THE BUCHAREST UNIVERSITY OF ECONOMIC STUDIES

DATABASE SYSTEMS JOURNAL

Vol. V, Issue 3/2014

LISTED IN

RePEc, EBSCO, DOAJ, Open J-Gate,
Cabell's Directories of Publishing Opportunities,
Index Copernicus, Google Scholar,
Directory of Science, Cite Factor,
Electronic Journals Library

BUSINESS INTELLIGENCE

ERP

DATA MINING

DATA WAREHOUSE

DATABASE

ISSN: 2069 – 3230
dbjournal.ro

Contact

Calea Dorobanţilor, no. 15-17, room 2017, Bucharest, Romania
Web: http://dbjournal.ro
E-mail: editordbjournal@gmail.com; editor@dbjournal.ro

CONTENTS

Labeling Consequents of Fuzzy Rules Constructed by Using Heuristic Algorithms of Possibilistic Clustering

Dmitri A. VIATTCHENIN[1,2], Stanislau SHYRAI[1], Aliaksandr DAMARATSKI[2]
[1]Belarusian State University of Informatics and Radio-Electronics, Minsk, Belarus
[2]United Institute of Informatics Problems of the National Academy of Sciences of Belarus, Minsk, Belarus
viattchenin@mail.ru, ashaman410@gmail.com, a.damaratski@gmail.com

The paper deals with the problem of automatic labeling output variables in Mamdani-type fuzzy rules generated by using heuristic algorithms of possibilistic clustering. The labeling problem in fuzzy clustering and basic concepts the heuristic approach to possibilistic clustering are considered in brief. Labeling consequents procedure is proposed. Experimental results are presented shortly and some preliminary conclusions are made.
Keywords: *Data Mining, Clustering, Fuzzy Rule, Consequent, Labeling*

1 Introduction

Fuzzy classifiers play an important role in different data mining approaches. Thus, the problem of generation of fuzzy rules is one of more than important problems in the development of fuzzy classifiers.

There are a number of approaches to learning fuzzy rules from data based on techniques of evolutionary or neural computation, mostly aiming at optimizing parameters of fuzzy rules. From other hand, fuzzy or possibilistic clustering seems to be a very appealing method for learning fuzzy rules since there is a close and canonical connection between fuzzy clusters and fuzzy rules. The fact was shown in [1].

Let us consider in brief some basic concepts. We assume that the training set contains n data pairs. Each pair is made of a m_1-dimensional input-vector and a c-dimensional output-vector. We assume that the number of rules in the fuzzy inference system rule base is c. So, Mamdani and Assilian's [2] fuzzy rule l within the fuzzy inference system is written as follows:

$$\begin{array}{l} If\ \hat{x}^1\ is\ B_l^1\ and\ \ldots and\ \hat{x}^{m_1}\ is\ B_l^{m_1} \\ then\ y_1\ is\ C_1^l\ and \ldots and\ y_c\ is\ C_c^l \end{array}, \quad (1)$$

where input variables $\hat{x}^{t_1}$, $t_1 = 1,\ldots,m_1$ are antecedents and output variables y_l, $l = 1,\ldots,c$ are consequents of fuzzy rules, $B_l^{t_1}$, $t_1 \in \{1,\ldots,m_1\}$ and C_l^l, $l \in \{1,\ldots,c\}$ are fuzzy sets that define an input and output space partitioning. A fuzzy classifier which is described by a set of fuzzy classification rules with the form (1) is the multiple inputs, multiple outputs system.

The principal idea of extracting fuzzy classification rules based on fuzzy clustering was outlined in [1] and the idea is the following. Each fuzzy cluster is assumed to be assigned to one class for classification and the membership grades of the data to the clusters determine the degree to which they can be classified as a member of the corresponding class. So, with a fuzzy cluster that is assigned to the some class we can associate a linguistic rule. The fuzzy cluster is projected into each single dimension leading to a fuzzy set on the real numbers. An approximation of the fuzzy set by projecting only the data set and computing the convex hull of this projected fuzzy set or approximating it by a trapezoidal or triangular membership function is used for

the fuzzy rules obtaining.

The idea of extracting fuzzy classification rules based on possibilistic clustering [3] is similar to the idea of deriving fuzzy rules based on fuzzy clustering.

On the other hand, a heuristic approach to possibilistic clustering was outlined in [4] and the approach was developed in other publications. Moreover, a method of the rapid extracting fuzzy rules based on results of the heuristic possibilistic clustering of the training data set was also proposed in [4] and the method is very effective in comparison with the method based on fuzzy clustering results. The idea of deriving fuzzy classification rules from the training data can be formulated as follows: the training data set is divided into homogeneous group and a fuzzy rule is associated to each group.

However, names should be assigned to each output variable y_l, $l = 1,\ldots,c$. The process of assigning names to output variables is connected with the problem of interpretation of classification results and a labeling procedure in clustering.

The main goal of this paper is a consideration of an approach to automatic labeling consequents of fuzzy rules generated by heuristic algorithms of possibilistic clustering. The contents of this paper is as follows: in the second section a labeling problem in fuzzy clustering is described, in the third section basic concepts of the heuristic approach to possibilistic clustering are considered, in the fourth section a labeling procedure for fuzzy rules consequents is described, in the fifth a numerical example of application of the proposed procedure to fuzzy rules generated from the Anderson's Iris data set are given, and some final remarks are stated in the sixth section.

2. Related works

The most widespread approach in fuzzy clustering is the optimization approach. Most optimization fuzzy clustering algorithms aim at minimizing an objective function that evaluates the partition of the data into a given number of fuzzy clusters.

All objective function-based fuzzy clustering algorithms can in general be divided into two types: object versus relational.

The object data clustering methods can be applied if the objects are represented as points in some multidimensional space $I^{m_1}(X)$. In other words, the data which is composed of n objects and m_1 attributes is denoted as $\hat{X}_{n\times m_1} = [\hat{x}_i^{t_1}]$, $i = 1,\ldots,n$, $t_1 = 1,\ldots,m_1$ and the data are called sometimes the two-way data [5]. Let $X = \{x_1,...,x_n\}$ is the set of objects. So, the two-way data matrix can be represented as follows:

$$\hat{X}_{n\times m_1} = \begin{pmatrix} \hat{x}_1^1 & \hat{x}_1^2 & \ldots & \hat{x}_1^{m_1} \\ \hat{x}_2^2 & \hat{x}_2^2 & \ldots & \hat{x}_2^{m_1} \\ \ldots & \ldots & \ldots & \ldots \\ \hat{x}_n^1 & \hat{x}_n^2 & \ldots & \hat{x}_n^{m_1} \end{pmatrix}. \quad (2)$$

So, the two-way data matrix can be represented as $\hat{X} = (\hat{x}^1,\ldots,\hat{x}^{m_1})$ using n-dimensional column vectors $\hat{x}^{t_1}$, $t_1 = 1,\ldots,m_1$, composed of the elements of the t_1-th column of $\hat{X}$.

The traditional optimization methods of fuzzy clustering are based on the concept of fuzzy c-partition [1]. The initial set $X = \{x_1,...,x_n\}$ of n objects represented by the matrix of similarity coefficients, the matrix of dissimilarity coefficients or the matrix of object attributes, should be divided into c fuzzy clusters. Namely, the grade u_{li}, $1 \le l \le c$, $1 \le i \le n$ to which an object x_i belongs to the fuzzy cluster A^l should be determined. For each object x_i, $i = 1,\ldots,n$ the grades of membership should satisfy the conditions of a fuzzy c-partition:

$$\sum_{l=1}^{c} u_{li} = 1,\ 1 \le i \le n,\ 0 \le u_{li} \le 1,\ 1 \le l \le c. \quad (3)$$

In other words, the family of fuzzy sets $P(X) = \{A^l \mid l = \overline{1,c}, c \le n\}$ is the fuzzy c-partition of the initial set of objects

$X = \{x_1, ..., x_n\}$ if condition (3) is met. Fuzzy c-partition $P(X)$ may be described with the aid of a partition matrix $P_{c \times n} = [u_{li}]$, $l = 1, \ldots, c$, $i = 1, \ldots, n$. The set of all fuzzy c-partitions will be denoted by Π. So, the fuzzy problem formulation in cluster analysis can be defined as the optimization task $Q \to \underset{P(X) \in \Pi}{extr}$ under the constraints (3), where Q is a fuzzy objective function.

The best known optimization approach to fuzzy clustering is the method of fuzzy c-means [6]. The FCM-algorithm is based on an iterative optimization of the fuzzy objective function, which takes the form:

$$Q_{FCM}(P, \overline{\mathrm{T}}) = \sum_{l=1}^{c} \sum_{i=1}^{n} u_{li}^{\gamma} \left\| x_i - \overline{\tau}^l \right\|^2 , \quad (4)$$

where u_{li}, $l = 1, \ldots, c$, $i = 1, \ldots, n$ is the membership degree, x_i, $i \in \{1, \ldots, n\}$ is the data point, $\overline{\mathrm{T}} = \{\overline{\tau}^1, \ldots, \overline{\tau}^c\}$ is the set fuzzy clusters prototypes, and $\gamma > 1$ is the weighting exponent.

The purpose of the classification task is to obtain the solutions $P(X)$ and $\overline{\tau}^1, \ldots, \overline{\tau}^c$ which minimize equation (4). Some other similar objective function-based fuzzy clustering algorithms are considered in [1], [5] and [6] in detail.

However, the condition of fuzzy c-partition is very difficult from essential positions. So, a possibilistic approach to clustering was proposed by Krishnapuram and Keller in [3] and developed by other researchers. Major algorithms of possibilistic clustering are objective function-based procedures.

A concept of possibilistic partition is a basis of possibilistic clustering methods and membership values μ_{li}, $l = 1, \ldots, c$, $i = 1, \ldots, n$ can be interpreted as the values of typicality degree. For each object x_i, $i = 1, \ldots, n$ the grades of membership should satisfy the conditions of a possibilistic partition:

$$\sum_{l=1}^{c} \mu_{li} > 0 , \; 0 \le \mu_{li} \le 1 . \quad (5)$$

So, the family of fuzzy sets $\mathrm{Y}(X) = \{A^l \mid l = \overline{1, c}, c \le n\}$ is the possibilistic partition of the initial set of objects $X = \{x_1, ..., x_n\}$ if condition (5) is met. Obviously that the conditions of the possibilistic partition (5) are more flexible than the conditions of the fuzzy c-partition (3).

In order to be applying the found cluster prototype as classifiers, they need to be given reasonable names. One can then use these names as column titles of the membership matrix when using the recall function of the FCM-algorithm. This helps in the interpretation of the results.

The process of assigning class names to cluster prototypes is called labeling. A labeling method for the fuzzy c-means method is to inspect the cluster prototypes and their respective membership values of the various attributes, and to assign a label manually.

However, usually, it is already known when training a classifier which objects belong to which classes. This information can be taken into account to use so as automatic the fuzzy c-means labeling process. The corresponding labeling procedure is described in [7] in detail. The principal idea of the procedure is to present a sample of objects to the FCM-classifier whose class membership are known in the hard form of 0 or 1 values and have also been calculated by the procedure. By means of the given cluster membership values for each fuzzy cluster prototype, the fuzzy cluster prototypes can be associated with their respective classes.

On the other hand, all objective function-based fuzzy clustering algorithms are iterative procedures and the initial fuzzy c-partition $P(X)$ is initialized randomly. So, coordinates of fuzzy clusters prototypes and values of membership functions will be different in each experiment for the same data set, because the result of classification is sensitive to initialization. Moreover, major objective function-based fuzzy

clustering algorithms are need for using some validity measures [1] for determining the most "plausible" number c of fuzzy clusters in the sought fuzzy c-partition $P(X)$. So, a problem of rapid automatic labeling is arises.

The effective labeling procedure for heuristic algorithms of possibilistic clustering was proposed in [8] and the procedure is the basis of the labelling procedure for consequents of derived fuzzy rules. However, basic definitions of the heuristic approach to possibilistic clustering should be considered in the first place.

3. Basic concepts of the heuristic approach to possibilistic clustering

Let us remind the basic concepts of the heuristic method of possibilistic clustering [4]. Let $X = \{x_1,...,x_n\}$ be the initial set of elements and $T: X \times X \to [0,1]$ some fuzzy tolerance on X with $\mu_T(x_i, x_j) \in [0,1]$, $\forall x_i, x_j \in X$ being its membership function. Let α be the α-level value of the fuzzy tolerance T, $\alpha \in (0,1]$. Columns or rows of the fuzzy tolerance matrix are fuzzy sets $\{A^1,...,A^n\}$ on the universal set X. Let A^l, $l \in \{1,\dots,n\}$ be a fuzzy set on X with $\mu_{A^l}(x_i) \in [0,1]$, $\forall x_i \in X$ being its membership function. The α-level fuzzy set $A^l_{(\alpha)} = \left\{(x_i, \mu_{A^l}(x_i)) \mid \mu_{A^l}(x_i) \geq \alpha, x_i \in X\right\}$ is fuzzy α-cluster. So, $A^l_{(\alpha)} \subseteq A^l$, $\alpha \in (0,1]$, $A^l \in \{A^1,\dots,A^n\}$ and $\mu_{A^l}(x_i)$ is the membership degree of the element $x_i \in X$ for some fuzzy α-cluster $A^l_{(\alpha)}$, $\alpha \in (0,1]$, $l \in \{1,\dots,n\}$. The membership degree will be denoted μ_{li} in further considerations. The membership degree of the element $x_i \in X$ for some fuzzy α-cluster $A^l_{(\alpha)}$, $\alpha \in (0,1]$, $l \in \{1,\dots,n\}$ can be defined as a

$$\mu_{li} = \begin{cases} \mu_{A^l}(x_i), & x_i \in A^l_\alpha \\ 0, & otherwise \end{cases}, \qquad (6)$$

where $A^l_\alpha = \{x_i \in X \mid \mu_{A^l}(x_i) \geq \alpha\}$, $\alpha \in (0,1]$ is the α-level of a fuzzy set A^l and the α-level is the support of the fuzzy α-cluster $A^l_{(\alpha)}$, $A^l_\alpha = Supp(A^l_{(\alpha)})$. The value of α is the tolerance threshold of fuzzy α-cluster elements.

Let $\{A^1_{(\alpha)},..., A^n_{(\alpha)}\}$ be the family of fuzzy α-clusters for some α. The point $\tau^l_e \in A^l_\alpha$, for which

$$\tau^l_e = \arg\max_{x_i} \mu_{li}, \ \forall x_i \in A^l_\alpha, \qquad (7)$$

is called a typical point of the fuzzy α-cluster $A^l_{(\alpha)}$, $\alpha \in (0,1]$, $l \in [1,n]$. A set $K(A^l_{(\alpha)}) = \{\tau^l_1,\dots,\tau^l_{|l|}\}$ of typical points of the fuzzy cluster $A^l_{(\alpha)}$ is a kernel of the fuzzy cluster and $card\left(K(A^l_{(\alpha)})\right) = |l|$ is a cardinality of the kernel. If the fuzzy cluster have an unique typical point, then $|l| = 1$.

Let $R^\alpha_z(X) = \{A^l_{(\alpha)} \mid l = \overline{1,c}, 2 \leq c \leq n\}$ be a family of fuzzy α-clusters for some value of tolerance threshold α, which are generated by a fuzzy tolerance T on the initial set of elements $X = \{x_1,...,x_n\}$. If condition

$$\sum_{l=1}^{c} \mu_{li} > 0, \ \forall x_i \in X, \qquad (8)$$

is met for all $A^l_{(\alpha)}$, $l = \overline{1,c}$, $c \leq n$, then the family is the allotment of elements of the set $X = \{x_1,...,x_n\}$ among fuzzy α-clusters $\{A^l_{(\alpha)}, l = \overline{1,c}, 2 \leq c \leq n\}$ for some value of the tolerance threshold α. It should be noted that several allotments $R^\alpha_z(X)$ can exist for some tolerance threshold α. That is why symbol z is the index of an allotment.

Obviously, the definition of the allotment among fuzzy clusters (8) is similar to the definition of the possibilistic partition (5). So, the allotment among fuzzy clusters can be considered as the possibilistic partition

and fuzzy clusters in the sense of (6) are elements of the possibilistic partition.

Allotment $R_I^\alpha(X) = \{A_{(\alpha)}^l \mid l = \overline{1,n}, \alpha \in (0,1]\}$ of the set of objects among n fuzzy clusters for some tolerance threshold $\alpha \in (0,1]$ is the initial allotment of the set $X = \{x_1,...,x_n\}$. In other words, if initial data are represented by a matrix of some fuzzy T then lines or columns of the matrix are fuzzy sets $A^l \subseteq X$, $l = \overline{1,n}$ and α-level fuzzy sets $A_{(\alpha)}^l$, $l = \overline{1,c}$, $\alpha \in (0,1]$ are fuzzy clusters. These fuzzy clusters constitute an initial allotment for some tolerance threshold α and they can be considered as clustering components. If some allotment $R_{c(z)}^\alpha(X) = \{A_{(\alpha)}^l \mid l = \overline{1,c}, c \le n\}$ corresponds to the formulation of a concrete problem, then this allotment is an adequate allotment. In particular, if a condition

$$\bigcup_{l=1}^{c} A_\alpha^l = X,$$

(9)

and a condition

$$card(A_\alpha^l \cap A_\alpha^m) = 0, \forall A_{(\alpha)}^l, A_{(\alpha)}^m, \quad l \ne m, \alpha \in (0,1]$$

(10)

are met for all fuzzy clusters $A_{(\alpha)}^l$, $l = \overline{1,c}$ of some allotment $R_{c(z)}^\alpha(X) = \{A_{(\alpha)}^l \mid l = \overline{1,c}, c \le n\}$ for a value $\alpha \in (0,1]$, then the allotment is the allotment among fully separate fuzzy clusters.

Fuzzy clusters in the sense of definition (6) can have an intersection area. If the intersection area of any pair of different fuzzy clusters is an empty set, then conditions (9) and (10) are met and fuzzy clusters are called fully separate fuzzy clusters. Otherwise, fuzzy clusters are called particularly separate fuzzy clusters and $w \in \{0,\dots,n\}$ is the maximum number of elements in the intersection area of different fuzzy clusters. For $w = 0$ fuzzy clusters are fully separate fuzzy clusters.

Thus, the conditions (9) and (10) can be generalized for a case of particularly separate fuzzy clusters. So, a condition

$$\sum_{l=1}^{c} card(A_\alpha^l) \ge card(X), \quad \forall A_{(\alpha)}^l \in R_{c(z)}^\alpha(X), \quad , \quad \alpha \in (0,1], card(R_{c(z)}^\alpha(X)) = c$$

(11)

and a condition

$$card(A_\alpha^l \cap A_\alpha^m) \le w, \forall A_{(\alpha)}^l, A_{(\alpha)}^m, \quad l \ne m, \alpha \in (0,1]$$

(12)

are generalizations of conditions (9) and (10). Obviously, if $w = 0$ in conditions (11) and (12) then conditions (9) and (10) are met. The adequate allotment $R_{c(z)}^\alpha(X)$ for some value of tolerance threshold $\alpha \in (0,1]$ is a family of fuzzy clusters which are elements of the initial allotment $R_I^\alpha(X)$ for the value of α and the family of fuzzy clusters should satisfy the conditions (11) and (12). So, the construction of adequate allotments $R_{c(z)}^\alpha(X) = \{A_{(\alpha)}^l \mid l = \overline{1,c}, c \le n\}$ for every α is a trivial problem of combinatorics.

Allotment $R_P^\alpha(X) = \{A_{(\alpha)}^l \mid l = \overline{1,c}\}$ of the set of objects among the minimal number c, $2 \le c \le n$ of fully separate fuzzy clusters for some tolerance threshold $\alpha \in (0,1]$ is the principal allotment of the set $X = \{x_1,...,x_n\}$.

Several adequate allotments can exist. Thus, the problem consists in the selection of the unique adequate allotment $R_c^*(X)$ from the set B of adequate allotments, $B = \{R_{c(z)}^\alpha(X)\}$, which is the class of possible solutions of the concrete classification problem. The selection of the unique adequate allotment $R_c^*(X)$ from the set $B = \{R_{c(z)}^\alpha(X)\}$ of adequate allotments must be made on the basis of evaluation of allotments. In particular, the criterion

$$F(R_{c(z)}^\alpha(X), \alpha) = \sum_{l=1}^{c} \frac{1}{n_l} \sum_{i=1}^{n_l} \mu_{li} - \alpha \cdot c,$$

(13)

where c is the number of fuzzy clusters in the allotment $R_{c(z)}^\alpha(X)$ and $n_l = card(A_\alpha^l)$,

$A^l_{(\alpha)} \in R^{\alpha}_{c(z)}(X)$ is the number of elements in the support of the fuzzy cluster $A^l_{(\alpha)}$, can be used for evaluation of allotments. Maximum of criterion (13) corresponds to the best allotment of objects among c fuzzy clusters. So, the classification problem can be characterized formally as determination of the solution $R^*_c(X)$ satisfying

$$R^*_c(X) = \arg \max_{R^{\alpha}_{c(z)}(X) \in B} F(R^{\alpha}_{c(z)}(X), \alpha),$$

(14)

The problem of cluster analysis can be defined in general as the problem of discovering the unique allotment $R^*_c(X)$, resulting from the classification process and detection of fixed or unknown number c of fuzzy clusters can be considered as the aim of classification.

Thus, the problem of cluster analysis can be defined as the problem of discovering the unique allotment $R^*_c(X)$, resulting from the classification process and detection of fixed or unknown number c of fuzzy α-clusters can be considered as the aim of classification.

Direct heuristic algorithms of possibilistic clustering can be divided into two types: relational versus prototype-based. A fuzzy tolerance relation matrix is a matrix of the initial data for the direct heuristic relational algorithms of possibilistic clustering and a matrix of attributes (2) is a matrix for the prototype-based algorithms. In particular, the group of direct relational heuristic algorithms of possibilistic clustering includes

- D-AFC(c)-algorithm: using the construction of the allotment among given number c of partially separate fuzzy clusters;
- D-PAFC-algorithm: using the construction of the principal allotment among an unknown minimal number of at least c fully separate fuzzy clusters;
- D-AFC-PS(c)-algorithm: using the partially supervised construction of the allotment among given number c of partially separate fuzzy clusters.

On the other hand, the family of direct prototype-based heuristic algorithms of possibilistic clustering includes

- D-AFC-TC-algorithm: using the construction of the allotment among an unknown number c of fully separate fuzzy clusters;
- D-PAFC-TC-algorithm: using the construction of the principal allotment among an unknown minimal number of at least c fully separate fuzzy clusters;
- D-AFC-TC(α)-algorithm: using the construction of the allotment among an unknown number c of fully separate fuzzy clusters with respect to the minimal value α of the tolerance threshold.

It should be noted that these direct prototype-based heuristic possibilistic clustering algorithms are based on a transitive closure of an initial fuzzy tolerance relation.

On the other hand, a family of direct prototype-based heuristic possibilistic clustering algorithms based on a transitive approximation of a fuzzy tolerance is proposed in [9].

So, the matrix of memberships $R^*_c(X) = [\mu_{li}]$, the value α of the tolerance threshold and the set of kernels $\{K(A^1_{(\alpha)}), \ldots, K(A^c_{(\alpha)})\}$ are results of classification. The results will be constant in each experiment for the same data set, because the sought clustering structure $R^*_c(X)$ of the set of objects X is based directly on the formal definition of fuzzy cluster and the possibilistic memberships are determined directly from the values of the pairwise similarity of objects.

The training data matrix (2) and clustering results are a basis for constructing of Mamdani-type fuzzy rules (1). The corresponding methodology described in [4] in detail.

4. A labeling procedure for fuzzy rules consequents

Fuzzy classifier can be generated directly by some heuristic algorithm of possibilistic clustering [4]. A fuzzy rule is associated to each fuzzy α-cluster of the obtained allotment, $R_c^*(X)$. So, a number of fuzzy rules is equal to a number of fuzzy α-clusters and equal to a number of output variables y_l, $l=1,\ldots,c$.

The results obtained from heuristic algorithms of possibilistic clustering are stable. The set of kernels $\{K(A_{(\alpha)}^1),\ldots,K(A_{(\alpha)}^c)\}$ and the set of labels $\{label\ 1,\ldots,label\ c\}$ are inputs for a labeling procedure [8]. We assume that a condition $card\left(K(A_{(\alpha)}^l)\right)=1$ is met for each kernel $K(A_{(\alpha)}^l)$, $l=\overline{1,c}$. In other words, the set of typical points $\{\tau^1,\ldots,\tau^c\}$ is given.

Each output variable $\{y_1,\ldots,y_c\}$ corresponds to a fuzzy α-cluster of the obtained allotment, $R_c^*(X)$. There is a two-step procedure which can be described as follows:

1. Perform the following operations for each typical point τ^l, $l=\overline{1,c}$ and each label $label\ m$, $m=\overline{1,c}$:
   1.1 Let $l:=1$ and $m:=1$;
   1.2 Check the following condition:
   if τ^l corresponds to $label\ m$ **then** the label $label\ m$ is the label for the typical point τ^l and go to step 1.3 **else** $m:=m+1$ and go to step 1.2;
   1.3 Check the following condition:
   if the typical point τ^l is labeled **then** $l:=l+1$ and go to step 1.2 **else** go to step 1.4;
   1.4 Check the following condition:
   if all typical points τ^l, $l=\overline{1,c}$ are labeled **then** go to step 2.
2. Perform the following operations for each output variable y_l, $l=\overline{1,c}$ and each typical point τ^l, $l=\overline{1,c}$:
   2.1 Let $l:=1$;
   2.2 A label of typical point τ^l should be assigned to output variable y_l;
   2.3 Check the following condition:
   if a condition $l<c$ is met **then** $l:=l+1$ and go to step 2.2 **else** stop.

That is why the proposed labeling procedure for consequents of fuzzy rules can be considered as an extended version of the procedure for labeling fuzzy α-clusters [8].

5. An illustrative example

The Anderson's Iris database [10] is the most known database to be found in the pattern recognition literature. The data set represents different categories of Iris plants having four attribute values. The four attribute values represent the sepal length, sepal width, petal length and petal width measured for 150 irises. It has three classes Setosa, Versicolor and Virginica, with 50 samples per class. Examples of records in the database are presented in Table 1.

Table 1. Examples of records in the Iris database

Numbers of objects	Attributes				Labels of classes
	Sepal length	Sepal width	Petal length	Petal width	
…	…	…	…	…	…
18	5.1	3.3	1.7	0.5	SETOSA
…	…	…	…	…	…
48	5.5	2.6	4.4	1.2	VERSICOLOR
…	…	…	…	…	…

108	5.6	2.8	4.9	2.0	VIRGINICA
…	…	…	…	…	…

The Anderson's Iris data form the matrix of attributes $\hat{X}_{150\times 4}=[\hat{x}_i^{t_1}]$, $i=1,\ldots,150$, $t_1=1,\ldots,4$, where the sepal length is denoted by $\hat{x}^1$, sepal width – by $\hat{x}^2$, petal length – by $\hat{x}^3$ and petal width – by $\hat{x}^4$. The data was normalized as follows:

$$x_i^{t_1}=\frac{\hat{x}_i^{t_1}}{\max\limits_i \hat{x}_i^{t_1}}. \quad (15)$$

So, each object can be considered as a fuzzy set x_i, $i=1,\ldots,150$ and $x_i^{t_1}=\mu_{x_i}(x^{t_1})\in[0,1]$, $i=1,\ldots,150$, $t_1=1,\ldots,4$, are their membership functions. The matrix of coefficients of pair wise dissimilarity between objects $I=[\mu_I(x_i,x_j)]$, $i,j=1,\ldots,150$ can be obtained after application of some distance to the matrix of normalized data $X_{150\times 4}=[\mu_{x_i}(x^{t_1})]$, $i=1,\ldots,150$, $t_1=1,\ldots,4$. In particular, the normalized Euclidean distance [11]

$$e(x_i,x_j)=\sqrt{\frac{1}{m_1}\sum_{t_1=1}^{m_1}\left(\mu_{x_i}(x^{t_1})-\mu_{x_j}(x^{t_1})\right)^2}. \quad (16)$$

was applied to the normalized data.

The matrix of fuzzy tolerance $T=[\mu_T(x_i,x_j)]$ was obtained after application of complement operation

$$\mu_T(x_i,x_j)=1-\mu_I(x_i,x_j), \quad (17)$$

to the matrix of fuzzy intolerance $I=[\mu_I(x_i,x_j)]$, $i,j=1,\ldots,150$ obtained from previous operations. The labeled training data is shown in Fig. 1.

By executing the D-AFC(c)-algorithm [4] for $c=3$ using the normalized Euclidean distance (16), we obtain that the typical point of the first class τ^1 is the object x_{23}, the typical point of the second class τ^2 is the object x_{95}, and the typical point of the third class τ^3 is the object x_{98}. The clustering result is presented in Fig. 2. The set of labels is {SETOSA, VERSICOLOR, VIRGINICA} and these labels were assigned to corresponding output variables. The performance of the generated fuzzy classifier is shown in Fig. 3 in which $l=1,\ldots,3$ is the number rule.

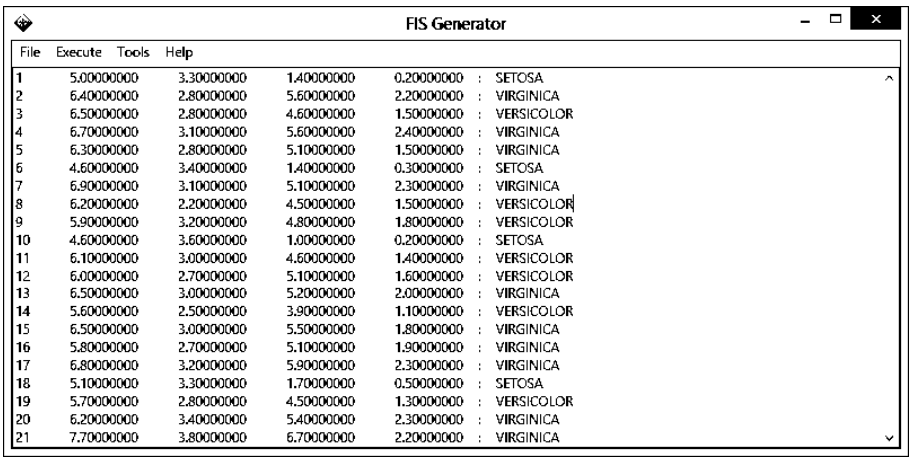


Fig. 1. The training data set

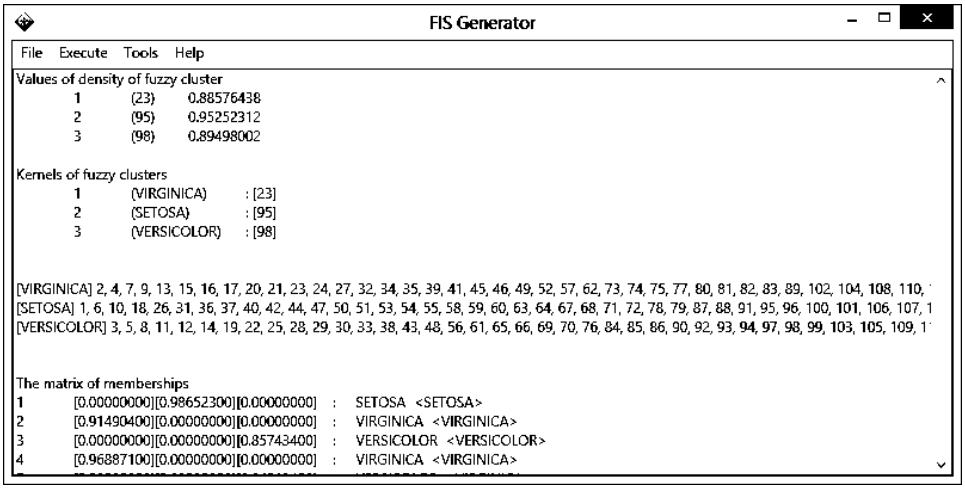


Fig. 2. The clustering result

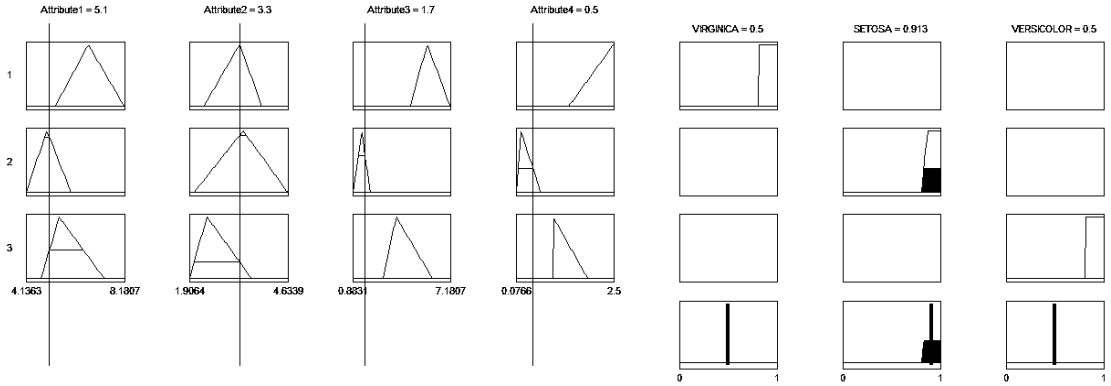


Fig. 3. Performance of the fuzzy classifier which was generated from Anderson's Iris data

So, the result of application the proposed labeling procedure seems to be satisfactory.

6 Conclusions

The fast procedure for labeling consequents of fuzzy rules generated by using heuristic possibilistic clustering results is proposed in the paper. Stability of results of heuristic possibilistic clustering is a basis of the developed procedure. The proposed procedure can be very useful in process of adding new records to database. For the purpose, all records in a database can be classified by using heuristic possibilistic clustering and fuzzy classifier can be generated on a basis of clustering results. That is why the corresponding label can be assigned to a new record which added to the database. These perspectives for investigations are of great interest both from the theoretical point of view and from the practical one as well.

Acknowledgment

This investigation was done in framework of a project of the scientific program "Monitoring-SG" of the Union State of Russia and Belarus.

References

[1] F. Höppner, F. Klawonn, R. Kruse and T. Runkler, *Fuzzy Cluster Analysis: Methods for Classification, Data Analysis and Image Recognition*, Chichester, Wiley Intersciences, 1999.

[2] E.H. Mamdani and S. Assilian, "An Experiment in Linguistic Synthesis with a Fuzzy Logic Controller", *International*

Journal of Man-Machine Studies, Vol. 7, No. 1, 1975, pp. 1-13.

[3] R. Krishnapuram and J.M. Keller, "A Possibilistic Approach to Clustering", *IEEE Transactions on Fuzzy Systems*, Vol. 1, No. 2, 1993, pp. 98-110.

[4] D.A. Viattchenin, *A Heuristic Approach to Possibilistic Clustering: Algorithms and Applications*, Heidelberg, Springer, 2013.

[5] M. Sato-Ilic and L.C. Jain, *Innovations in Fuzzy Clustering*, Heidelberg, Springer, 2006.

[6] J.C. Bezdek, *Pattern Recognition with Fuzzy Objective Function Algorithms*, New York, Plenum Press, 1981.

[7] *DataEngine: Tutorials and Theory*, Aachen, MIT GmbH, 1999.

[8] D.A. Viattchenin, A. Damaratski, E. Nikolaenya and S. Shyrai, "An Outline for an Approach to Automatic Labeling for Interpretation of Heuristic Possibilistic Clustering Results," *Proceedings of the 12th Int. Conf. on Pattern Recognition and Information Processing*, pp. 290-294, Minsk, Belarus, 28-30 May 2014.

[9] D.A. Viattchenin and A. Damaratski, "Direct Heuristic Algorithms of Possibilistic Clustering Based on Transitive Approximation of Fuzzy Tolerance", *Informatica Economica Journal*, Vol. 17, No. 3, 2013, pp. 5-15.

[10] E. Anderson, "The Irises of the Gaspe Peninsula", *Bulletin of the American Iris Society*, Vol. 59, No. 1, 1935, pp. 2-5.

[11] A. Kaufmann, *Introduction to the Theory of Fuzzy Subsets*, New York, Academic Press, 1975.

Dmitri A. VIATTCHENIN was born on 13.12.1972 in Moscow, Russia. In 1994 he graduated the Department of Mathematical Modelling and Data Analysis of the Faculty of Applied Mathematics and Informatics in Belarusian State University. He defended his PhD in the field of philosophy of sciences and technology in 1998. He got the title of Associate Professor in the specialty informatics and computer engineering in 2013. He is currently an associate professor in the Department of Software Information Technology of the Belarusian State University of Informatics and Radio-Electronics and a leading researcher in the Laboratory of Information Protection Problems of the United Institute of Informatics Problems of the National Academy of Sciences of Belarus. His research interests include techniques of fuzzy and possibilistic clustering, fuzzy control, intelligent decision making systems, simulation of complex systems, and philosophical aspects of artificial intelligence. He is the author of over 135 publications including papers in international and domestic journals and conference proceedings, 3 monographs and 1 edited volume. He is also a reviewer various journals, for example Fuzzy Sets and Systems, Neurocomputing, Journal of Information and Organizational Sciences.

Stanislau SHYRAI was born on 28.09.1991 in Stavropol, Russia. He graduated from the Faculty of Computer Systems and Networks of the Belarusian State University of Informatics and Radio-Electronics in 2014. At present he is a lecturer in the Department of Software Information Technology of the Belarusian State University of Informatics and Radio-Electronics. His interest domains related to computer science are: software development methodologies, intuitionistic fuzzy sets, fuzzy cluster analysis, fuzzy classifiers, information systems and databases.

Aliaksandr DAMARATSKI was born on 29.11.1983 in Mogilev, Belarus. He has graduated the Faculty of Computer Design of Belarusian State University of Informatics and Radio-Electronics in 2005. He is currently a junior researcher in the Laboratory of Information Protection Problems of the United Institute of Informatics Problems of the National Academy of Sciences of Belarus. His work focuses on possibilistic clustering, image processing, decision making, developing software, and computer networks analysis. He is the author of more than 25 papers in journals and conference proceedings.

BI solutions for modern management of organizations - Cognos

Ruxandra OPREA, Ionuț LĂCĂTUȘ, Cătălin HUȚANU
[1, 2, 3]University of Economic Studies, Bucharest, Romania
rux.oprea@yahoo.com, ionutlacatus@gmail.com, catalin@hutanu.ro

Business Intelligence has evolved in the last decade increasingly relying on real-time data. Business analysis has become essential. This involves actions in response to results analysis and instant change of business processes parameters making BI beneficial for several reasons.
Keywords: *Business Intelligence, Performance Management, Cognos*

1 Introduction

Business Intelligence (BI) has two different basic meanings related to use of the term *intelligence* [1]. The first meaning, less frequent, is referring to the capacity of human intelligence in business or in activities. Business Intelligence is a new and powerful area of investigation of the applicability of human cognitive abilities and artificial intelligence technologies in management and decision support in various business issues. The second meaning displayed in Fig.1 refers to intelligence as information valued for its use and relevance.

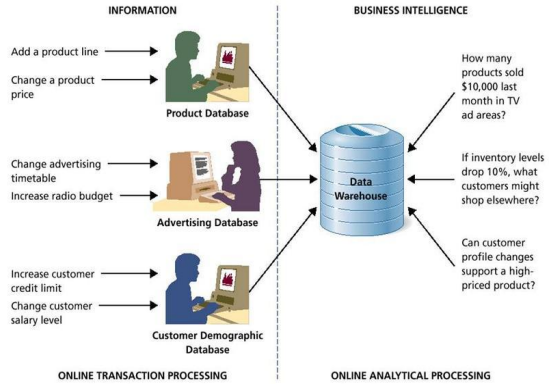


Fig.1 Information and business intelligence [2]

Application. BI applications include decision support systems, query and reporting tools, online analytical processing and also forecasting and data mining systems. Ultimately, the final results of BI implementations are depth analysis, refining and concentrating a large number of business information in concrete performance indicators and finally the organizational knowledge. BI implementations are efforts involving multiple issues, from organizational strategy to organizational processes and management, from application management to information infrastructure changes. BI projects do not aim to teach managers how to make the right decisions; instead they help them make decisions based on facts and figures, not on assumptions. Companies collect vast amounts of data through transactional systems (e.g. ERP, CRM, and SCM) that have implemented over the years and which they use daily to perform a variety of corporate functions. Before the notion of BI was launched, there was no concept that allowed the use of large amounts of data by integrating and transforming them into information [3]. Development of BI concepts and technologies creates a management environment where current and new data can be used to improve the quality in the process of decision making. In addition, the existence of large volumes of transactional data, especially transactional data with a high degree of specificity and particularity, creates opportunities for management to improve forecast accuracy.

BI Tools:

- Spreadsheets
- Query software and OLAP (Online analytical processing)

- Digital Dashboards (an executive information system user interface is designed to be easy to read)
- Data mining (process of extracting patterns from large volumes of data by combining methods from statistics and artificial intelligence with those of the database management)
- Decision Engineering (framework that unifies a number of best practices for organizing decision making)
- Process mining (extracting knowledge from the events recorded by the information system) [3]
- Business performance management is a set of management and analytic processes that enable management of an organization's performance in order to achieve one or more goals.

Fig.2 Product capabilities [4]

Traditional BI capabilities are extended beyond reporting, analysis and creating dashboards and scorecards, which are represented in figure 2. Those who use BI can also plan and conduct scenario modelling, real-time monitoring and predictive analytics. Managers want to be able to perform their own analyzes and view data in real time [6]. They want to connect predictive analytics and social media analysis with traditional BI reports and dashboards to get a complete view of the business, customers, competitors and market.

Applications for social media make it possible to analyze the attitude of customers, products and associated brands and emerging issues on the organization or market segment.

IBM Cognos Business Intelligence provides reports, analysis, dashboards and scoreboards to help support the way people think and work when they are trying to understand business performance. You can freely explore information, analyze key facts and quickly collaborate to align decisions with key stakeholders [5].

- **Reports** equip users with the data they necessitate to make fact-grounded conclusions [4].
- **Dashboards** avail users access, interact and personalize content in a way that assists on how they make decisions help users access, interact and personalize content in a way that supports how they make decisions.
- **Analysis capabilities** provide access to data from multiple angles and perspectives so you can watch and study it to make informed decisions [4].
- **Collaboration** capabilities include communication tools and social networking to fuel the exchange of thoughts during the decision-making process.
- **Scorecarding** capabilities automate the capture, management and monitoring of business metrics so you can compare them with your strategic and operational targets [5].

2. IBM Cognos Business Intelligence

Cognos Incorporated was an Ottawa, Ontario-based company making business intelligence and performance management software. “Founded in 1969, at its peak Cognos employed almost 3,500 people and served more than 23,000 customers in over 135 countries until being acquired by IBM on January 31, 2008. While no longer an independent company, the Cognos name continues to be applied to IBM's line of business intelligence and performance management products” [11].

IBM Cognos Business Intelligence is an integrated business intelligence suite that offers a broad range of functionality to help understanding an organization's information. Everyone in the system can use IBM Cognos 8 to view or create business reports, examine information, and monitor outcomes and metrics so that they can prepare efficient business decisions [8].

IBM Cognos 8 integrates the following business intelligence activities, shown in table 1, in one Web-based solution.

Table 1 Components and activities [8]

IBM Cognos Connection	Publishing, managing, and viewing content
IBM Cognos Insight	Managed workspaces
IBM Cognos Workspace	Interactive workspaces
IBM Cognos Workspace Advanced	Ad hoc querying and data exploration
IBM Cognos Report Studio	Managed reporting
IBM Cognos Event Studio	Event management and alerting
IBM Cognos Metric Studio	Scorecarding and metrics
IBM Cognos for Microsoft(tm) Office	Works with IBM Cognos BI content in Microsoft Office
IBM Cognos Query Studio	Ad-hoc querying
IBM Cognos Analysis Studio	Data exploration

"**IBM Cognos Connection** is the Web portal for IBM Cognos Business Intelligence." It is the beginning spot to access BI information and the functionality of IBM Cognos BI. The Web portal can be used to publish, find, manage, organize, and view organization's business intelligence content, such as reports, scorecards, and agents. Possessing the necessary permissions, various studios can be accessed from the portal and this can be used for content administration, including programming and distributing reports, and creating businesses. In **IBM Cognos Insight**, the user can analyze information, explore scenarios, and influence decisions by creating personal or managed workspaces. The interactive workspaces can be used to pass the results to managers. Because Cognos Insight supports write-back, the user can also use these workspaces to collect and consolidate management targets, commitments, and forecasts. "IBM Cognos Insight is provided with IBM Cognos BI. IBM Cognos Connection Installer for Cognos Insight can be used in order to install provisioning software on Cognos BI servers. This software allows multiple users to download and install IBM Cognos Insight on their computers from the Cognos Connection interface."

In **IBM Cognos Workspace**, the user can create sophisticated interactive workspaces using IBM Cognos content, as well as external information sources such as TM1 Websheets and CubeViews, according to specific information needs. The user can view and open favourite workspaces and reports, manage the content in the workspaces, and e-mail the workspaces. Also, can be used comments, activities and social software like IBM Connections for collaborative decision making.

With **IBM Cognos Workspace Advanced**, the user can perform advanced data exploration and create simple reports. When the user is in a workspace in IBM Cognos Workspace and wants to perform deeper analysis and report authoring, he can seamlessly graduate to Cognos Workspace Advanced, where he can perform more advanced data exploration, such as adding additional measures, conditional formatting, and advanced computations [7]. The user can also launch Cognos Workspace Advanced from the IBM Cognos Connection portal. With Cognos Workspace Advanced, the user can create reports with

relational or dimensional data sources, and that shows the data in lists, crosstabs, and charts. He can also use his external data source. With this feature, a report authored in IBM Cognos Report Studio can be accessed; the objects that can be inserted only in Report Studio can be seen (such as map). However, he cannot modify these objects.

Using **Report Studio**, report authors create, edit, and distribute a wide series of professional reports. They can also define corporation standard report templates for use in Query Studio, and edit/modify reports created in Query Studio or in Analysis Studio.

Report Studio can be used for reports which are intended for a large audience, exist long enough to require maintenance for updateing requirements and data or require a thorough check on the appearance. Report Studio provides powerful features, such as transferring in burst, prompts, maps and create advanced graphics and offers many ways to customize reports.

Event Studio is an action oriented agent which notifies the users when an bussiness event occurs. Agents can publish details to the portal can deliver alerts by e-mail, can run and distribute reports based on events and can monitor the status of events.

In **Metric Studio**, the user can create and deliver a customized scorecarding environment for monitoring and analyzing metrics throughout organization. Users can monitor, analyze, and report on time-critical information by using scorecards based on cross-functional metrics.

In **IBM Cognos for Microsoft Office**, the user can work with secure IBM Cognos Business Intelligence content in his familiar Microsoft Office environment. He can extract report content from a variety of IBM Cognos applications, including IBM Cognos BI and IBM Cognos PowerPlay. IBM Cognos for Microsoft Office provides access to all IBM Cognos report content, including data, metadata, headers, footers, and charts. The user can use predefined reports, or create new content using IBM Cognos Query Studio, IBM Cognos Analysis Studio, or IBM Cognos Report Studio. By importing content into Microsoft Excel spreadsheet software, the user can work with the data and leverage Microsoft Excel's formatting, calculation, and presentation capabilities. The user can also use the formatting and charting features of Microsoft Excel. By importing content into Microsoft PowerPoint and Microsoft Word, various reports and charts can be included to enhance the presentations and documents.

Using **Query Studio**, users with little or no training can quickly design, create and save reports to meet reporting needs not covered by the standard, professional reports created in Report Studio [8].

In **Analysis Studio**, users can explore, analyze, and compare dimensional data. Analysis Studio provides access to dimensional, OLAP (online analytical processing), and dimensionally modelled relational data sources. Analyses created in Analysis Studio can be opened in Report Studio and used to build professional reports [12].

IBM Cognos Transformer is a multi-dimensional data modelling component designed for use with IBM Cognos Business Intelligence. This component is useful to create a multi-dimensional model: a business presentation of the information in one or more different data sources that share common data [8]. After the user add the needed metadata from IBM Cognos Business Intelligence packages, reports, and other various data sources, model the dimensions, customize the measures, and apply IBM Cognos BI secured views with dimensional filtering, he can create IBM Cognos PowerCubes based on this model. These cubes can be deployed in order to support OLAP reporting and analysis around the globe.

IBM Cognos Business Intelligence Administrators ensure that IBM Cognos BI

runs without interruption and at optimal performance. They can:

- Define connections to data sources of the organization
- Define security permissions for users and groups in the organization
- Specify distribution lists, contacts and printers
- Manage servers and dispatchers, and can finely adjust performance for IBM Cognos BI
- Predefine links to an entire package that authors can easily add it to their reports
- Customize the appearance and functionality of IBM Cognos BI

IBM Cognos BI is secured by setting permissions and by enabling user authentication. When anonymous access is enabled, IBM Cognos BI can be used as a specific user without authentication. In IBM Cognos BI, administrators define permissions so that users can access functions. For example, to edit a report using IBM Cognos Report Studio, security permissions and the appropriate license are required. In addition, each entry in IBM Cognos Connection is secured by defining those who can read it, edit and run [8].

IBM Cognos BI comes with the following capabilities:

1. **Reports**

- Cognos Business Intelligence features professional report generation capabilities that are easy to use and help minimize the effort to build them.
- Users can create their own reports (queries) or edit existing reports.
- Collaborative reporting features built into Cognos Business Intelligence helps users to communicate between them to make decisions and to gain additional insights.
- Users can access reports on mobile devices and can already interact with the report while downloading - can view the downloaded pages and interact with them without the need to wait for complete download of the report.

2. **Dashboards**

- With the ability to create dashboards, Cognos Business Intelligence allows users to create and customize their ways of viewing dashboards that they can shape the needs.
- Historical data with variable data and what-happens-if scenarios provides extended insight of performance.
- Portable dashboards can be created that can support decisions on when and where to create

3. **Analysis**

- With analysis capabilities of Cognos Business Intelligence, users can use analytical reporting, trend analysis, statistical analysis and some other intuitive tools that can easily analyze information.
- Financial and business analysis can estimate the short-term business through advanced analytics and predictive statements what-happens-if.
- Support for operational and strategic decision cycles may engage the right people at the right time for analysis.
- Cogons Business Intelligence comes with options for mobile devices, analysis and offline interactivity along with versions for Microsoft Office to the appropriate user needs.

4. **Collaboration**

- The ability to form communities, to capture annotations, views and share opinions help streamline and improve decision-making groups.
- Workflow and ability to handle tasks are designed to connect users and to enhance activities coordination.
- Using IBM Connections and its social capabilities, users can share their opinions and request ideas.

5. **Scorecarding**

- Using measurement capabilities inside (scorecarding) of Cognos Business Intelligence, organization can compare

performance indicators with strategic goals.
- Facilities available that can analyze the organizational strategy can help departments and employees set their priorities.
- Model based on web interface increase management efficiency and encourages adoption by users.

3. Example – Performance Monitoring

With IBM Cognos Metric Studio, you can track the performance of an organization compared with its objectives. In a quick glance, decision-makers at every level of the organization can see the status of the organization and then react or make a plan. A metric is a key indicator of measures that compare actual results with the results of the target. An index of measurement registers also who is responsible for the outcome and impact index [8]. A scorecard is a collection of performance metrics and projects that reflect the strategic goals of a department in an organization. The following example shows how you can:

- examine a strategy map for a visual representation of the strategy and objectives of this strategy for an organization
- examine and understand the performance of a metric on a scorecard
- create an action on an index measuring
- add a metric to watch list

Suppose that we are a sales manager for a region in Sample Outdoors Company. Regularly consult a scorecard that contains metrics for the company's sales. Measurement indices provides us a quick comparison between actual sales and the company's objectives it. The amount returned is one of the indices of measurement. To perform this exercise, you need to have licensing and security permissions appropriate for this functionality.

Things to note:

- Measurement indices appear as **weak** (red), **medium** (yellow) or **excellent** (green).
- For each metric, we can see if the situation improves, stays the same or gets worse.
- When we position the pointer over the title of a metric, we can see a chart with historical performance metric. It extends a floating menu when you pause the cursor over a metric or above diagram.

Procedure

1. Opening IBM Cognos Connection from the web browser and access the URL obtained from the administrator. URL is like this: http://*nume_server*/cognos
2. On the **Welcome to IBM Cognos**, choose **Manage my measurement indices**.
3. Choose **Samples**, **Models**, **GO Metrics**.
4. Viewing a map of strategies: On the left pane, choose **Scorecards** tab and then **GO Consolidated**. In the right pane, select **Diagrams**. In **GO Strategy map** (figure 3), we see how we can quickly assess the performance of each purpose of the organization. Major indicators show the status of a specific strategy. Smaller indicators that appear in strategy show trend. Measured values of indices that do not register a good performance will appear in red. Indicators of status and trends, which appear in red, indicate possible problem areas. For example, in **Production and Distribution** function of strategy map, in **Control product quality** measuring index, **Return quantity %** is red, representing a measure of poor performance.

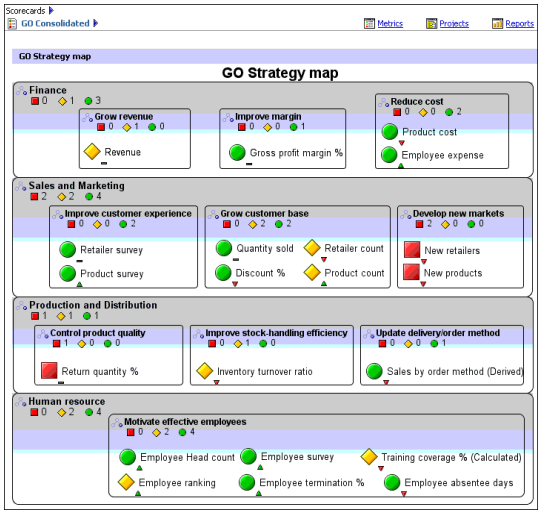


Fig.3 GO Strategy map

5. Exploring scorecard: In the left pane, click **Scorecards** and expand **GO Consolidated**. Note that Sample Outdoors Company has scorecards for each of the four functions of the company: Finance, Sales and Marketing, Production and Distribution and Human Resource. Expand **Production and distribution**, choose **Asia Pacific** and then choose **Metrics** tab showing all indices of measurement associated with the Asia Pacific region. Note that **Returns by reason** index in **Asia Pacific % wrong product ordered** is in red. To understand the cause of the problem, we further explore this metric: press **Asia Pacific - Amount refunded %**, then the **Diagrams** tab (figure 4). Expand measurement index in chart by clicking on the arrow next to the metric **Asia Pacific Returns by reason in % Orders failed**. We can see what data can be found in the index measuring **Quantity returned %**. We can also see which of the return reasons is a major problem.

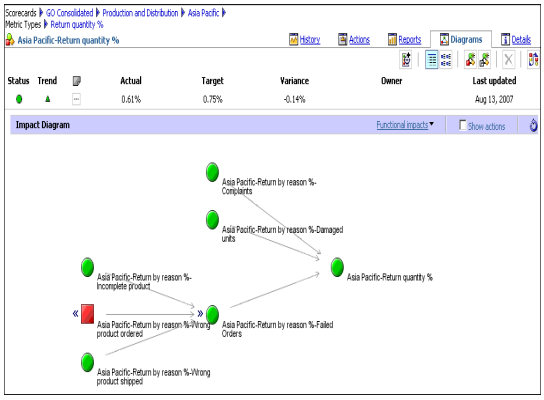


Fig.4 Diagram tab

To get more information on this metric, pause the cursor over **Asia Pacific - Quantity returned %**. (figure 5)

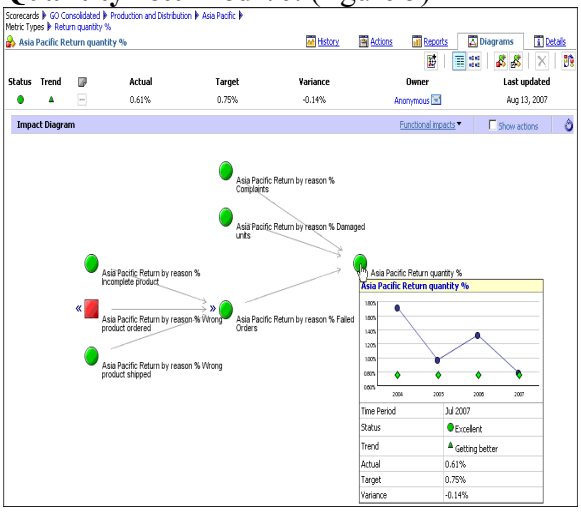


Fig.5 Metric tooltip

In the left pane, under **production and distribution**, press **Asia Pacific** and then on **Metrics** tab (figure 6) to view the metric related to the amount refunded. Note that the metric condition **Asia Pacific - Return by reason % wrong product ordered** is poor.

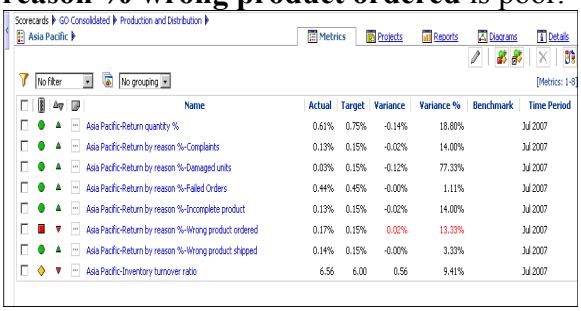


Fig.6 Metric tab

6. We click on the metric **Asia Pacific - Return by reason % - Wrong product ordered**. In Metric Studio, tabs represent

questions that we might put, as we try to solve a problem or understand the information.

7. In the right pane, the **History** tab (figure 7), we click on the list. The information in the History tab, answer the question "When?". We see the actual values on the target metric data for previous periods. Also see half-yearly and annual summaries.

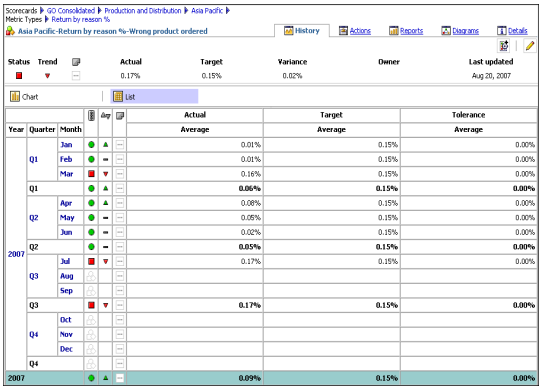


Fig.7 History tab

8. We click on the **Diagrams** tab. Diagrams tab information helps us understand the question "How?" by comparing our indices measuring other metrics. Impact diagrams show the relationship between the metric **Return by reason %** and other indices.

9. Create action for one of our business analysts, in which he investigates the performance of products returns: click the **Actions** tab. Choose new action. **Tip**: You must select a metric to see new action icon. In the **Name** box, type Kazumi Uragome the person who will own action. To announce the Kazumi what to do, in the **Description** box, type Please investigate. Near **Planned End** box, we click the Calendar icon and select a date a week away from today. **Hint**: You may need to scroll down to see **Planned End** box, and then we click on OK.

10. Add index measurement **Asia Pacific - Returns by reason % - wrong product ordered** on our list of supervisors to monitor it in the future easily in the following way: in the left panel, the scorecard P**roduction and distribution**, we click on **Asia Pacific** and then tab M**etrics**; we click the measuring index **Asia Pacific - Returns by Reason % - wrong product ordered** and on the superior toolbar choose **Add the watch list** (figure 8), and press **OK**. In the left panel, we click on **My Folders** on the **Watch List**. Index **Asia Pacific - Returns by Reason % - wrong product ordered** appears on our list of supervisors

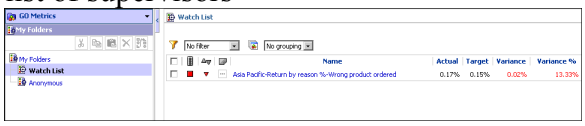


Fig.8 Watch list

4. OBIEE vs. COGNOS

In Business Intelligence Tools 2014 survey was intended to compare various Business Intelligence tools, a comparison that ignores the supplier. Comparison of Obie vs. fourteen other tools Cognos Business Intelligence in more than 100 selection criteria. The criteria we are talking here are essential for obtaining success with Business Intelligence [9]. For example: OBIEE support or Cognos-based reporting roles and reporting components can be reused in the implementation of business intelligence, whether they are used in web reports, dashboards, analysis, intranet or mobile version?

The survey results revealed that Obie and Cognos have scored almost equal if all criteria are taken into account. But in different categories, each one has different score, and also the costs each are different [9].

References

[1] - Business Intelligence: Concepts, Components, Techniques and Benefits, http://www.techrepublic.com/resource-library/whitepapers/business-intelligence-concepts-components-techniques-and-benefits/

[2] - Building Business Intelligence, http://exonous.typepad.com/mis/2004/03/building_busine.html

[3] - BUSINESS INTELLIGENCE, http://ie2.wikispaces.com/Business+Intelligence
[4] - Making Business Intelligence a Part of Your Organization, http://socialmediatoday.com/docmarketing/1711106/making-business-intelligence-part-your-organisation
[5] - Business intelligence software with reporting, analysis, dashboards and scorecarding capabilities, http://www-03.ibm.com/software/products/sv/business-intelligence
[6] - What is Business Intelligence, http://www.selectbs.com/products-general/what-is-business-intelligence
[7] - IBM® Cognos® 8 Business Intelligence, http://public.dhe.ibm.com/software/data/cognos/documentation/docs/en/8.4.0/wig_cr.pdf
[8] - IBM® Cognos® 8 Business Intelligence Getting Started, http://publib.boulder.ibm.com/infocenter/c8bi/v8r4m0/topic/com.ibm.swg.im.cognos.wig_cr.8.4.0.doc/wig_cr.html
[9] - OBIEE VS COGNOS, http://www.passionned.com/business-intelligence/business-intelligence-tools/business-intelligence-tools-comparison/obiee-vs-cognos/
[10] - cognosvsobiee-120628035521-phpapp01, http://www.share-pdf.com/c47886b35f4e47739ab9efa9e82c3e53/cognosvsobiee-120628035521-phpapp01.htm
[11] - Cognos, http://en.wikipedia.org/wiki/Cognos
[12] - IBM Cognos Business Intelligence, http://en.wikipedia.org/wiki/IBM_Cognos_Business_Intelligence

Ruxandra OPREA graduated from the Faculty of Mathematics and Informatics of the Ovidius University, Constanța, in 2012. She is currently graduating from the Databases for Business Support master program organized by the Academy of Economic Studies of Bucharest.

Ionuț LĂCĂTUȘ graduated from the Faculty of Mathematics and Informatics of the Ovidius University, Constanța, in 2012. He is currently graduating from the Databases for Business Support master program organized by the Academy of Economic Studies of Bucharest.

Cătălin HUȚANU graduated from the Faculty of Mathematics and Informatics of the Ovidius University, Constanța, in 2012. He is currently graduating from the Databases for Business Support master program organized by the Academy of Economic Studies of Bucharest.

Business Intelligence overview

Dragoş AGIU, Vlad MATEESCU, Iulia MUNTEAN
University of Economic Studies,Bucharest, Romania

agiu.corneliudragos@yahoo.com, vlad.mateescu30@gmail.com,
muntean.iulia12@gmail.com

The need for information has always been present, but what defines the present time is the big amount of data available everywhere and the need to have all the answers we need in a very short time. Because of the agglomeration of markets and the dynamics of economic environment, the ability to collect data and turn it into useful information for decisional process may be the element which makes the difference. Thus, the term of "business intelligence" was first introduced by Gartner Group in the mid of 90's, but it still exists since the 70's, when the reporting systems were static, two dimensional, with no analytical skills.

Keywords: *Business Intelligence, management, Data Mining, CRM, ERP*

1 The concept of BI

Business Intelligence can be defined theoretically as the use of high class software or business applications or the use of values to make better decisions for the company, as IBM confirms.

Technical and practical, Business Intelligence are tools for collecting, processing and analyzing data. This way, the company can evaluate the results and interpret them .

Based on the newest technologies, BI systems are essential for the decisional level efficiency, but also for improving relations with the clients, employees and suppliers by: facilitating the decisional process, increasing productivity of employees, lower costs, increasing the relationship with the partners and business development.

To understand the importance of the BI trend, it's important to know what are the most relevant benefits after using this type of systems, as follows:

1. BI systems enable effective risk management: the value brought to managers by using BI systems is given by monitoring the risks that may threaten the organization's strategic objectives and financial losses.
2. BI solutions integrate all data for analysis: standard relational data, exports of texts, Microsoft Excel data up to XML data streams, which are stored in data warehouse or operational systems. This way, they are always available for the use in BI applications.
3. The costs for implementing a BI system are low, large investments are not required in hardware equipment and the training for the next users can be done in a short time, all of them being minor investments, which will be recovered in first moths.
4. Decrease the influence of power games in the company
5. Avoiding decisional problems.

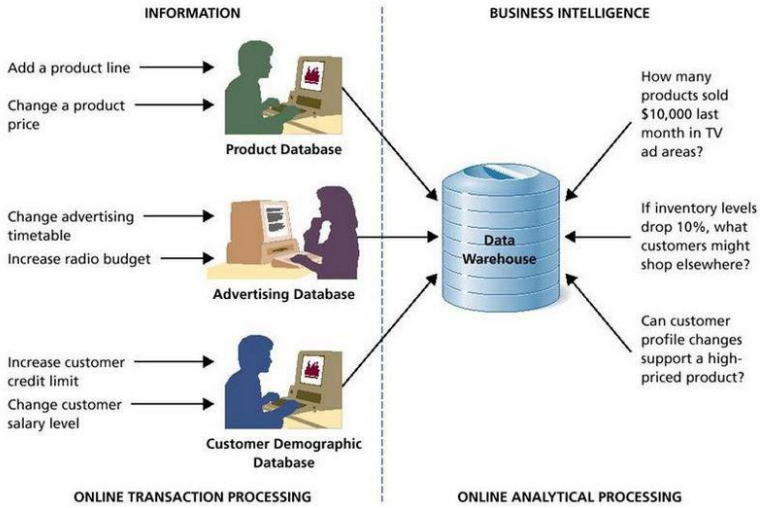


Fig. 1 Business Intelligence Process

2. Choosing a BI solution

It's already known that today's economic activity can generate a lot of data volumes. Each one represents a tiny part of the business and it is found in various locations or departments, or often in different geographical regions.

With BI, as we've already mentioned, all data are collected, processed in information, information that will be well analyzed and used for the next decisions.

Under the current business environment, the quality and timeliness of information is not just a choice between hard profit and loss, but a matter of survival and training.

The benefits of a BI system are obvious – analysts are optimistic, showing that the coming years millions of people will become regular users of the system. Around this concept however is now seeking the most appropriate strategy for choosing a business intelligence system and of course, this strategy will be customized to the needs of each organization.

A BI system without a clear end goal can certainly give a yield, however will never guide the organization where desired, because no one knows the ultimate goal. So, the first step in building a Business Intelligence strategy is to know where the company wants to reach out and which are the effects expected after implementation.

To obtain the necessary information and to present them "user-friendly" is an important step to pass in desire to solve the problem of BI in an organization. For example, data difficult to understand may discourage users from using them.

Here, it should be also noted that "dirty" data in the system must be identified, removed and replaced.

It is essential that the organization should not forget the benchmarks for performance – measuring process will require both the company past and future. Visions must be very clear, so the measure progress must be easy to do.

Once established where the company is in the chosen moment for analysis, what does it want to be done, it's time to decide how to reach the established goal. The standards can be set, BI Competence Centers can be implemented, or BI in cloud or framework can be used, all according to the needs of the company.

3. BI solutions for organization management

Business intelligence applications include decision support systems, query and reporting tools, online analytical processing (OLAP) and data mining and forecasting systems. But ultimately the final results of implementing business intelligence are in depth analysis, refining and concentrating a large number of business intelligence in concrete performance indicators and ultimately, organizational knowledge. The business intelligence projects do not aim to teach managers how to make the right decisions - but they help to understand the figures and the decisions they take in their analysis.

Basically companies collect vast amounts of data through transactional systems (eg ERP, CRM, SCM) that have been implemented over time and which they use to perform daily a variety of corporate functions. The existence of such large volumes of transactional data creates opportunities for management and improves forecast accuracy.

Business Intelligence wants to eliminate assumptions and uncertainties in decision-making processes, both tactical and strategic level. At the tactical level, BI helps optimize business processes or product lines by identifying trends, changes or behaviors which need improving management and control functions. Regarding the strategic level, BI can provide significant value increased by different alignment of business processes and product lines with the strategic objectives of the organization through an integrated performance management framework and systematic analysis. It is essential to consider that Business Intelligence has tended to shift from reporting past events to forecast and prediction.

Defining BI solutions begins with strategic performance measurement requirements of the organization and not necessarily with the technical details related to it. In general, performance evaluation can be defined as key performance indicators (KPI) for an organization or business in the form of questions that require specific answers and are based on facts. Some examples of key performance indicators include: the evolution of income, profits earned by a line of business, and cost projections regarding business questions including management problems with numerical answers: "how much are we buying from a supplier in one year?" and so on.

In defining business intelligence solutions within an organization, the focus should fall on functional analysis, the design solution and the outcome or information. One of the best strategies in the establishment of such a solution is the strategy of "top-down", which starts from the Executive Management that the company and their needs for information and ends with the information technology and integration of multiple data sources to meet the information needs of management.

However, research reveals that more than half of Business Intelligence projects hit a low degree of acceptance or fail. What factors influence the implementation negative or positive?

The most important factor is definitely the project sponsor, and this variable will be bet on the functions of the organization's needs. A Business Intelligence application is divided into three levels: operational reporting, analysis and strategy. If the beneficiary is defining realistic expectations that are based on clear objectives and requirements, they create prerequisites for a successful implementation. In addition, there are limitations related to human resources. Knowledge of business will be provided by the client. These projects involve a team from the implementer, and one from the customer consisting of user’s key functional areas (finance, sales, middle management and top management) which, besides regular duties, should work with consulting solution implemented.

In terms of the defining trends of BI solutions in companies in the current year is important to note the following:

- Companies choose mobility - inflexible switching solutions that create static reports to user-driven BI solutions will accelerate as more organizations recognize the importance of obtaining data that each user be able to base their decisions. Therefore BI solutions adapt the way people work, giving them access to information wherever they are at any given time. For both the large companies and small ones the speed of the business world means that mobile business intelligence solutions are commonly used and not just occasionally. Business users want to be able to access information in the context of the natural flow of their day and not just when they are physical in their offices. Devices such as a PC Tablet for example, are ideal for mobile BI applications and they support and encourage collaborative decision-making process. The data brings out its secrets - is obvious that those organizations who use data obtained from daily activities to make decisions are more successful, while those companies that do not, see their market position threatened. If traditionally the data analysis process is the responsibility of an "analyst" expert, today this process counts more often as a responsibility of a regular user of business skills. The objective of the most successful Business Intelligence solutions today is to offer any user, regardless of their technological skills, the power to harness the valuable information hidden in large amounts of data at its disposal to make critical business decisions.
- Data from social media environment become useful tools for the business – more and more, social media is becoming an essential component for the presence of a company on the market and contributes to a more direct and authentic communication with the public. At the same time, however, beyond the number of likes or followers a company earns on various platforms of social networking, social media becomes a relevant source of information that helps measuring the reputation of the company in online market or it can reveal unexpected insights considering the composition of the audience.
- Moving to cloud – a few years ago, the cloud system was an interesting subject in almost all the segments of the software except for BI. 2014 will be, most likely the year when the cloud will become an important element of BI solutions. The change will be determined by the international market maturation and by the companies which realize that the costs of thir own infrastructure are bigger than a cloud solution. Because the BI solutions are limited, the organizations prefer their people to spend time exploring data, testing new ideas and highlight business insights instead of being stuck with infrastructure activities and upgrading software.

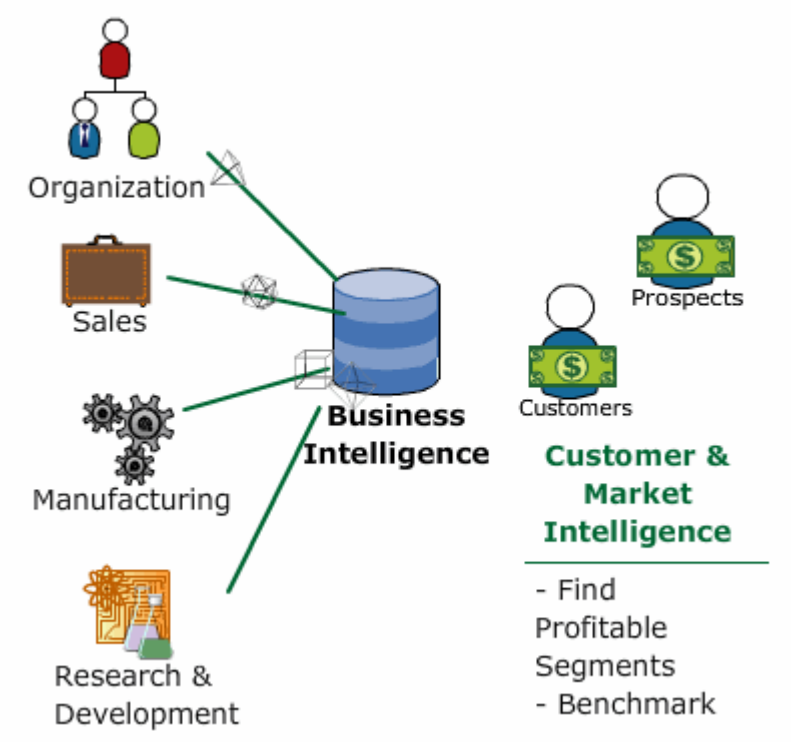


Fig. 2 Customer&Market Intelligence

4. Life cycle of a BI system

Business Inteligence is a strategic initiative which helps the organizations to measure the effectiveness of their plans on the market.
A succesful company must know how to plan and how to address a BI strategy so that the project or projects implicated in the process to have maximum profitability. Company managers with each project manager should adopt a specific methodology based on the needs they know they have.

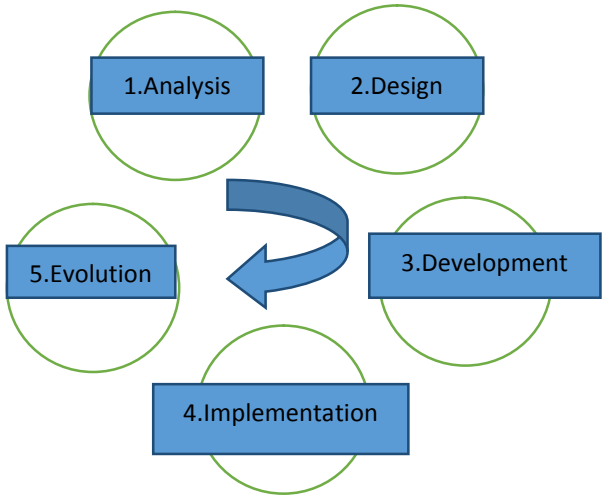


Fig. 3. The five main phases of a BI system

4.1. The analysis stage

The aim is one of the most important things in any project, but sometimes it should be put aside and addressed in a different stage than the one in the beginning. Any project must be compared from a technical and an organizational point of view, but it must coincide with the primary aim of the company. This will be able to give a clearer view over what it is expected to be done during the following period of time.

Any BI project has to clearly justify the cost and the benefits of solving a business problem. The analysis is carried out by a predefined set of key performance indicators (KPI) and it is requested by the end users.The analysis stage produces the design of different components of the solution with the relevant information sources. Because of the dynamic nature of BI projects, the changes in the objectives, people, estimates, technologies, users and sponsors can have a serious impact in the success of the project.

4.2. The design stage

Based on the complexity of the solution, the BI technologies are carefully chosen.One of the usual methods of trying out solutions is building prototypes. This way, it is possible to adjust the requests based on the expectations.

The key performance indicators should be defined without taking account of the current informations, the aim is to capture the business needs, even though the support for these needs is unavailable at the time being. The plan should include a high level design of different components of the solution, including the relevant information sources. Then, the team members of the project, including the key managers and the IT department should all agree in a formal way over the plan and the success criteria.

The projectation stage should include the adequate selection of BI technologies, based on the needs of the users and the complexity of the implementation. The selected instruments should be jointly managed by the Center of Excelence (COE) and the IT department (according to the BI architectural standards) and it should include the active participation of the end users, to make sure that the expectations are the same with the needs.The plan should also decide what information sources are necessary to support the solicited key peformance indicators, including their quality, as well as any other necessary transformations for the analysis.

4.3. The development stage

In this stage, all the information flow from the organization must be modeled. It is critical to create the prototypes and a testing environment to verify and compare the target objects of the company.

The data infrastructure could count as much as 70% in the effort and the costs of this stage. The previous stage, as well as this one, are usually the ones who consume the most time and resources during the development cycle. The workload involved in building the solution afterall the data is in place depends on the complexity of the project. Only a simple configuration could be necessary or a full customization may be needed.

The specification that gives us what kind of data we need for the development and must be stored can be grouped in a metadata model. Further more, the specification needed for the delivery of the metadata to the clients must be analysed.

If a metadata deposit is acquired, this will have to be extended with characteristics solicited by the BI applications. If a metadata deposit is built, the database must be designed based on this deposit. The database design schema must coincide with the access specification of the business.

According to how precise the data and the requests for data transofrmation during the analysis, an ETL instrument may or may not be the best solution. In both cases, preprocessing the data and writing the

extensions for the instruments are frequently necessary.The real challenge for the BI applications comes from the BI that is hidden in the data of the organization, that can only be discovered with data mining instruments. Developping metadata deposits becomes a subproject of the main BI project.

4.4. The implementation stage

After all the BI components were thoroughly tested, the application is implemented at user level. Regardless of the technology used, the success of a project will depend on the training done at the user level and the support of a dedicated team, especially in the first stage of implementation. This phase requires an iterative approach with extra training sessions so that customer needs are met. This step also requires a preliminary development of specific reports and analysis for the business users type. This will create a foundation for future advanced analysis.

All support and guidance operations will be provided by the IT team and the users will serve as consultants during the execution of this stage. Of course, the IT team will have to train the technical support department for a better guidance of the end user.

4.5. The evolution stage

This stage can be called, contrary to the title, the consumption stage because the user uses the information received to change the business and make decisions.

The main goals of this stage are:

- Measuring the success of a project
- Extending the application to the enterprise level
- Increasing the exchange of information between business and functional, both internally and externally.

Business-oriented aspects are often ignored at this stage - since the BI solution is technically capable, the user is free to make any decision. Many projects stalled in this phase because responsibility of passing to the user is not officially sent to a business manager. Therefore a good study is dividing this stage in several sub-phases to show more clearly how this methodology can lead to the end of a life cycle of BI.

Shown below it is presented the schema of the evolution stage subdivisions:

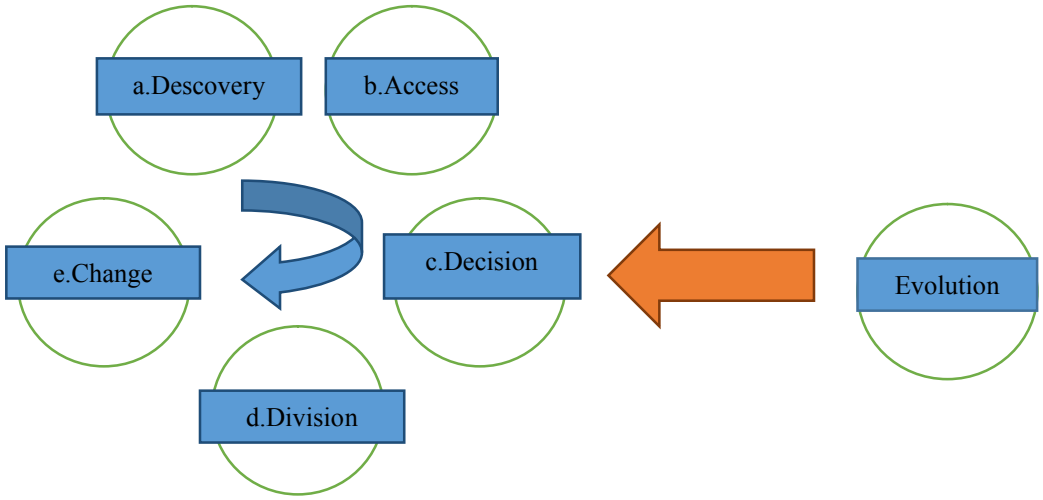


Fig. 6 Business cycle.

a. The discovery stage

Often the organization does not understand how the solution can lend itself to the exact needs of the customers. End users along with the first version of the solution should create the main environment for using an effective solution, but for there has to be a mutual agreement over a basic system where the solution can run at first.

b. The access stage

Having identified the indicators and the valuable information during the discovery stage, the end users begin to follow, understand and manage the information leading to deeper perspectives in the chosen business solution. The users can ask for assistance or contact the organization for further clarification.

c. The decision stage

The end users make definite decisions based on the newly obtained information. The Center of Excellence (COE) may be involved in verifying the solution so that the user will make a better decision.

d. The division stage

The decisions and analyzes made are distributed in the company to better analyze whether the BI solution given to the user was indeed a good one, and if something did not come out as expected organizational changes can occur.

e. The changing stage

Changes made can trigger a fundamental process of reengineering. On this stage all technical resources together with COE shall review the issues and help resolve them.

Any BI cycle that ends should start again from the first stage, but the methodology approach will have a new level of focus:

- The analysis stage
- The reevaluation stage
- The modifying stage
- The optimization stage
- The adaptability stage

This determines that the benefits of the experiences should be added again to a process and maintain the relevant BI cycle. Using a BI methodology helps the organization to understand and produce a sequence of steps to develop and implement a successful Business Intelgence site. Some methodologies may serve as a guide for consuming the resources effectively and/or funding from other successful companies.

5. BI Components: OLAP, Data Mining

The relational model is the most common model in the representation of databases. Although it has its strengths, it does not handle complex queries on large data sets. The concept of Business Intelligence can be divided into three parts, data collection, data analysis and reporting. The main tool for data analysis is the cube, which is a multidimensional data structure built on a data warehouse. The cube is used for data clustering on multiple dimensions and selecting a subdomain of interest. The selected data can be interpreted with data mining tools, a concept used to find trends and patterns in data structures.

In the following sections we will analyze an approach that uses cubes, OLAP (Online Analytical Processing) as well as the data mining concept.

a. OLAP

According to [5], OLAP is an interactive technology that allows the user to make the following operations:

- Fast and dynamic analysis of aggregate data;
- Viewing the information from multiple perspectives and dimensions;
- Analysis of trends during significant time periods;

An OLAP application is designed to allow users to browse, retrieve and present specific data of the business. The tables in the relational model use only one or two dimensions, which does not correspond to

the complex data of a company. In the OLAP concept a cube is a model specifically designed to work with data stored on multiple dimensions. The concept of a dimension used with the OLAP technology refers to a characteristic of the data, not an actual dimension in space.

For a data model to be defined, it is necessary to define the model structure (the objects of the model and the relations between them), the operators (those acting on the structure), and the integrity constraints (rules and constraints imposed to ensure the correctness of the model). The structure of the model consists of dimensions (structures spanning different hierarchical levels that the data are grouped by) and tables of facts (containing measures and foreign keys to dimension tables). A time dimension, for example, may include days of the week, months, or even years.

Generally in Business Intelligence there are used three types of schemes: star schema, snowflake schema and multidimensional cube. In a star schema type (Fig. 5), there is a table of facts, of which the dimension tables are connected to. Dimension tables are not connected among them, otherwise the scheme would be called snowflake. A star schema does not correspond to the third normal form, since the dimension tables are made up of several adjacent tables. But this is preferable, due to the loss of performance when the schema is in third normal form and running operations on large data sets. If it is wanted a variation in the third normal form, a snowflake schema can be used. [6]

Given the fact that the users who use the OLAP technology want quick answers to questions like "How many bikes have sold to customers in Amsterdam in the last four months?" several queries are performed on large databases and united, in order to provide an answer. For this reason, it is preferred to use a multidimensional model, the Cube. A cube is consists of as many dimensions as are needed for the business model. For dimensions with large number of records, hierarchical dimensions are made. They use a field named "parent" which groups all fields of the same type. For example, if a store sells shirts with three colors, red, green and purple, instead of having one field named "Color", there will be a parent-field called “Shirt” and the child-fields will be "Red", "Green" and "Violet".

There are two main models for implementing cubes, MOLAP (Multidimensional OLAP) and ROLAP (Relational OLAP). MOLAP assumes that a cube is multidimensional and has the data stored in a multidimensional way. Thus, the data is copied from the data-store to the data-store of the cube, and aggregates of different size combinations are pre-calculated and stored in the cube at a given vector. This means that the response time will be very short, but there is a possibility that explosion of data could occur when all combinations of size aggregations are stored in the cube. ROLAP, as the name suggests, uses the relational model and considers that it is for the best to keep the data stored in the data warehouse. As in the MOLAP, the data can be aggregated and pre-calculated using materialized views. [6]

b. Data-Mining

Data Mining is the process of finding hidden patterns and associations, constructing analytical models, achieving classification and prediction and presenting the obtained results. In other words, the process of data mining can be viewed as examining past data to find useful information that can be used as a guide for the future. Thus, data mining gets the answers to questions like "What happened in the past?", "Why it happened?" and "What is likely to happen again in the future?”.

To be effective, the process of data mining needs "clean" and organized data, therefore preferring them to be stored in data warehouses, because they have

already been cleaned, processed and organized by ETL processes.

According to [7], data mining process is divided into the following steps:

1. Defining the problem - as in any Business Intelligence process, first up, is the fully understanding of the problem and the questions need answers. It is recommended to focus on what we want to ask, not how we will respond.
2. Preparing data - once the necessary information is well defined, the data is being prepared. Data mining algorithms are very intelligent in solving the problem, but will fail if they receive the data in a foreign format. As mentioned, it is recommended that the data should be retrieved from a data warehouse.
3. Exploring Data - to understand the result of the data mining process, the data must first be covered and studied. This is usually done using one of the two methods :
    a) Means and extremes: on different sets of media numbers, the means are calculated and the standard deviation from the mean too. Also, the minimum and maximum values are determined. This process helps to understand the data range used for working.
    b) Basic statistics: different data sets are taken and simple statistical operations are performed.
4. Creating the data mining model - queries are created using DMX language (Data Mining Extensions) or using a software wizard.
5. Exploration and validation of the model – once the model is build, it should be explored to see the trends and patterns that the model produces and validate it, to ensure that trends and patterns found, answer the questions defined in the first stage .
6. Launching and updating the model – in the last step, the model is launched in execution so that the users can use it. While in use, it is possible that a new set of questions which need to be answered could occur and the model must be modified.

The functions that a system based on data mining process performs, are:

- Classification - maps data into predefined groups and classes
- Clustering - groups similar data into clusters
- Analysis of linkages - discovers relationships linking certain data
- Finds data that misbehaves from the general behavior of the data.

A natural question is "why data mining?". Data mining can be successfully used in data analysis processes and support of decision making. For example, in a system of marketing and management analysis, the first elements that are studied, are the modes of the origin of data (transactions, payments, gift vouchers and so on). The second step is to find clusters of customers who have common features (ways of spending the money, income, etc.). Further, operations marketing analysis can be performed, establishing associations and correlations between product sales so that predictions based on these assumptions can be made. Also, by clustering and classification, the types of customers that buy certain products can be determined and may make assumptions about what factors will attract new customers. [8]

A data mining system is generally the architecture shown in Figure 7.

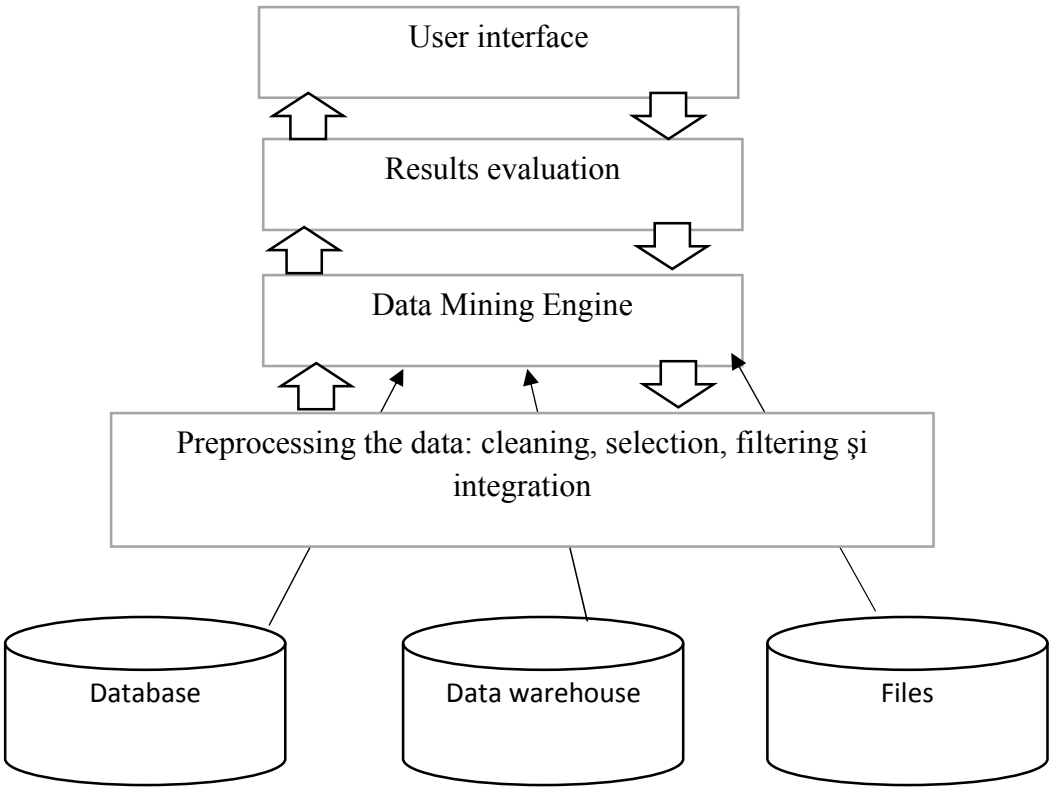


Fig 7 The architecture of a data mining system

6. Building a Successful BI

There is no certain recipe or template for success on how to build a business, as well as building a model of Business Intelligence does not have a classified number of standard steps. This depends very much on the company, market, people involved, and defining the problem. According to [8], there is still a lot of ideas that can help a company to go right into creating a successful BI.

1. Choosing good key performance indicators (KPI - Key Performance Indicators) - start from the question "How do you measure a company's success?". KPIs are metrics and measurements that can indicate whether a company is successful or not. It is known that if we started with the right KPIs selected, the business will benefit in the long term.
2. Adjusting recipe - there are plenty of tips, or even complete BI solution, ready to be used. But, so far, there hasn't been invented a size that can fit for everyone. Each company is complex and has its own terms and objectives.
3. Understanding complexity - although it may seem easy on paper, a Business Intelligence solution is complex and this must be understood from the beginning, in order to avoid unpleasant situations that may arise later.
4. Creativity – sometimes it is needed to look at things from another perspective. This is especially important in BI, where a book cannot ensure success. If a team can come up with new ideas that do not comply with a pattern defined by someone else, this should be encouraged.
5. Choosing a good team - as a sequel to those mentioned in the previous paragraph, choosing a team that has

a good grasp of knowledge about the data problem, is a step towards to finding the best solution for the company.

6. Study technologies - Business Intelligence is a rapidly evolving field, there are numerous books, studies, or blogs about this area. It is encouraged to keep up with new appearances.
7. Learning from mistakes - sometimes there are pitfalls and mistakes that cannot be avoided. The key is to remember what happened, how it happened, why and how the problem can be avoided in the future. This applies to any field, but is especially
8. important in Business Intelligence, is avoiding to repeat the error.
9. Understanding the company - every company is different and before starting the development of Business Intelligence is necessary to know in detail how the company works. How the company is split, how decisions are made, how the employees work, are things that must be included in the planning process.
10. Creation of well-defined phases - opportunities for Business Intelligence solutions are endless. Executive management has less patience for huge projects and prefers to see small results over time, rather than at the end. Therefore, the vision of business intelligence should be divided into several phases. Each phase must contain a well-defined purpose, achievable goals and benefits that can be demonstrated to all.
11. Support from the executive - as Business Intelligence solutions touch several departments of the organization, this may be seen with skepticism. An influential person in the company who understands the importance of the solution can convince those skeptical persons of the importance of implementation.

7. Conclusions

Organizations today can store and access a big piece of information gathered over the last decade. Business Intelligence is the ability to bring to account the true value of this information and this is becoming increasingly important. Unfortunately, not all companies today have developed this part of the organizational strategy.

Following the steps mentioned, having a center of excellence, plus a BI strategy based on certain standards and methodologies chosen correctly, companies can reach the top. They can recover their investment through different suppliers, they can reduce costs and they can especially learn how to use the company's capital.

As possibilities in business intelligence get to be a top management priority, companies will need more flexible IT solutions that meet the needs of the BI, not only in short term, but also with the company's growth they need more complex solutions.

There are three main factors to be taken into account when selecting a business intelligence solution:

- Involvement of business intelligence across the enterprise - BI solutions that make work easier for all employees must be found, not just for a few, in order to work, partake, understand and interpret the data and information in the organization.
- Integration with other systems and applications - BI solutions must be sought to integrate well with the rest of the enterprise business systems in order to be used in the future.
- Flexibility of business intelligence - when requirements change or when new functionality is added, BI solutions that mold on such changes must be researched.

You don’t have to make the main purpose of a company to implement Business Intelligence, but this should be the way to

guide the company towards an ideal of success.

REFERENCES

[1] http://www.gartner.com/technology/about.jsp
[2] http://www.webopedia.com/TERM/B/Business_Intelligence.html
[3] http://incomemagazine.ro/articole/7-tendin-e-definitorii-pentru-business-intelligence-in-acest-an#
[4] http://www.cio.com/article/40296/Business_Intelligence_Definition_and_Solutions
[5] Introducing Students to Business Intelligence:Acceptance and Perceptions of OLAP Software, Mike Hart, Farhan Esat, Michael Rocha, and Zaid Khatieb, Department of Information Systems University of Cape Town, South Africa
[6] Business Intelligence: Multidimensional Data Analysis, Per Westerlund
[7] Microsoft's Business Intelligence, Ken Withee
[8] Business Intelligence for Dummies, Swain Scheps
[9] Data Mining & Business Intelligence: Practical Tools, Motaz K. Saad

Dragoş AGIU - As a person I developed abilities for studying mathematics and informatics, so as for my education I studied the "CA Rosetti" High School - mathematics-informatics profile, 2005-2009. After that I went at the Bucharest Academy of Economic Studies, Department of Cybernetics, Statistics and Economic Informatics, Economic Informatics specialization, 2009-2012, Bachelor Degree. To complete my studies in this range of knowledge I attended also the Academy of Economic Studies, Department of Cybernetics, Statistics and Economic Informatics the Database Master, 2012-2014 in Bucharest. For 2 years I've been working in a IT company and my skills in computer science are growing every day.I intend to reach a high level in a IT company or anywhere I can get the chance. I want to prove to myself that I can give all that is needed when it comes to involvement in a company. For my hobbies I practice martial arts, swimming and motorcycle riding.

Vlad MATEESCU - My passion for informatics has started in high school and it continued after graduation, as I've attended The Faculty of Automatic Control and Computers within the Politehnica University of Bucharest (between 2008 and 2012) and the Business Support Databases master program within the University of Economic Studies (between 2012 and 2014). I am currently working at an IT company as a software developer, cultivating new skills every day, both professionally and personally. Although my job consists in creating software applications, I am interested in all the stages of a project and would like to work as a software analyst one day. In my free time, I enjoy travelling, photography and live music.

Iulia MUNTEAN - The passion I have for IT has started 2 years ago, during a project I made for school. That's the moment when I decided I would love to work in this domain. Since then, I started to learn everything about programming. I am currently working as a web programmer and I love this job. Every day is a new day to apply what I learned theoretical in the past and see it's working.In my spare time, I like reading historical books, travel and don't miss any chance to learn something new.

Comparative study on software development methodologies

Mihai Liviu DESPA
mihai.despa@yahoo.com

Bucharest University of Economic Studies

This paper focuses on the current state of knowledge in the field of software development methodologies. It aims to set the stage for the formalization of a software development methodology dedicated to innovation orientated IT projects. The paper starts by depicting specific characteristics in software development project management. Managing software development projects involves techniques and skills that are proprietary to the IT industry. Also the software development project manager handles challenges and risks that are predominantly encountered in business and research areas that involve state of the art technology. Conventional software development stages are defined and briefly described. Development stages are the building blocks of any software development methodology so it is important to properly research this aspect. Current software development methodologies are presented. Development stages are defined for every showcased methodology. For each methodology a graphic representation is illustrated in order to better individualize its structure. Software development methodologies are compared by highlighting strengths and weaknesses from the stakeholder's point of view. Conclusions are formulated and a research direction aimed at formalizing a software development methodology dedicated to innovation orientated IT projects is enunciated.
Keywords: *software development, project management, development methodology*

1 Introduction

The management process of a software development projects follows the basic rules of project management but also includes particular features. A software development project manager has to deal with challenges and setbacks that are proprietary to the IT industry. Also in software development there are benefits and strong points that help ease the burden of management.

Software development projects are notorious for frequently changing initial planning and specifications. In order to identify the main reasons for changing specification during the development stage of a software product debates were started on LinkedIn project management groups. Debates were initiated on 11 LinkedIn groups starting the discussion with the following introduction: *Software development projects are notorious for frequently changing initial planning and specifications. What do you believe are the main reasons for changing specifications during the development stage?* Project managers responded on 3 out of the 11 groups. The 3 groups are: PMO – project management office, Agile Project Management Group and International Society for Professional Innovation Management. Based on the information collected from the LinkedIn discussions and on the author's own experience as software development project manager specifications change due to the fact that:

- the project owner identifies new business opportunities and decides to integrate them into the software being developed;
- due to the technical nature of software development projects there is a lack of shared understanding of

expected outcomes;
- original planning was based on specifications that were misinterpreted by the project manager or poorly illustrated by the project owner;
- project team is unable to implement planned functionalities due to lack of expertise or technological limitations;
- the context in which the software is going to be used changes thus generating the need for the software to change;
- new technology or software product is launched on the market.

Changing specifications has a negative impact on the project management process as it reduces predictability and exercises pressure on the budget and deadlines.

The software development field is characterized by high dynamics of technology and standards. Programming languages evolve, new frameworks arise and fall with astonishing speed, user interfaces become more and more diverse as software is required to work on a larger array of devices. PHP server-side scripting language registered 16 releases on new and improved versions in 2014 alone [1]. The software development community widely accepted Phalcon and Laravel, as two of the most powerful PHP frameworks. Both were released in 2012. The project manager has to keep up with the latest trends in software development in order to meet the project owner's requirements and in order coordinate effectively the project team.

Software development projects involve project teams that are made up of highly skilled and highly trained individuals with predominantly technical backgrounds. Highly skilled and highly trained individuals will require significant compensation for their work and that will translate into high man-hour or man-day rates. So the project manager is under considerable pressure to provide accurate time estimates in terms of required man-days as every inconsistency will generate significant additional costs. Also highly skilled individuals don't integrate well in a team as they have a tendency of being arrogant and self-absorbed. The project manager should be able to exploit their ego in the best interest of the project and mitigate disputes and opinion clashes that occur between team members.

Software development projects are often implemented by teams that have members distributed all over the globe. Building software does not require formal face-to-face communication. Task assignment and task tracking is done by using online management tools like Pivotal Tracker, Basecamp or Producteev. Code version control is ensured by using versioning tools like SVN or GIT. File sharing is accomplished by using tools like Dropbox, Google Drive or Box. Online meetings can take place using Skype video conferences.

Table 1. Software development projects characteristics

Characteristic	Positive Impact	Negative Impact
frequently changing specifications	-	jeopardize deadlines
		results in exceeding the project budget
		causes stress and discontent for the development team
high dynamics of technology and standards	generates new opportunities in terms of design and codding	software can become obsolete by the time it hits the market
		software developers have to invest a lot of time in

		researching new technologies
skilled workforce	increases the likelihood of achieving innovative results	high cost generated by human resources
globally distributed teams	work can be performed around the clock	monitoring and control becomes more difficult
	cultural diversity nurtures creativity	integrating new code is more challenging

Table 1 summarizes the impact that software development characteristics have on the project management and implicitly on the project team. The only characteristic that does not have a positive impact is *frequently changing of specifications*.

2. Software development stages

Building a software product is a process consisting of several distinct stages. Each stage has its own deliverables and is bound by a specific time frame. Depending on the project, certain stages gain additional weight in the overall effort to implement the software product.

Research is the stage where the project owner, the project manager and the project team gather and exchange information. The project owner is responsible for formulating requirements and passing them on to the project manager. In order to properly formulate requirements the project owner has to first define a set of goals. Then he has to envision the way a software product will help him achieve those goals. In the research stage the project owner will try to find people or companies with similar goals and document the way those people or companies acted upon fulfilling their goals. The project manager is responsible for receiving the requirements from the project owner, evaluating them and passing them to the project team as technical specification. The project manager has to be able to evaluate the requirements from both a business perspective and a technical perspective. The project manager has to research market characteristics and user behaviour patterns. The project team is responsible for evaluating the requirements from a technical perspective. The project team will have to research the frameworks, API's, libraries, versioning tools and hosting infrastructure that will be required in order to build the software product.

Planning is the stage where all the elements are set in order to develop the software product. Planning starts with defining the overall flow of the application. Next step is to breakdown the flow into smaller, easier to manage subassemblies. For each subassembly a comprehensive set of functionalities has to be defined. Based on the required functionality a database structure is designed. Taking into account the overall flow of the application, the subassemblies, functionalities and database structure, the project manager together with the project team have to choose the technology that will be employed to develop the application. Also the project manager should decide on the best suited management methodology and the proper work protocol for the project at hand.

Design is the stage where the layout of the application is created. Web applications and mobile applications tend to grant more impotence to layout than desktop applications. Depending on the nature of the application designs can range from rough and functionality driven to complex and artistic. An accounting application will only require basic graphic design but an online museum will require high end design work. In an accounting application design has to emphasize and enhance functionality whereas in an online museum

functionality has to be tailored in order to fit the design. The graphic design can overlap with the planning and with the programming stage. The graphic design stage is important because it will display to the project owner a preview of the application before it is actually built. At this stage usually the project owner comes up with new requirements that have to be summited to research and planning.

Development is the stage where code is written and the software application is actually built. The development stage starts with setting up the development environment and the testing environment. The development environment and the test environment should be synchronized using always the same protocol. Code is written on the development environment and uploaded on the test environment using the synchronization protocol. Another important aspect of the development stage is progress monitoring. The project manager has to determine actual progress and evaluate it against the initial planning. The project manager should constantly update the project owner on the overall progress. When writing code the software developers should also perform debugging operation in order to upload clean and bug free updates on the testing environment. Software developers should also comment their code so that they can easily decipher later or make it easy to understand for other developers.

Testing is the stage where programming and design errors are identified and fixed. Programming errors are scenarios were the application crashes or behave in a way it was not supposed to according to the designed architecture. Programing errors also consist in security or usability issues. If the application is vulnerable to attacks and can therefore allow attackers access to private data, then that is regarded as a programming error. If users have problem with slow response time from the application than that also is a programming issue. Design errors are actually inconsistency between what the project owner requested and what the project team ended up implementing. Design errors occur in the planning stage, have a significant impact on the project and are usually harder to fix. Identifying design errors is considerably more efficient when the project owner is involved as he is the one that formulated the application requirements.

Setup is the stage where the application is installed on the live environment. The setup stage precedes the actual exploitation of the software product. The setup entails configuring the live environment in terms of security, hardware and software resources. Back-up procedures are defined and tested. The actual setup of the software product includes copying the source code, importing the database, installing third party applications if required, installing cron-jobs if required and configuring API's if required. Once the application is installed it will go through another full testing cycle. When testing is completed content is added to the application.

Maintenance is the stage that covers software development subsequent to the application setup and also the stage responsible for ensuring that the application is running within the planned parameters. Ensuring that the application is running properly is done by monitoring the firewall, mail, HTTP, FTP, MySQL and SSH error logs. Also monitoring traffic data will provide valuable input on potential issues that may affect the application's performance. An important part of the maintenance stage consists of systematically testing functionalities for errors that were not identified in the testing stage or for issues that are not displayed in the error logs. The maintenance stage also provides the opportunity to add new features of functionality to the software application. Adding new code or changing the old code will have to be submitted to

research, planning, programing, testing and setup.

The above mentioned stages are generally agreed by the software development community as being the cornerstones of every software development project. Depending of software development methodology they may be found under different naming conventions, they may be overlapping changing order or missing altogether.

3. Current software development methodologies

A software development methodology is a set of rules and guidelines that are used in the process of researching, planning, designing, developing, testing, setup and maintaining a software product. The methodology also includes core values that are upheld by the project team and tools used in the planning, development and implementation process. This paper reviews 20 of the most popular software development methodologies and highlights their core characteristics. The analysis includes specifying the scale of the project the methodology is suited for, the stage project owner feedback is delivered and a graphic representation of the methodology. For coherence reasons stages defined in the second section of the article are also used in depicting the graphic representations of the methodologies.

Waterfall is the first methodology generally acknowledged as being dedicated to software development. Its principals are for the first time described by Winston W. Royce even though the actual term waterfall is not used in the article [24]. It emphasizes meticulous planning and it outputs comprehensive documentation. The Waterfall methodology is linear-sequential process where every stage starts only after the previous has been completed. Each stage has its own deliverables. The Waterfall methodology is predictable and values rigorous software planning and architecture.

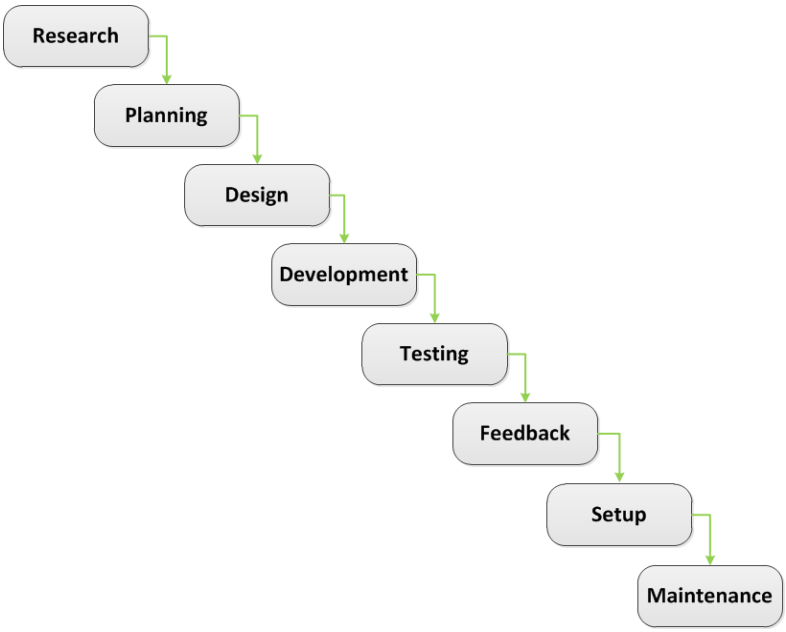


Fig. 1. Waterfall methodology

The project owner's feedback is received after the software application is completely developed and tested. The Waterfall methodology is suitable for small scale software development projects where requirements are clear and detailed planning can be easily drafted for the entire project.

Prototyping is a methodology that evolved out of the need to better define specifications and it entails building a demo version of the software product that includes the critical functionality**.** Initial specifications are defined only to provide sufficient information to build a prototype. The prototype is used to refine specifications as it acts as baseline for communication between project team and project owner. The prototype is not meant to be further developed into the actual software product. Prototypes should be built fast and most of the times they disregard programming best practices [2].

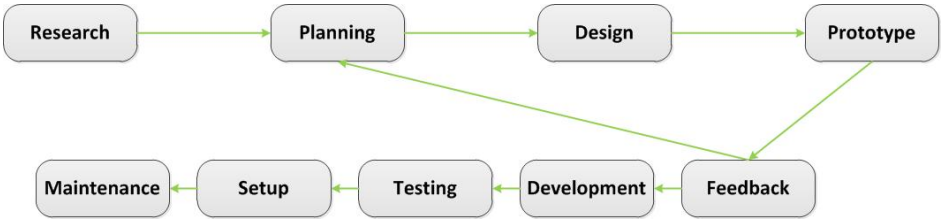


Fig. 2. Prototyping methodology

The project owner's feedback is received after the prototype is completed. The Prototyping methodology is suitable for large scale projects where is almost impossible to properly define exhaustive requirements before any actual codding is performed. Prototyping methodology is also suitable for unique or innovative projects where no previous examples exist.

Iterative and incremental is a methodology that relies on building the software application one step at the time in the form of an expanding model [3]. Based on initial specification a basic model of the application is built. Unlike the prototype, the model is not going to be discarded, but is instead meant to be extended. After the model is tested and feedback is received from the project owner specifications are adjusted and the model is extended. The process is repeated until the model becomes a fully functional application that meets all the project owner's requirements.

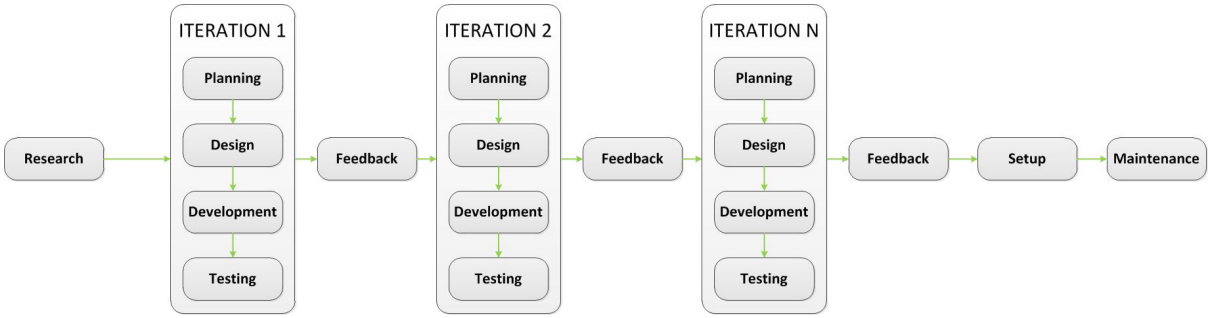


Fig. 3. Iterative and incremental methodology

The project owner's feedback is received after each iteration is completed. The Iterative and incremental methodology emphasizes design over documentation and is suitable for medium and large projects.

Spiral is a methodology that focuses on identifying objectives and analysing viable alternatives in the context well documented project constrains [4]. The Spiral methodology has 4 major phase: planning, risk analysis, development and evaluation. Project will fallow each phase multiple times in the above mentioned order until the software application is ready to be setup on the live environment. The Spiral methodology emphasizes risk analysis and always evaluates multiple alternatives before proceeding to implementing one.

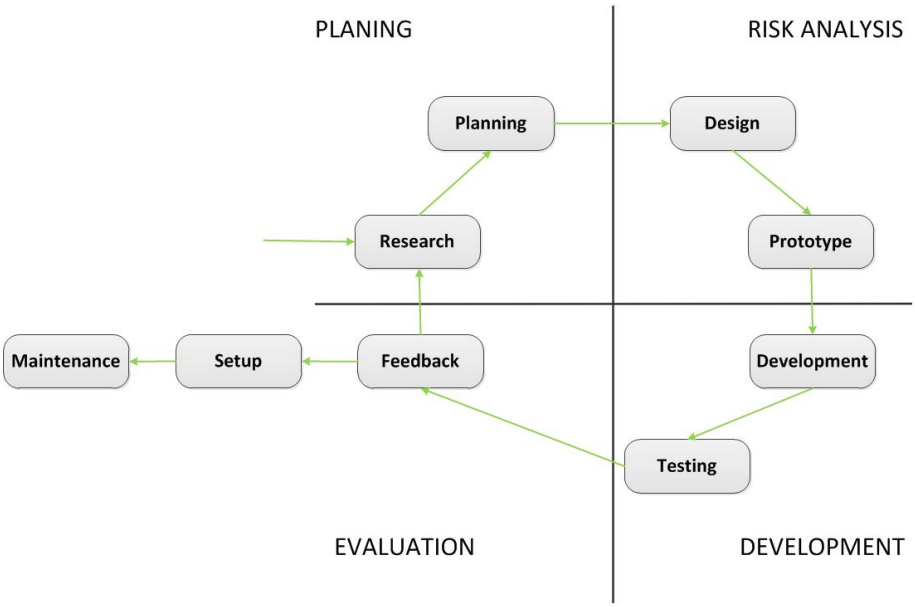


Fig. 4. Spiral methodology

The project owner's feedback is received after the first iteration of the spiral is completed. The Spiral methodology is suitable for medium and large scale projects. It has also proven more effective in implementing internal projects as identifying risks proprietary to your own organization is easier.

Rapid application development is a development lifecycle designed to give much faster development and higher-quality results than those achieved with the traditional methodologies. It is designed to take the maximum advantage of powerful development software [15]. Rapid application development imposes less emphasis on planning tasks and more emphasis on development. Development cycles are time boxed and multiple cycles can be developed at the same time.

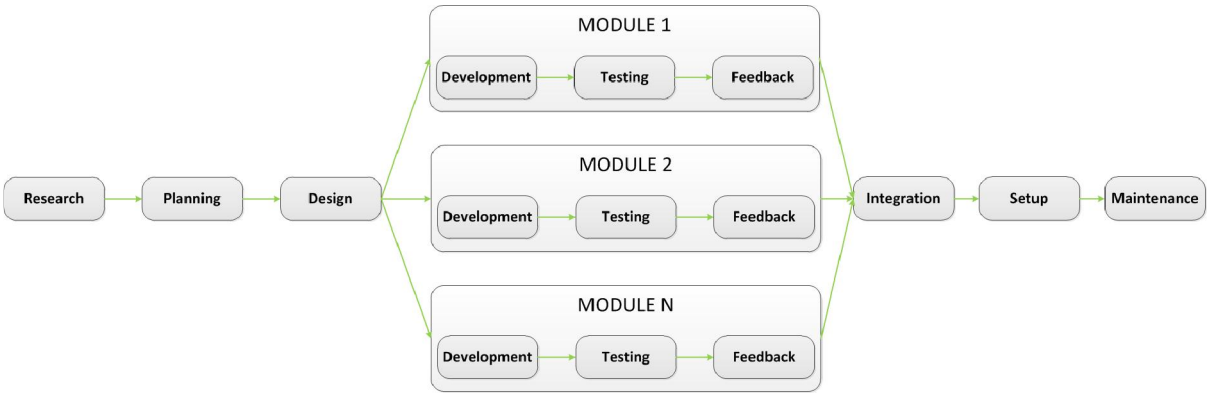


Fig. 5. Rapid application development methodology

The project owner's feedback is received after each module is completed. The Rapid application development methodology is suitable for small, medium and large scale projects with the constraint that projects have to be broken down into modules.

Extreme programming breaks the conventional software development

process into smaller more manageable chunks. Rather than planning, analysing, and designing for the entire project at once, extreme programming exploits the reduction in the cost of changing software to do all of these activities a little at a time, throughout the entire software development process [5]. It enforces pair programming where two developers use the same computer. One is writing code and the other is supervising. They change roles at regular intervals. For reducing the number of errors it relies heavily on unit testing and developers are required to write tests before writing the actual code. There is a collective ownership code policy where any developer can change any code sequence even if it was not written by him. The project owner is the one that decides the priority of the tasks.

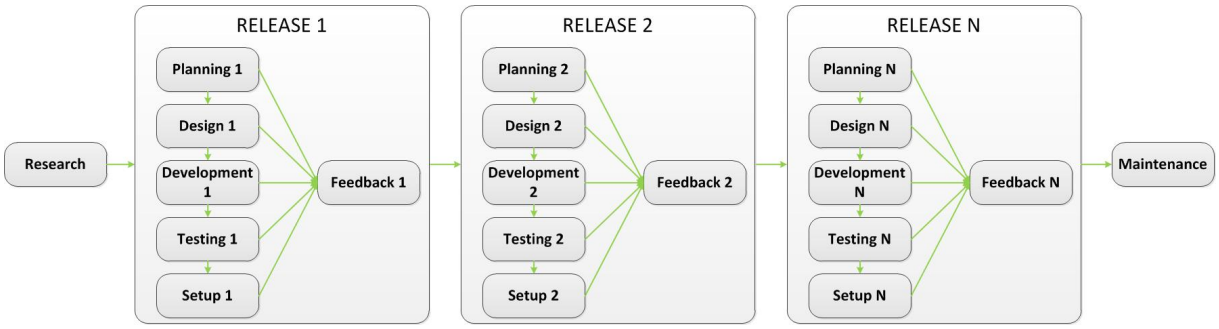


Fig. 6. Extreme programming methodology

Extreme programming methodology requires that a represented of the project owner is always with the development team in order to have access to continuous and relevant feedback. The Extreme programming methodology is suitable for small, medium and large scale projects.

V-Model methodology is a software development process which is an extension of the waterfall model [16]. It emphasizes thorough testing by pairing each software development stage with a matching phase of testing.

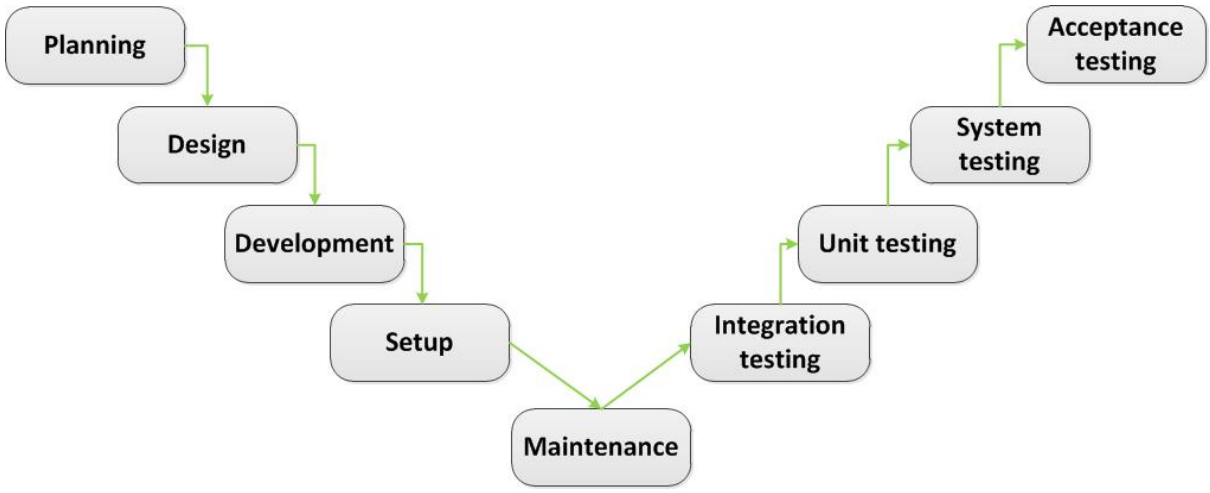


Fig. 7. V-Model methodology

The project owner's feedback is received in the form of acceptance testing after the entire application is completed. The V-Model development methodology is suitable for small and medium scale projects.

Scrum is a methodology for incrementally building software in complex

environments [6]. Software requirements are formulated and prioritized by the product owner and are called stories. All the stories make up the Product Backlog. The Scrum methodology adopted a time box approach where development cycles known as Sprints take no more than 4 weeks and end with a working version of the application. All the stories of a sprint make up the Sprint's Backlog. Progress is assessed in daily meetings that are confined to 15 minutes and are known as Daily Scrum. Task assignment is not done by the project manager or by any other individual. Scrum development teams are self-organized and task assignment is a process where every team member is involved. The team efforts are kept on track by a Scrum Master.

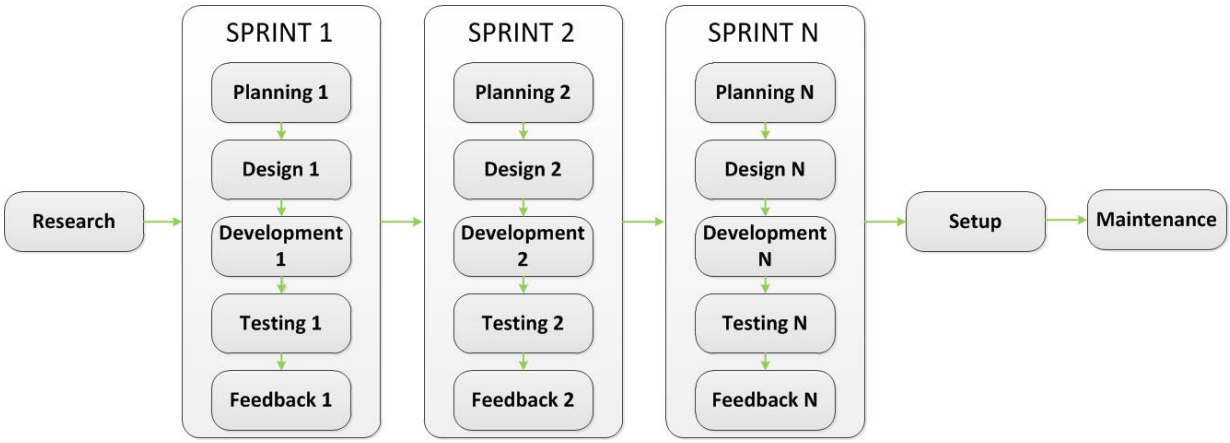


Fig. 8. Scrum methodology

The project owner's feedback is received at the end of each sprint. The Scrum methodology is suitable for small, medium and large scale projects.

Cleanroom is a methodology that focuses on defect prevention. The motivation for this approach is the fact that defect prevention is much less expensive than defect removal. The goal of the Cleanroom methodology is to construct software with no defects during development [7]. Cleanroom methodology relies on a box structure method to design the software product. Quality control is performed by using mathematic models and it also introduces a statistical approach to testing. Developers do not test the code that is the testing team's responsibility.

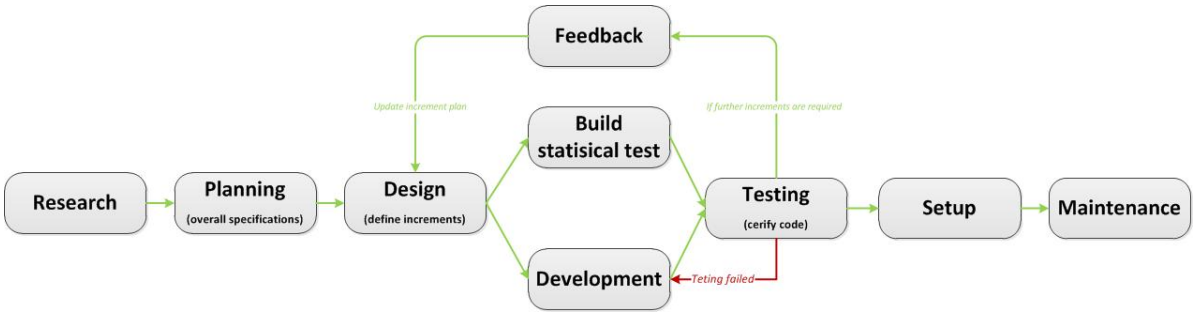


Fig. 9. Cleanroom methodology

The project owner's feedback is received at the end of each increment. The Cleanroom methodology is suitable for small, medium and large scale projects.

Dynamic systems development methodology is focused on developing application systems that truly serve the

needs of the business [17]. Dynamic systems development methodology is an iterative development model that uses a timebox approach and MoSCoW task prioritization. It defines strict quality standards at the beginning of the project and it sets non-negotiable deadlines. Testing is done early and continually throughout the entire development cycle.

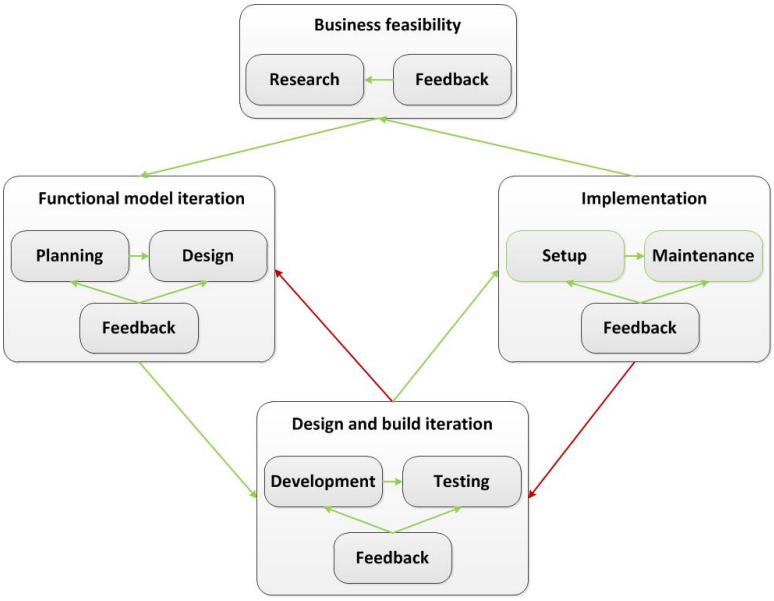


Fig. 10. Dynamic systems development methodology

Project team and project owner share a workplace, physical or virtual, in order to facilitate efficient feedback at every stage of the project. The Dynamic systems development methodology is suitable for medium and large scale projects.

Rational unified process methodology provides a disciplined approach to software development. It comes with several out-of-the-box roadmaps for different types of software projects and it provides guidance for all aspects of a software project. It does not require the project team to engage in any specific activity or produce any specific artefact. It provides guidelines that help the project manager tailor the process if none of the out-of-the-box roadmaps suits the project or organization [8].

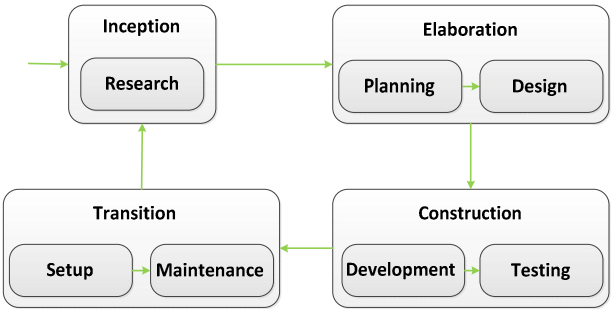


Fig. 11. Rational unified process methodology

The project owner's feedback is received as agreed at the start of the project as the methodology does not enforce a rule on project team – project owner collaboration. The Rational unified process methodology is suitable for small, medium and large scale projects.

Lean software development is a product development paradigm with an end-to-end

focus on creating value for the project owner and for the end user, eliminating waste, optimizing value streams, empowering people and continuously improving [9]. Value is defined as something that the project owner would pay for. Anything that is not adding value is considered waste and has to be discarded. Lean software development delivers early iterations of working code. It also insures that team members are motivated by empowering them to make significant decisions regarding the application. The Lean software development methodology does not enforce a certain process in terms of conducting the project. Project manager and team members are free to use any process they see fit as long as it stays true to core lean development principals.

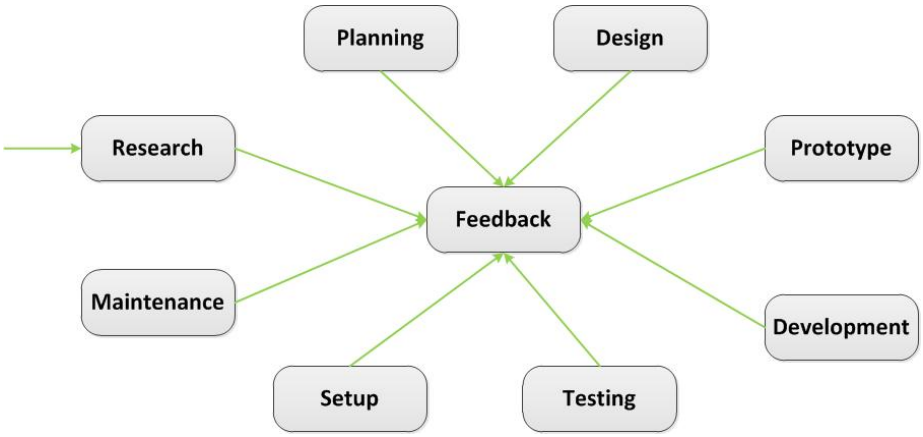


Fig. 12. Lean software development methodology

The project owner's feedback is central to the Lean software development methodology. The Lean software development methodology is suitable for small, medium and large scale projects.

Test-driven development is a methodology developed around unit testing. Before writing any actual code the developers write automated test cases for new functionality. If the tests work then there is no need to write any code as the functionality already exists, it was just not know to the developer. This scenario is often encountered when dealing with legacy code. If tests do not compile then the developer will write the code and run the tests again. The process is repeated until all requirements are met [10].

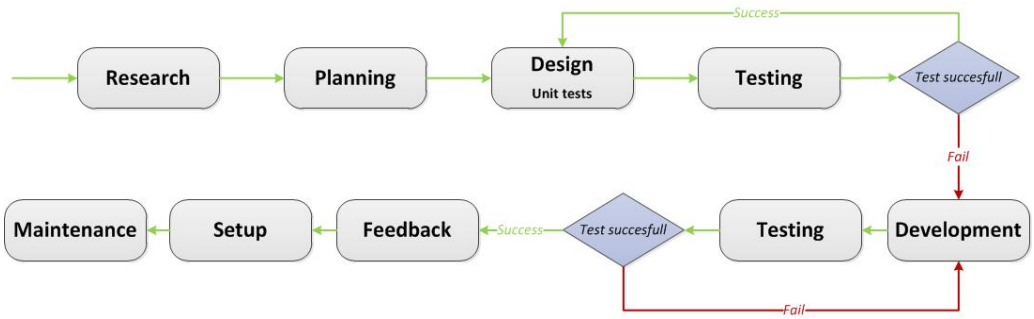


Fig. 13. Test-driven development methodology

The project owner's feedback is received after the code tested successfully. The Test-driven development methodology is suitable for medium and large scale projects. It is also recommended for scenarios where developers have to work with legacy code.

Behaviour-driven development methodology is developed around acceptance testing. The project owner writes the requirements in the form of acceptance tests using a standard format. Requirements are defined as user stories and include a title, a narrative part and acceptance criteria [18]. Based on the acceptance test scenarios developers will implement functionality. When functionality in developed it is tested using the same acceptance testing scenarios. If it passes the tests code is moved on the live environment. The entire process is repeated until all requirements are met.

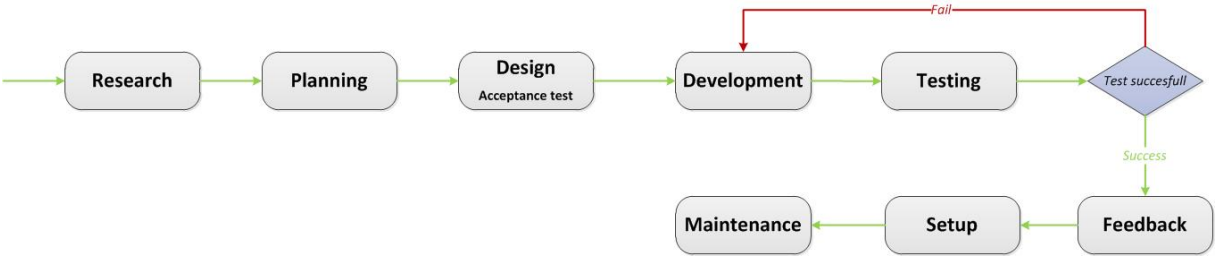


Fig. 14. Behaviour-driven development methodology

The project owner's feedback is received after the code tested successfully. The Behaviour-driven development methodology is suitable for small, medium and large scale projects. It is also recommended for scenarios where developers have to work with legacy code.

Feature-driven development is a methodology focused on actual functionality. Each feature in the Feature-driven development methodology reads as a requirement which is understandable by the project owner, it has true business meaning and describes true business value [11].

Fig. 15. Feature-driven development methodology

The project owner's feedback is received after the application is already setup but constant interaction between development team and project owner takes place throughout the entire length of the project. The Feature-driven development methodology is suitable for small, medium and large scale projects.

Model-driven engineering is a complex methodology that uses domain models as ways of handling requirements [12]. Based on project owner's requirements a metamodel is define. The metamodel is actually a platform independent model that can be migrated onto any environment. UML is usually used for building the metamodel. The metamodel is then converted into a model that is specific to a certain platform. Based on the model the actual code is then generated.

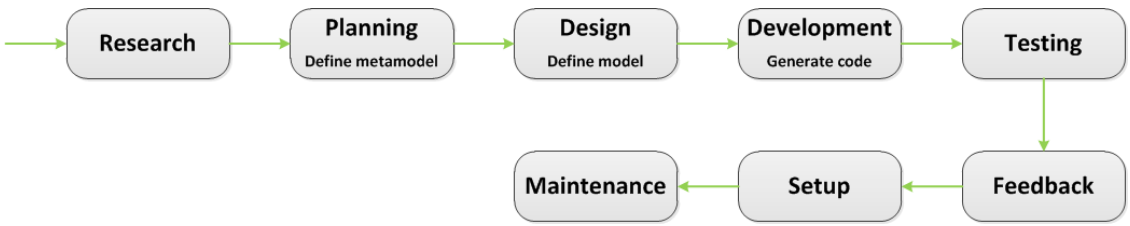


Fig. 16. Model-driven development methodology

The project owner's feedback is received after the code tested successfully. The

Model-driven engineering methodology is suitable for small, medium and large scale projects. It is also recommended for projects that will have long exploitation period as metamodels can be easily migrated and adapted to new technologies.

Crystal Methods is a family of methodologies that developed around the theory that people, and not tools or process, are the most important factor in any software project. Crystal Methods is a myriad of methodology elements and does not tackle every project in the same manner but instead uses custom tailored processes and tools depending on the project's profile and scale. Large or safety critical projects require more methodology elements than small non-critical projects. With Crystal Methods, organizations only develop and use as much methodology as their business needs demand [19]. Crystal uses an iterative approach but does not enforce a release with every iteration.

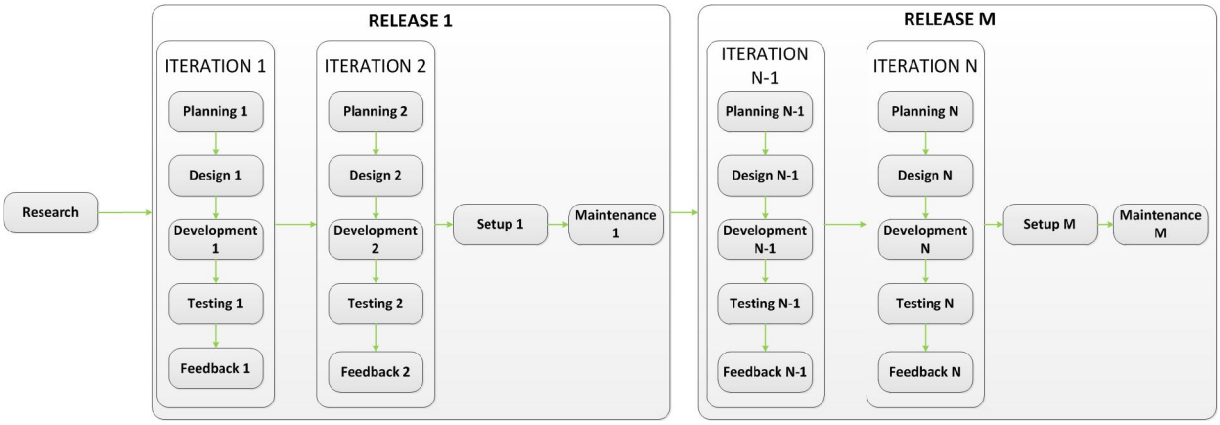


Fig. 17. Crystal Methods methodology

The project owner's feedback is received after each iteration is finished. The Crystal Methods methodology is suitable for small, medium and large scale projects. It has a different approach depending on the scale of the project.

Joint application development is a methodology that focuses on system requirement determination by involving end users, project owner and project team in a series of freely interacting meetings [13]. End user and project owner are also heavily involved in the design and development stages. Joint Application Development methodology uses prototyping as the basis for the actual software development.

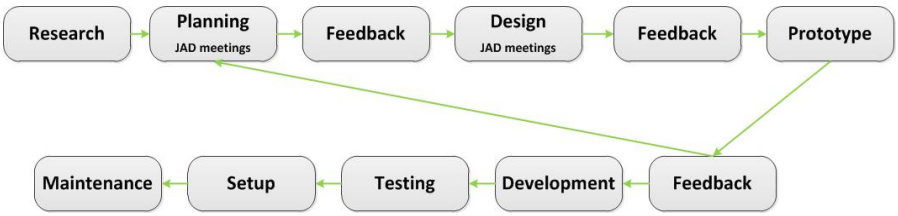


Fig. 18. Joint application development methodology

The project owner's feedback is received at every JAD meeting and after the prototype is finished. The Joint application development methodology is suitable for medium and large scale projects.

Adaptive software development is a methodology that was built as a response to an economy that is increasingly changing and evolving [14]. Adaptive software development is based on iterative development and is oriented on the project's mission. It is a timeboxed model that values delivering features and accepts changes in all stages of the project. Adaptive software development responds accepts risks and handles them efficiently.

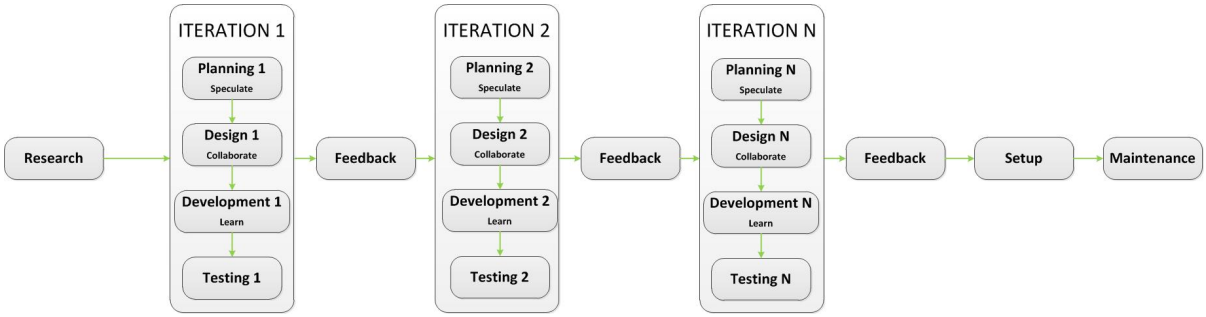


Fig. 19. Adaptive software development methodology

The project owner's feedback is received after each iteration is finished. The Adaptive software development methodology is suitable for small, medium and large scale projects.

Open source software development is a decentralized methodology with no central authority, project owner, no compensation for the project team, no accountability and yet with a high success rate [20]. Open source software development defies traditional economic theory as thousands of programmers work on writing, debugging and testing software without expecting any direct compensation. Most open source software developers will never meet face to face and yet find a way to collaborate in harmony. Open source software development has the advantage of comprehensive testing as code is reviewed by a large number of developers and also benefits from around the clock work on the project as developers are geographically scattered all around the globe.

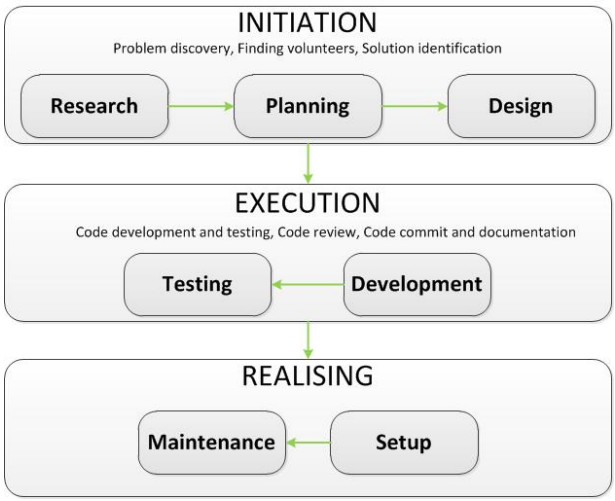


Fig. 20. Open source software development methodology

In open source software development there is no project owner to provide feedback. The methodology is suitable for small, medium and large scale projects.

Microsoft Solutions Framework is a deliberate and disciplined approach to technology projects based on a defined set of principles, models, disciplines, concepts, guidelines, and proven practices from Microsoft [21]. Microsoft Solutions Framework methodology has versions for both lightweight and heavyweight implementation so it can be applied in an agile manner or using a waterfall approach. It fosters open communication and empowers team members but at the same time establishes clear accountability and shared responsibility.

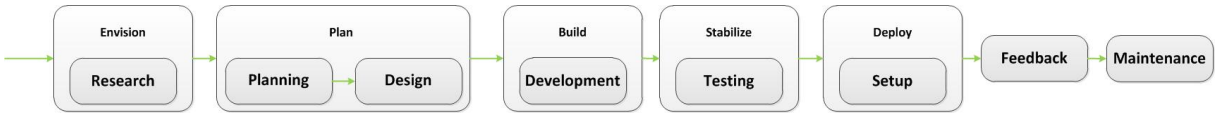


Fig. 21. Microsoft Solutions Framework methodology

The project owner's feedback is received after deployment. The Microsoft Solutions Framework is suitable for small, medium and large scale projects.

Apart from the above mentioned methodologies a further research of the topic might consider analysing the following methodologies:

- Dual Vee model – a variation of the V model;
- Evolutionary project management – forerunner of the current agile methodologies [23];
- Agile Unified Process – a simplified version of the Rational Unified Process;
- Essential Unified Process – a variation of the Rational Unified Process;
- Open Unified Process - a variation of the Rational Unified Process [22].

They were not included in the current comparative study as they are variation of other methodologies and their characteristics were already submitted to analysis.

4. Strengths and weaknesses

Software development methodologies follow one of two paths: heavyweight or lightweight. Heavyweight methodologies are derived from the waterfall model and emphasizes detailed planning, exhaustive specifications and detailed application design. Lightweight methodologies are derived from the Agile model and promote working software, individuals and interactions, acceptance of changing requirements and user feedback. The Microsoft Solutions Framework includes variations for both heavyweight and lightweight philosophies.

Table 2. Software development methodologies characteristics

Methodology	Characteristics	Strengths	Weaknesses
Waterfall	- comprehensive documentation - meticulous planning - linear-sequential process - each phase has its own deliverables	- easy to manage - easy to understand for the project owner and team	- working code is delivered late in the project - does not cope well with changing requirements - low tolerance for design and planning errors
Prototyping	- build one or more demo versions of the software product - project owner is actively involved - prototypes are meant to be discarded	- accurate identification of application requirements - early feedback from the project owner - improved user experience - early identification of missing or redundant	- leads to unnecessary increase of the application's complexity - increased programming effort - costs generated by building the prototype

	- writing code is valued over writing specifications	functionality	
Iterative and incremental	- build an initial model that is meant to be extended in successive iterations - emphasizes design over documentation - project owner is actively involved	- continuous feedback from the project owner - multiple revisions on the entire application and on specific functionality - working code is delivered early in the project	-each iteration is a rigid structure that resembles a small scale waterfall project
Spiral	- focuses on objectives, alternatives and constrains - has 4 major phase: planning, risk analysis, development and evaluation - emphasises risk analysis - evaluates multiple alternatives before proceeding to the planning stage	- working code is delivered early in the project - minimizes risk - strong documentation	- costs generated by risk handling - dependent on accurate risk analysis
Rapid application development	- less emphasis on planning tasks and more focus on development - timebox approach	- applications are developed fast - code can be easily reused	- poor documentation - high development costs - code integration issues - application has to be broken into modules
Extreme programming	- pair programming - unit testing - fast consecutive releases - collective ownership - on-site project owner - open workspace - project owner decides the priority of tasks	- application gets very fast in the production environment - frequent releases of working code - reduced number of bugs - smooth code integration - continuous feedback from the project owner	- lack of documentation - developers reluctance to pair programing - developers reluctance to write tests first and code later - requires frequent meetings - lack of commitment to a well-defined product leads to project owner reluctance
V-Model	- introduces testing at every development stage - highlights the importance of maintenance	- low bug rate -easy to understand and use	- vulnerable to scope creep - relies heavily on the initial set of specifications
Scrum	- iterative development - timebox approach known as Sprints - daily meetings to assess progress known as Daily Scrum - self organizing development team - tasks are managed using backlogs; product backlog and sprint backlog	- deliver products in short cycles - enables fast feedback - rapid adaptation to change	- lack of documentation - requires experienced developers - hard to estimate at the beginning the overall effort required to implement large projects; thus cost estimates are not very precise
Cleanroom	- iterative development - box structure method - using mathematic models in quality control - statistical approach to testing	- considerable reduction in bug rate - higher quality software products	- increased development costs - increased time to market for software product - requires highly skilled highly experienced developers
Dynamic systems development	- iterative development - MoSCoW prioritisation of	- focusses on addressing effectively the business	- requires large project teams at it has multiple

method	tasks - timebox approach - non-negotiable deadlines - strict quality standards set at the beginning of the project - project team and project owner share a workplace (physical or virtual) - test early and continually	needs - post project implementation performance assessment - complete documentation - active user involvement	roles to cover - requires very skilled developers
Rational Unified Process	- iterative development - prioritize risk handling - adequate business modelling - change management - performance testing	- accurate and comprehensive documentation - efficient change request management - efficient integration of new code - enables reuse of code and software components	- requires highly qualified professionals - development process is complex and poorly organized
Lean software development	- iterative development - discards all components that do not add value to the product - amplify learning - customer focus - team empowerment - continuous improvement	- reduced project time and cost by eliminating waste - early delivery of working code - motivated project team	- project is highly dependable on individual team members - a team member with strong business analysis skills required
Test-driven development	- unit testing -testing scenarios are developed before actual coding - repeated short development cycles - suitable for debugging legacy code developed with other techniques	- less time spent on debugging - higher quality code - by designing tests the developer empathizes with the user - less defects get to the end user	- tests are focused on syntax and overlook actual functionality - requires more code than most methodologies - the developer is actually the one doing the testing - writing unit tests increases costs
Behavior-driven development	- unit testing - focuses on business value - genuine collaboration between business and development	- easy to maintain - usability issues are discovered early - reduced defect rate - easy to integrate new code	- project owners are reluctant to write behaviour scenarios
Feature-driven development	- iterative development - application is broken down into features - no feature should take longer than two weeks to implement - uses milestones to evaluate progress	- multiple teams can work simultaneously on the project - scales well to large teams - good progress tracking and reporting capabilities - easy to understand and adopt	- individual code ownership - iterations are not well defined
Model-driven engineering	- uses domain model - models are automatically transformed into working code - knowledge is encapsulated in high level models - emphasizes reuse of standardized models	- high degree of abstraction - increased productivity - delivers products with a high degree of compatibility and portability - shorter time to market	- requires considerable technical expertise - documentation is readable only by domain experts - is difficult to implement version control on modelling environment

		- lowers maintenance costs	
Crystal Methods Methodology	- focusses on people and skill and not on process - more than one iteration in a release - different approaches depending on the projects size and criticality	- easy to implement - frequent delivery of working code - developers have dedicated timeslots to reflect on possible code improvements	- critical decisions regarding the architecture of the application are made by individuals and not by the entire team
Joint Application Development	- emphasises system requirement determination - involves the project owner and end user in the design and development - JAD meetings - prototyping	- accelerates design - enhances quality - promotes teamwork with the customer - creates a design from the customer's perspective - lowers maintenance costs	- relies heavily on the success of the group meetings - does not have a documented approach for stages that follow system requirements determination and design
Adaptive software development	- iterative development - focusses on the final goal of the project - feature based - timeboxed - risk driven	- effective handling of change and scope creep - easy to understand and implement - enables innovation	- low risk handling - uses assumptions and predictions - lacks tangible documentation
Open source software development	- iterative development - geographically distributed teams - collaborative work	- low costs - highly motivated and dedicated developers - comprehensive testing as code is reviewed by a large number of developers	- low accountability for submitted code - no central management authority - unstructured approach to development
Microsoft Solutions Framework	- has versions for both lightweight and heavyweight implementation - fosters open communication - empower team members and establishes clear accountability and shared responsibility	- supports multiple process approaches - solid risk handling policies - built to respond effectively to change - reduces team size	- difficult to setup and configure

Out of a total of 20 software development methodologies that were analysed 6 were based on the heavyweight model and 13 were based on the lightweight model. One methodology offered support for both a lightweight and a heavyweight approach. In order to choose the appropriate software development methodology for a project one should consider: project owner profile, developer's technical expertise, project complexity, budget and deadlines. An innovative software development project is difficult to match with one of the existing development methodologies. Innovative software development projects require writing comprehensive documentation in order to patent any original output that might result. It also requires a considerable degree of flexibility as changes occur often.

5. Conclusions

Software development methodologies follow two major philosophies: heavyweight and lightweight. Heavyweight methodologies are suitable for projects where requirements are unlikely to change and the software complexity allows for detailed planning. Heavyweight methodologies are easy to understand and implement. They provide solid documentation and appeal to project owners because they are well structured and showcase tangible deliverables for every stage of the project. With heavyweight methodologies the project

manager can easily perform tracking, evaluation and reporting. The project owner is considerably involved only in the research and planning stages. Lightweight methodologies are suitable for projects were specifications are unclear or are likely to change due to project internal or external factors. Lightweight methodologies are based on an incremental approach were software is delivered in multiple consecutive iterations, all of them being functional versions of the application. Lightweight methodologies provide great flexibility and can easily adapt to change. They promote early delivery of working code, self-organizing teams and adaptive planning. The project owner is heavily involved in all the stages of the project as its input and feedback is critical for the success of lightweight methodologies. When choosing a software development methodology project owner profile, developer's technical expertise, project complexity, budget and deadlines must be taken into account. Often no methodology will fit perfectly the profile of a specific project. Then the best matching methodology should be used or in the case of experienced project teams and project managers a combination of methodologies could be introduced. In the case of innovative software development projects a new methodology is required. This topic is a subject for further research in the software development field.

6 Acknowledgment

This paper was co-financed from the European Social Fund, through the Sectorial Operational Programme Human Resources Development 2007-2013, project number POSDRU/159/1.5/S/138907 "Excellence in scientific interdisciplinary research, doctoral and postdoctoral, in the economic, social and medical fields -EXCELIS", coordinator The Bucharest University of Economic Studies.

References

[1] http://php.net/ChangeLog-5.php

[2] J. E. Cooling, T. S. Hughes, "The emergence of rapid prototyping as a real-time software development tool", *Proceedings of the Second International Conference on Software Engineering for Real Time Systems*, 18-20 Sep. 1989, Cirencester, UK, Publisher: IET, 1989, pg. 60-64

[3] C. Larman, V. R. Basili, "Iterative and Incremental Development: A Brief History", *Computer*, vol. 36, no. 6, pg. 47-56, 2003, doi:10.1109/MC.2003.1204375

[4] B.W. Boehm, "A spiral model of software development and enhancement", *Computer*, vol. 21, no. 5, pg. 61-72, 1988, doi: 10.1109/2.59

[5] K. Beck, "Embracing change with extreme programming", *Computer*, vol. 32 , no.10, pg. 70 – 77, 1999, doi: 10.1109/2.796139

[6] L. Rising, N. S. Janoff, "The Scrum Software Development Process for Small Teams", *IEEE Software*, vol. 17, no. 4, pg. 26-32, 2000, doi:10.1109/52.854065

[7] A. Spangler, "Cleanroom software engineering-plan your work and work your plan in small increments", *IEEE Potentials*, vol.15, no. 4, pg. 29 – 32, 1996, doi: 10.1109/45.539962

[8] G. Pollice, *Using the Rational Unified Process for Small Projects: Expanding Upon eXtreme Programming*, Rational Software White Paper, 2001

[9] C. Ebert, P. Abrahamsson, N. Oza, "Lean Software Development", *IEEE Software*, vol. 29, no. 5, pg. 22-25, 2012,

[10] E. M. Maximilien, L. Williams, "Assessing test-driven development at IBM", *Proceedings. 25th International Conference on Software Engineering*, ICSE 2003, 3-10 May 2003, Portland, USA, Publisher: IEEE, 2003, doi: 10.1109/ICSE.2003.1201238, pg.564-569

[11] D. J. Anderson, *Feature-Driven Development*, Microsoft Corporation, Oct. 2004

[12] J. Bezivin, *Model Driven Engineering: An Emerging Technical*

Space, International Summer School, *Summer School on. Generative and Transformational Techniques. in Software Engineering*, 4-8 Jul. 2005, Braga, Portugal, Publisher: Springer Berlin Heidelberg, pg. 36-64, doi: 10.1007/11877028_2.
[13] E. W. Duggana, C. S. Thachenkaryb, "Integrating nominal group technique and joint application development for improved systems requirements determination", *Information & Management*, vol. 41, no. 4, pg. 399–411, 2004, DOI: 10.1016/S0378-7206(03)00080-6

[14] D. Riehle, "A comparison of the value systems of Adaptive Software Development and Extreme Programming: How methodologies may learn from each other", *Proceedings of the First International Conference on Extreme Programming and Flexible Processes in Software Engineering*, XP 2000, 21-23 Jun. 2000, Cagliari, Italy, pg. 35-50
[15] J. Martin, *Rapid application development*, Publisher: Macmillan Publishing, 1991, pg. 788, ISBN:0-02-376775-8
[16] S. Mathur, S. Malik, "Advancements in the V-Model", *International Journal of Computer Applications*, vol. 1, no. 12, pg. 29-34, 2010, doi: 10.5120/266-425
[17] J. Stapleton, P. Constable, *DSDM, dynamic systems development method: the method in practice*, Publisher: Cambridge University Press, 1997, pg. 192, ISBN-10: 0201178893
[18] M. Soeken, R. Wille, R. Drechsler, "Assisted behavior driven development using natural language processing", *Proceedings of the 50th International Conference on Objects, Models, Components, Patterns*, TOOLS 2012, 29-31 May 2012, Prague, Czech Republic, Publisher: Springer Berlin Heidelberg, 2012, doi: 10.1007/978-3-642-30561-0_19, pg. 269-287
[19] J. A. Livermore, "Factors that Impact Implementing an Agile Software Development Methodology", *Proceedings of IEEE SoutheastCon*, SECON 2007, 22-25 Mar. 2007, Richmond, USA, Publisher: IEEE, 2007, doi: 10.1109/SECON.2007.342860, pg.82-86
[20] G. Madey V. Freeh R. Tynan, "The open source software development phenomenon an analysis based on social network theory", *Proceedings of the 8th Americas Conference on Information Systems*, AMCIS, 9-11 Aug. 2002, Dallas, USA, 2002, pg. 1806-1813.
[21] G. Lory, D. Campbell, A. Robin, G. Simmons, P. Rytkonen, *Microsoft Solutions Framework version 3.0 Overview*, White Paper, June 2003.
[22] R. Balduino, *Introduction to OpenUP (Open Unified Process)* http://www.eclipse.org/epf/general/OpenUP.pdf
[23] S. Woodward, "Evolutionary project management", *Computer*, vol. 32, no. 10, pg. 49-57, 1999, doi: 10.1109/2.796109
[24] W. W. Royce, "Managing the development of large software systems: Concepts and techniques", *IEEE WESCON*, vol. 26, no. 8, pg. 1-9, 1970

Mihai Liviu DESPA has graduated the Faculty of Cybernetics, Statistics and Economic Informatics from the Bucharest Academy of Economic Studies in 2008. He has graduated a Master's Program in Project Management at the Faculty of Management from the Bucharest Academy of Economic Studies in 2010. He is a PHD Student at the Economic Informatics PHD School and he is currently Project Manager at GDM Webmedia SRL. His main field of interest is project management for software development.

Social BI. Trends and development needs

Corneliu Mihai PREDA
Academy of Economic Studies, Bucharest
mihaicpreda@gmail.com

This paper tries to show that the need to implement BI solutions that integrate social networks data is increasingly more acute. Companies need to understand that efficient decision making has to take into account the importance that social medial has nowadays. In addition, the software developing companies have to keep up to date with new trends generated within these communication channels and adapt their solutions to this dynamic environment.

***Keywords**: Business Intelligence, Social Network, Social BI, Market Trends*

1 Introductory notes

Social BI is a management technique that integrates the sharing of social resources in order to improve existing company projects, products or business processes. This type of BI is usually handled by software products that offer the possibility to analyze social media indicators while, at the same time, managing projects like a traditional solution. Also, the main characteristic of this type of BI is that it welcomes the involvement of the client from the starting point of business processes and long time before the product development end with its market release.

Social BI is seen as an ideal way to take full advantage of human capital, both inside and outside the company. From the employee's point of view, this kind of application offers transparent information distribution across project assignments, no matter if they are involved or would like to be involved in these projects. Also, it creates a widely accepted environment for merit acknowledgement. On the other hand, outside stakeholders (mainly clients) receive the means to actively give feedback on product/services development, ensuring themselves that
they will have a higher level of satisfaction of their needs [5.

Currently, through the contribution of social networks to the scope of BI, the focus is increasing on their ability to bring a lot of customers for small companies that are producing for niche markets. In addition to this, they provide analysis of trends and statistics with greater relevance to the company and the end customer that is involved in creating the products for which he is the target of. The next figure allows us to see the features outlined by social media networks into the BI domain [2].

As networks and internet applications are becoming increasingly complex, given increasingly more detailed data can be collected about users/consumers and sent to companies operating in narrow markets, which they can use for commercial purposes (e.g. loyalty campaigns, promotions, etc.).

At the same time, user interaction can lead to favorable information dissemination about the company which can become major waves, this being among the most effective ways for market advertisement today. Also, this information can turn into reliable sources for consumers who want to document about a specific product that they want to purchase. Propagation speed of this medium is constantly increasing, actively using friend networks and acquaintances of users.

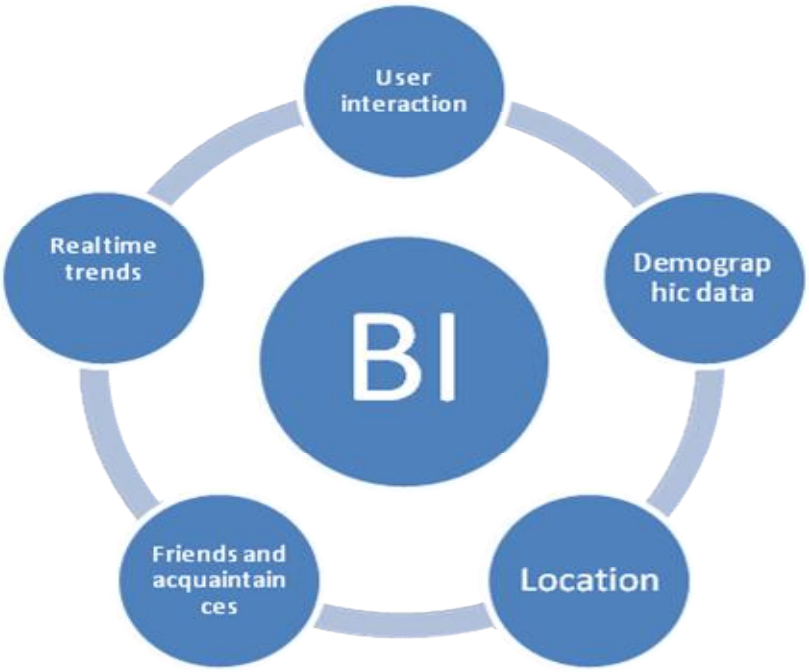


Fig. 1. BI – Social media

Companies need to recognize the need for virtual existence in the social environment because the potential for loyalty and sales increase is considerable. Even if the purpose is just notifying the end user that the company exists is huge and at the same time, the effects are measurable and the measurements reliable.

The fact that both companies that are trying to sell products or services and also social networks have as a common goal the meeting of the needs of the same customer/end user, will lead over time to the strengthening of relations of interdependence. This will actively be reflected in new products for BI techniques and processes. These tools will provide the link that will connect the customer to the companies that serve him and bring increased efficiency and competitive advantage for businesses that will invest in these solutions.

BI areas of influence and social media were, since their appearance, the subject of continuous development and research. While BI aims to help decision-making at company level by providing reports containing relevant data, social media aims to develop into a significant source of personal data at the individual, opinion, relationship and interests level. As a result, the lasting parallel existence of these two concepts, led inevitably to the compund concept of Social BI.

Social BI, conceptually wants to extract relevant information that would lead to effective decisions, all based on data from social media. Also, it seeks the participation in the design of BI solutions that contain tools for analysis and measurement of social metrics. In the present context, in which more companies succeed in implementing BI systems, the addition of this new component leads to accentuated competition and a major benefit for the end customer [2].

2 BI Interaction Management

Management Systems of social networking as an integral part of BI, is a group of applications or methods used to manage and follow business processes in a distributed environment. These systems can be manually or automatically maintained and they allows managers to track, aggregate, publish and take part in

many social media channels using a single software product.[4]

The operating processes have three main features:

- Connection to social media channels;
- Ability to quickly publish from anywhere inside these channels being able to adjust or program the occurrence of one or more messages;
- Ability to manage social data. The system allows managers to view aggregated data and have various analytical reports with different levels of complexity and detail.

Fig. 2. BI interactions

Each of these basic functions can be extended to a very high level of detail. Social media platforms are proving to be an important component of any marketing strategy. At the same time, marketing has come to use many of the BI tools to reach for the target audience, often extracting data directly from the internal business processes. Consequently, we see the need for marketing to track customer response in real time.

CRM (customer relationship management) focuses on providing services, loyalty techniques and sales analysis that address actual client needs in a more direct way that before on what they consider to be the main issues of company-customer relationship (prices, warranties, shipping, etc.).

CRM is fundamentally different from social media components in that it is much less immediate. This is the mechanism through which contracts and products reach the end user. Interacting with social media, CRM uses information provided by it in order to publish messages that should reach both current customers as well as prospective customers. By integrating these two components, potential customers will know what to expect if they choose a product provided by a certain company.

Inside the social media component, gaining credibility is a matter not only of content, but also of a good social interaction. A common mistake is to approach social media as platforms to send some mass e-mails and not as some communication processes. They should be considered more like publishing tools through participation.

As mentioned previously, CRM integration with social media result in a long term marketing strategy. The problem here is the huge amount of data coming through these two channels. This is the time when BI techniques must intervene in order to interpret the data into relevant reports

available for the decision-making factor. What BI cannot do, however, is improve the quality of input data. External consultants are needed to standardize the data stream so that it becomes valuable for the company. The existence of these expensive consultants does not preclude the use of free, but very simple services (e.g. the number of brand occurrence over a given period of time). Top market share service providers in this area are SAP, Oracle and IBM which provide huge processing capabilities and complex analysis.[3]

3 Social media dynamics and associated metrics

Regarding the most important changes that social media has been through since its appearance, we can summarize the following in Table 1.[2]

Table 1. Evolution of social media

Beginning of social media	**Present social media**
Social Media Guru: the man who has numerous communication strategies and is an expert in creating accounts. A very short phase in which sales people tried to artificially increase their level of apparent technical training.	Social Media Analytics, BI. Given that data are widely available through various services and interfaces is difficult to shape and monitor communication strategies.
Web Tools for distribution to multiple social media sites. Useful for a while, but many sites offer plug-ins that allow postings.	Distribution coupled with analytics. These have as a consequence products that include embedded analytics service.
Web Analytics. Although web analytics are still in use they do not provide information on visitor opinions.	Social Media Analytics. These scan comments about a particular brand or product not only to refrain opinions, but also to track trends.
Many social sites. The area of social networks was more diverse in the past. The success of a website is not determined by technology, but by its users.	YouTube, Twitter, LinkedIn, Facebook. Strongest sites have persisted and currently have the most numerous users. Although there are possible failures or hindered daily traffic, users' loyalty is unquestionable.
Spam and numerous users. Earlier Facebook or Twitter users prided themselves with many friends and large volumes of messages sent.	Target audience. Since then, marketing is trying to attract audiences using content relevant to them, and by using means of increasing ingenuity.
Public image. Having access to large quantities of information there was a need for more detailed material.	Content. People are now more motivated to seek customized content.
Quick and impulsive sales. The motto "buy fast" or "limited stock" does not apply online. It is easy for users to search for similar items.	Greeting the client. Consumers expect to be greeted with a concept before the acquisition that will integrate him in a community.
Unmanageable data. Previously, the data came in a large variety of formats and there were very few methods of integration.	Big Data. Today, data is formatted and available for distribution via various interfaces. There are also many instruments of conversion from one data format to another.

Regarding the most important social media metrics that companies have to be responsive to, consultants identify the following:

1. Monitoring views of a target demographic in a certain geographical area by filtering used keywords in natural language to identify opinions of satisfaction or negative feelings associated with a product or company. This process can prevent a viral explosion of negative opinions. If negative opinions are needed direct intervention is needed to solve valid issues of the most influential individuals.
2. Identifying cases where negative opinions could turn into a PR crisis is achieved through effective monitoring of social media to identify key individuals before the problem gets proportions.
3. Identify individuals, blogs and web applications that are most influential to a brand should lead to discussions on topics with which they have shown interest.
4. Real time tracking of the evolution of views and opinions, especially about new products and services leads to shortening the time between the signaling of the problem and its solution.
5. Measuring the link between marketing efforts and customer response at certain times of day and divided into geographical areas leads to the customizing level that marketing campaigns need to be more efficient.
6. Determining the type of media and platforms that have the most success allows the company to focus its efforts. For example, a company may divert resources to video campaigns because this means of communication has proved most effective in attracting clients.
7. Monitoring competitors must be done using all available means to redeploy their tested business formula into the company's own processes.
8. Documenting the industry sector in which the company operates leads to rapid identification of growth opportunities and this can be achieved by the same methods of filtering online content.

4 Development possibilities

Characteristics of social media together with design elements of BI currently contribute as a foundation for any social BI strategy. As areas of development for the above mentioned tools, we mention the following:

Users and customers: the inclusion of social media related set of data available for analysis can attract new BI users within an organization. The importance attached to social interaction between users is an example of the new design requirements of BI. Also, if the BI system allows and encourages interaction with customers, this will generate new technical needs.

Products and services: along with the availability of data from social media, new products and services within the BI can be offered for use. Inside this context, development directions include the discovery of methods to integrate traditional BI results and analysis related to the structure of social networks or opinions trends. In addition, the limitations regarding the quality and security of data, will constantly update the need for appropriate tools.

Processes: some BI processes should be adapted for the integration of social media functionality. However, the research and development requirements are relatively low in this case, contrary to the situation in which the BI system is used directly in the interaction with the clients' calls via live channels.

Data: nearly all of the social media impacts data management architecture. There will be a need for a broad spectrum approach to BI tools if they are to process both traditional transactional data and data from social media. Social data type is extremely dynamic and therefore cannot guarantee an accurate mapping over a long

period of time. There is a need for continuous information for BI solution developers on use cases that may arise. In terms of quality and/or legality of data sources, at first glance, social media provides individualized, but non-standard and poor quality data. Therefore, the need to develop methods of aggregation will lead to insight into the trends observed at the population level. Also, the Wikipedia example convinces us that making common and mutual control over user-generated content can lead to a continuous increase in data quality.

IT&C: huge data volumes and their frequent updates involved in social media have a greater and greater impact on IT&C. These unstructured data require specific software to be processed. Currently the most popular management concept for this is the "Big Data", but this is a domain anticipated to give rise to further developments.

Techniques: similar characteristics between different social environments on the high volume of data, unstructured data, quickly updated content, result in a major need for research on analysis techniques. Unlike IT&C, which requires *de novo* development of software products, social BI needs to improve the current, widely used ETL process.

Governance: social media and BI systems can generate the emergence of new roles and responsibilities within organizations. It can identify a clear impact on organizational rules and define the compliance with those rules. In parallel with the use or distribution of information through these social channels, problems related to the legal aspects of global interactions may pose a few problems.

Strategy: BI strategy, as part of the company strategy, aims to provide support for optimal decisions on the market. Effective use of social data provides many ways to contribute to the overall objectives such as customer satisfaction. One direction of research and development could be measuring the contribution social BI in particular the performance of the organization. [1]

5 Conclusions

The importance and practical application of BI solutions that integrate ways to measure and analyze data from sources in the field of social networking, records an upward trend. As a result, we see an increase of awareness regarding the importance of these tools from both companies that are competing in different markets and also from the developers involved in research and development.[4]

We conclude that the domain of "Social BI", although recently emerged as a concept, is one of great growth potential, representing the future of any effective marketing campaign.

References

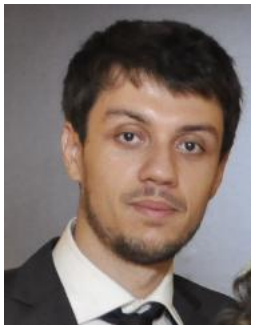
[1] Dinter, Barbara; Lorenz, Anja “Social Business Intelligence: a Literature Review and Research Agenda”, 33rd ICIS 2012 (International Conference on Information Systems)

[2]”Social Media and Business Intelligence” First Edition, CIO Whitepapers, Ventus Publishing ApS 2012

[3] “Social Media and Business Intelligence:Creating the IntegratedCustomer Hub”, Oracle White Paper, September 2012

[4] “Unlock the value of social media data – Social Intelligence”, Viewpoint Paper, Hewlett-Packard Development Company, October 2013

[5] http://www.techopedia.com/



Corneliu Mihai PREDA graduated from the Faculty of Finance, Insurance, Banking and Stock Markets of the Bucharest University of Economic Studies in 2008 and the Capital Markets Master's program in 2010 with a paper called „Testing the informational efficiency of stock markets". He is currently a final year student at the Database Business Support Master's program at the Faculty of Cybernetics, Statistics and Economic Informatics. His past working experience include Business Planning and Financial Analisys and also SAP/ABAP Technical Development.