ESTIMATION OF SOFTWARE PROJECT RANGE

Stanislava Šimonová[a)], Aleš Hudec[b)]

[a)]University of Pardubice, Faculty of Economics and Administration, [b)]Metering Service s.r.o., Praha

Abstract: *Organizations use information systems for support of their entrepreneur activities, thus as support for fulfilling their entrepreneur goals. Information systems are subject to frequent changes, such changes are realized through software projects. The goal of project management is to ensure required functionality within set time period and with relevant cost. Success rate of SW project is in interest of both customer and supplier, while the supplier can be either external (by means of outsourcing) or internal (own IT department within organization); the customer can also be either external (customer from different organization) or internal (user within organization). The customer decides, based on values determined by estimation, whether to realize SW project and whether to realize the project in given range; i.e. the customer is interested in the project not being overestimated. The supplier of IT services plans project elaborateness according to availability of their capacity, i.e. there is the interest of the supplier in project not being underestimated. The text addresses the topic of designing a method for creating estimation of SW project range by means of expert estimation modification with consideration of historical data.*

Keywords: *Software Product, Software Project, Methods of Estimation of SW Project Range, Expert Estimation.*

JEL Classification: *M11, M15.*

1. Introduction

Various organizational levels of the company requires specific type of information and specific method of information processing, while various data levels are recognized within organization [17] [18]:

- Operational data level: requires processing of data considering routine corporate agenda; information systems react on performance of daily tasks and monitor transaction flow across the organization;
- Knowledge data level: contains not only client applications of corporate information system, but also personal information tools; these applications supports growth of knowledge base of the organization and manage mainly document flow.
- Managing data level: requires information required for fulfilling of administrative tasks and decision-making support,
- Strategic data level: requires systems which support top management in course of identification of long-term trends; managing and strategic level are based mainly on analytical data, or in other words they result from operative data transformed to analytical data.

Information systems or their applications are always aimed at certain data type, on certain group of users or on certain functionality. Changes in conditions related to entrepreneur activities must necessarily reflect the requirements for organization information environment; information systems therefore change / develop according to the needs of the organization.

Changes in information support within an organization, i.e. changes in information systems, are realized by means of software projects [11]. The goal of project management is to ensure required functionality within set time limit and with relevant costs [12] [7].Determination of software project range, especially information about realization time and project costs, is a complex process of creation of estimations in various project phases. A great volume of information which influences characteristics and functions of resulting software product needs to be processed. Estimation of software project is further burdened by unpredictable external events; project team can get an order to prefer other/key project, changes of assumptions on functionality [14]. The price of a SW product is then influenced by, besides range, other aspects, such as usability [9].

Software development is a complex process while a number of factors directly or indirectly influence the success rate of the project. Within The Standish Group company [16] a long-term study has been elaborated which addresses the evaluation of software projects success rate (it is about IT projects realized in the USA and information about projects is obtained by means of interviews or workshops with project managers); results are published annually within the CHAOS Report [5]. It is clear from the results that successful projects constitute approximately one third of all projects; these are projects finished in planned time, with planned costs and where planned goals were satisfied. Two thirds of projects are either challenged (project was finished with usable results, but time schedule was not fulfilled, costs were higher than planned or not all requested functionalities were implemented) or failed (project was either terminated before finishing or its results were never implemented).

Observation of problem of success rate of software projects is in the best interest of both customer and supplier, while the supplier can be external (by means of outsourcing) or internal (own IT department within the organization), and, consequently the customer can be either external (customer from different organization) or internal (user within the organization). For the supplier the estimation of software project is a tool for HR planning. The customer, on the other hand, is interested in delivery date or purchase price. The text is aimed at estimation possibilities performed at the start of a SW project.

2. Statement of a problem

SW project is based on methods or methodologies of information system development. These methods can be classified for example by range, scope, detail level of the method, development approach and so on. [2] [8] [18]. One of those is the RUP (Rational Unified Process) [10], where the iterative development, visual modeling, object approach and communication is stressed; for modeling and documentation purposes the UML standard [6] is applied. Within the RUP a SW project goes through following phases – Inception (with the aim to define boundaries of the project and to identify key requirements for finite product), Elaboration (specification of requirements for purpose of analysis and design), Construction (programming itself, creation of database structure and GUI), Transition (transition of system from development environment to customer's production environment; final testing and installation).

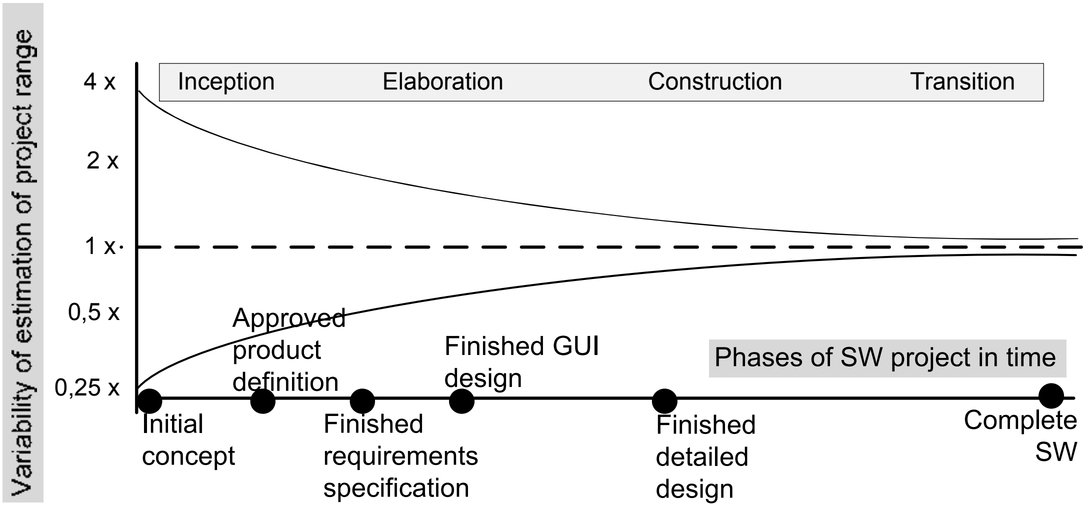

Fig. 1: Uncertainty cone expresses variability (increasing precision) in estimation of SW project within SW project steps (work, costs, characteristics)

Source: [14], [16]; *own adaptation*

Quick and precise estimation of SW project is an important piece of information for the purpose of deciding about realization of SW project and consequently it influences success rate of whole project. Quickness and precision of estimation of SW project are mutually conflicting. To perform estimation the fastest possible, i.e. in the beginning of a project, means to use minimal amount of information. The user defines input requirements and his/her requests for information system, but these requests tend to be further specified in more details or even change. That way elements of further phases of information system development change. That means that initial estimation is very much burdened with all the changes made, which will manifest gradually. This situation is expressed in figure 1.

The most exact "estimation" can be done at the end of the project, when all information about work realized and costs within the project are at disposal. Estimation performed in any other phase of the project development of SW project can differ from reality for example two times (real costs are in the graph expressed by hatchet line with description "1 x"). The user and supplier are both interested in estimation performed at the beginning of the SW project, but there the risk of insufficient amount of information is the greatest, which results in deviation of estimation from real and finite costs and time consumed – in sense of either overestimation or underestimation. Various estimation methods are suitable for various development phases of a SW project.

2.1 Estimation methods

Methods of software estimations can be divided into two main groups; the first group consists of methods based on experience and comparison while the other group is based on usage of “historical data“ by means of mathematic models. Utilization of these methods can be measured by various aspects [14]:

- estimated characteristics: for instance volume, extent of work,
- project volume: small projects (up to five workers, duration time in weeks or months), middle sized projects (5 to 25 workers, duration time from 3 to 12 months),large projects (over25 workers, duration time 6 to 12 months, possibly even longer),
- development phases: stems from used method for software development,
- development phases: gradual (required finishing one phase before proceeding to next one) or iterative (phases could be repeated several times),
- reachable precision: for example classification to low, average and high.

Suitable methods for estimation of the range of a SW project are considered the following ones– Expert Estimation, Analogy, Use Case Point and Functional Points method [14].

EXPERT ESTIMATION METHOD is based on experience of an expert. It is a relatively simple method. Subjectivity of the method can be eliminated by participating of more experts. The expert estimation is suitable for smaller projects and for situations when there are no historical data about previous projects within organization. The ANALOGY METHOD is a modification of the expert estimation method by information about projects created in the past.

USE CASE POINT METHOD (UCP) evaluates project by points. Determination of points results from use case model and from classification of points by their complexity. Points obtained are then multiplied by the number of hours which are required in order to process one point while the recommended value is 20 hours per one point [3]. Calculations are based on four variables - Technical Complexity Factor (TCF), Environment Complexity Factor (ECF), Unadjusted Use Case Points (UUCP) and Productivity Factor (PF).

$$UCP = TCP * ECF * UUCP * PF \quad (1)$$

The process of the method consists of following steps – Determine and compute the Technical Factors, Determine and compute the Environmental Factors, Compute the Unadjusted Use Case Points, Determine the Productivity Factor, Compute the product of the variables. The method is suitable for any project size, necessary ground are customer defined requirements within use case models.

FUNCTIONAL POINTS METHOD (FP) measures the range of SW projects by the number of application functions and amount of data; the method requires detailed analysis. Transaction functions are related to - external inputs (EI), external outputs (EO) and external enquiry (EQ); data functions are related to internal logical files (ILF) and external interface files (EIF).Each group is assigned with weight (W) and the number of elements (N) in group is considered. Based on these values the Unadjusted Function Point Count (UFPC) is determined.

$$UFPC = W_{EI} * N_{EI} + W_{EO} * N_{EO} + W_{EQ} * N_{EQ} + W_{ILF} * N_{ILF} + W_{EIF} * N_{EIF} \quad (2)$$

Next, general system characteristics are valuated (reliability, accessibility, response time and others) and this valuation would reflect the UFPC; from this process we will obtain finite function points. Then it is necessary to determine price and elaborateness of one function point; the result is then total cost and total elaborateness of SW project.

Further methods are based on the source code volume. The method called The Constructive Cost Model (COCOMO) [4] can serve as an example. Nevertheless, the subject matter of this text is estimation performed in initial phases of SW project, when source code does not exist yet.

2.2 Requirements of customers

Requirement is a concrete verifiable behavioural function of the system which is defined and verified within Service Level Agreement (SLA). Basic characteristics of the requirement are verifiability and feasibility; analysis of requirements is done in order to comprehend subjective problem of a customer. Specified requirements are a presumption for unification of customer's and supplier's conception of a SW project. Analyst, based on available information, defines the target of the project, functional and non-functional system requirements (functional requirement is system functionality; non-functional requirement is for example safety issue). In order to gain and specify requirements, interviews with customers are done, prototypes are created and GUI is modeled. Key role in specification of customer's requirements and also risk of SW quality related to that is expressed in figure 2.

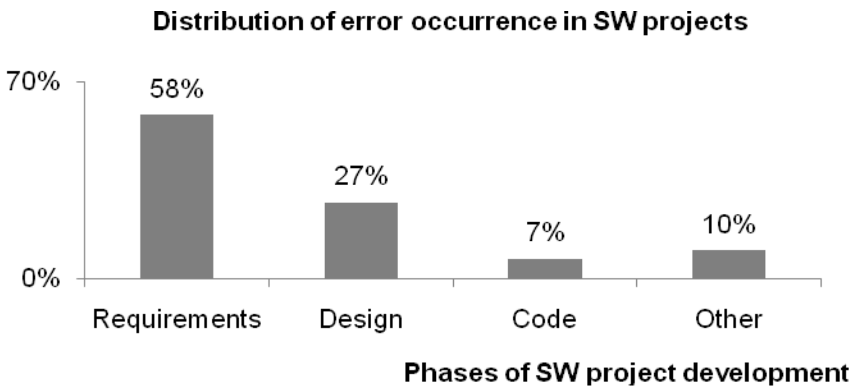

Fig. 2: Customer's requirements as main sources of errors in SW projects

Source: [1], [13]; *own adaptation*

The graph in the picture stresses out the risk that more than a half of errors in SW projects are a result of inaccurately specified and analyzed customer requirements. Causes can be – incorrect, confused or ambiguous requirements, eventually omitted requirements.

3. Problem solving

The aim was to determine a method for creation of elaborateness estimations of SW projects for selected area. The method should serve to maximize precision of estimations performed in the initial phases of the SW project. Following initial conditions are stated:

- organization represents suppliers of application services, which are determined for external customer (customer uses SW services by means of outsourcing); these application services are implemented through software projects (in this case there are twelve software projects with the names Project A to Project L); organization is in "common" situation, i.e. it is a highly expert development team and consequently there are topicality and range reserves in documentation,
- historical data – data about estimations – (estimations within categories – analysis, design, programming, testing, documentation) is archived and it is, therefore, possible to confront them with information about real project range (see figure 3);
- the estimation of the range in the organization is performed by means of subjective estimation of an expert and the value of the estimation is a significant part of dealing with a customer; cooperation between supplier and customer is then lead within SLA.

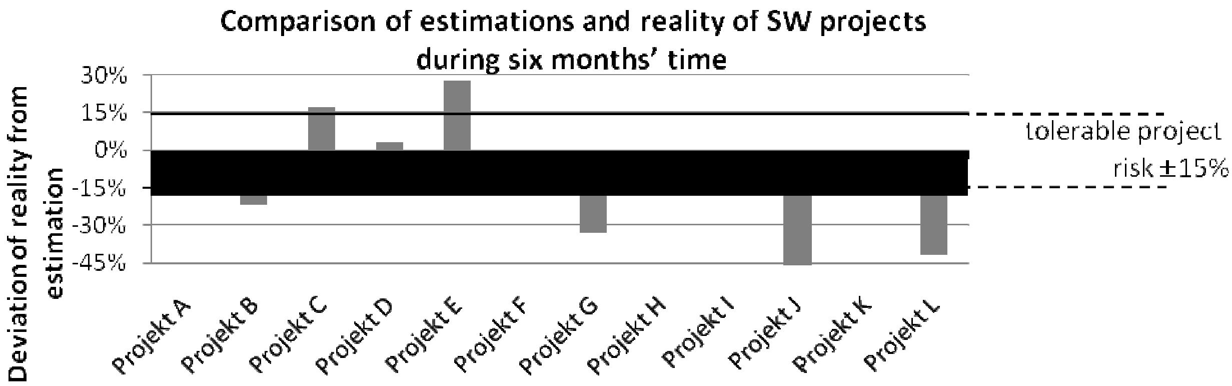

Fig. 3: Deviation of real ranges of SW projects from initial estimation

Source: (own adaptation)

The graph in figure 3 expresses a comparison of elaborateness estimation with real elaborateness, while SW projects within six-month period were compared. A tolerance of ±15 % is used in SW project estimation; it is calculated as project risk. It is obvious from the graph that only three projects (25 % from total number) within the monitored time period were in the ±15 % tolerance (D, F, K), seven projects were underestimated (A, B, G, M, I, J, L) under the tolerance level and two projects were overestimated over the tolerance level (C, E). That means that only in 25 % of projects the reality adhered with the estimation (within tolerated project risk) and 75 % had incorrect estimation, while there was tendency of underestimating. In the monitored organization the method of expert estimation was used for estimation. From monitoring of projects performed (see figure 3) it is obvious that projects tend to be underestimated (58 % from total of 12 projects monitored).It is therefore necessary to specify a SW project.

Table 1 compares utilization of the methods considered – expert estimations, use case point and function point. On one hand there is a request for utilization of objective methods, on the other hand there is a request to perform estimation in initial phase of SW project. Because of the second request, utilization of UCP and FP methods seems inappropriate, since in order to efficient application analysis has to be performed beforehand; use case models for UCP method and functional requirements models for FP method.

Tab. 1: Comparison of usability of methods

	Expert estimate	**UCP**	**FP**
Data about previous projects	Does not use	Uses	Uses
Application within SW project	Possible at the beginning	Not possible at the beginning, because use case models have to be done	Not possible at the beginning, because it is necessary to have functional requirements models elaborated
Disadvantages / requirements	Subjective estimation	Partial analysis has to be performed	Detailed analysis has to be performed

Source: (own adaptation)

Design of modified estimation method is based on following reflections:

- to specify the estimation method mainly on the fact that present method led in most of projects to underestimation of elaborateness,
- base estimation of range on calculation according to the PERT method (Program Evaluation and Review Technique),
- method extended by utilization of historical data (data about previous projects) which would ensure correctness of expert estimations.

The starting point is the elaborateness estimation according to the PERT method calculation [15]:

$$T_E = (T_O + 4 * T_M + T_P)/6 \text{, where} \quad (3)$$

T_E is estimated time necessary for activity, T_O is optimistic estimate of activity time, T_M is the most probable estimate of activity time, T_P is pessimistic estimation of activity time. Classification of following parameters was designed (ranges of elaborateness in man-days):

- R: value of real range of given type of activity (gained from historical data),
- R_{AVG}: average value of real range of given type of activity (gained from historical data),
- E_{MIN}: estimate of the best (the smallest) value of given activity type range,
- E_{MAX}: estimate of the worst (the greatest) value of given activity type range,
- E_{EXP}: estimate of expected activity range.

- T: resulting extent of the project (in T_A, T_B, T_C alternatives – see below for description).

The calculation of elaborateness estimate will not exist in three alternatives.

The first alternative is calculation according to PERT.

The second alternative is modified calculation with consideration of R_{AVG} value.

The third alternative is modified calculation, where R_{AVG} value is not available; in such case influence of E_{MAX} value is "reinforced" – mainly in order to eliminate tendency of underestimation of project range:

$$T_A = (E_{MIN} + 4 * E_{EXP} + E_{MAX})/6 \qquad (4)$$

R_{AVG} value is available:

$$T_B = (E_{MIN} + 3 * E_{EXP} + R_{AVG} + E_{MAX})/6 \qquad (5)$$

R_{AVG} value is not available:

$$T_C = (E_{MIN} + 2 * E_{EXP} + 2 * E_{MAX})/6 \qquad (6)$$

The calculation method was the first verified on previous / realized software projects (projects A to L), see table 2. Subjective / initial estimation of elaborateness was determined by count of partial estimations on these SW project phases – analysis and design, implementation (programming), testing, user acceptance tests and project management. It is apparent from the table that all three types of calculations (T_A, T_B, T_C) have brought improvement of elaborateness estimation against reality while the most precise alternative was T_C. From the viewpoint of tolerable project risk±15 % it can be stated that alternative T_C did not fulfill this tolerance only in one case (project J), alternative T_B did not fulfill this tolerance in three cases (projects G, I, J).

Tab. 2: Deviations of new estimation determination of elaborateness against real range

Software projects	A	B	C	D	E	F	G	H	I	J	K	L	Average value (absolute values)
R (real elaborateness, man-day)	45	78	48	33	72	121	236	247	298	162	42	85	
S (subjective estimation of elaborateness, man-day)	38	64	57	34	100	112	178	210	255	110	39	60	
T_A (manday)	40	69,8	42,2	29,8	71	108	200	227	247	125	37,3	81,9	
T_B (manday)	43,6	68	44,3	31,2	74,9	110	202	227	254	127	39,1	85,3	
T_C (manday)	44,5	73,3	45,8	31,5	76,8	117	213	241	265	134	41,3	88,4	
Deviation of S from R	-18%	-22%	16%	3%	28%	-8%	-33%	-18%	-17%	-47%	-8%	-42%	18%
Deviation T_A from R	-13%	-12%	-14%	-11%	-1%	-12%	-18%	-9%	-21%	-30%	-13%	-4%	13%
Deviation T_B from R	-3%	-15%	-8%	-6%	4%	-10%	-17%	-9%	-17%	-28%	-7%	0%	10%
Deviation T_C from R	-1%	-6%	-5%	-5%	6%	-3%	-11%	-2%	-12%	-21%	-2%	4%	6%

Source: (own adaptation)

After the verification on realized projects, the process was applied on new projects. For collection of estimation it was it was stated that the expert estimation value will be determined as first and subsequently estimations by new process will be created. The reason for that is to avoid affecting expert estimation. Practice had shown that when determining values E_{MIN} and E_{MAX}, workers consider possible effects on the given project. Results are shown in table 3. It is obvious from the values that all three calculation methods (T_A, T_B, Tc) meant more precise estimations even in further verification on new projects (projects M to Z); while the most precise was again alternative Tc. When considering ±15 % tolerated project risk, we can state that all three alternatives (T_A, T_B, Tc) fulfilled this tolerance in all projects (M to Z).

Tab. 3: Deviations of new elaborateness estimation determination (alternatives T_A, T_B, T_C) from real project range; applied on new projects (projects M to Z)

	Software projects													Average value (absolute values)
	M	N	O	P	Q	R	S	T	U	V	X	Y	Z	
	Values in %													
Deviation S from R	-3	-15	-18	10	11	-7	0	14	-5	-12	-13	-8	13	9%
Deviation T_A from R	-8	-13	-10	5	7	-3	4	4	-7	-8	-8	-8	-7	6%
Deviation T_B from R	-6	-9	-2	11	13	-3	13	5	-8	-9	-7	-7	0	7%
Deviation T_C from R	-2	-5	-1	10	15	3	12	9	-2	-1	0	0	-1	5%

Source: (own adaptation)

4. Conclusion

Estimation of a SW product elaborateness is an important piece of information not only for the customer, but also for the supplier. In environment of a supplier IT company there is an efficient IT team with experience in development of SW products; many SW products repeat (customers ask for similar products), which contributes to more qualified estimations of "usual" SW products. In such environment it is usual to use the expert estimation method, which is based mainly on experience. In the monitored organization during analysis of historical data from previous projects we could state that most of SW projects were underestimated, i.e. the real time period and costs of SW projects were higher for the supplier than the customer paid for according to the signed contract. It was caused by certain routine in estimations when creators of the estimation did not consider other / limiting factors that much.

For design of new process following was stated following; the organization has its own IT team; historical data about previous projects are available, current estimations of experts mostly led to underestimation of range. That is why requirements were determined – to create method for specification of estimations, to utilize historical data, to eliminate tendency of project range underestimation. Three methods of calculation were determined; the first method (T_A) results from the PERT method; the second method (T_B) extends this calculation by value gained from historical data; the third method (T_C) extends this calculation by enforcing higher valuation of range. These three methods were confronted with currently performed expert estimation and with real SW project range. All three calculation alternatives were more precise than expert estimation. It is necessary to state that even expert estimation had increased its precision, while its creators started to consider various limiting factors. It is apparent from values gained that the most precise values are obtained from alternatives T_B and T_C. Alternative T_C is suitable for such organizations, where current expert estimations showed often underestimation under tolerated -15 % in most of SW projects. Alternative T_B is suitable for organizations, where current expert estimations showed "equal" underestimation and overestimation over ±15 % tolerance.

References

[1] BORLAND. *Testování založené na požadavcích* [online]. [cit. 2010-06-28]. Available from WWW <http://www.borland.cz/resource/requirements_based_testing.html>

[2] BUCHALCEVOVÁ, A. *Metodiky vývoje a údržby informačních systémů*. Praha, Grada, 2005. 164 s. ISBN 80-247-1075-7.

[3] CLEMMONS, R. *Project Estimation with Use Case Points*. [online]. [cit. 2010-06-08]. Available from WWW <http://www.codeproject.com/KB/architecture/usecasep.aspx>

[4] COCOMO II - *Constructive Cost Model* [online]. [cit. 2010-06-02]. Available from WWW <http://sunset.usc.edu/csse/research/COCOMOII/cocomo_main.html>

[5] DOMINGUEZ, J. *The Curious Case of the CHAOS Report 2009* [online]. [cit. 2010-06-01]. Available from WWW < http://www.projectsmart.co.uk/the-curious-case-of-the-chaos-report-2009.html >

[6] ERIKSSON, H. et al. *UML 2*. Indiana: Wiley Publishing, 2004. 511 p. ISBN 0-471-46361-2.

[7] GÁLA, L., POUR, J., TOMAN, P. *Podniková informatika*. Praha: Grada, 2006. 484 s. ISBN 80-247-1278-4.

[8] HOFFER, J., GEORGIE, J., VALACICH, J. Modern Systems Analysis and Design. New Jersey: Prentice Hall, 2004. 683 s., ISBN 0-13-145461-7.

[9] HUB, M., ZATLOUKAL, M. Towards Establishing a Score of Usability Evaluation. In E+M: Economics and Management, 2009, roč. 12, č. 2, s. 156-168. ISSN 1212-3609.

[10] IBM. *The Rational Unified Process (RUP)* [online]. [cit. 2010-06-07]. Available from WWW <http:// www.01.ibm.com/ software/awdtools/rup/>

[11] JIRAVA, P. System development life cycle. In Scientific Papers of the University of Pardubice Series D. Pardubice: Univerzita Pardubice, 2004. P. 118-125, ISSN 1211-555X.

[12] KIMLIČKA, Š. *Informačné systémy: teoretické východiská, princípy, metódy projektovania*. Martin : SNK, 2006. 116 p. ISBN 80-89023-88-6.

[13] MARTIN, J. *An Information Systems Manifesto*. Prentice Hall, 1984. 300 p. ISBN 9780134647692.

[14] McCONNELL, S. *Odhadování softwarových projektů*. Brno: Computer Press, 2006. 320 p. ISBN 80-251-1240-3

[15] NĚMEC, V. *Projektový management*. Praha : Grada Publishing, 2002. 182 s. ISBN 80-247-0392-0.

[16] PERGL, R. *Metody odhadu pracnosti založené na modelu*. In *Konference ČSSI*. Ostrava: ČSSI, 2006, s. 1-8.

[17] SODOMKA, P. Informační systémy v podnikové praxi. Brno: Computer Press, 2006. 351 s. ISBN 80-251-1200-4.

[18] STAIR, R., REYNOLDS, G. Principles of Information Systems. Boston: Thomson Learning, 2006. 758 s. ISBN 0-619-21525-9.

[16] *The Standish Group* [online]. [cit. 2010-06-02]. Available from WWW < http://www1.standishgroup.com/ about/index.php >.

Contact Address

Ing. Stanislava Šimonová, Ph.D.
University of Pardubice, Faculty of Public Administration, Institute of System Engineering and Informatics
Studentská 84, 532 10 Pardubice, Czech Republic
E-mail: Stanislava.Simonova@upce.cz
Phone number: +420 466 036 009

Aleš Hudec
Metering Service s.r.o.
Plzeňská 3185, 150 00 Praha 5, Czech Republic
E-mail: ales.hudec@landisgyr.com
Phone number.: +420 606 620 648