to which you wish to put them, i.e., in the context of the specific area in which you wish to work with them.

Once the appropriateness of the sample had been assured, it was possible to begin to obtain conclusions. A statistical study was therefore performed with the intention of determining the DQ characteristics that are most important for different types of users according to their various demographic aspects.

5 DESCRIPTION OF THE PARTICIPANTS AND THE SAMPLE

In this section we analyse the responses related to the demographic aspects, which were obtained from the useful surveys. Details of the results obtained are shown:

With regard to gender, the percentages are equally distributed between men and women, of which 50.5% are men and 49.5% are women.

With regard to age range, in our case the majority of the sample was in the range between 25 and 35 (see Figure 2), and in general, all the user ranges of under 55 years of age are representative.

In the case of the level of studies of the participants surveyed, those participants with University studies are widely represented.

Finally, as regards the type of organisation to which the user is linked, the most representative are 'Education' and 'Service Sector'.

6 DISCOVERING USER PROFILES

We shall now analyse the results obtained from the surveys. This will be done by following the three steps, details of which are provided in the following subsections.

6.1 Step 1: Descriptive Statistical Analysis

The first step involved carrying out a descriptive statistical analysis in order to obtain the minimum, maximum and mean values of each DQ characteristic. This allowed us to determine whether all of them are important to users of this type of Web portal. It can be concluded that they all are important for the analysis of the DQ in the Contextual category for 'Commercial Interaction' Web portals because the set of DQ characteristics as a whole was evaluated with a mean score of over 7.2.

6.2 Step 2: Factorial Analysis

In the second step, the initial set of DQ characteristics from the Contextual category was used to create homogeneous groups of these DQ characteristics (denominated as factors). These factors summarise and synthesise the information since they reduce the quantity of initial DQ

characteristics. These groups are formed of those DQ characteristics that have a considerable correlation with others. Moreover, each group should be independent from the others.

We decided to form three factors. Furthermore the Cronbach's alpha was calculated for each of the factors obtained in order to estimate the reliability of the results. Factor 1 obtained a Cronbach alpha value of 0.925, Factor 2 obtained a Cronbach alpha value of 0.881 and Factor 3 obtained a Cronbach alpha value of 0.788. This signifies that the values obtained are good, and that the results are therefore reliable. Table 3 shows the DQ characteristics in each factor.

Table 3: Factorial analysis.

Factor 1	Factor 2	Factor 3
Scope	Flexibility	Novelty
Reliability	Applicability	Timeliness
Validity	Value-added	Relevancy
Traceability	Usefulness	Precision
Compliance		
Specialization		
Efficiency		
Effectiveness		

6.3 Step 3: Cluster Analysis

In the last step, the results obtained from the factorial analysis were used to carry out a cluster analysis in order to group the factors by resemblance or similitude (denominated as clusters). These clusters determined the importance placed on the DQ characteristics in each factor. Three clusters were also obtained in this case, each of which contained one single factor. The results are shown in Table 4.

Table 4: Factors in each cluster.

Cluster		
1	2	3
Factor 3	Factor 1	Factor 2

The clusters were then related to the demographic aspects by using the contingency tables. These tables allowed us to determine the set of DQ characteristics that was most important according to the cluster to which each variable of each demographic aspect belonged.

Table 5 shows a summary of the contingency tables, in which the clusters are related to the demographic aspects.

As was mentioned previously, the contingency tables allowed us to determine the set of DQ characteristics for each variable of each

demographic aspect. This was done by using the results from Table 5 and following the procedure shown below:

Table 5: Relationship between demographic aspects and clusters.

Demographic Aspect	Variable	Cluster (%)		
		1	2	3
Gender	Male	46	56	57
	Female	54	44	43
Age range	Under 25	18	35	7
	Between 25 and 35	28	20	39
	Between 35 and 45	22	13	25
	Between 45 and 55	27	26	18
	Over 55	5	6	11
Level of Studies	High School	18	24	11
	Vocational Training	20	9	11
	University	51	59	61
	Postgraduate	11	8	17
Type of Organization	Education	41	39	39
	Industrial-commercial-financial	5	15	4
	Service Sector	31	24	39
	Other	23	22	18

First: We determined which variable had the greatest value for each demographic aspect and each cluster. For example, Cluster 1 and the demographic aspect 'gender' give us a value of '54' which corresponds with the variable 'female', while the demographic aspect 'level of studies' gives us a value of '51' which corresponds with the variable 'University', which is also in Cluster 1.

Second: The value obtained was compared with the other values in this variable for the other clusters. In the example, the values are '44' and '43' for gender aspect in Clusters 2 and 3, respectively, and the values are '59' and '61' for the level of studies in Clusters 2 and 3, respectively.

Third: We chose the greatest of the values obtained for this variable. In the example, we selected the value '54', which is in Cluster 1, for the variable 'female', and the value '61', which is in Cluster 3, for the variable 'University', and this variable was discounted in Cluster 1 (the values shown in bold type in Table 5).

Fourth: For those variables which did not yet have a selected value, we chose the highest value in their row. For example, for the 'Vocational Training' variable, whose values are '20', '9', and '11' in Clusters 1, 2 and 3, respectively, we chose the value '20' which is in Cluster 1 (the values shown in italics in Table 5).

This procedure was followed to obtain the shaded values shown in Table 5. These results

allowed us to place each variable of each demographic aspect in one of the three clusters. The set of demographic aspects of each cluster will allow us to define a user profile. A summary of these results is shown in Table 6.

6.4 Limitations of this Study

This work has been carried out in a systematic manner. Nevertheless, we are conscious that it has certain limitations. In this paper we have limited ourselves to studying Web portals of the 'Commercial Interaction' type, and have obtained those DQ characteristics which are most relevant according to the users' different demographic aspects. However, all the users who responded to our survey are from Europe and Latin America, since we have surveyed users within a known environment. The geographical zone to which the users belong will be extended in future surveys by using the snowball method.

In a future work we shall analyse the results for other types of Web portals.

7 TOWARDS THE GUIDELINES DEFINITION

Since we have defined the factors and clusters, and have discovered the user profiles. In this section, we show the method used in order to create guidelines for designers and developers so that they will know which DQ characteristics are most important according to the type of user to which the 'Commercial Interaction' Web portals are oriented and which they intend to develop or modify. This is done as follows:

First: We identify the type of user towards whom the 'Commercial Interaction' Web portal that is going to be created or modified is targeted. The type of user will be determined by the demographic aspects of gender, age range, level of studies and type of organisation with which they are linked (for study or work purposes). The following examples will allow the results to be analysed:

- Example 1: Men between 35 and 45 with Postgraduate studies belonging to service organisations.
- Example 2: Women under 25 with university studies belonging to educational organisations.

Second: The profile to which each demographic aspect belongs are obtained (see Table 6).

- From Example 1:
 - Men belong to Profile 3.
 - Those between 35 and 45 are in Profile 3.

- Those with Postgraduate studies are in Profile 3.
- Service organisations belong to Profile 3.
- From Example 2:
 - Women belong to Profile 1.
 - Those under 25 are in Profile 2.
 - Those with University studies are in Profile 3.
 - Educational organisations belong to Profile 1.

Third: Finally, it is necessary to consider the DQ characteristics in each factor. The designers and developers will put special emphasis on the DQ characteristics in the factor that appears most often. If none of the factors are repeated and various factors are repeated the same amount of times, they will consider the DQ characteristics of those factors.

▪ From Example 1: most attention should be paid to the DQ characteristics in Profile 3 (see Table 6), which indicates that these users are interested in data that can be adapted to users' needs (Flexibility), and are also interested in data that are oriented towards a destination community (Applicability), permit advantages to be attained (Value-added) and are useful (Usefulness).

▪ From Example 2: most attention should be paid to the DQ characteristics in Profile 1 (see Table 6), which is the predominant factor and indicates that these users are interested in the fact that the data are new (Novelty), are obtained in the least possible amount of time (Timeliness), are applicable and innovative (Relevancy) and are exact (Precision).

8 CONCLUSIONS AND FUTURE WORK

The objective of this document is to analyse the DQ characteristics for 'Commercial Interaction' Web portals in order to verify whether some are more important than others according to various user profiles which are determined by the demographic aspects of gender, age range, level of studies and type of organisation to which the users are linked.

We carried out a survey containing questions concerning the DQ characteristics identified and the demographic aspects of the users of this type of Web portals.

The results obtained from the surveys were then analysed and they allowed us to determine three user profiles, in addition to the most important DQ characteristics for each of these user profiles.

We have therefore verified that some DQ characteristics are effectively more relevant than others when considering gender, age range, level of studies and the type of organisation to which the users are linked. For example, men place

Table 6: Demographic aspects in each Profile.

Profile / Cluster	Factors	DQ Characteristics	Demographic aspects			
			Gender	Age	Level of Studies	Type of Organization
1	3	Novelty, Timeliness, Relevancy, Precision	Female	Between 45 and 55	Vocational Training	Education Other
2	1	Scope, Reliability, Validity, Traceability, Compliance, Specialization, Efficiency, Effectiveness		< 25	High School	Industrial-commercial-financial
3	2	Flexibility, Applicability, Value-added, Usefulness	Male	Between 25 and 45 > 55	University Postgraduate	Services

importance on the fact that data are oriented towards a destination community (Applicability), and users between 25 and 35 years of age consider it relevant that the data satisfy the users' needs (Usefulness).

This paper also describes the criteria needed to establish guidelines that will allow Web portal designers and developers to discover which DQ characteristics are most important for 'Commercial Interaction' Web portal users according to their user profiles.

As future work we shall analyse the other types of Web portals, and the guidelines for the designers and developers will be programmed into a free software tool that will be available for use.

ACKNOWLEDGEMENTS

This research has been funded by the following project: GEODAS-BC project (Ministerio de Economía y Competitividad and Fondo Europeo de Desarrollo Regional FEDER, TIN2012-37493-C03-01).

REFERENCES

- Cappiello, C, Francalanci, C and Pernici, B., 2004, 'Data quality assessment from the user's perspective', in *Proceeding on International Workshop on Information Quality in Information Systems (IQIS2004)*, Paris, France. ACM, pp. 68-73.
- Durndell, A and Haag, Z., 2002, 'Computer self efficacy, computer anxiety, attitudes towards the Internet and reported experience with the Internet, by gender, in an East European sample', *Computer in Human Behavior*, vol. 18, pp. 521-535.
- Hupfer, ME and Detlor, B., 2006, 'Gender and Web information seeking: A self-concept orientation model.', *Journal of the American Society for Information Science and Technology*, vol. 57, no. 8, pp. 1105-1115.
- Katerattanakul, P and Siau, K., 1999, 'Measuring Information Quality of Web Sites: Development of an Instrument', in *20th International Conference on Information System*, pp. 279-285.
- Kitchenham, B and Pfleeger, S., 2002, 'Principles of survey research part 2: designing a survey', *SIGSOFT: Software Engineering Note*, vol. 27, no. 1, pp. 18-20.
- Komathi, M and Maimunah, I., 2009, 'Influence of gender role on Internet usage pattern at home among academicians', *The Journal of International Social Research*, vol. 2, no. 8.
- Lenhart, A, Purcell, K, Smith, A and Zickuhr, K., 2010, 'Social Media and Mobile Internet Use Among Teens and Young Adults', *Pew Internet and American Life Project*, <http://pewinternet.org/Reports/2010/Social-Media-and-Young-Adults.aspx>.
- Moraga, C, Moraga, M, Calero, C and Caro, A., 2009, 'SQuaRE-Aligned Data Quality Model for Web Portals', in *9th International Conference on Quality Software (QSIC 2009)*, pp. 117-122.
- Şahin, C., 2011, 'An analysis of Internet addiction levels of individuals according to various variables', *TOJET: The Turkish Online Journal of Educational Technology*, vol. 10, no. 4, pp. 60-66.
- Shanks, G and Corbitt, B., 1999, 'Understanding Data Quality: Social and Cultural Aspects', in *10th Australasian Conference on Information Systems*, Wellington, New Zealand, pp. 785-797.
- Strong, D, Lee, Y and Wang, R., 1997, 'Data Quality in Context', *Communications of the ACM*, vol. 40, no. 5, pp. 103-110.
- Wang, RY and Strong, DM., 1996, 'Beyond accuracy: What data quality means to data consumers', *Journal of Management Information Systems*, vol. 12, no. 4, pp. 5-34.