Orthogonally Based Digital Content Management Applicable to Projects-bases

Ion IVAN[1], Andrei SANDU[1], Daniel MILODIN[1], Cosmin LUGOJI[2]
[1]Academy of Economic Studies, Bucharest, Romania
[2]Polytechnics University, Bucharest, Romania,
ionivan@ase.ro, andreisandu@yahoo.com, mc_daniel2004@yahoo.com,
cosmin.lugoji@gmail.com

There is defined the concept of digital content. The requirements of an efficient management of the digital content are established. There are listed the quality characteristics of digital content. Orthogonality indicators of digital content are built up. They are meant to measure the image, the sound as well as the text orthogonality as well. Projects-base concept is introduced. There is presented the model of structuring the content in order to maximize orthogonality via a convergent iterative process. The model is instantiated for the digital content of a projects-base. It is introduced the application used to test the model. The paper ends with conclusions.
Keywords: *digital, quality, orthogonality, projects-bases*

1 Digital Content

Digital content represent the implementation within the computer memory of information regarding a certain aspect of the surrounding world.
Digital content is built up as follows:
- direct recording of data with the help of devices created for capturing images, sounds;
- physical parameters such as temperature, pressure, humidity, etc;
- inserting from keyboard arrays of characters;
- processing of the image of documents resulted via scanning, taking a photo, recording, etc.

Digital content is stored in files. The files have different characteristics:
- total length in bytes;
- extension type: txt – text files; cpp – C++ file, pas – Pascal source file, asm – Assemble file, jpg – image files created by using the JPEG standard;
- creation date or last update;
- identification name;
- description of the structure in order to have direct access to the requested information.

If the file has been created using a process P, there must be defined the process P^{-1} which allows the access to the digital content of the file. In the case of collectivity C formed by the elements C1 C2…CN for which there has been defined a pattern:

```
struct pattern{
       Type1 name1;
       Type2 name2;
       …..
       Typek namek;
} art[N];
```

Initializing all N components art[i] of the structure of the vector is performed for each field namej using a type conversion from base type typej, and the process P includes as well writing the vector art[N] in the file *art.det*.
The process P^{-1} must include in a symmetrical way reading of the values art[N] from the file and performing those conversions which allow the printing of the contents of the file and processing the fields from the art[i] elements of the vector, according to the objective established in the problem to be solved.
If the file movie.avi contains a movie which is compressed using the P algorithm, first of all there has to exist the operation of reading the different articles from the file and then there has to exist the algorithm P^{-1} for decompression.
Another example in this direction is com-

pression applicable to digital photos. No matter if it is performed in format png, jpeg or gif, before applying the extraction algorithm there has to be performed the operation of reading the data from the corresponding files.
Creation of digital content is the process via which there are defined the conditions, structure and name of the file and afterwards there are generated and stored arrays of symbols which are in correspondence with arrays of bytes.
Creation of files is performed by:

- imputing information the keyboard and writing it to the file;
- generating data and writing it to the file;
- acquisition of data and writing it to the file.

Creation requires:

- establishing the format for presenting the data;
- specifying instructions for writing the data;
- defining the type of file.

File access is performed by reading the data. Some functions that can be used:

- *int fprintf (FILE * stream, const char * format, …);* writes in a specified file a sequence of data using the format of the specified arguments;
- *ostream& put (char c);* writes the c character in the output buffer at the current insert position and increments the position pointer in order to indicate the next character;
- *ostream& write (const char* s , streamsize n)*; writes the data block indicated by the pointer s, dimension n, in the output buffer. The characters are written sequentially until the character on position n is reached;
- *int fscanf (FILE * stream, const char * format, …)*; reads information from a file and stores it according to the parameter *format* in the location indicated the additional arguments provided;
- *int get()*; extracts a character from a file and returns its value as an integer;
- *istream& read (char* s, streamsize n)*; reads a block of data with the size of n characters and stores it in an array indicated by the pointers.

It is important to know the instruction used for writing the information in order to use the symmetrical one for recovering the info.

2 Content Management Requirements

Defining the content represents the main objective for the team which deals with content management. We have different scenarios when performing content management:

- sub defining the content, by choosing less characteristics, and sometimes less elements from the collectivity which results in the impossibility of repeating the data collection process and in most cases compromises the problem solving process;
- over defining the content by performing repeated measurements for the same characteristic or taking into account characteristics which are not directly linked to problem solving;
- correctly defining the content in the sense that there is collected information only regarding the characteristics which enter in the process of processing and the data reflects statistics applicable to all elements of the collectivity.

Defining the content is an iterative process. There are performed simulations on samples. There is used the experience on similar problems. Only after solid research it is then switched to data collection and creation of digital content regarding a collectivity C and problem P which is required to be solved.
Usage of the content is one of the essential problems of content management.
The digital content must be conserved. Using it incorrectly leads to destruction of the content, temporary destruction or even permanent destruction. In order to mitigate such risks it is suggested to copy first the content. There must be put in place a configuration management schema for the digital content copies, in order not to mix them up.
If an initial digital content CD_0 receives modification, it becomes CD_1. If copies are performed, there will be CD_1_01, CD_1_02…CD_1_20, where CD_1_20 is the 20^{th} copy, identical with the others. If copy

CD_1_15 gets modified there will result a new version of the copy called CD_1_15_01. Therefore there has to be considered a mechanism for avoiding the generation of very long names for versions of the digital content [2].
If we deal with multiple files F1, F2, F3, … Fn, those files being identical, there will exist just one physical object. For manipulating the names of the files there is used a name dictionary. When one of the copies gets modified there is created a new physical object which stores the new information.

Updating the content means:
- adding texts, images, sound, values of physical parameters for a process;
- changing one array of bytes with another;
- eliminating an array of bytes from the files.

It is important that the management of digital content:
- restricts access to updates, allowing the right to only certain people to perform updates, according to table 1:

Table 1. Model of separation of access rights

	Addition	Change	Deletion
Person 1	X		
Person 2		X	
Person 3			X
…	…	…	…

- to identify the person that has updated a field, an article, a file, the place and time as well as the performed operation;
- to limit the effects of the modifications so that everything reduces to addition of information; it is indicated the operation and value that is required to be considered as new content, according to figure 1.

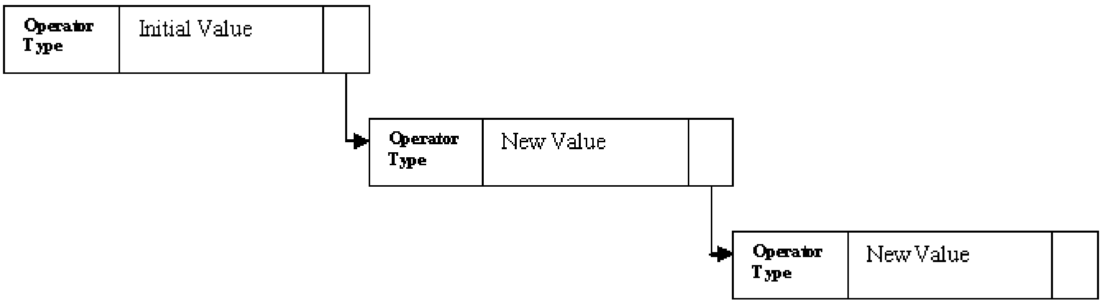

Fig. 1. Model for updating the information

By following the above updating indications the information content can be reconstructed at any time the by starting from a previous version, maintaining at the same time a smaller capacity for storage purposes.

3 Quality Characteristics of digital content

Structuralism is the characteristic stating that the digital content is decomposed in parts and there is established relation between those parts.
Linear structure of a digital content shows that parts A1, A2…Ak are placed one after the other and the interpretation of the content is obtained by crossing from component Ai to Ai+1, where i=1,2,….k-1.
Arborescence structure correspondents to placing components on different levels of detail and understanding the details requires going deeper through the levels.
The components of a book A, chapters B, C, D and E are formed from subchapters B11, B12, B13, C11, C12, C13, D11, D12, D13 and E11, E12, E13, E14, E15,.
The subchapters as well include the detail represented by the detailed text. The structur-

ing goes to the level of paragraph, phrase and sentence.
Graph structure corresponds to the implementation of the principle of composing constructions from a complex digital context, which are referring components which were previously defined, according to figure 2.

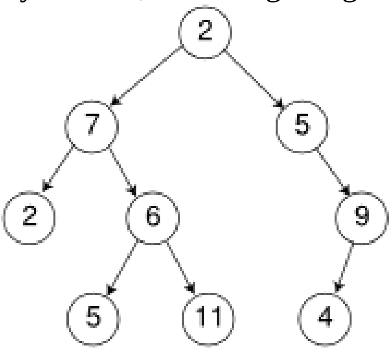

Fig. 2. Graph structure

The above graph is a simple example of oriented graph which can be implemented for building the digital context. One will observe that it is compose of numbered nodes. Each node can contain pieces of digital information. By crossing the graph in one of the directions one will obtain coherent information which can be interpreted. This way, in order to obtain the digital content we are interested into it is enough to cross the corresponding graph.
Completeness represents the digital content quality characteristics which shows the degree by which at the moment of coding and storing of the digital information there have been taken into account all the components requested by the final users. A relevant example is that of a foreign movie which will be subtitled, so that as many as possible final users will be able to access and understand the video material. Lack of colors, sound or subtitle from the digital content will decrease the level of completeness and will generate additional costs in order to attend the required level of the requirements.
The level of completeness will be considered 1 if there are no additional costs. The more the level of completeness is closer to 0, the more it will become obvious that the responsible with delivering the digital content has not taken into account some activities, has incorrectly estimated the durations or has not correlated correctly the dependences between the activities.
Correctitude, as it refers to digital content, means respecting the imposed standards. Correctitude refers for instance to the editors used, the filters, the algorithms and ultimately to respecting author rights and patents. Therefore, in the case of a video material, besides obtaining the consent of the participants to the movie, there must be procured licenses for using certain editors, filters and encrypting algorithms. In the case of texts, correctitude consists in respecting the industry standards and the professional ethics code, according to which all that has previously been published must be cited and referred in the bibliography.
Another quality characteristic applicable to digital content is **consistency** [6], which refers to respecting all logical interdependencies regarding type and order of comprised information. Therefore it's not consistent to have the cast of actors of a movie placed in the middle of the film or it is not consistent to have a text that refers elements, which have not been previously defined or it is not consistent to a have an audio file with the name of a song, but which contains a different song.
Accuracy, as applicable to digital content, refers to playing as close as possible to the quality of the message recorded at the source for the elements of image, sound and text.
This way a sound file that is stored in a format with a low compression ratio will have a higher degree of accuracy as compared to the source of origin. For texts, even if we apply compression algorithms, the accuracy will not be affected, whereas if we perform additions or deletions of characters or words, the accuracy as compared to the original will decrease.
The quality of the content refers to the reality it reflects. Compared to data quality, which results from measurements, the quality of the content makes reference both to the quality of the data as resulted from measurements as well as the existing connections between the data and the meaning they generate.

4 Content Orthogonality

The vocabulary is composed from a series of words C1,C2,C3,…Cn. A text is a series of words. The length of the text is given by the number of words from which it is composed. Words are separated by an alphabet symbol chosen as separator, which is not part of the vocabulary.

Orthogonality of two texts is based on the following elements [3]:

- length of the texts expressed as number of words;
- vocabularies of the texts Vi and Vj;
- appearance frequencies for the words in vocabularies;
- intersections of vocabularies;
- taking into account of the words that belong to the standard terminology of the domain referred by the digital content.

A very simple implementation of orthogonality [8] is:

$$ORTO = 1 - \frac{NW_{both}}{NW_{\max}}$$

where:

- NW_{both} - number of words that are part of the standard domain terminology and that are found in both texts that are being analyzed;
- NW_{max} - the higher of the two projects related to the count of words that are part of the standard terminology of the corresponding domain.

Two texts are perfectly orthogonal, meaning that the indicator ORTO equals 1, if they don't have anything in common and the vocabularies are disjunctive.

The orthogonality of the content is given by the texts comprised. In order to transform the video and audio information into text there is established a set of conventions, as follows:

- the sound undergoes a filtering in order for the words to be recognized, so that the lines in the dialogue and the soundtrack get transformed in an array of words;
- video sequences are posted on a web site, where people specialized in movie production and directing, will have the task of making a review of the action; following this actions a specialized software product will identify a vocabulary which is common to all reviews of a certain video content and will generate only one review.

Taking into account the complex scenario of video materials that have video, sound and subtitle content, there will be created 3 text flows, which characterize the digital content.

For the digital content A1, A2, ….. At, it is said that Ai and Aj –figure 3 – are orthogonal if the orthogonality level H(Ai, Aj) is comprised in the interval [0,78;1].

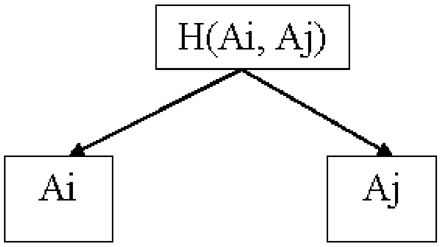

Fig. 3. Orthogonality of projects Ai and Aj, from the projects-base

There are the following orthogonalities applicable to digital content:

- H_v (A_{iv}, A_{jv}) ortogonality of video content;
- H_s (A_{is}, A_{js}) ortogonality of sound;
- H_t (A_{it}, A_{jt}) ortogonality of text.

A value closer to 1 for any of the 3 indicators represents that the two digital contents are orthogonal for the characteristics taken into account – video, audio and text.

$$H_v(A_{iv}, A_{jv}) = 1 - \frac{CAZ_{both}}{TCAZ_{\max}}$$

where:

- CAZ_{both} - total number of action zones from each frame, which are common to both digital contents;
- $TCAZ_{max}$ - the maximum number of action zone from all frames from the total identified for each digital content.

The action zone represents the base granularity as it is considered by specialists, which identifies for each frame the reference points. Therefore for static frames we will have only one action zone, whereas for frames with several participants, we will have one action zone for each participant.

$$H_s(A_{is}, A_{js}) = 1 - \frac{CSL_{both}}{TCSL_{max}}$$

where:

- CSL_{both} - total number of sound layers, which is common to both materials;
- $TCSL_{max}$ - the highest of the total of sound layers of each digital content.

The digital sound content is decomposed until a basic layer level. This way, an instrumental song with one instrument will have only one layer. For the song of singer who is accompanied by an orchestra, there will be one layer for the singer and one layer for each instrument.

$$H_t(A_{it}, A_{jt}) = 1 - \frac{CSW_{both}}{TCSW_{max}}$$

where:

- CSW_{both} - total number of words, which is common to both digital contents;
- $TCSW_{max}$ - the highest of the total count of words for each of the two digital contents.

Orthogonality of the two digital contents, which relies on video, sound and text orthogonality as well, H(Ai, Aj) will take values in the interval [0,1], where 1 means that the two digital contents are perfectly orthogonal.

$$H(Ai, Aj) = \alpha H_v(A_{iv}, A_{jv}) + \beta H_s(A_{is}, A_{js}) + \gamma H_t(A_{it}, A_{jt})$$

where:

- α - coefficient of importance for the video component of the digital content;
- β - coefficient of importance for the audio component of the digital content;
- γ - coefficient of importance for the text component of the digital content;
- $\alpha + \beta + \gamma = 1$.

Orthogonality between digital contents translates to building a collection of digital components CDB1 CDB2 … CDBH. When it is being evaluated if the component CDBH+1 is to be included or not, the first check is for H (CDBi, CDBH+1)>0.78, where i=1,2,….,H. Only upon successful completion of all tests will ensure that the orthogonality of the collection is maintained.

5 Structured Homogenous Digital Content

There are considered the collectivities C1, C2, …. Ct described with the characteristics Q1, Q2, … QR. For each of those characteristics there is defined a vocabulary V1 V2… VR and $Vi \cap Vj = \emptyset, \forall i,j = 1,2,\ldots,R.\ i \neq j$.

There is built up the digital content for solving the problem P – Table 2.

Table 2. Analysis of digital content

	Q1	…	Qi	…	Qj	…
C1						
C2						
…						
Ci					αi,j	
…						
Cj						
…						

αi,j represents the word from the Vj vocabulary which defines the level of the Qj characteristics for the Ci element of the collectivity. Each vocabulary Vi has its own blank word Ci0.

In order to the digital content the following steps are followed:

- there are built procedures for putting in place measurements and for putting in correspondence the level of the characteristics Q1, Q2 … QR with the components identifiers C1, C2…Ct, and with α1, α2, … αt as identifiers;
- it is then verified if a digital content has been obtained which fulfills quality characteristics related to completeness, correctitude, structuralism and so on.
- The digital content is considered operational, if and only if:
- it includes information related to all elements of the C collectivity;
- there are sufficient elements which allow

the split of the C collectivity in sub collectivities based on the words used for defining the Qi characteristic based on a set of characteristics Qi1, Qi2, … Qim;

- there are algorithms which require the parsing of the content while extracting words that have a certain distance between them or a certain significance in order to compute meaningful indicators which allow the interpretation of the contribution to the whole collectivity.

Once we have obtained an operational digital content, then the retrieval time of information will be highly ameliorated.

6 Applying orthogonality of digital content to projects-bases

One area of applicability of orthogonality computation for digital content is that of databases. Projects-bases are defined collections of projects, which are enrolled in process of obtaining financing, therefore having a common scope and framework [4] [7].
From this point forward we will consider the entire projects-base as being the digital content, whereas the projects represent the components of the digital content.

It is considered the C collectivity of projects, taking into account the description characteristics Q1, Q2, … QR, for all αi,j words which are linked to the Ci element of the collectivity and the Qj characteristics, which belongs to the vocabulary associated to the description characteristics.
A matrix is constructed starting from the vocabularies of characteristics which contain the position of the words within those vocabularies. There are summed up the frequencies of words usage. There are interchanges performed between the columns in order to obtain descending frequencies. There are performed interlocks between the lines in order to obtain homogenous utilizations.
There are built up subsets of characteristics. The subsets get separated. The subsets get referred to via indexes created within the chapters. The process is iterative.
Our intent is that of obtaining an internal orthogonality as high as possible, meaning that the projects will get grouped in chapters, which will be orthogonal taken randomly any two of them.

Table 3. Analysis of description characteristics

	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8
C1	1	Banking	Retail	Avg.	Co-financing	60	Engineer	200K €
C2	2	Banking	Retail	Avg.	Co-financing	50	Engineer	150K €
C3	3	Banking	Corporate	Avg.	Co-financing	55	Professor	250K €
C4	4	Banking	Investment	Avg.	Co-financing	50	Economist	150K €
C5	5	Banking	Investment	Avg.	Financing	65	Retired	100K €
C6	6	Insurance	Life	Avg.	Financing	40	Jurist	300K €
C7	7	Insurance	Life	High	Co-financing	30	Business Owner	500K €
C8	8	Insurance	Health	Avg.	Co-financing	31	Business Owner	600K €
C9	9	Banking	Corporate	Avg.	Co-financing	65	Retired	150K €
C10	13	Banking	Retail	Avg.	Co-financing	56	Engineer	250K €
C11	17	Banking	Investment	Avg.	Financing	58	Economist	230K €
C12	21	Insurance	Health	Avg.	Financing	29	Manager	750K €
C13	40	Banking	Corporate	Avg.	Co-financing	45	Economist	220k €
C14	41	Unspecified	-	-	Co-financing	40	Economist	300K €
C15	42	Unspecified	-	-	Financing	30	Jurist	350K €
C16	45	Unspecified	-	-	Financing	44	Jurist	370K €

Therefore we consider the collectivity C1, C2, … C16 and the description characteristics Q1- ID ,Q2- Domain, Q3- Sub domain, Q4- Complexity, Q5- Co-financing Status, Q6- Age of the PM, Q7- Primary Skill set, Q8- Project Budget, according to table 3.
Based on the available information the vocabulary is built. The length of the vocabulary is given by the number of words, the ties between the collectivities being performed with the help of keys.
Any of the characteristics can serve the role of a key for describing the elements from the table. There are two types of keys: primary and secondary. Primary keys are unique and not null, whereas secondary keys are used for different classifications. An object can be identified and referred to with the help of its primary key [5].
Sorting the digital content can be performed in several ways:

- based on the primary key (Q1);
- based on secondary keys (Q2-Q8).

The digital content gets sorted after one of the characteristics - primary key or secondary key. If the sort key is Q2 – Domain, we have the output presented in table 4.

Table 4. Sort by domain

	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8
C1	1	Banking	Retail	Avg.	Co-financing	60	Engineer	200K €
C2	2	Banking	Retail	Avg.	Co-financing	50	Engineer	150K €
C3	3	Banking	Corporate	Avg.	Co-financing	55	Professor	250K €
C4	4	Banking	Investment	Avg.	Co-financing	50	Economist	150K €
C5	5	Banking	Investment	Avg.	Financing	65	Retired	100K €
C9	9	Banking	Corporate	Avg.	Co-financing	65	Retired	150K €
C10	13	Banking	Retail	Avg.	Co-financing	56	Engineer	250K €
C11	17	Banking	Investment	Avg.	Financing	58	Economist	230K €
C13	40	Banking	Corporate	Avg.	Co-financing	45	Economist	220K €
C6	6	Insurance	Life	Avg.	Financing	40	Jurist	300K €
C7	7	Insurance	Life	High	Co-financing	30	Business Owner	500K €
C8	8	Insurance	Health	Avg.	Co-financing	31	Business Owner	600K €
C12	21	Insurance	Health	Avg.	Financing	29	Manager	750K €
C14	41	Unspecified	-	-	Co-financing	40	Economist	300K €
C15	42	Unspecified	-	-	Financing	30	Jurist	350K €
C16	45	Unspecified	-	-	Financing	44	Jurist	370K €

If we sort using the Q5 key – own finance means or co-financing, we reach the data in table 5.

Table 5. Sort by financing means

	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8
C1	1	Banking	Retail	Avg.	Co-financing	60	Engineer	200K €
C2	2	Banking	Retail	Avg.	Co-financing	50	Engineer	150K €
C3	3	Banking	Corporate	Avg.	Co-financing	55	Professor	250K €
C4	4	Banking	Investment	Avg.	Co-financing	50	Economist	150K €
C9	9	Banking	Corporate	Avg.	Co-financing	65	Retired	150K €
C10	13	Banking	Retail	Avg.	Co-financing	56	Engineer	250K €
C13	40	Banking	Corporate	Avg.	Co-financing	45	Economist	220K €
C7	7	Insurance	Life	High	Co-financing	30	Business Owner	500K €
C8	8	Insurance	Health	Avg.	Co-financing	31	Business Owner	600K €
C14	41	Unspecified	-	-	Co-financing	40	Economist	300K €
C5	5	Banking	Investment	Avg.	Financing	65	Retired	100K €
C6	6	Insurance	Life	Avg.	Financing	40	Jurist	300K €
C11	17	Banking	Investment	Avg.	Financing	58	Economist	230K €
C12	21	Insurance	Health	Avg.	Financing	29	Manager	750K €
C15	42	Unspecified	-	-	Financing	30	Jurist	350K €
C16	45	Unspecified	-	-	Financing	44	Jurist	370K €

This way the information can be grouped according to users' needs, in such a way that the access is done the easiest way possible.
For each grouping performed there are considered newly created chapters that include projects with the same secondary key, which has been used as sort criteria.
For each grouping orthogonality is computed between the resulted vocabularies of each chapter, except for the secondary key used for grouping criteria. Therefore, if Q2 has been used as a grouping criterion – like in table 4 - , we obtain 3 different chapters, called *Banking*, *Insurance* and *Unspecified*.
For columns Q3, Q4 … Q8, which are part of each chapter, it is being counted the number of appearance for each distinct word. Figures get excluded from such an analysis. The resulted synthesis is presented in table 6.

Table 6. Count of appearances of words

Chapter	Word	No. occurrences	Total cumulated
Banking	Retail	3	3
	Corporate	3	6
	Investment	3	9
	Avg.	9	18
	Co-financing	7	25
	Financing	2	27
	Engineer	3	30
	Professor	1	31
	Economist	3	34
	Retired	2	36
Insurance	Life	2	2
	Health	2	4
	Avg.	3	7
	High	1	8
	Co-financing	2	10
	Financing	2	12
	Jurist	1	13
	Business Owner	2	15
	Manager	1	16
Unspecified	Co-financing	1	1
	Financing	2	3
	Economist	1	4
	Jurist	2	6

For the records in table 7 it is computed H_{AVG} – average orthogonality.

$$H_{AVG} = (H_A + H_B + H_C) / 3 = 0.81.$$

As a result, the average orthogonality of the identified chapters is a good indication that the components of the digital content have been distributed correctly. The only question mark arises for the chapter *Unspecified*, where orthogonality is lower than 0.78. This can be explained by the fact that this chapter comprises elements that did not have all elements thoroughly specified.

Table 7. Orthogonality computation

Code	Ai	Aj	Common words	Maximum no. words	H(Ai,Aj)
H_A	Banking	Insurance	7	36	0.8
H_B	Banking	Unspecified	4	36	0.88
H_C	Insurance	Unspecified	4	16	0.75

The same steps are repeated for every type of scenario at a time, when a different secondary key is chosen as grouping criterion. In the end the variant with the highest orthogonality is considered the best grouping criterion implementation.

From the above demonstration it can be depicted that the management of digital content is performed by maximizing orthogonality. This guarantees the homogeneity of the digital content and facilitates the creation of orthogonal projects-bases.

7 DiMaO Software Application Overview

DiMaO – Digital Matrix Othogonality – software application implements a method of structuring the information from a given matrix, through a series of operations of reversing lines and columns, with the purpose of obtaining a block matrix which contains or-

thogonal zones of information.
The architecture of DiMaO application is described in figure 4.

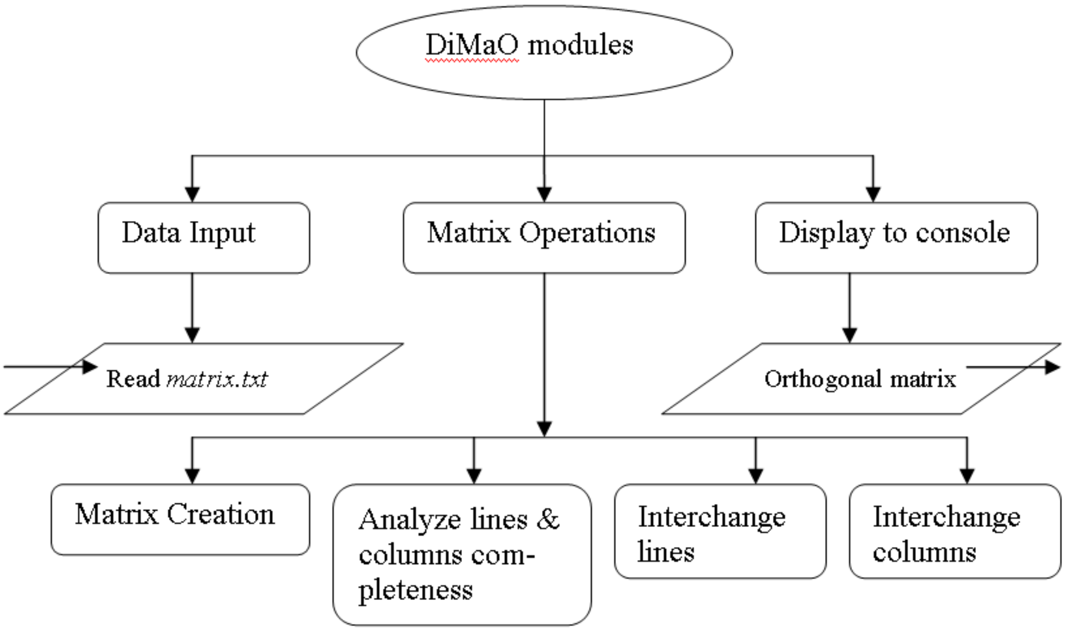

Fig. 4. Modules of application DiMaO

The application receives the data via a text file *matrix.txt*, which is stored in a predefined resources folder. The file contains words which are distributed randomly within a matrix structure, having as delimitation the blank character. The character that is used as a convention for a blank word or missing data in a cell is ‚-'.
In a first stage there are used two vectors which memorize the number of words which are non blank for both the lines and the columns of the matrix. The two vectors are then sorted, using the bubble sort method, which is modified accordingly in order to invert the needed lines and columns.
The data from the file is saved initially in a matrix of Strings, using the method *WriteFileToMatrix*. After populating the matrix, it is then used for generating the needed vectors *horVector* and verVector. The results of the two sort operations, bubbleSortLines and *bubbleSortColumns*, are stored in the matrix *orderMatrix*. The matrix is also displayed in the console for easy visualization of results.
The application is developed in Java 1.4, using Eclipse as IDE. The Java classes used to implement the algorithm are easy to integrate in a package, which can be reused within another application, no matter if it's a Web or stand-alone application. This is a strong point as the application proves to be both scalable and versatile.

8 Conclusions

In order to bring value to the contents of the current paper and in order to continue the research it is needed for the presented software solution to be extended. This way the presented algorithm will be generalized and extended from just projects-bases to any collection of elements with digital content.
By analyzing the digital content it is determined the internal orthogonality and there are given indications for increasing othogonality. Some immediate benefits are:

- reduction of necessary memory space;
- reduction of redundancy caused by repeating pieces of the same information;
- reduction of the time needed to retrieve information;
- distinct approach of sub-problems by focusing directly on the digital content part which includes a sub list of descriptive characteristics.

This approach is useful for structuring databases, structuring synthesis papers and encyclopedias.

Another implementation is that of organizing unstructured digital content. It is considered a system which generates unstructured digital contents, which is then stored and which then goes through a process of structuring. The steps undertaken are:

- words are identified and vocabularies are built;
- frequencies of appearance for each word are stored;
- sub vocabularies are built;
- it is analyzed the distribution between words within the same sub vocabulary;
- a matrix of records is build;
- orthogonality computed;
- records are identified;
- tables are identified and the database is structured.

The objective of pursuing the steps described above is that of creating a database structure efficient from an orthogonality point of view.

References

[1] C. N. Bodea, *Handbook of Project-Based Management*, Economic Printing House, Bucharest, 2002.

[2] I. Ivan, D. Milodin, M. Popa and C. Lugoji, “Baseline Entities,” *Economia magazine, Management series*, 8th year, No.1, Bucharest, 2005, pp. 84 – 97.

[3] I. Ivan and D. Milodin, “Data Orthogonality Metrics,” *Workshop Information Systems and Operations Management*, Romanian American University, Bucharest, 2006, pp. 326-341.

[4] I. Ivan and A. Sandu, “Projects Hierarchy Based on Duration and Complexity,” *22nd IPMA World Congress Proceedings - Project Management to Run*, Roma, Italy, 9-11 November 2008.

[5] C. M. Matei, C. I. Silvestru and D. Ş. Silvestru, “Internet Banking Integration within the Banking System,” *Economy Informatics Magazine*, No. 2 (46), 2008, pp. 55-59.

[6] M. Popa, I. Ivan, G. Nosca and S. Capisizu, *Quality Management of Software Applications*, ASE Printing House, Bucharest, 2006.

[7] R. Gareis, *Happy Projects!*, 2nd edition, ASE Printing House, Bucharest, 2006.

[8] A. Sandu, “Indicators for Measuring the Quality of Projects-base Components,” *4th International Conference for Applied Statistics*, 20-22 November, Bucharest, NIS Printing House, 2008.

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.

Andrei SANDU has graduated the Faculty of Cybernetics, Statistics and Economic Informatics in 2005. He holds a Masters degree in Project Management. He held positions as project manager, technical team leader and software developer. Currently, he is portfolio project manager in banking field. Also, he is PhD student enrolled in a PhD program on Computer Science in Economics. His key skills refer to the following: project management, business management skills, banking skills, coaching team members, technical team leadership and IT&C specialist in business oriented software development and maintenance.

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. Currently he is a PhD candidate within the Academy of Economic Studies, specialization "Economic Science", with the theme "The orthogonality of the structured entities".

Cosmin LUGOJI has graduated from the Faculty of Computer Science, "Politehnica" University of Bucharest in 2009. Currently he is working for a software company which produces applications for financial administrations for Banking Companies like "Banca Comerciala Romana" or "Raiffeisen Bank". He is interested in information technology, especially web applications and web portals.