

Digital Content Audit – Collaborative Process

Prof. Ion IVAN, PhD

Academy of Economic Studies, Bucharest, Romania
ionivan@ase.ro

Daniel MILODIN, PhD

Academy of Economic Studies, Bucharest, Romania
daniel.milodin@ase.ro

Mihai GEORGESCU

Academy of Economic Studies, Bucharest, Romania
mihaita.georgescu@gmail.com

***Abstract:** It is presented the concept of the digital content. Are listed the steps of achieving the digital content. In order to become operational, the digital content is subjected to the auditing process. Are presented the requirements of the auditing process. Are developed the auditing metrics of the digital content. It is proposed a software structure meant to validate the auditing process of the digital content. It is presented the concept of collaborative process.*

***Keywords:** digital content, metric, audit, software auditing, collaborative process.*

1. Digital content

The digital content is represented by all the forms of data storage from the informatics applications. The files with data numeric, alphabetic, alphanumeric, sounds, evidence retrieval, documents form the digital content. The files with documents, with source texts, with programs executable form digital content. The portals, the websites, the web applications of any kind operate with digital content. The digital content is characterized through:

- **homogeneity** - the feature of digital content of being composed by data of the same type; of structuring and storing the data according to the same model; the processed data using the digital content are stored based on a model, respecting the model and improving further data processing; the homogeneity feature concerns :
 - o using the same units of measurement to express the data value;
 - o using the same procedures for data collection;
 - o data storing based on the same structures as part of the same digital content;
 an important role in meeting the characteristic of homogeneity belongs to the data processing module; data certification when processing ensures respecting the conditions imposed by the structure of the digital content, the terms of order, value, type of data, the processing manner being an extension of the limits imposed on the data given by the user;
- **completeness** - characteristic of the digital content that shows the extent to which its fields contain values; the digital content is completed when all its fields have been loaded; the completeness shows the ability of the digital content to provide information that normally should be found stored; the completeness ensures the existence of complete data regarding a subject's treatment; the completeness concerns the extent to which the attributes of a digital content have assigned values; the lack of the attribute's value has different meanings depending on the attribute's character:

- necessary, optional or inapplicable; the completeness concerns the coverage of the standard specifications of the reference domain, the absence of one or more elements leading to incomplete implementation approach;
- **correctness** - the feature used to certify if the data are related to a certain context, an organization, a category; the data are values assigned to records, staff, bank accounts, stock transactions, payments; viewed globally and simplistic, the data are considered as part of a working structure; for a coherent representation are required more structures that are connected to each other based on some keys with common values; this work modality applies in the case of relational theory; the correctness certifies the data belonging to a field and respecting the quality conditions imposed by it;
 - **orthogonality** - studies the degree of similarity of the digital content; through this quality characteristics it is determined the extent to which a digital content is different from other digital content; the concept of orthogonality comes from the mathematics domain; the orthogonality is studied on the basis of orthogonality criteria; with these criteria are outlined the characteristics that have the same value for the studied digital contents and are determined the levels of similarity between them;
 - **comparability** expresses the quality characteristic that allows determining the digital content with lower or higher values depending on the purpose of the comparison operation; the characteristic of comparability establishes an order of the digital contents considered under the imposed criteria and the recorded values; the digital contents' ordering is done by their disposal starting from the lower values and continuing with the higher ones, or conversely, depending on the mode of comparison; the operators that ensure the comparability are $<$, $>$, $=$, applied to the data stored as part of the digital content, the application method differing according to the data specific; through comparability is ensured the digital contents' arrangement according to the values of the key field and is decreased the time needed for other operations: through sorting is ensured an optimal time necessary for a content's retrieval; the optimization of the comparability characteristic improves the operations that use sorted values of the digital content: retrieval, storage, modification, updating, deletion, all based on retrieving the wanted item; the comparison is done using the stored values; a condition to compare is related to the type of data: the compared data must be of the same type; a further condition is that the data must have valid and comparable values;
 - **graduality** is the feature through is ensured the progressive approach of the used concepts; graduality ensures the definition of concepts by using concepts that have already been presented; are like this avoided the situations where concepts are defined by self reference;
 - **opportunity** of a digital content refers to its availability and use at the right time, when the user needs; the opportunity is a function of actuality and volatility; a formula to calculate the opportunity level is [IVNOT99] : $OP_R = \sqrt{AT_L \cdot VL_T}$ where AT_L represents the actuality and has values between 0 and 1; the 0 value indicates that the entity has just been created or updated; VL_T represents the volatility and takes values between 0 and 1; the 0 value means no change has been made; the opportunity value, OP_R , is comprised between 0 and 1, where 0 value is the best value and the value 1 the least good; through opportunity is identified the characteristic of the entities to respond promptly to the customer ,s requirements by implementing a superior management of the issue's approach.

The digital content is organized depending of the storage specific of the processed formats. Thus, are encountered the following manner of organization:

- the article, the digital content being structured on content lines, specific for working with texts; in this way is build a proper background for stocking and processing the text units;
- the block, builds homogeneous structures, consisting of several articles, organized on identical types or different types; are used to create ways of working with data characterizing large and various collectivities;
- databases are generalizations of the block type structures, presenting superior forms of organization and processing.

Depending on the specific of each digital content are chosen and allocated its materialization forms, taking into account aspects that depend on the procesarii efficiency and improving the

Depending on the specifics of the digital content its materialization forms are chosen and allocated taking into account aspects of processing efficiency and improving the work with users.

2. Development cycle of digital content

To obtain digital content, the stages of a development cycle should be followed.

Defining digital content involves specifying the C collection consisting of C_1, C_2, \dots , CN elements and the characteristics and complete description of each element of the community.

The characteristics of the C collection are Q_1, Q_2, \dots, Q_M .

For the digital content of the description of a country's population the following characteristics are taken into account:

- personal identification number – is an identification number, unique; the mechanism of building the code should ensure its unique feature and validation according to the required criteria;
- last name and first name – are texts to identify persons; name and surname are parts of an unique identification code;
- address – represents the official address where the person having the above characteristics lives;
- serial number of ID - to uniquely identify the person having the above characteristics;
- administrative-territorial unit to which the person is subordinated.

In the case of structured entities, the digital content shows specific models of definition.

In [1] it is presented the mechanism of building the digital content both structured and with links.

Consider the set of books that describes the C++ programming.

Each book is characterized by structures like: introduction, chapter 1, chapter 2, ..., chapter n , as shown in figure 1.

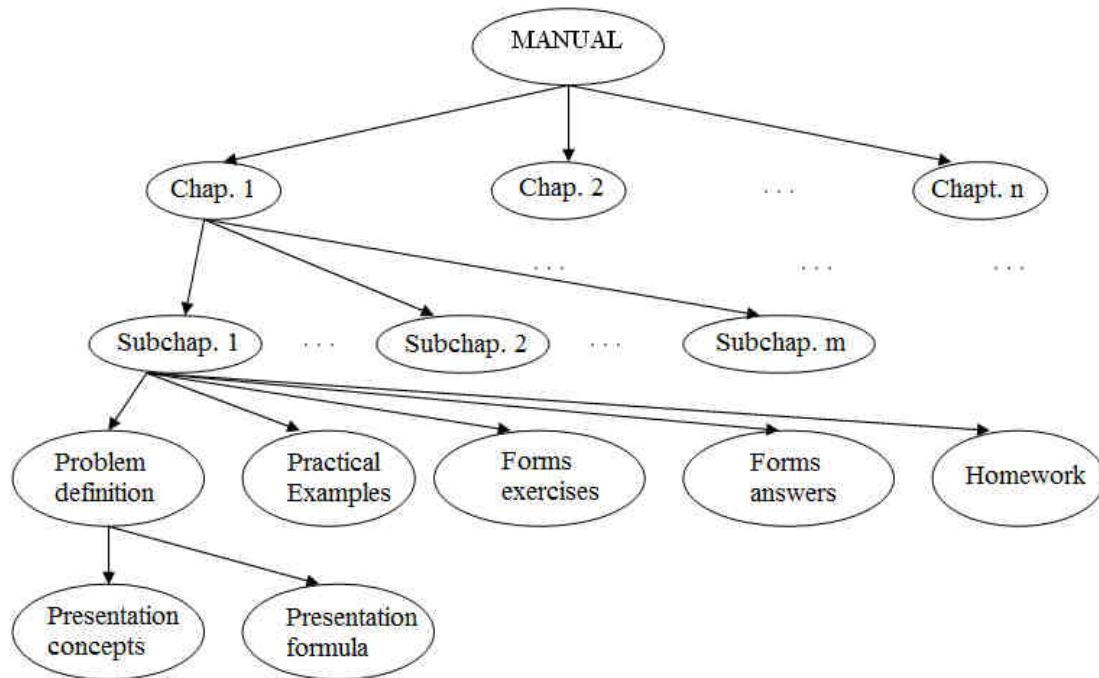


Fig. 1. The manuals' structure arrangement

The chapters are organised in subchapters. The subchapters are divided into paragraphs.

In order to edit a manual, it should follow a predefined structure, a certain required template.

The C++ manuals have the information organized starting with the basic concepts and continues with advanced concepts. An essential characteristic of this principle is that over the presentation of material all the complex concepts become basic concepts in the light of explaining new concepts.

This will ensure learning through learning.

In this way, a relation between concepts develops, called precedence relation which explains the dependence between concepts.

Assimilation of existing knowledge in the books leads to creation of chains of dependence.

The required structure of C++ programming manuals stored with the help of structured entities branches by raising the connection degree between the consisting levels and sublevels.

The precedence link builds the precedence graph related to the concepts inside each chapter.

Using the precedence graph is established the order of presenting the knowledge for ensuring the fundamentals required for its assimilation.

In the graph of precedence are analysed the implications and connections necessary to establish a network as similar to the chain of everyday life knowledge.

In figure 2 it is presented a precedence graph for the knowledge defined in the C/C++ programming manual [4].

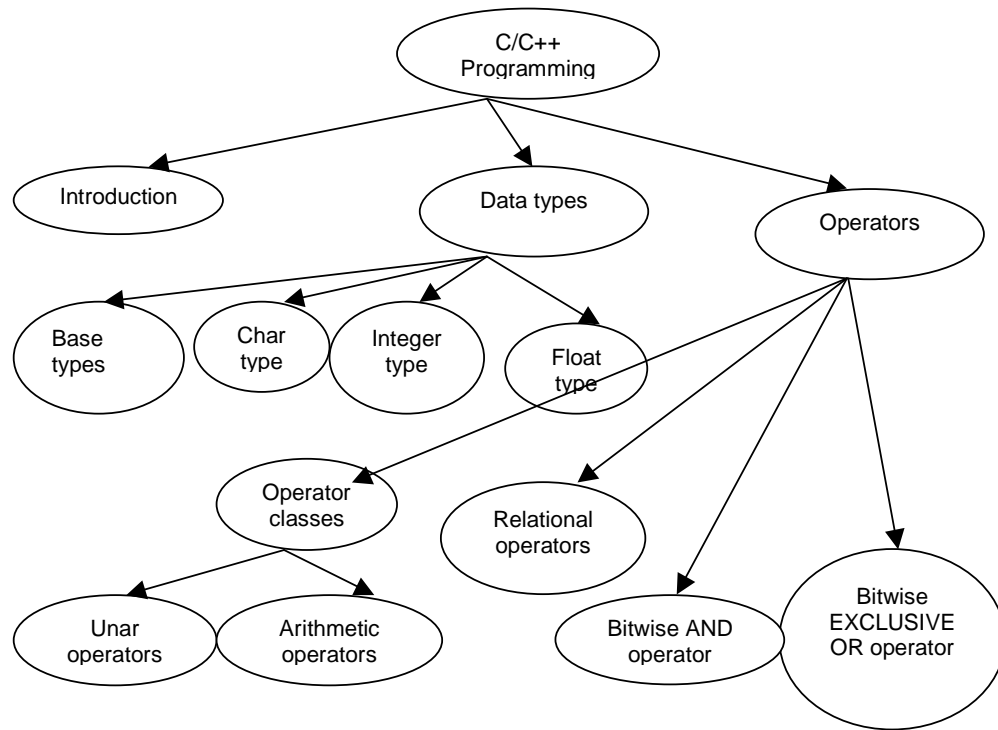


Fig. 2. The precedence graph of knowledge from the programming manual

From the precedence graph in figure 2, the corresponding matrix of precedence is built. Based on it, precedence computations are done related to the number of days necessary to present all the knowledge, to the order of presenting the knowledge, to the identification of the critical knowledge whose time limit is reduced. Also it is done a schedule with the presentations that takes into account the existing dependencies.

Based on the template of the manuals it is done a structuring that customizes this concept for each programming manual. In figure 3 it is implemented the structuring of the programming manuals for different years of study.

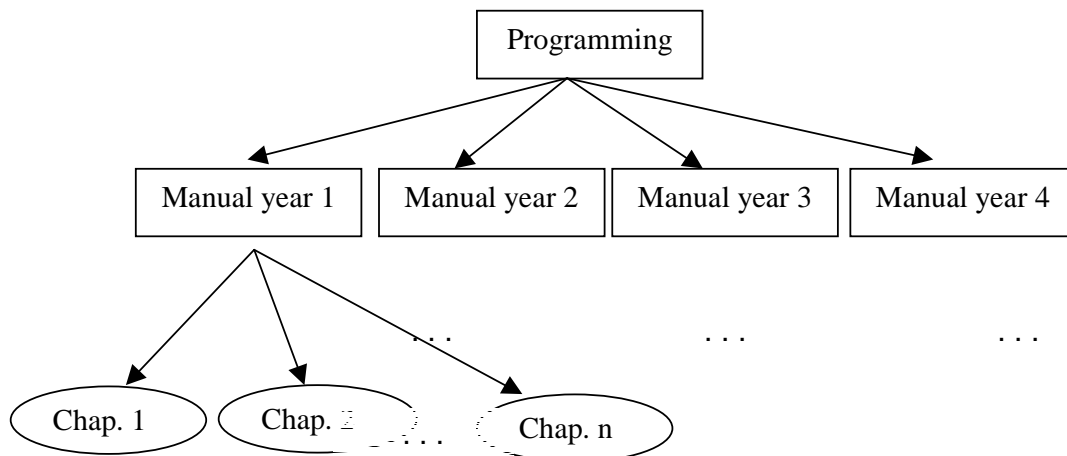


Fig. 3. Programming manual structure

Figure 3 shows the organization of the manuals, respectively the allocation of knowledge based on the difficulty level.

The differentiation of the structures is determined by the particularities of the manuals. There are manuals that cover the area of exact sciences: mathematics, physics, chemistry and that require acquiring basic concepts in order to achieve the transition to new notions explained through the basic concepts and textbooks covering the area of arts: music, drawing, in which the concepts have chronological meaning and not of dependence.

Designing books is a collaborative [5] process, taken into consideration in order to obtain a controlled level of redundancy.

Building the digital content is achieved according to the following specifications:

- defining the target group, namely the identification of individuals who are interested in the data stored by the digital content, determining the persons who provide data and informations, namely the ones that uses these informations;
- development of specifications that define the area of each field, the data acquisition, setting specific errors that occur, ways of validating and correcting of data sets, both by specialized users and third parties without specialized training regarding the use of data and their validation;
- creation of structures for storage of digital content; the structures for formatting the database structure are imposed; the composing fields of the digital content are implemented;
- loading the digital content; are taken into account aspects related to validation, to define the required memory for storing the digital content; this step ensures that the implemented quality requirements are achieved regarding the stored data, thus ensuring high efficiency in processing and handling them;
- assesses the quality of the digital content;
- conducting operations on the digital content: hierarchy, reunion, intersection, insertion, comparison, parsing, counting, reading, selection, concatenation; the operations are focused on the processing of individual elements and on the whole digital content, based on some criteria; the management of performed operations on the digital content should ensure implementation of a control key which should ensure the correct implementation of operations;
- building a system of auditing to certify the correct implementation and efficiency of modules for working with digital content;
- ensuring the maintenance process who will allow adding new informations to the digital content and who will also allow the expansion of digital content by defining fields used to store alternative values, enabled or disabled based on the need of using them;
- implementing the concept of engineering software, meaning the use of components from the digital content by which new variations of the digital content are developed.

The implementation of quality principles in the processes of acquiring, storing and processing ensures the efficiency of working with components that form the digital content.

3. Quality features of the digital content

Digital content is mainly composed of data in various forms of storage. The concept of data quality has different meaning depending on the area in which is used. If the data are used to achieve statistical situations, it is important that the data are representative, to be homogeneous and their spreading should follow certain rules. When referring to the data stored by an organization during its peak sales period, they should inform the personnel about

the stock being wound up, about the top demanded categories of products, in general about the trend of sales.

Quality is an attribute of a product or a service based on their ability to meet customer requirements. Starting from the wide range of products and especially customers, there is a wide variety of concepts in terms of data quality.

Quality assurance of the data determined the introduction of data control systems.

Because the transmitting and processing of data can generate errors, the reception confirmation procedure was introduced. Control is achieved by checking data at input time or acquisition time by validating them. The membership to a set is checked, to a class of numbers in order not to acquire irrelevant information. The way data is copied is controlled. The possibility of copying of information for a very large number of times leads to the inability of controlling the correctness of data and involves respecting legal aspects of paternity of the work.

Solutions for ensuring a good quality of data should follow these guidelines:

- remove the data whose origin is uncertain or little information is known, meaning that they do not derive by computations from other data but are taken directly;
- strict checking of data taken directly from uncertain sources by imposing criteria regarding size, source formatting, use of standards required for writing sources;
- elimination of unnecessary copies by checking the degree of differentiation between those and the stored sources;
- use of qualified staff who know the significance of the data they work with;
- train of staff and a good presentation of the application for knowing the role the data have in the organization and to know what data should be entered in the acquisition fields;
- standardize the way data is stored and verified before saving.

The source of data is very important. Data are taken using many ways, but the data to be used in computations, forecasts, should have an exact source. In order to ensure consistency and correctness, the data are subject to statistical processes to eliminate exaggerate values and who do don't fall into a certain margin of error.

The data are important for decision making. In order to provide consistent data values in terms of decision making, methods for collecting, analyzing and interpreting data and numerical information are created. Data should be delimited from the source, must be designed in such a way that it will serve for a decision purpose.

A key feature of quality of data is *representativeness*. Starting from this is extrapolated from a small number of data to a large number of data, retaining the characteristic quality because of the data from which the processing starts summarizes the values of quality for the entire collection. Using representative data is a key factor in increasing the performance of the organization. Starting from qualitative informations, data analysis in turn also provides qualitative results.

Another key feature of the quality of data is *variability*. This is seen from the fact that the data are results of a relatively large number of factors of the same nature but different directions.

Quality of the data means switching from large data, individual, to results specific to a set of data, obtained by exploiting individual data. The data are used by organizations in different ways and purposes. The main purpose of data is to supply the necessary knowledge in decision making process. Starting from this premise, data quality is defined as that feature that takes into account aspects of the value of data. A data that is not statistical qualitative but who offers enough knowledge is in some cases considered a qualitative data when it is profitable.

The process of data processing is done using statistical methods. A commercial store which wants to increase sales based on history of prior periods decide which is the most profitable and invest in advertising, creating better conditions for viewing the goods. The decision in this case is made using statistical instruments, like mean, dispersion, deviation or correlation. In other cases, such as information about the layout of enemy troops on the battlefield, the statistics does not interfere with processing of data. Data quality is given by the accuracy, reliability and accuracy of them.

The concept of data quality varies depending on the area where applicable. In decision making, the measures for ensuring data quality are summarized in the following areas:

- prevent processing with low representativeness, with a high variation;
- identify data that are *imperfect* and the cases who generated them;
- actions to be taken to correct the data and to ensure consistency.

An approximation method for consistent data is the application of global standards, as collaborative processes.

Data quality involves how data are used, but also how they are structured. The usefulness of the data is the first test that directly checks the quality of data. Data quality refers to data entry errors. Altering data occurs also by redefining fields. By redefining are changed field types. If age was initially saved as a numeric field, then it is redefined as a text field. New types of management systems of databases, both relational and objectual, do not encourage such a practice. Change of fields has serious implications on the accuracy of data if the procedures to process these fields are also changed. Switching from numerical to characters does not require large differences in the way of processing, because the character is seen as a numeric field. Differences occur in storage and when that storage capacity is exceeded. All this combined with the fact that the end user has no effective access to the error message or the source of error is leading to a disruption in the working environment. Transferring data to a new database is done either by following the structure of tables in which data are taken either by implementing a conversion algorithm that transforms data and provide statistics in order to identify the resulted errors.

Data quality inconsistencies processed by using digital content are presented:

- *incorrect data* – in order for a date to be valid it must belong to the domain of validity; months must be between 1 and 12, one month must be between 1 and up to 31 days, correct data depends on how validation is completed;
- *inaccurate data* - a date is correct and not accurate, inaccuracy comes from the not taken into account all aspects, a higher profit rate is incomplete without specifying the level of elements involved in the calculations, the accuracy of the interconnected data is difficult to obtain and controlled and that is why additional conditions of membership should be created;
- *breaching of rules relating to the practical aspects of data processing* - data processing rules require validation conditions related to membership to a certain range of values, compliance with a range of values and the introduction of unduly repeated the same value, setting *expiration date field* to a certain value makes any higher value placed on that field to be incorrect;
- *inconsistent data* - uncontrolled data redundancy leads by default to inconsistency data; with this problem are facing organizations and is especially seen in the data related to customers; in the situation when the record is not uniquely identified and not it is not checked in the data acquisition process, if they stored duplicate data occur; the causes that can generate such situations of inconsistency are related to:
 - o inaccurate request of information;
 - o presenting incomplete requirement of informations that needs to be provided by the user;

- inconsistency of the organization in terms of value of information supplied by customers;
- building client interfaces that generate confusion and do not present in a clear way the requirement for data that must be entered;
- *incomplete data* - each department within the organization intends to take over first the information that is directly concerning that department, omitting to take the remaining information, by which another department is interested; a bank is taken mostly information on work or income, rather than those related to individual client;
- *non-integrated data* - data stored in the organization over time, are transferred to new systems; primary keys are not the same type, the same length, the new systems are not always designed so as to take over the old data, so old data become unusable or at best the old and new system is used in parallel.

Quality characteristics of digital content are given by:

- *accuracy* expresses the extent to which digital content is useful and accurate; accuracy depends on the order in which digital content is used; for digital content that stores numerical data, accuracy is determined as the difference between the value obtained and the standard value; if the digital content is text, accuracy is derived from comparing the content of a work with the standard specifications of the areas addressed in that paper; by standard specifications means all concepts, hypotheses, models defined and elaborated in an area and which were accepted by the scientific community as correct; there are cases when the introduced concepts bring innovations not covered in standards; in this case it should proceed to their analysis, and if the conclusion is that the elements that they work with do not contradict reality the standards are updated;
- *relevance* shows the ability of the digital content to express significant information; the importance of digital content derives from its use to produce decisions; the use of a correctly digital content, well selected, which is seen as a reference for the field covered, contribute to an efficient analysis; use of a digital content that is inaccurate, irrelevant, incorrect, not corresponding to reality and which does not correctly reflects the performance of the organization leads to the propagation of errors throughout all future processes;
- *standardization* of digital content enforce a very important quality characteristics: portability, meaning the ease with which digital content is transferred from one medium to another; through standardization the digital content is created so as to be used in all the systems that were constructed by implementing the same module of standardization;
- *flexibility* of digital content must be imposed as a key requirement in the processing systems, storage and transmission of data; characteristic of flexibility is driven by applications that process digital content; the interface and procedures that form the processing software are the ones that present digital content as the user asks; arranging, processing, arrangement of digital content based on demand, providing answers to questions raised are some features that software product must meet to provide digital content flexibly to emerging situations, digital content is flexible trough presentation and processing;
- *validity* of the digital content refers to data components belonging to a range of values; the validity is verified, particularly when loading values by testing data regarding their inclusion in a set of data considered accurate;
- *relativity* of digital content is produced by the acceptance by a community of digital content in relation to the knowledge of its members.

Quality management identify areas that meet quality requirements in a lesser extent, determine causes and provide solutions to eliminate inconsistencies. Digital content is dependent on users. Similarly, aspects of the quality of digital content are dependent on its use by users. There are times when digital content is handled by categories of users for whom it is not intended. Protecting digital content must be effectively, so that possible mistakes should not appear in the results. The most common way to protect against unauthorized users is to create accounts with different access levels, through which connection to the requested services is done according to a domain and a certain level of access.

4. Quality metrics for digital content

Software metric is a mathematical model that contains one or more equations or inequalities and has one or more target functions and its role is to describe the associated system state [3]. Metrics are implemented models to test the quality entities, taking into account the factors that are influencing the level of the measured characteristic.

The need to use metrics is given by the following considerations:

- allow setting targets for improving implemented entities;
- ensure an effective way to achieve these objectives;
- allow identify the causes that adversely affect the quality characteristics of a structured entities;
- identify requirements that must be followed to develop models of structured entities that are qualitatively superior;

Metrics test the differences between the entities resulting from implementation of a structured model and the expected results and determine the causes that led to these differences.

Metric is a definition, an algorithm or a mathematical function used for quantitative assessment of the tested product.

The measure is a numerical value obtained by direct or indirect determinations. The measure is also an input or an output value of a metric.

Metrics are management tools used to estimate cost and resources necessary for the project.

A metric is built to meet the following objectives:

- quantifying the characteristics;
- determine the influence of indirect factors;
- aggregation of values;
- comparative analysis.

Depending on the type of the characteristic used, metrics are classified into:

- metrics that are directly measurable - are associated with quantitative characteristics; the length of a structured entity, whatever the nature of expression, involves counting the components that have the same unit;
- metrics that are indirectly observable or measurable – are associated to qualitative characteristics that involves taking into consideration both quantitative indicators and subjective issues related to the mental representation of the one who uses the entity; a quality characteristic of a structured entity of qualitative nature is aiming to contribute in expanding the field; the evaluation of a metric associated with this characteristic takes into account numerical aspects (number of concepts, techniques and methods, methodologies and new technologies that are introduced) as well as subjective elements, by making associations between these issues and their perception of the real world by the reader.

Considering the complexity, the metrics are classified into:

- complex - associated models of the metrics involve the use of a large number of variables, and their implementation involves using extensive knowledge of mathematics; also determining variable values requires applying preparatory operations: normalization, grouping, aggregation;
- simple - models have a simple structure and the number of variables is small.

Depending on the model associated to the metric, there are metrics that are determined based on linear or nonlinear functions. Including metrics in different classes ensures completeness for the analysis done. Also, the nature of the analysis is highlighted by the number of metrics used in various categories.

The metric associated with digital content is directly influenced by the complexity of data components calculated using the formula:

$$C = \sum_{i=1}^k f_i * \lg_2 f_i$$

where:

- k is the number of types of words found in the digital content, like verbs, nouns, adverbs, adjectives, prepositions, conjunctions, pronouns, numbers, interjections;
- f_i is the frequency of apparitions for word i;

Let's consider a number of n fields, m vocabularies on which membership of words to vocabulary is calculated. Thereby determine how many words from each field belong to each identified vocabulary. Starting from the elements identified, the heterogeneous degree is calculated based on the formula:

$$NH = \frac{t}{n * b}$$

where:

- t is the number of words that does not belong to the vocabulary associated with the field;
- n is the number of fields;
- b is the number of items in the set.

Based on the formula of computing heterogeneous, the homogeneity level of a digital content, is determined as:

$$H = 1 - NH$$

Based on the quality metrics, quality models are constructed that test and certify the opportunity of the digital content used as a form of storage and information processing. Criteria with economic impact on the organization are considered.

5. Auditing processes

The audit represents the transfer of credibility of the team that performed the testing process of digital content.

The audit involves:

- digital content;
- steps;
- reports;
- the conclusion, based on which it is certified whether after auditing of the digital content, the requirements of the objectives for which it was built are met or not met,

because there are significant differences between the specifications of digital content and what exists in reality.

ORTOES application is tested taking into account sets of data from a community of users, the data are received at different times, thus resulting in changes in the number of users involved.

The initial community of users consisted of 276 users, of whom a total of 259 users load the first form of digital content, and between them a total of 246 users load the second form of testing the digital content. In this case, the digital content consists in the text stored in several forms.

The ORTOES application identifies whether the two sets of texts retain the required characteristics, namely membership to the same area.

The degree of similarity of digital content is represented in Table 1.

Table 1. Membership of the level of orthogonality on ranges of values

Digital content	interval [0;0.75)	interval [0.75;0.85)	interval [0.85;1.00]
CD ₁	227	0	32
CD ₂	41	0	205

In this case, the application validates the retention of the characteristics required for sets of digital content.

The low level of similarity of sets is influenced by:

- acquisition of text from one user to another;
- definition of similar texts applicable in a strictly defined area.

Table 2 texts that are most frequently loaded by users are presented:

Table 2. The frequency of occurrence of the texts

Titlu	Frecvență
Database administration	5
Business on Internet – E-commerce	3
Phonebook	2
Databases	8
Electronic commerce	3
Development of electronic bussines	3
Development of electronic business in commercial environment	3
Developing an application using Java in Oracle	4
Developing an informatics system using business intelligence technologies	4
Tracking members of a sport club	3
Virtual store	10
Data structures operations	5
Designing a website	4
Developing a computer system	5
Database security	5
Management system for hotel	5
Web technologies	9

As shown in Table 3 the loaded digital content sets have similarities, leading to the conclusion that the characteristics required from one set to another are kept, in this case the planned texts match closely with the effective texts.

Table 3. Similarity levels of digital content sets

interval [0;0.75)	interval [0.75;0.85)	interval [0.85;1.00]
104	23	149

By establishing the similarity between the sets of digital content it can be concluded regarding the way of required criteria of quality is respected.

So, if the degree of similarity is higher, the quality criteria are met more effectively, audit digital content indicating that the required quality thresholds are met.

6. Software auditing of digital content

For auditing digital content the ORTOES software product was developed.

ORTOES application is implemented to allow users to establish a number of differences between obtained results and the ones planned related to an issue dealt with, being spread across several modules, as shown and in Figure 4.

Each module is processing the different types of digital content.

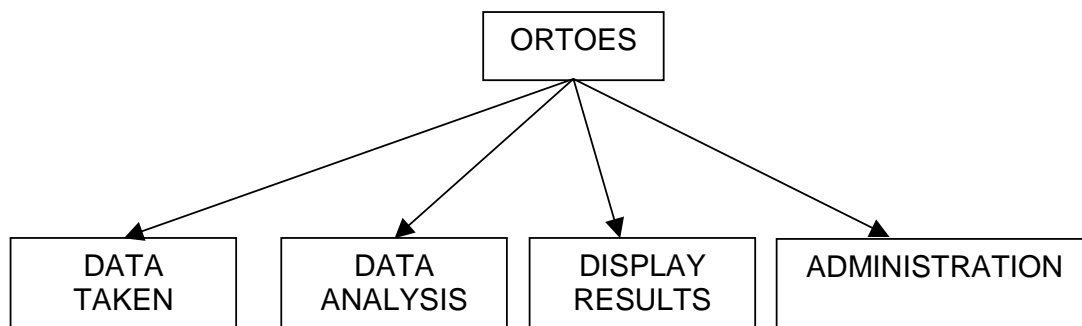


Fig. 4. ORTOES structure

The modules are composed of several functions, implemented to ensure processing and analysis of existing data, using several models.

For digital content based on texts the software product analyses texts using several criteria:

- Identification of common words that are found in the texts;
- Identification of the degree of diversity of words used;
- Identification of the frequency of terms' occurrence in the text.

In order to achieve superior text processing functions have been implemented which compare texts or take source files and transfer text components in the dynamic structures.

The functions of texts over taking take into consideration the category of symbols that separates the terms of texts. These symbols are stored in text files which are downloaded and stored by the product program. The source texts are processed character by character and depending on the character's value the loading of words is made in the dynamic structures or continues the process of forming the word.

The modules are implemented using the programming environment Visual C++, PHP language, and for building and accessing the database it is used MySQL management system.

The modules are functional in nature, consisting of functions with general character, which are called for acquisition and processing of various source files containing structures complying arrangement of information retrieval criteria specified in the file.

The modules have functional behaviour, consisting of functions with general character, which are called for acquisition and processing of various source files containing structures complying the arrangement of information specified in the file acquiring criteria.

The program takes data from a text file in a queue-type structure, and using the implemented functions results a series of working out:

- Determining the number of distinct words;
- Determine the number of words which are repeating;
- Determining the total words that appear in the article;
- Determining the frequency of the words' occurrence;
- Calculating orthogonality indicators.

Application's administration requires:

- Placing the list of users;
- Introduction of accounts and passwords associated to the list of users, each user chooses an account and password for accessing the application;
- Assignment of unique numeric codes for each user, the allocation is made when entering data about users, not when he creates his account;
- Managing messages sent by the users and displaying them either individual or in the form of lists containing the final results;
- Managing different application's versions for different time periods when users access application in order to load digital content;
- Managing the user's accounts;
- Generating new passwords where desired;
- Deleting a user's account;
- Identifying entities introduced by an user and viewing them.

User's managing requires:

- Defining user's account and password;
- Viewing individual and final results.

To ensure the security of online application that provides the facility to establish the orthogonality of structured entities generated by users, were used facilities offered implicitly by the MySQL used database (creation of users with unique credits, store their passwords in an encrypted format) and secondly the use of working sessions.

For each user who opens the online application starts a separate session, this provides protection from the unauthorized attempts to access the application resources. Thus, the session is terminated when the user leaves the current account or when an elapsed time greater than 5 minutes (configurable) past without the user using the personal project.

Another method implemented to ensure the application' security is the function `mysql_real_escape_string (string unescaped_string [, resource link_identifier])`. It is used to assure data before being sent as a request to the MySQL server by eliminating potentially dangerous character of the string.

Removal is done by adding backslashes (`\ x00`, `\ n \ r \ , ' , "`and `\ x1a`), which forces MySQL to interpret the characters as characters with a single apostrophe only as part of SQL statement.

The ORTOES application is developed in order to automate the interaction with users. First, the administrator of the application loads the users who have access to the application functions. Based on these records the users create accounts, load solutions, see the quality of

their digital content. Passwords are generated automatically by the application and transmitted to the users' e-mail addresses.

7. Conclusions

The digital content auditing process is complex, achieved within an audit team. The starting point is established by setting the audit objective. Starting from the objective, analysis procedures are built, the results are classified into categories of acceptance. The digital content is analyzed based on these procedures and are prepared reports containing examinations and indicators.

Based on the intermediate reports, the final audit report is generated in which the proposed solutions are explained, namely whether the digital content is accepted or not, audited in terms of quality, taken into account the specified results and actual results based on the data stored.

The audit of the digital content is subject to specific operations for distributed applications:

- the audit of manuals guarantees the accuracy of formula, completeness, consistency and concordance between the complexity of texts and the ability of understanding students;
- the population database provides a complete picture of population age, number of inhabitants, studies conducted;
- the medical database provides medical information on population, data on diagnoses, about illness, about treatment applied, about allergies, and about health policy considered on a community of people.

Auditing is required primarily for digital content categories that require online maintenance. They provide and ensure the quality of processing.

Only highly reliable digital content should be subject to audit.

By building the methodologies required for the auditing of digital content, the premises of storing and processing are ensured in such a way that the quality criteria required to base decisions are followed, these are affecting the entire population over a medium or long period of time.

References

- [1] C. D. Milodin, *The orthogonality of the structured entities, PhD Thesis*, Academy of Economic Studies, Bucharest, 2009.
- [2] I. Ivan, C. Ciurea and D. Milodin, "Collaborative Educational System Analysis and Assessment," *Proceedings on The Third International Conference on Advances in Computer-Human Interactions, ACHI 2010*, Saint Maarten, Netherlands Antilles, February 10-16, 2010.
- [3] I. Ivan and C. Boja, *Metode statistice în analiza software*, Editura ASE, Bucuresti, 2004.
- [4] I. Smeureanu, M. Dârdală and A. Reveiu, *Visual C#.NET*, Editura CISON, București, 2004.
- [5] C. Ciurea, "A Metrics Approach for Collaborative Systems," *Revista Informatica Economica*, Vol. 13, No. 2, 2009.

[6] I. Ivan, C. Ciurea and D. Milodin, „Autotestarea aplicatiilor informatice distribuite cu continut oferit,” *Revista Română de Informatică și Automatică*, Vol. 19, No. 3, 2009.

Authors



Ion IVAN has graduated the Faculty of Economic Computation and Economic Cybernetics in 1970. He holds a PhD diploma in Economics from 1978 and he had gone through all didactic positions since 1970 when he joined the staff of the Bucharest Academy of Economic Studies, teaching assistant in 1970, senior lecturer in 1978, assistant professor in 1991 and full professor in 1993. Currently he is full Professor of Economic Informatics within the Department of Computer Science in Economics at Faculty of Cybernetics, Statistics and Economic Informatics from the Academy of Economic Studies. He is the author of more than 25 books and over 75 journal articles in the field of software quality management, software metrics and informatics audit. His work focuses on the analysis of quality of software applications. He has participated in the scientific committee of more than 20 Conferences on Informatics and he has coordinated the appearance of 3 proceedings volumes for International Conferences. From 1994 he is PhD coordinator in the field of Economic Informatics. His main interest fields are: software metrics, optimization of informatics applications, developments and assessment of the text entities, efficiency implementation analysis of the ethical codes in informatics field, software quality management and data quality management.



Daniel MILODIN graduated the Economic Informatics department of the Faculty of Economic Cybernetics, Statistics and Informatics in 2005. He followed the master program “Project’s Computerized Management” between 2005 and 2007. He has published several articles in specialized magazines, articles focused mainly on studying and implementing the concept of orthogonality: the orthogonality of the alphabets, of the Arabic digits, of the Latin alphabet, of the distributed informatics applications and of the structured entities. The dissertation paper within the master is placed on the same line of research, “The projects’ orthogonality, conditions of entry into the assessment program”, proposing to define the projects’ concept of orthogonality, methods for determining the degree of similarity between two projects, and also to develop software products for identifying similar projects. He is a PhD graduate of the Academy of Economic Studies, specialization “Economic Science”, with the thesis “The orthogonality of the structured entities”.



Mihai GEORGESCU has graduated the Faculty of Electronics, Telecommunications and Information Technology – Politehnica University Bucharest in 2005 and also completed a master program at the Faculty of Economic Cybernetics, Statistics and Informatics from the Bucharest Academy of Economic Studies in 2007. He has experience in the development of software products using various programming languages like C/C++, C# both Windows Forms and Web.