

PhD Thesis Review on The orthogonality of the structured entities

by Constantin Daniel MILODIN

Keywords: orthogonality, structured entities, quality, orthogonality metrics, alphabet, vocabulary, software.

The orthogonality of the structured entities was elaborated by **Constantin Daniel MILODIN**, under supervision of PhD, Univ. Prof. Ion IVAN, in the Academy of Economics Studies in Bucharest, and it was presented for public criticism on the November 27, 2009, with the permission of the Cybernetics, Statistics and Informatics Faculty.

The thesis' main objective is to set up a set of techniques and methods for evaluating the structured entities, with practical application

The PhD thesis consists of eight chapters that cover both theoretical aspects, and a practical application.

In Introduction, Mr. Milodin presents the necessity to approach such kind of theme, the thesis' objective, defines the orthogonality concept, and presents all the aspect included in the thesis.

In the chapter *Current stage of the researches in the domain of orthogonality* are given the results in the literature, both on the internal and external plan. Are also identified the research directions. Books, articles published in specialized magazines or in volumes for conferences, scientific communications relevant for the approached issues were studied.

The chapter *Structured text entities* presents the basic concepts used, defines the concept of structured text entity, presents criteria and modalities for the text entities' classification, means for the concatenation of the entities. Are identified the operations performed with the text entities and are proposed solutions for their improvement. Are also studied the links between the component concepts of text entities and the way of building the text entities based on the used concepts.

The structured entities create and implement storage and processing forms which are used for the management of the information. The organization of the information is performed according to the characteristics that define and customize it. The entities 'organization reflects the existent associations between data. The structured text entities are used to organize the text entities.

The organization is made on different fields of activity, the definition of the text entities being made under the criteria of structuring and the operating information. The structured entities are effective ways of storing and collecting information.

The determination of similarity is based on the structure analysis of the level. This is calculated on each level and on the basis of partial indicators it is determined the global indicator of similarity of two or more structures.

The implementation of techniques for studying the orthogonality of the structured entities requires a very good knowledge of the applicability domain, same as the recognition of algorithms and techniques for studying the orthogonality.

The concept's implications are related to customizing the reference areas using structures that redefine the concepts used at macro level in micro concepts, implemented through the structured entities.

The operations of creating, classification, concatenation, comparison, extraction, dissemination and suppression made on the structured entities contribute to improving the performances of using this type of entity.

The quality of structured entities is the chapter in which is defined the system of implemented quality features. Also, the used quality characteristics are customized on the texts. In order of being qualitative the data must meet the quality criteria. For this purpose it is defined the quality characteristics system. Besides the quality criteria, for accuracy, the data must be created, retrieved or processed so that does not contain errors. Are also considered the negative implications that a low level of the quality data has.

The solutions for ensuring a better quality of the data are related to:

- removing the data about whose origin are known little information, meaning that they are don't derived by calculations of other data but are taken over;
- strict checking the data taken directly from unsecured sources, by imposing the criteria of dimension, source formatting, use of standards required for writing the sources;
- eliminating the unnecessary copies by checking the degree of differentiation between them and the stored sources;
- using qualified staff who know the significance of the data they work with;
- training the staff and a good description of the application for the knowing the implication the data have within the organization and to know what data should be loaded in the fields of acquisition;
- standardizing the manner of stocking and verifying the data before saving.

It is placed a great emphasis on the data source. These are taken over using many ways, but the data that will be used in calculations, previsions, must have an exact source.

In order to ensure coherence and correctness, the data are subject to statistical processes to eliminate the aberrant values and which do not fall within a certain margin of error.

The data are important for decision-making. In order to provide data with coherent values in terms of decision making are constructed methods for the collection, analysis and interpretation of data and numerical information. Data should be bounded by the source, must be designed so that the decision to serve a decisional purpose.

The data quality characteristics are: representativeness, variability, uniqueness, cardinality, accuracy, orthogonality, relevance, completeness, correctness, reliability, adaptability, consistency, flexibility, validity. The implementation of the quality characteristics system ensures the data quality testing and filters the data used depending on the considered quality characteristics.

The compliance with the quality characteristics should be tested at periods, as the quality requirements evolve, and the data are always subject of change of the implemented software.

The capacity of the entities to meet the quality requirements defines the entities' *compliance* with the quality characteristics system. At the bottom of studying the quality is the work of identifying the correct data and eliminates the errors.

The structured entities are the entities that include a layout data structure, a tree representation of the concepts. The entities' quality is a concept that encloses quality aspects of the data itself, the texts, the entities and the structured entities.

The quality of the entities does not include the data quality, meaning that if an entity complies with the quality characteristics required for entities, then automatically the entity data are qualitative.

The quality attributes of the entities are given by the opportunity, comparability, homogeneity, completeness, graduality, abstraction.

The quality characteristics of the texts reunite the criteria applied to the texts.

Are taken into account the quality characteristics relating to the text's affiliation to a particular domain, the use of the keywords in a predetermined dimension, the degree of use of a vocabulary words.

Also, through the quality characteristics is verified that the text contains a certain type of separators or does not contain separators from a multitude considered inappropriate.

The modelling of the texts' quality is made by taken into account structural aspects and their underlying aspects.

The quality of structured entities is the result of quality system implementation and construction of individual entities that individually meet the quality criteria.

Within each entity, the concepts are built individually and according to strict standards of the quality targets. By defining and taking over the concepts from lower levels to higher levels are embedded the quality characteristics.

Respecting the quality standards within the entities ensures achieving the objectives for which they were built and leads to results delivery underlying organizational decisions.

In the chapter *The orthogonality quality characteristic of the text entities* is made a definition of the concept of orthogonality, are presented the methods for identifying the degree of orthogonality of the entities, are defined the indicators used in studying the orthogonality, and are presented ways to increase the entities orthogonality. Are presented the modalities of building the indicators used in the studying the orthogonality.

Through the orthogonality study are identified the authenticity elements of the text entities, and also the degree of similarity for the non-orthogonality data.

The orthogonality studies the degree of similarity between two or more entities. Using this quality characteristic is determined the degree in which the entities differ from each other.

Whether the E_1 entity, defined with the C_1, C_2, C_n characteristics. In order to be studied the orthogonality of two or more entities, as a first condition that must be respected, they should have the same structure that is to be defined by means of the same features.

Starting from the characteristics of the entities, are defined their quality indicators. Based on these indicators is constructed an indicator that takes into account the values of the component indicators.

One area in which the orthogonality is very important is programming. Programming languages are projected so that their orthogonality is maximum, meaning that the implemented concepts, terms and keywords used are unique not to create confusion, both among users, and especially not to put the compilers in the position of not knowing what a line code supposes, because of several possible options.

Comparing two entities is reduced to reporting one entity to the other that is to identify the common parts and parts that differ. Are thus compared the corresponding characteristics of the two entities.

In order to study the orthogonality it is defined a normalized indicator of orthogonality in the range $[0, 1]$, which takes the following values:

- 1, if the elements are orthogonal, i.e. have nothing in common;
- 0, the elements are identical, i.e. have no different values for any feature.

If the indicator's value tends to 1 means that the data sets contained by the two entities tend towards orthogonality, and if the indicator's value is close to 0 means that the data sets have many identical elements.

The texts orthogonality is determined as internal and external orthogonality. The internal orthogonality sets the measure in which, as part of the text, the words used are repeating, the existent links between words as part of the text.

The external orthogonality shows the existent differences and similarities between texts.

It is noticed that the object oriented programming uses concepts that help structuring the concepts underlying the development of software products. Moreover, the main concern is represented by the existent links between data, the means of creating links between them so that the messages between objects to move effectively.

Starting from key concepts, such as variables or functions, the object oriented programming allows the achievement of superior concepts by encapsulation and inheritance. Through encapsulation is realized a structuring of a class' content and through inheritance is realized the structuring between the classes. Are created in this way, starting from a simple class, complex classes that are based on taking-over the characteristics of the simpler classes. The implementation of these concepts leads to the creation of complex applications that take modules already created by defining the derived notions. The application is structured in this way on modules operating independently and that lies at the basis of the development of new functions within the basic application.

The structured entities become therefore functional and above all become a starting point for developing and improving existing solutions.

The similarity between the manner of structuring the content of a class and the entities contributes to the construction of efficient methods for determining the orthogonality of the structured entities.

Studying the orthogonality takes into account the history of program-products, the orthogonality indicator being calculated for a large period of time.

This requires building a deposit of text entities based on which the orthogonality is calculated.

The deposit of entities is based on the need of stocking the text entities and calculating the level of orthogonality for both a new entity, compared with the already stocked entities, and for the entities stocked within the deposit.

Creating a deposit of entities is done by taking into account the work objectives and its character of generality.

Approaching new areas increases the orthogonality degree by combining new texts with methods of applying them in the area of interest.

The development of texts that present subjects that were not treated primarily involves the construction of new criteria for the assessment of the entities' value using a deposit of entities.

The orthogonality of the texts is directly influenced by the words' degree of diversity. By diversity we understand the number of different words used as part of the text, and the words' frequency of occurrence.

The lower the frequency of occurrence is and the bigger the number of words, the level of orthogonality of the text increases.

By diversity we also understand the level of affiliation of the text to a vocabulary, respectively how many words from the vocabulary are found in the text. Also in this case, the orthogonality is directly affected by the values of the text's diversity.

To increase the programs' orthogonality are implemented solutions such as:

- the specialization of programs by types of issue;
- the creation of libraries containing specialized programs and procedures that are called according to the problem's particularity;
- the avoidance of unnecessary use of the resources and their call in accordance with the problem's particularity.

The increase of the text entities' orthogonality composed of source programs ensures a higher quality of the source programs. This aims to identify new areas that require software products for the activities' automation.

Creating new software that offers results that meet the criteria of evaluation helps improving the orthogonality indicator.

Since as part of the programs are used free procedures stored in the store of entities, the orthogonality is influenced by the manner in which these procedures are called within the programs and the manner in which they are integrated.

The increase of the orthogonality of the text entities leads to the increase the quality of the texts' indicators. A good level of the quality indicators ensures an orthogonality high level, since the link between orthogonality and other quality indicators is mutual.

The orthogonality underlies the concepts of originality and reuse, which helps defining texts superior from the qualitative and functional point of view.

It is build a system of indicators in the chapter *Orthogonality metrics*, a system that is used for testing the quality of the text entities. The defined indicators are analyzed in terms of sensitivity, representativeness, and catastrophicity.

Also, in this chapter, using the orthogonality metrics and their indicators are studied the software products' orthogonality and databases' orthogonality. In this chapter's summary are defined the orthogonality's categories of indicators used within the text entities and is made an illustration of their use.

Are implemented techniques used to identify the degree of orthogonality between more software products and more databases.

In *The analysis of the orthogonality of structured entities* are presented results of studying the orthogonality of the elements underlying the concept of structured entity: alphabets, symbols, article, dictionary, texts, and projects, same as establishing the orthogonality level of a text in relation with the specific vocabulary.

In the chapter *The assessment of the orthogonality of homogeneous lots of structured entities* are presented criteria for building the homogeneous lots and the experimental results related to application testing. It is also defined the problem for whose solving is made the ORTOES software product, is structured in modules the software product, according to the solving issue, are presented the used resources and the modules of the software products, are defined solutions for solving the problem and is implemented the software product that is based on the presented solutions.

The chapter *The dissemination of the original solutions* includes concrete elements of establishing the author's achievements as part of the research approach. Are presented modalities through which the works were disseminated, and the author contribution in researching the orthogonality of the structured entities, emphasizing the original methods proposed, the obtained results.

The paper ends with *Conclusions* where are embodied the results of the study realized, and the aspects for following to improve the results.

The thesis includes the original and valorous results of a young researcher and it is a mark for his future as a scientist.

Gheorghe NOSCA, PhD

r_g_nosca@yahoo.com

Association for Development through Science and Education