

ANALYZING STRUCTURED TEXT ENTITIES USING A CITIZEN ORIENTED APPLICATION

Bogdan Vintila¹

Abstract

The concept of citizen oriented informatics application is presented in the context of the knowledge society. The differences between these applications and the traditional applications are highlighted. The diversity of problems the citizens has lead to a high diversity of application structures that is described in the paper. Usual applications are taken into discussion and comments are made on their citizen orientation. Quality standards for informatics applications are described. An application for the analysis of the structured entities is presented. The methods that were used to orientate it towards the citizens are described. The procedure for the score computing is described. The performance of the application measured by automatic means is analyzed. Performance improvements are discussed. Future work directions and improvements are discussed.

Keywords: citizen oriented applications, knowledge, analysis, structured text entities.

1. Citizen oriented informatics applications

In the context of the knowledge based society and of higher citizen requirements the appearance of a new category of informatics applications is necessary. The citizen orientated applications bring a new orientation as the citizen is considered to be the central element. These are different by the classic applications through:

- these are developed to solve the problems of the citizens, not the problems of the organization for which are developed;
- the target group is very large and very divers being formed by all the citizens;
- the applications are always available online;
- the citizen oriented applications aren't dependent on the hardware or software platform;
- the cost of use is very low or null;
- the quality requirements are much more strict than for traditional applications;
- localization assumes having the dialog with the user in his own language;
- the use of the applications doesn't assume previous training of the users [1];
- are very often updated to reflect the changes in the environment;
- adaptation to offer the citizens a greater degree of satisfaction.

The structure of the citizen oriented informatics applications differ on the offered functionality and the domain they are created for. The citizen oriented informatics applications are with:

- simple linear structure; these are applications that, for problem solving, assume the following of a number of steps, in a preset order, without the possibility to go back to a previous step; the first step of the sequence starts up the processing and the last one returns the results for another application; for an informing

¹ Bogdan Vintila, University of Economics, Bucharest, Romania, email: vb@vintilabogdan.ro

application, the information is structured in a logic sequence of the steps that must be followed; Figure 1 exemplifies this type of structure;



Figure 1 - Simple linear structure

- linear structure and simple links between components; these assume the possibility of going back to the previous steps; these are applications for which the possibility of modifying data from the previous steps or repeating them is a must; for the applications with a high number of steps is unacceptable for the user to redo all the steps just because he made something wrong in the end;
- linear structure and multiple links; assumes the existence of links between components and the navigation is made between any of the connected components respecting limitations imposed for the correct functioning of the application; the navigation towards a step is not allowed without the fulfillment of the prerequisites; for the informing applications of this type, the navigation has no restrictions beside the logical links;
- tree structure and simple links; these are applications for which from a step the user can move in many directions; the simple links between modules allow the advancement only on vertical as the user is going away from the tree root; these type of applications is suitable for showing information on the basis of selection criteria; Figure 2 shows a tree structure with simple links;

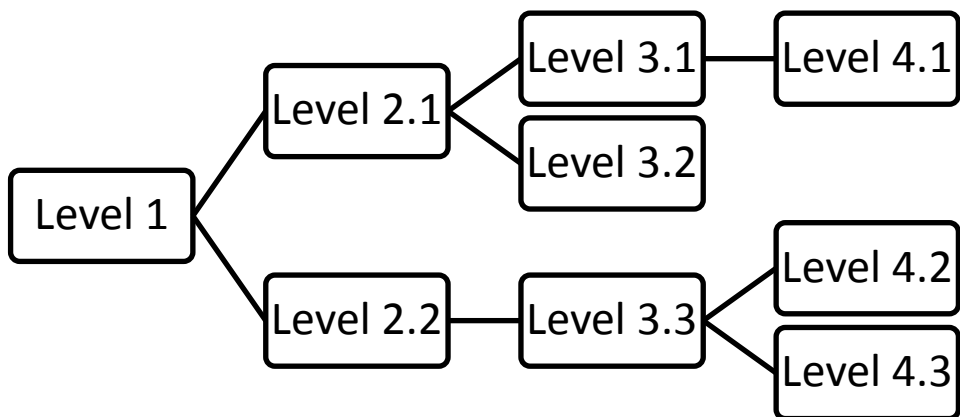


Figure 2 - Tree structure with simple links

- tree structure and double links; these assume the existence of bidirectional links between the components to browse the tree structure both top-bottom and bottom-top; double links ensure the possibility to go back to previous steps;
- tree structure and multiple links; the pass from a component to another is made only in the limit of the good functioning given by the logic of the processing which the applications make; the tree structure with multiple links is the most complex of them all; this allows the development of complex citizen oriented applications; Figure 3 exemplifies the tree structure with multiple links.

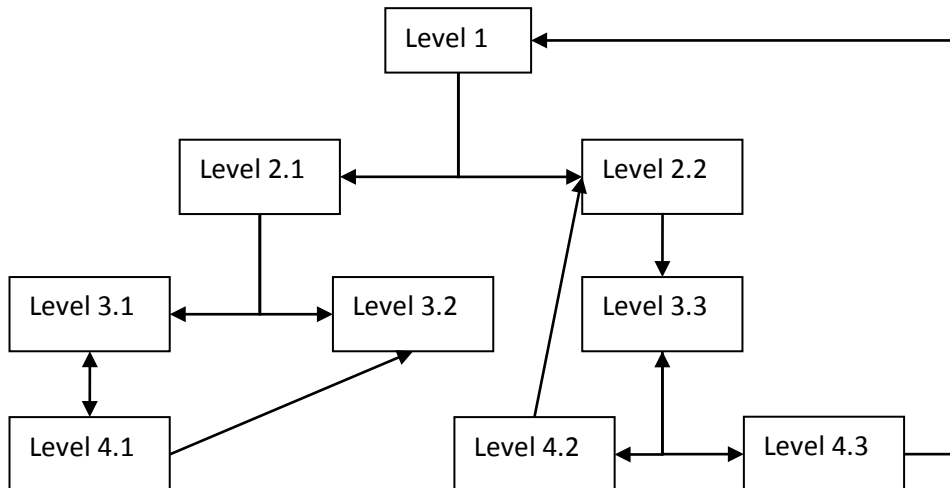


Figure 3 - Tree structure with multiple links

The applications for virtual campus training must satisfy the requirements of the persons that access the educational system. For this, these must be flexible, maintenance free, secure, accessible, platform free, without additional costs, always available, adaptable. Considering the very dynamic character of the virtual campus training domain, the requirements of the users quickly change and the applications, in order to be competitive, must evolve to fulfill them.

E-commerce is a very popular form of commerce as it has some clear advantages on the traditional commerce:

- no more stocks; virtual stores don't have stocks, or, if they do, these are very small [2];
- geographic borders are eliminated;
- very low running costs;
- users' comfort;
- automation.

The e-governing applications are used in the relation of the state with the citizens for solving different situations in which they are partners [3]. Issue of certificates and forms is made automatically. The e-voting applications are also very used in the e-governing process. These must be accessible from as many geographical points as possible.

Informing applications are those that guide the users to obtain information regarding a certain domain, state, process, object, phenomena. This type of application must be characterized by a clear structure allowing the user to reach, in as few steps as possible, to the desired result. The informing applications must not have using costs.

2. Quality models for informatics applications

The rapid evolution of technologies and approaches of the events leads to the creation of new standards. Quality requirements are higher and higher. Standardization allows objective assessments of products and services to be made. Without quality standards, each evaluates the quality from his own point of view and ranks its product as the best in its class. It is also necessary to standardize procedures for measuring the levels of various characteristics surveyed. In the absence of the measurement procedures, data analysis results are not comparable and are compromised.

McCall quality model developed in 1977 has eleven software quality factors [4]. These factors are grouped into three categories:

- factors on product operation: correctness, reliability, efficiency, integrity, usability;
- factors on product revision: maintainability, flexibility, testability;
- factors of product transition: portability, reuse, interoperability.

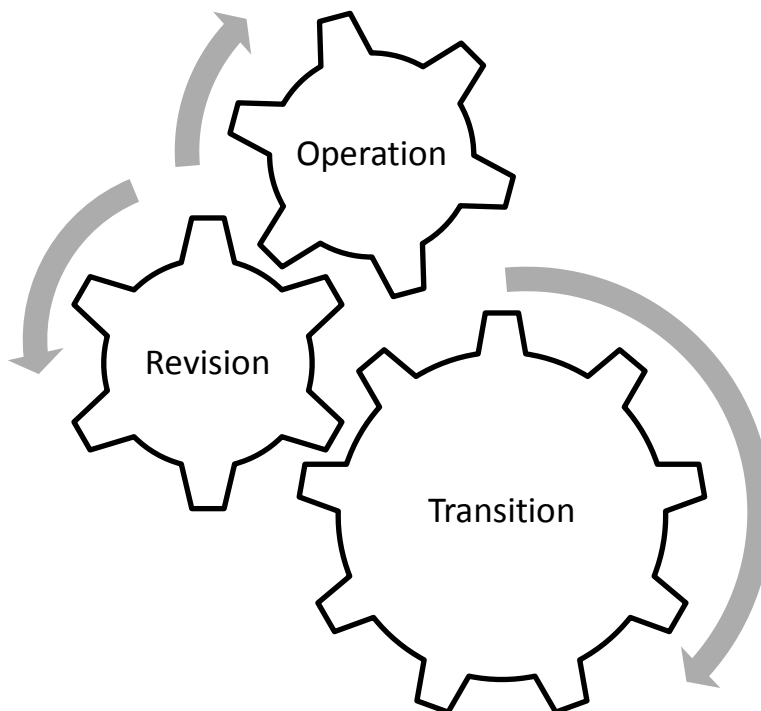


Figure 4 - Influence factor grouping for McCall model

Figure 4 notes the model categories of influence factors for software quality defined by the McCall quality model. Even if the McCall quality model is quite old, it is still the base for software quality as the models developed later didn't bring any revolutionary improvements.

Some improvements attempts were made after the McCall model. The two most important alternative models that have emerged after the McCall model are:

- the model of Evans and Marciniak 1987;

- the model of Deutsch and Willis in 1988.

These two models bring new influence on software quality. Both models exclude testability from the McCall model. The Evans and Marciniak model has twelve factors that are grouped into three categories: design, performance and adaptation. The Deutsch and Willis model has fifteen factors grouped into four categories: operating performance and change management. Two of the factors suggested by the new models are very similar to the model factors McCall (survivability, expandability). The three new factors are verifiability, safety and manageability.

Verifiability requirements define design and programming features that enable efficient verification of design and programming.

Safety requirements are designed to eliminate hazard conditions for operators of equipment as a result of errors in process control software. These errors can degenerate into inappropriate reactions to dangerous situations or failure to give warnings at the detection of dangerous situations. Safety is a very important factor for the products and services nowadays.

Manageability requirements are related to administrative tools that support software changes during the development and maintenance period. This ensures shorter development and maintenance times and also lower resource consumptions.

IEEE and ISO standards for software quality are made on the basis of quality models. ISO defines four software components:

- software;
- procedures;
- documentation;
- data required for running the software system.

ISO 9126 is an international standard for software quality assessment [5]. The fundamental objective is to address human error which may affect the development of a software project. The standard is divided into four parts:

- quality model;
- external metrics;
- internal metrics;
- metric use quality.

The quality is established in the first part of the standard ISO 9126-1. It classifies software quality in a set of features as follows:

- functionality;
- reliability;
- usability;
- efficiency;
- maintenance;
- portability.

Each feature is divided into subfeatures. Each of these subfeatures is divided into attributes. Each attribute can be verified or measured for a considered software product. We won't discuss any of these details in this paper. Attributes not defined in the standard may therefore be different for different software products. The standard provides a framework for organizations to define a quality model for software.

External metrics are applicable to software running.

Internal metrics do not depend on running the software. They are applied statically. Quality in use metrics is available only when the final product is used in field conditions. Ideally internal quality determines the external quality.

ISO 9126 is replaced in 2005 by ISO 25000.

The standards are constantly updated to meet all user requirements better. Adapting to standards is a vital activity for any organization performing software.

3. Software structure for the analysis of structured text entities

Structured entities are those that have a fixed structure defined by internal building rules [6]. These can have a variable number of components, but only if they respect the base structure. The entities are verified using a series of rules and indicators that measure the degree in which the entities respect the rules are calculated. According to the information from Figure 6 the messages given to the users are:

- localized; localization allows users from different geographical locations to access the application and the messages they receive are in their own language;
- clear; the messages indicate the nature of the problem that appeared so that the users know what they must do to improve the situation;
- concise; shows the location of the problem that appeared and indicates ways of improving quality to the users;
- give suggestions on modifications to be made on the analyzed entity to

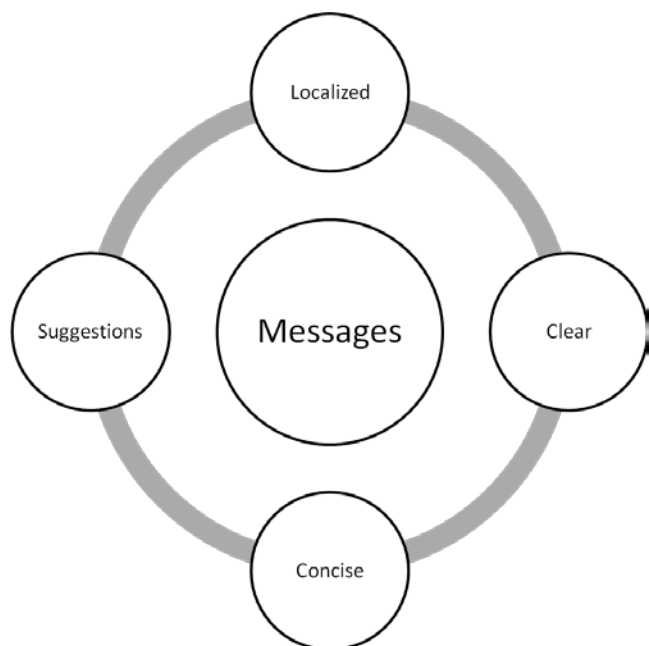


Figure 5 – Messages characteristics

satisfy the quality requirements.

It is also offered the possibility of saving the messages and entities. The application implements a mechanism for data collection regarding users' behavior measuring durations, frequencies and actions. On the basis of the collected data improvements are permanently made to the application consisting in restructures, completing functionalities, completing incomplete modules, modifying or adding messages.

The application uses the class structured entity that stores the user's entity and ensures all the methods and functions necessary for the analysis. The graphical interface of the application allows the transfer of data between the user and the structured entity class. Also, after analyzing the input data, the structured entity class gives the user messages through the graphical interface. The application gives a score from 0 to 10. The score is obtained after the analysis of the following factors:

- the lexical density is defined as a report between the number of different words and the total number of words; the lexical density has average values in the interval [0.4, 0.6]; for a value greater than 0.5, the application gives the maximum score for this component; for smaller values, the score is weighted;
- the number of citations in the text reported to the size of the text is also a very important component; the application gives the maximum score only if all the entities from the bibliography are cited in the text; in the contrary case, the score is weighted with the number of cited works;
- the relevance of the keywords reported to the text of the entity is also analyzed; the keywords are the ones synthesizing the essence of the entity and thus, these must be relevant for it;
- the correctness of the entity is determined through the comparison of the entity's words with the standard dictionary of the language it is written in; as the entities can have elements defined by the users, a degree higher or equal to 80% is accepted for the maximum score; if the correctness level is in the interval [0.5, 0.8] the application gives a weighted score; if the correctness of the entity is lower than 0.5, the application gives the score 0 as the entity is not correct;
- the number of references reported to the length of the entity is very important to ensure that the author documented for its realization;

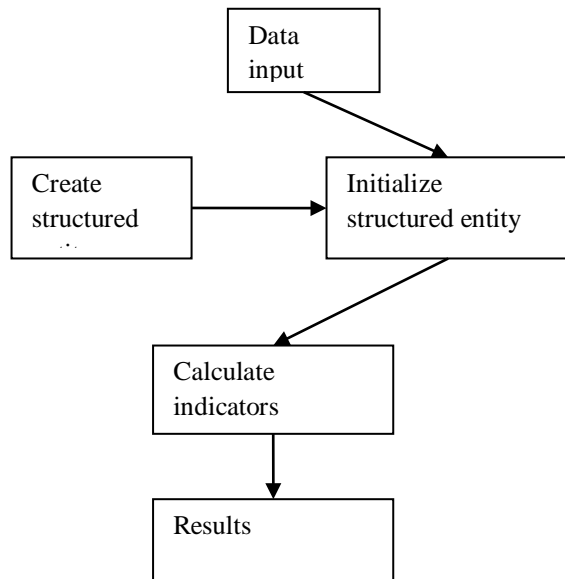


Figure 6 – The functioning of the application for the analysis of structured entities

- the equality of the chapters is also a measure used by the application to ensure a unitary entity with parts relatively equal; for this it is used the medium value of the size of chapters from which these can derive with a predefined value;
- the figure and table citations indicate their necessity in the entity; if a figure or table is not cited in text, this means it is not necessary and it can be eliminated; the application measures the degree in which the figures and tables are cited in text.

The score accorded to the entity is obtained as a weighted average of the influence factors. According to its importance, each factor has a greater or lower contribution to the final score. Figure 6 highlights the information flux and processing done by the application for the analysis of the structured entities.

The data processing durations are very important for the users. As the bandwidth is variable from user to user and the developer can't influence this aspect it is very important that the time necessary for the application to process data and show results to be very short. For the application for the analysis of the structured entities, available online, the durations necessary for different analysis are given in Table 1. All durations are given in milliseconds.

Table 1 – Processing durations for the application for the analysis of structured entities

	Final indicator	Indicator 1	Indicator 2	Indicator 3	Indicator 4	Indicator 5
Minim	42	20	2	5	9	6
Maxim	105	68	7	9	12	9
Medium	55.7	28	3.4	6.8	10.5	7

The processing durations are automatically recorded by the application for the analysis of structured entities. The five considered indicators take part in the calculus of the final indicator. For the indicators the situations in which these can't be computed because of the data input or the logic of the algorithm are ignored. Only datasets for which the complete set of indicators can be computed are taken into account. The durations depend greatly on the dimension and complexity of the structured entities the users insert.

It can be observed that the lowest duration necessary for the application to compute the final indicator is 42 milliseconds. This response time from the server component is very small. The maximum time of 105 milliseconds is also unnoticeable for the client. The data are based on the information recorded after the analysis of 255 structured entities. It can be observed that the average duration for the computing of the indicators is low, as the final indicator has an average value of 55.7 milliseconds.

In order to improve even more the durations necessary for the computing of the indicators, it is necessary the optimization of the algorithms or the use of a programming language of higher performance. The improvement of these durations is not necessary unless an analysis of a very large set of entities is required. For a database with millions

of records, a difference of a few milliseconds for each record can lead to hours for the whole operation.

4. Conclusions

Citizen oriented applications appeared in the context of the knowledge society as a result of the users' requirements. These differ from the traditional applications in many aspects. Most of them highlight the citizen as the central element and stress the development of the application on fulfilling the citizens' requests. The target group for these applications is very large, usually made of all the citizens of a country. This leads to a wide range of structures for citizen oriented applications. Structures of citizen oriented informatics applications are simple linear, simple linear with backwards links, simple linear with multiple links and tree structures with simple, double and multiple links. The quality of the applications is assured through the compliance to quality standards. The oldest and most known is the McCall quality standard. Another two models were later developed. The improvements they brought were not of capital importance. The application for the analysis of the structured entities helps the users in developing such entities by guiding them through clear and concise messages. The application gives a score for the analyzed structured text entity. This is obtained through the weighted mean of values obtained for different indicators. The application is fast and the users don't notice the processing time for just one analyzed entity. Future improvements aim to give the users the possibility to supply the application with different files as input. This will shorten the time needed for the users to solve their problem and also improve their experience.

5. Acknowledgements

This article is a result of the project „Doctoral Program and PhD Students in the education research and innovation triangle”. This project is co funded by European Social Fund through The Sectorial Operational Program for Human Resources Development 2007-2013, coordinated by The Bucharest Academy of Economic Studies.

6. Bibliography

- [1] C. Y. Yoon, "Measures of perceived end-user computing competency in an organizational computing environment," *Knowledge-Based Systems*, vol. 22, no. 6, pp. 471-476, Aug. 2009.
- [2] L. J. Harrison-Walker, "The measurement of a market orientation and its impact on business performance ," *Journal of Quality Management*, vol. 6, no. 2, pp. 139-172, 2001.
- [3] C.-S. Ong and S.-W. Wang, "Managing citizen-initiated email contacts," *Government Information Quarterly*, vol. 26, no. 3, pp. 498-504, Jul. 2009.
- [4] G. Daniel, *Software Quality Assurance From Theory to Implementation*. Addison Wesley, 2004.
- [5] Wikipedia. (2010, Feb.) ISO 9126. [Online]. http://en.wikipedia.org/wiki/ISO_9126
- [6] I. Ivan, B. Vintila, and D. Palaghita, "Types of Citizen Oriented Informatics Applications," *Open Education Journal*, no. 6, 2009.