Techniques of Building and Validating Metrics for Collaborative Systems Applied in Economy

Cristian CIUREA
Economic Informatics Department,
Academy of Economic Studies, Bucharest, Romania,
cristian.ciurea@ie.ase.ro

The most encountered types of collaborative systems in knowledge-based economy are formalized and their characteristics are identified. Each collaborative system is analyzed from the point of view of inputs, states and outputs. Metrics for assessing the resources allocation in a banking system are built and validated by testing on real data. The users' behavior in the Collaborative Multicash Servicedesk application is evaluated and indicators are defined for this matter. A genetic algorithm is implemented for building and validating metrics of collaborative systems. The idea of multimedia techniques aggregation in collaborative systems is frequently found and mediafilia is such a result of aggregation.
Keywords: *Collaborative Systems, Metrics, Building, Validating, Genetic Algorithms, Mediafilie*

1 Formalization of collaborative systems in the knowledge-based economy

In [1] is considered that in the knowledge-based economy, collaboration means more than two agents working together. It requires defining a shared goal and, in order to achieve this goal, the agents should create an agreement upon their ways of actions. The implementation of collaborative informatics systems in different activity fields facilitates the resolution of many citizens' problems.
A collaborative system is represented by:

- a lot of input messages, $I = \{i_1, i_2, ..., i_m\}$, which determine the conduct of activities and the development of process stages;
- a lot of states, $X = \{x_1, x_2, ... x_s\}$, that require actions, consumption of resources, operations, equipment, people;
- a lot of output messages, $E = \{e_1, e_2, ... e_n\}$, which accompany finished products and services.

The input and output messages consists of subsets of homogeneous messages in relation to one or more criteria.
A collaborative system is defined by the following elements [2]:

$$\Sigma = (T, S, R, I, \Omega, X, E, \Gamma, F, \varphi, \eta),$$

where:
T – the time, represented by the lot of moments in which the system operates;
S – the space, represented by the set of locations where the system operates;
R – the resources, the lot of human, material and energy resources that contribute to activity achievement;
I – the set of values of input variable i;
Ω – the class of temporary evolutions allowed, $\Omega = \{\omega : T -> I\}$, $\omega = \{i(t) / t \in T, i(t) \in I\}$, $\Omega \neq \phi$;
X – the space of states, represented by the set of values of state variable x;
E – the set of values of output variable e;
Γ – the class of possible outputs, $\Gamma = \{\gamma : T -> E\}$, $\gamma = \{e(t) / t \in T, e(t) \in E\}$;
F – the work flows, the set of values of flow variable f;
φ – the transition function of the system, $\varphi : TxTxXx\Omega -> X$, $x(t) = \varphi(t;\tau;x;\omega)$;
η – the output function of the system, $\eta : TxX -> E$, $e(t) = \eta(t; x(t))$.
Figure 1 presents the components of a collaborative system and the relationships between them:

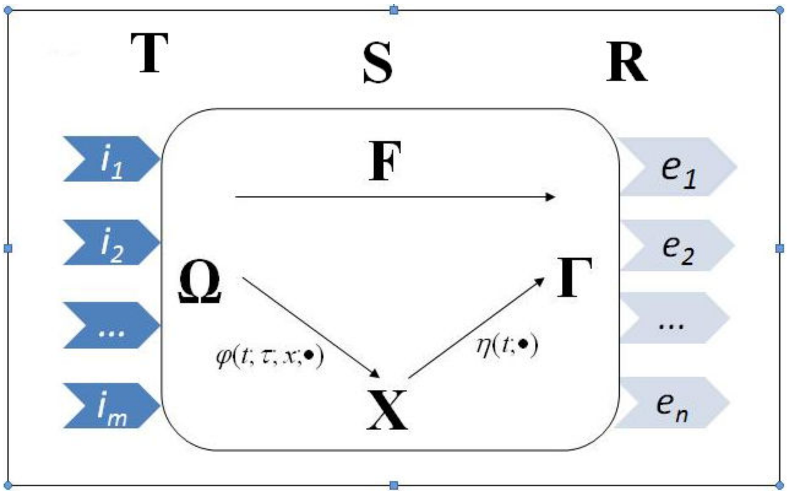

Fig. 1. The components of a collaborative system

In the knowledge-based economy, there are many types of collaborative systems, the must important being the collaborative banking systems, collaborative educational systems and collaborative systems in production.

In the machine-building industry are encountered the followings elements that define a collaborative production system: raw materials types, machine types, types of operations, types of finished products, and number of beneficiaries.

The collaborative production system, *SCP*, from the machine-building industry is represented by the following quintuple:

SCP = (T, S, R, NMAT, MMAS, KOPE, X, PPRO, BBEN, F, φ, η),

where:

T – the set of time moments when the system operates;

S – the set of locations where the system operates;

R – the lot of human, material and energy resources that contribute to activity achievement;

NMAT – the list of raw materials;

MMAS – the list of machines types;

KOPE – the set of operations;

X – the space of states, represented by the set of values of state variable;

PPRO – the set of finished products;

BBEN – the list of beneficiaries;

F – the work flows;

φ – the transition function of the system;

η – the output function of the system.

The collaborative side of the production system concern the existence of databases for raw materials, workers, products, equipment, operations, beneficiaries, calendar, with which are built entries of planning, programming, monitoring and production optimization models.

The concept of *mediafilie* refers to collaborative construction elements, which are materialized using two results from two totally different areas in order to get a new product with new features that enhance the features of the initial products. Thus, a static product, as a representation, is aggregated with a multimedia product that refers to the same object, static represented, which will lead to achieving a collaborative product with high complexity level [3].

Figure 2 presents the structure of a collaborative production system in which two sections operates for producing the P1 and P2 products [4].

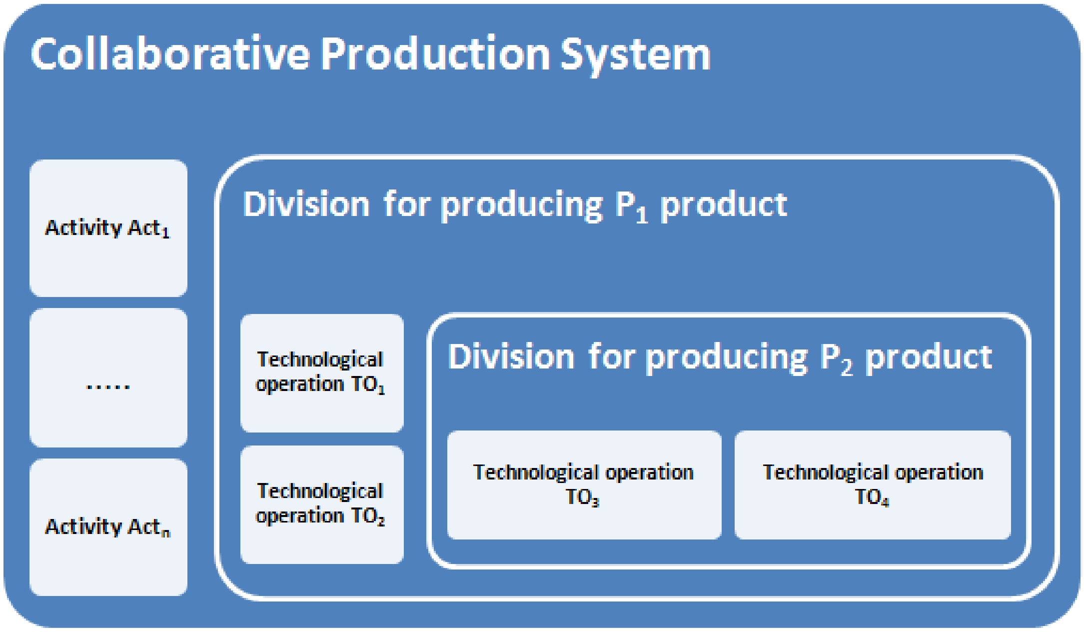

Fig. 2. Collaborative production system with two divisions

In production processes the following combinations of operations, machines and products are encountered:

$1 - 1 - 1$, meaning that an operation is running on a machine to manufacture a product;

$q - 1 - 1$, meaning that q operations are executed on a single machine for one product development;

$1 - m - 1$, an operation is executed on m machines to obtain a single product;

$1 - 1 - p$, an operation is running on a single machine for making p products;

$q - m - p$, in the sense that q operations are executed on m machines to obtain p products.

The collaborative side involves elements of compensation, such as pluses in one place and minuses in another place. Compensation elements are balanced so that flagrant violations come to management to be corrected by allocating additional resources.

A collaborative banking system, *SCB*, seen as finite state machine, is formalized as:

$SCB = (BC_1, BC_2, BC_3, BC_4, BC_5, BC_6)$,

where:

BC_1 – the input alphabet, containing an array of input messages that determine the conduct of activities and the coverage of process steps;

BC_2 – the output alphabet, defined as messages that accompany the services offered by the bank;

BC_3 – a finite and non-empty array of states that require action, consumption of resources, operations, equipment, people;

BC_4 – the initial state, element of BC_3;

BC_5 – the transition function $BC_5 : BC_3 \times BC_1 \rightarrow BC_3 \times BC_2$;

BC_6 – the output function.

Collaborative systems require the existence of such activities that need to be automated, in order to streamline the workflow within an organization. In a collaborative banking system, any change in the workflow must be found in the corresponding rules and procedures.

2 Building metrics for assessing the resources allocation in a banking system

Metrics of collaborative systems represent mathematical models developed around an equation. Metrics use analytical expressions having the form: $y = f(x, z, w)$, where x, z and w are variables of influence factors, and y is the result variable [5]. A metric of collaborative systems must be characterized by the following properties: sensitivity, not compensatory character, not catastrophic character, representativeness [6].

In a collaborative banking system, the agents are instructed to carry out other activities than those they currently realize. If the case in which a banking officer is on leave, his duties are handled by the customer account manager so that the system operate normally.
It is considered the indicator for assessing the level of specialization of an agent in a collaborative banking system, *ESA*, defined as follows:

$$ESA = AFP \times pe_1 + NFP \times pe_2,$$

where:
AFP – the number of activities from the job description, performed by an agent within a working day;
NFP – the number of activities not covered in the job descriptions that were made by the agent in a working day;
pe_1, pe_2 – weights associated to the number of activities realized ($pe_1 = 0.6$, $pe_2 = 0.4$).
In a collaborative system, the indicator for assessing the level of specialization of an agent has a value close to 1, since the agents are instructed to carry out activities other than those defined in the job description. In the classical system, each agent realizes only what is provided in the job description and is not trained to perform other tasks.

The decision makers responsible for the efficient allocation of human resources from a bank are the managers and directors of departments. There are situations in which certain jobs from the bank departments remain uncovered temporarily, and managers require the support from other departments regarding the allocation of existing human resources, until are found and hired people to fill those positions. There are identified relevant people within the bank departments that have the necessary knowledge and skills and are allocated on vacancies jobs.
There are considered *TS* - the set of bank employees and *TE* - the set of open positions in the bank. Associations between elements of *TS* and *TE* are shown in Figure 3. Depending on the knowledge and skills of each employee, the associations are of followings types:

- *1 to 1*, meaning that an employee meets the requirements of a single job;
- *1 to M*, an employee is able to meet the requirements of many jobs;
- *M to 1*, in the sense that the requirements of a job are highly complex and require the knowledge and skills of many employees.

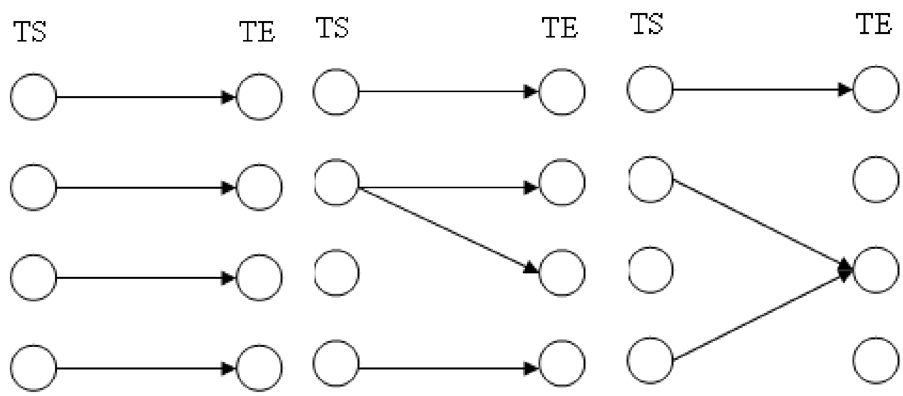

Fig. 3. The associations between employees and positions in a bank

Figure 2 presents the correct allocation of employees on vacancies jobs within a bank. The employee *AA* is assigned to the post *AA* that fits his skills and knowledge. Even if the job responsibilities are inadequate with those of the employee, the allocation is correct, based on the premise that no employee is fully compatible with the position held. The same correct allocation is presented in the case of the employee *BB*.

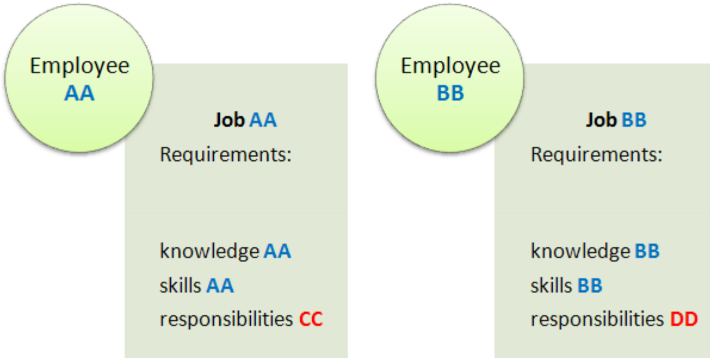

Fig. 2. The correct allocation of human resources in a collaborative banking system

In Figure 3 is shown the situation in which the employee *AA* was improperly assigned to the job *XX*, because the knowledge, skills and responsibilities of the job are quite different from those of the employee.

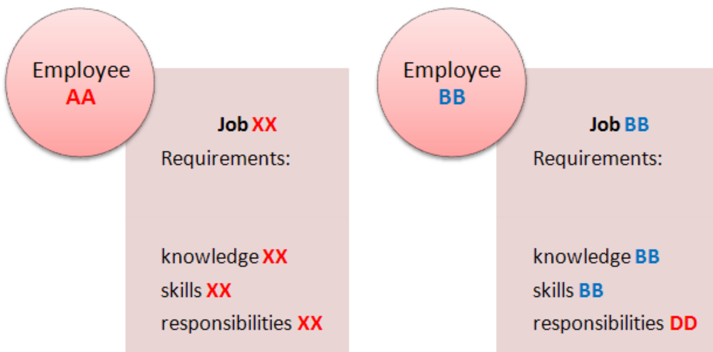

Fig. 3. The incorrect allocation of human resources in a collaborative banking system

The allocation should be made so that moving a person from a job to another does not adversely affect the bank's activity. The incorrect allocation of human resources on different jobs has dramatic consequences for the proper operation of the bank, but also on personal and professional development of individuals.

An indicator of efficiency is represented by the degree of increasing the level of staff training. Considering the training period of five years, the minimum number of qualifications that an employee gets is one in five years and the maximum number is one qualification per year or five qualifications in five years.

The degree of increasing the level of staff training, *GCNP*, is determined according to the relationship:

$$GCNP = \frac{NTCC}{NPCC} \times 100$$

where:

NTCC – the total number of trainings or qualifications supported by the bank in five years;

NPCC – the number of employees qualified in five years.

If are taken into consideration the training durations, namely the minimum duration of a training and the maximum duration of a training, expressed in months, then the total number of trainings supported by the bank within five years, *NTCL*, expressed in months, is given by:

$$NTCL = NA * NPCC * Dmax,$$

where:

NA – number of years;

Dmax – the minimum duration of training, expressed in months.

The degree of increasing the level of staff training is determined according to the same relationship, except that the number of employees qualified in five years is weighted with the durations of qualifications for each person. The indicator shows the situation where the collaborative banking system is found, in terms of level of staff training and the quality of human resources.

3 Metrics validation by testing on real data

The indicator for evaluating the efficiency of a bank officer, BO, is calculated as follows:

$$BO = \frac{ANP \times p_1}{AST \times p_2 + AHT \times p_3} \times AAM \times p_4,$$

where:
AST – average time for serving a client;
AHT – average time for waiting in line;
ANP – average number of clients waiting in line;
AAM – average amount of money traded by a client per unit time;
p_1, p_2, p_3, p_4 – weights, with $p_1 + p_2 + p_3 + p_4 = 1$.

For a bank in Romania the following results were obtained for the indicators *AST*, *AHT*, *ANP* and *AAM* in the case of 10 clients served:

Table 1. The values of indicators for a collaborative banking system

	Client 1	Client 2	Client 3	Client 4	Client 5	Client 6	Client 7	Client 8	Client 9	Client 10
AST	10	9	8.5	9.5	11	10.5	10	7.5	11.5	9.5
AHT	5	6	5.5	3.5	4	6.5	3	5	4.5	5.5
ANP	50	48	55	52	46	59	37	47	45	54
AAM	105	103	110	114	98	93	126	108	99	101

To calculate the indicator for evaluating the efficiency of a bank officer, there are considered the following values of the weights: $p_1 = 0.3$, $p_2 = 0.2$, $p_3 = 0.2$, $p_4 = 0.3$.
From Table 1 results that the average time for serving a client *AST = 9.7* minutes, the average time for waiting in line *AHT = 4.85* minutes, the average number of clients waiting in line *ANP = 49.3* and the average amount of money traded by a client per unit time *AAM = 105.7* RON. Results that the indicator for evaluating the efficiency of a bank officer, *BO = 161.16*, which means that the bank officer is better with *61.16%* than the level of efficiency required by the bank.
Metrics validation is a process of verification in practice of indicators built to determine their accuracy and ability to reflect the quality characteristics measured.
The data on which indicators are tested are collected from the banking system through the Collaborative Multicash Servicedesk – CMS application. Table 2 shows the categories of operations carried out by analysts using the application in the period April 1 to May 1, 2011.

Table 2. The number of operations performed by category

Operation	Total
Authentication	686
Search	47
Export in Excel	16
Export in text file	10
Close request	16
Register request	229

Table 2 shows that most analysts have conducted Authentication operations to check the status of requests registered in the application database and 30% of Authentication operations were followed by operations of registering new request. From the 229 requests recorded in April 2011, a number of 16 requests have required subsequent closure, not being resolved at the moment they were taken.
Table 3 presents the report regarding the operations performed by *Analyst 99* between April 1 and May 1, 2011.

Table 3. Number of operations performed by *Analyst 99*

Analyst name	Operation	Number
Analyst 99	Authentication	11
	Search <0002798866>	1
	Search <268>	1
	Close request <268>	1
	Register request <239>	1
	Register request <252>	1
	Register request <264>	1
	Register request <268>	1
	Register request <272>	1
	Register request <285>	1
	Register request <321>	1

As presented in Table 3, the *Analyst 99* has performed 11 *Authentication* operations, 2 *Search* operations, 7 operations of *Register request* and 1 operation of *Close request*.
It is considered the indicator number of requests not completed by analyst *k*, $NSNA_k$, defined as follows:

$$NSNA_k = \frac{NSINC_k}{NSINR_k},$$

where:
$NSINC_k$ – number of requests closed by analyst *k*;
$NSINR_k$ – number of requests registered by analyst *k*.

The frequency of realizing an operation by analyst *k*, $FROA_k$, is determined as follows:

$$FROA_k = \frac{NOP_{ik}}{NTOP_k},$$

where:
NOP_{ik} – number of *i* type operations carried out by analyst *k*;
$NTOP_k$ – total number of operations realized by analyst *k*.

The CMS application is used in a commercial bank in Romania to analyze the types of problems faced by users of internet banking service. Having available the database with all customer requests, are determined the strategies to address each client, depending on the history of problems he encountered.
The situation of requests registered on categories during the period February 15 to March 15, 2011, is presented in Table 4:

Table 4. Number of requests on categories during the period February 15 to March 15, 2011

Code	Category	Number of requests
501	Information regarding internet banking access page - 501	87
500	General information regarding internet banking - 500	42
519	Adding accounts - 519	31
513	Supervisors enrollment - 513	24
511	Set template for payment file import - 511	19
516	Payments rejected - 516	18
512	Operators enrollment - 512	18
504	Complaints regarding personal account - 504	16
509	View payment confirmation - 509	10
503	Enquiries regarding payments disbursement - 503	9
508	Unlock PIN token - 508	4
510	Delete payments - 510	4
517	View intraday statement - 517	3
502	Require confirmation of payment status - 502	3
514	Modify payments - 514	2

518	View rejections - 518	2
505	Unlock user at logon - 505	2
507	Modify transaction limits - 507	1
520	E-mail advice - 520	1
521	Reporting malfunction error - 521	1

From the analysis of data in Table 4 results that most requests are registered on the category Information regarding internet banking access page, as users call technical support service for assistance in the application authentication.

The daily volume of data from the database of CMS application, *VDz*, is given by the relationship:

$$VD_z = \sum_{i=1}^{NCAT} NSOL_i ,$$

where:

NCAT – number of problems categories on which requests have been registered;

$NSOL_i$ – number of requests registered on the category *i*.

The volume of data from the database of CMS application over a period of *h* days is:

$VD_{h\_days} = h * VD_z$,

where:

h – number of days;

VD_z – daily volume of data.

Based on data from Table 4 is determined the amount of data added during one month, as sum of the number of requests registered on each category.

Regarding the behavior of users from CMS application, indicators associated with performance criteria are defined to identify the types of operations performed by each analyst and the workstation from which they were made.

For each user are saved in the database the IP address of the computer from which the application was accessed, the browser type, the date and time of the operation. Table 5 presents a summary of actions realized in the application by the user *Analyst 99* during March 10 - May 10, 2011.

Table 5. Number of actions performed by an analyst for each workstation

User	IP address	Date and time	Number of operations
Analyst 99	10.231.0.15	28/04/2011 10:20:39	1
		28/04/2011 10:22:27	1
		29/04/2011 16:08:07	1
	10.231.0.15 Total		**3**
	10.231.0.18	10/03/2011 14:15:04	1
		10/03/2011 14:21:30	1
	10.231.0.18 Total		**2**
	10.231.0.2	07/04/2011 09:48:23	1
		07/04/2011 09:53:09	1
	10.231.0.2 Total		**2**
	10.231.0.21	02/05/2011 22:37:38	1
		03/05/2011 00:03:46	1
		03/05/2011 00:05:33	1
	10.231.0.21 Total		**3**
		Total	**10**

Based on data from Table 5 the following indicators are built:

- maximum number of operations performed on a workstation, *NMAXop*:

$$NMAXop = \max_{i=1}^{npl} \{Nop_i\} ,$$

where:

Nop_i – number of operations performed by the analyst from the workstation i;
npl – total number of workstations from which the analyst accesses the application.

- number of operations performed by an analyst during the month k, $Nopl_k$:

$$Nopl_k = \sum_{j=1}^{npl} noppl_{jk} \, ,$$

where:
$noppl_{ik}$ – number of operations performed at the workstation j in the month k;
npl – total number of workstations from which the analyst accesses the application.

In the case of daily transactions conducted in a bank, all characteristics are related to the workstation, the client account, the beneficiary account and the amount. The transactions are sorted by value. Then are determined the frequencies of amounts traded and the highest frequencies are chosen. The workstation is accessed in order to check if the transactions are made from the same computer. The people who operated the transactions are identified. The beneficiaries of the amounts traded are determined.
From the analysis results the strange situations, meaning that all the transactions are to a certain destination, or all the transactions are conducted from the same workstation. An explanation of these abnormal situations is searched to prevent any attempts of fraud.

4 Genetic algorithm implementation for building and validating metrics

The relative complexity of a collaborative system, CR, is determined as follows:

$$CR = \frac{\sum_{i=1}^{n} (f_i * \log_2 f_i)}{(\sum_{i=1}^{n} f_i) * \log_2 (\sum_{i=1}^{n} f_i)} \, ,$$

where:
f_i – the weight i associated to a quality characteristic of the system, with the property that $f_1 + f_2 + ... + f_n = 1$ and $f_i \in R$.

Collaborative informatics systems from the banking field differ one from each other by complexity of their components. The relative complexity of a banking informatics system, RC, is determined according to the relationship [7]:

$$RC = \frac{NCAI * \log_2 NCAI + NSIB * \log_2 NSIB}{(NCAI + NSIB) * \log_2 (NCAI + NSIB)}$$

,
where:
$NCAI$ – the number of components associated with software applications integrated in the banking informatics system, with the property that $NCAI > 0$ is a natural number, $NCAI \in N$.
$NSIB$ – the number of modules forming the banking informatics system, where $NSIB > 0$ is a natural number, $NSIB \in N$.

In Figure 4 is presented the 3D graphic of the function $RC(x, y) = \frac{x * \log_2 x + y * \log_2 y}{(x + y) * \log_2 (x + y)}$, where $x, y \in N$:

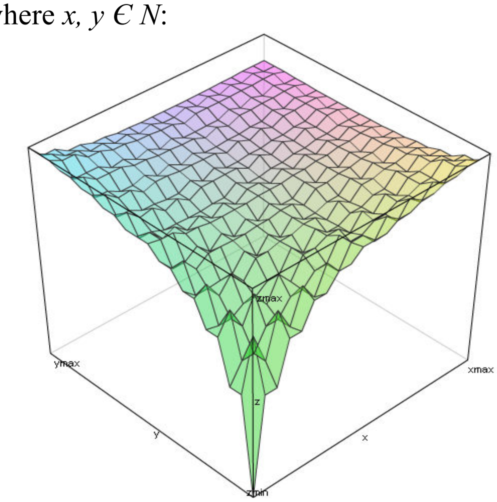

Fig. 4. The 3D graphic of the function *RC()*

A genetic algorithm was implemented in the CMS application, available at http://collaborative.ase.ro/teza, in order to determine the local minimum and maximum values of the $RC(x, y)$ function, where $x, y \in R$. This algorithm objective is to determine the number of modules and components associated with software applications integrated in the banking informatics system for which the relative complexity is minimum or maximum.

The genetic algorithm requires the following steps:

- creating new populations;
- selecting the best two individuals of the population and cross them to obtain children;
- replacing the old population with a new one;
- resumption of the previous steps until it reaches the optimal solution of the problem.

Different genomes are created with elements in the interval 0-1, but the values are improved at each generation, so that the final values obtained are much closer to the values of maximum and minimum points of the function.
In Table 2 are presented the values obtained in three successive generations of the genetic algorithm [8].

Table 2. The values obtained with the genetic algorithm

Number of generations	**Maximum point value**	**Minimum point value**	**RC() function value in the maximum point**
1	0.94760	0.68387	-0.00785
2	0.98459	0.55019	-0.00225
3	0.99667	0.30934	-0.00048

Figure 5 shows the values of maximum and minimum points in the three generations of the algorithm. As seen from the graphic, the maximum point tends near to 1 and the minimum point decrease near to 0.3.

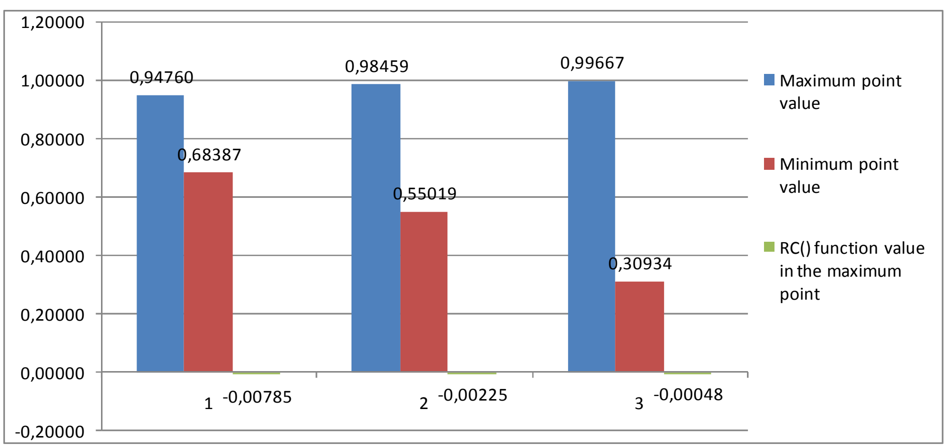

Fig. 5. The values of maximum and minimum points

The local maximum point value, *0.99667*, obtained with the help of the genetic algorithm, shows that the relative complexity of the banking informatics system is maximal when the number of components associated with software applications, *NCAI = 1*, and the number of modules forming the banking informatics system, *NSIB = 1*.
Integration of indicators obtained by genetic algorithms in metrics for collaborative systems help to automate current operations carried out in a banking information system, but also to provide strategic, tactical and operational information required in the decision-making process.

5 Conclusions

The metrics helps to make a quantitative analysis of the collaborative systems from various economic fields. In order to evaluate a collaborative system, several metrics must be defined and analyzed from the point of view of following properties: sensitivity, not compensatory character, not catastrophic character, representativeness.

The existing systems of indicators propose many models of metrics associated to quality characteristics of collaborative systems, presenting the working hypotheses and the description of exogenous variables. The question that arises is to use the appropriate indicators and to achieve significant results.

The real problem is to apply the metric and most important to validate it. The metrics of collaborative systems must be not too complicated, because will use lots of resources when implemented, and not too simple, because the measured levels will lose relevance.

Acknowledgements

This article is a result of the project POSDRU/6/1.5/S/11 „Doctoral Program and PhD Students in the education research and innovation triangle". This project is co funded by European Social Fund through The Sectorial Operational Programme for Human Resources Development 2007-2013, coordinated by The Bucharest Academy of Economic Studies, project no. 7832, Doctoral Program and PhD Students in the education research and innovation triangle, DOC-ECI.

References

[1] S. O. Oguz, A. Kucukyilmaz, T. M. Sezgin and C. Basdogan, "Haptic negotiation and role exchange for collaboration in virtual environments," *2010 IEEE Haptics Symposium*, 25-26 March, 2010, pp. 371-378.

[2] V. Ionescu, *Sisteme liniare*, Academiei Publishing House, Bucharest, 1973, 222 pg.

[3] I. Ivan, *Mediafilie*, Available at: http://www.ionivan.ro/MEDIAFILIE/MIHAI-VITEAZU.htm

[4] C. Ciurea, "Building and Validating Metrics of Collaborative Systems," *Proceedings of the Tenth International Conference on Informatics in Economy, IE 2011*, May 5-7, 2011, Bucharest, Romania, ASE Printing House.

[5] I. Ivan and C. Boja, *Metode statistice în analiza software*, ASE Publishing House, Bucharest, 2004, 341 pg.

[6] C. Ciurea, "Metrics of Collaborative Banking Systems," *Journal of Applied Collaborative Systems*, Vol. 2, No. 1, 2010, pp. 16-23.

[7] C. Ciurea, "Designing Collaborative Informatics Systems in Banking," *Journal of Applied Collaborative Systems*, Vol. 2, No. 4, 2010, pp. 48-55.

[8] C. Ciurea, "Using Genetic Algorithms for Building Metrics of Collaborative Systems," *Informatica Economică Journal*, Vol. 15, No. 1, 2011, pp. 80-91.

Cristian CIUREA is teaching assistant at the Economic Informatics Department from Academy of Economic Studies, Bucharest, Romania. He has a background in computer science and is interested in collaborative systems related issues. Cristian has graduated the Faculty of Economic Cybernetics, Statistics and Informatics from the Bucharest Academy of Economic Studies in 2007 and the Informatics Project Management Master in 2010. He is currently conducting doctoral research in Economic Informatics at the Academy of Economic Studies. Other fields of interest include software metrics, data structures, object oriented programming in C++ and windows applications programming in C#.