Plecka, Przemysław; Bzdyra, Krzysztof

Article

The software cost estimation method based on fuzzy ontology

Foundations of Management

Provided in Cooperation with:
Faculty of Management, Warsaw University of Technology

Suggested Citation: Plecka, Przemysław; Bzdyra, Krzysztof (2014) : The software cost estimation method based on fuzzy ontology, Foundations of Management, ISSN 2300-5661, De Gruyter, Warsaw, Vol. 6, Iss. 2, pp. 21-30, 
http://dx.doi.org/10.1515/fman-2015-0008

This Version is available at:
http://hdl.handle.net/10419/184572

Terms of use:
Documents in EconStor may be saved and copied for your personal and scholarly purposes.
You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.
If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.
https://creativecommons.org/licenses/by-nc-nd/3.0
THE SOFTWARE COST ESTIMATION METHOD BASED ON FUZZY ONTOLOGY

Przemysław PLECKA*, Krzysztof BZDYRA**

Faculty of Electronics and Information Technologies, Koszalin University of Technology, Poland

*e-mail: przemek.plecka@gmail.com
**e-mail: krzysztof.bzdyra@tu.koszalin.pl

Abstract: In the course of sales process of Enterprise Resource Planning (ERP) Systems, it turns out that the standard system must be extended or changed (modified) according to specific customer’s requirements. Therefore, suppliers face the problem of determining the cost of additional works. Most methods of cost estimation bring satisfactory results only at the stage of pre-implementation analysis. However, suppliers need to know the estimated cost as early as at the stage of trade talks. During contract negotiations, they expect not only the information about the costs of works, but also about the risk of exceeding these costs or about the margin of safety. One method that gives more accurate results at the stage of trade talks is the method based on the ontology of implementation costs. This paper proposes modification of the method involving the use of fuzzy attributes, classes, instances and relations in the ontology. The result provides not only the information about the value of work, but also about the minimum and maximum expected cost, and the most likely range of costs. This solution allows suppliers to effectively negotiate the contract and increase the chances of successful completion of the project.

Keywords: ERP, software cost estimation, metrics, fuzzy ontology, implementation, specification of requirements.

1 Introduction

First Enterprise Resource Planning (ERP)-class information systems were available only for very large companies. What appeared to be a barrier in their dissemination was the high cost of implementation and maintenance, due to the fact that software providers created tailor-made solutions for specific customers. Gaining experience from working with many clients, producers identified a set of functionalities that were frequently ordered and created a standard version of their products. The costs of implementation and maintenance were reduced so much that even medium-sized companies could afford such systems [1]. However, standardization causes a situation in which there are groups of business processes that are not supported by a given information system (IS). Hence, there is a need to adapt the system to the customer's company. The cost of such adaptation (modification) increases the value of the contract (implementation). A study by Standish Group [2] shows that in 2012, only 39% of information technology (IT) projects ended successfully. In the rest, the budget or the implementation deadline has to be extended or not all of customer’s requirements will be met. The cause of these failures is incorrect cost estimates prior to the onset of the project. Suppliers expect a more complete knowledge on the costs during contract negotiations. Information about the estimated value of the work is insufficient.

On this basis, the supplier is not able to assess whether the value of the contract may be reduced by, for example, 5 or 25% under pressure from the client, or a higher value of costs should be negotiated due to high risk of underestimation.

The methods that help to estimate software production costs are known and described in the literature, for example, by McConell [3]. Due to changes in IT and the conditions in IS market, the popularity of these methods is constantly changing. The use of algorithmic methods (Function Point Analysis, COCOMO) in the initial stages of projects is difficult. At this stage, the project documentation containing data necessary for estimating algorithms is still non-existent. Although examples of the use of these methods in early stages of the projects can be found in the literature [4, 5], suppliers’ experience shows that non-algorithmic methods are applied more frequently (Expert Evaluation, Estimation by Analogy) as faster, less-expensive and easier-to-implement solutions. Each of these methods provides only an estimated value of the cost. Additional information, which providers demand, can be obtained using the fuzzy versions of some methods. In the literature, one can find works on fuzzy COCOMO method, cf. Fei [6] and Attarzadeh [7] or fuzzy Function Point Analysis, c.f.: Xu [8] and Lima [9]. The disadvantage of the fuzzy version, as well as the original ones, is their low accuracy in the early stages of implementation.
The authors’ study concerns the search for an alternative method of estimating the cost at the stage of trade talks and pre-implementation analysis. The expected results are not worse than those that can be obtained using the existing methods, and additional information about the possibility of exceeding the cost (underestimation) or the possibility of a cheaper implementation (revaluation) is provided.

Section 1 describes previous research that led to the development of the method based on the ontology of implementation costs and verification of its usefulness which guided towards further exploration on the use of fuzzy attributes, classes, instances, and relationships. The section 2 defines the problem of fuzzing the component attributes of ontology implementation costs and reduces it to a specific class of ERP systems. Section 3 is a quick overview of knowledge about the use of fuzzy ontologies. While section 4 presents an improved method of estimating the costs by focusing on aspects related to the fact of fuzzy attributes. Next, an example of the application of the method, conclusions and suggestions for further research directions are presented.

2 Previous studies

In the course of the study, the efficiency of the already-known methods of estimating the cost of implementation based on previous projects was verified [10]. Significant errors were noted resulting from incorrect selection of methods to the level of data quality (requirements specification) obtained from the client. In the next step [11], an algorithm of selection method, according to the type and quality of information obtained from the client, was proposed. It guaranteed achieving higher accuracy of estimates than it was in the cases studied. However, this solution preferred algorithmic methods (Function Point Analysis, COSMIC Full Function Point, COCOMO), which require expertise from suppliers (and are more expensive) and can be used only from the pre-implementation analysis stage. For this reason, a method that can be used at the stage of trade talks was searched for. The use of business process model was proposed as the basis of recording knowledge on the IS and the customer's requirements [12]. The use of the BPMN standard (Business Process Model and Notation) [13] for this purpose allows for earlier dimensioning; it is easy to use and requires no technical knowledge. An additional advantage is the possibility to automate estimations of recording the model processes with the metrics done in Business Process Modelling Language (BPML). The disadvantage of this solution is a limited scope of the representation of data structures (entities). Therefore, another method of dimensioning was proposed—dimensioning based on the metrics included in the ontology of implementation costs (SICO - Software Implementation Cost Ontology) [14]. Ontological cost model allows for the representation of both data structures and processes. The low quality of the information obtained from the client in the initial stages of implementation hardly affects the accuracy of the estimation. The supplier prepares IS ontology and completes it gradually with the experience gained from successive implementations. Each occurrence and relationship in the ontology is assigned metrics relating to the costs, for example: cost of training, cost of parameterization, cost of adding, etc. Such an ontology is the reference ontology in relation to the ontologies created on its basis. During trade talks and pre-implementation analysis, the supplier modifies the reference ontology based on customer's requirements relating to the IS. In the next step, the difference between ontology containing additional information about customer's requirements and the reference ontology is determined. The result is the collection of unrelated objects, such as classes, relationships and instances. Then the record of differences is converted into a serial format (serialization) and grammatical structure is analyzed (parsing). These processes lead to the selection and aggregation of the values of the attributes relevant to the costs of implementation. In other words, it is assumed that in the context of trade talks, it is sufficient to specify and implement changes resulting from customer’s requirements into a well-known ontology to be able to determine the cost of a particular implementation on this basis. The steps of estimating costs with the SICO method are shown in Fig. 1.

At the stage of negotiation of the contract, the supplier would expect more complete information about the outcome of the estimation. If, for example, by estimating the costs, the supplier receives a value of 50,000 PLN, the value is increased by the risk margin, for example, 20%, determined on the basis of their experience from previous implementations, and negotiates with the client the value of the contract, e.g. 60,000 PLN.
If the negotiated amount is too small, the supplier loses; if it is too high, he will not be competitive with other suppliers and will lose the contract. Information on the most optimistic (minimum), the most likely and the most pessimistic (maximum) value of costs will allow the supplier negotiate a better contract. The proposed solution is the modified SICO method using fuzzy of attribute values in the ontology (fuzzy SICO). As a result, the supplier receives three or four values of the costs, for example:

- minimum value: 40,000 PLN,
- most-likely value: 50,000 PLN,
- maximum value: 60,000 PLN.

In this case, accepting the small risk of exceeding costs, the supplier negotiates a contract with a value of 55,000 PLN. He is more competitive then other suppliers negotiating with the customer. And if in other case, he receives the result of the estimation according to fuzzy SICO method, which is different from the previous result by only the maximum value - 75,000 PLN, the supplier still negotiates the amount of 55,000 PLN but accepts the high risk of cost exceeding. In such a case, the value of contract is not lower than 60,000 PLN.

Such deliberate negotiations with the supplier are possible when they are based on fuzzy estimation results.

### 3 Problem definition

It is assumed that the supplier has an ERP-class IS (Information System) with a specific set of functionalities and experience with the cost of implementation of that system. Supplier’s knowledge is stored in the ontology, where implementation costs are represented by fuzzy attribute values of concepts, instances and relationships. The requirements of the medium-sized company concerning the expected functionality of the implemented IS are known. On the basis of these requirements, the supplier changes the components of the ontology of implementation costs. The method of estimating the costs of implementation based on SICO ontology is also known. The question is whether the use of fuzzy attributes in the ontology of implementation costs gives estimation results not worse than using “sharp” attribute values. During the evaluation, the results of proposed method are compared with the results obtained using the method of “acute” - SICO and other well-known methods. The problem is limited to a specific class of medium-sized companies and their respective ERP systems.
4 Knowledge on ontologies in IT and the use of fuzzy ontology

A few examples of the idea of using fuzzy attributes of the components of the ontology and practical examples of their use can be found in the literature. For example, D. Parry proposed to use fuzzy ontologies for searching through medical records [15]. On the other hand, C. Lee [16] presented a review of the mechanisms of fuzzy ontologies. R. Lau presented the use of fuzzy ontology in e-learning, particularly in assessing the progress of learners [17]. A method for automatic extraction of attributes of concepts, leading to the automatic creation of ontologies was proposed by G. Cui [18]. Examples of the use of fuzzy ontology in decision-making were described by C. Carson [19]. In contrast, P. Alexopoulos proposed a method to convert an “acute” ontology in a fuzzy one [20]. Examples of the use of ontologies in the area of IT can be found in the work of C. Orłowski and A. Czarnecki for evaluating software [21] and the use of ontologies for modeling requirements [22].

5 An improved method for estimating the cost of implementation - fuzzy SICO

In the most popular and widely used method for estimating the costs by an expert or a group of experts [3], the problem is disordered knowledge of these experts. In the process of estimating, the experts “recall” from memory their own experience and “fit” it into current requirements. In this situation, it happens frequently that a considerable experience is omitted or used in the wrong context.

To prevent this, expert knowledge about the costs of implementation is attributed to relevant concepts (classes), instances and relations in the ontology [14]. For example, on the basis of his experience, the expert estimated that the addition of service document “consignment note” into the IS generates the cost $a$ and the user training in this area generates the cost $b$. These amounts are the values of the attributes of the concept of "consignment note": the cost of the addition and the cost of training. This procedure is known from SICO methods. During the subsequent implementation, experts notice that the cost of adding the document “consignment note” does not always equal $a$. There are implementations when the cost is a minimum amount of $a_1$, $a_2$, is the most frequent amount, but there are cases with the maximum amount of $a_3$ (case 1). Or there are cases of another IS, when adding the document “consignment note” generates a minimum amount of $a_1$, the most frequent is the amount between $a_2$ and $a_3$, but there are cases that the maximum amount of $a_4$ (case 2) is generated. Such experiences of experts suggest a way of representing the cost of implementation using a fuzzy model, which allows mirroring phenomena and concepts of ambiguous nature. The reason for fuzzing the attributes of the component ontology is the necessity to re-determine the method of estimating fuzzy SICO:

- representation of these attributes,
- method of propagation of attributes onto the superior components,
- methods of summing attributes.

In the theory of fuzzy sets, the degree of membership is expressed as a real number from the interval $[0, 1]$. A fuzzy set $A$ is defined on a space $X$ then the set:

$$A = \{x, \mu(x) : x \in X \},$$

where $\mu: X \rightarrow [0, 1]$ is the membership function of fuzzy set $A$. The value $\mu_A(x)$ is the number from the interval $[0, 1]$, and it is called the degree of membership of $x$ element to the set $A$.

The element $x$ belongs to the set $A$, if $\mu_A(x) = 1$, and it does not, when $\mu_A(x) = 0$. Between full membership and lack of membership, there is a smooth transition in the form of partial membership, whose degree is determined by the number of the range $[0, 1]$.

Membership function that describes the collection of a fuzzy set has most frequently a triangular shape (case 1), trapeze (case 2), as shown in Fig. 2. In order to record membership function in the ontology, in particular fuzzy SICO, characteristic values can be represented by a vector $M = [a_1, a_2, a_3, a_4]$, where $a_1 \leq a_2 \leq a_3 \leq a_4 \in R$. In particular cases, it may possible that: $a_2 = a_3$ or $a_2 = a_4$, or $a_3 = a_4$, which has influence only on the shape of membership function.

In the most popular and widely used method for estimating the costs by an expert or a group of experts [3], the problem is disordered knowledge of these experts. In the process of estimating, the experts “recall” from memory their own experience and “fit” it into current requirements. In this situation, it happens frequently that a considerable experience is omitted or used in the wrong context.

To prevent this, expert knowledge about the costs of implementation is attributed to relevant concepts (classes), instances and relations in the ontology [14]. For example, on the basis of his experience, the expert estimated that the addition of service document “consignment note” into the IS generates the cost $a$ and the user training in this area generates the cost $b$. These amounts are the values of the attributes of the concept of "consignment note": the cost of the addition and the cost of training. This procedure is known from SICO methods. During the subsequent implementation, experts notice that the cost of adding the document “consignment note” does not always equal $a$. There are implementations when the cost is a minimum amount of $a_1$, $a_2$, is the most frequent amount, but there are cases with the maximum amount of $a_3$ (case 1). Or there are cases of another IS, when adding the document “consignment note” generates a minimum amount of $a_1$, the most frequent is the amount between $a_2$ and $a_3$, but there are cases that the maximum amount of $a_4$ (case 2) is generated. Such experiences of experts suggest a way of representing the cost of implementation using a fuzzy model, which allows mirroring phenomena and concepts of ambiguous nature. The reason for fuzzing the attributes of the component ontology is the necessity to re-determine the method of estimating fuzzy SICO:

- representation of these attributes,
- method of propagation of attributes onto the superior components,
- methods of summing attributes.

In the theory of fuzzy sets, the degree of membership is expressed as a real number from the interval $[0, 1]$. A fuzzy set $A$ is defined on a space $X$ then the set:

$$A = \{x, \mu(x) : x \in X \},$$

where $\mu: X \rightarrow [0, 1]$ is the membership function of fuzzy set $A$. The value $\mu_A(x)$ is the number from the interval $[0, 1]$, and it is called the degree of membership of $x$ element to the set $A$.

The element $x$ belongs to the set $A$, if $\mu_A(x) = 1$, and it does not, when $\mu_A(x) = 0$. Between full membership and lack of membership, there is a smooth transition in the form of partial membership, whose degree is determined by the number of the range $[0, 1]$.

Membership function that describes the collection of a fuzzy set has most frequently a triangular shape (case 1), trapeze (case 2), as shown in Fig. 2. In order to record membership function in the ontology, in particular fuzzy SICO, characteristic values can be represented by a vector $M = [a_1, a_2, a_3, a_4]$, where $a_1 \leq a_2 \leq a_3 \leq a_4 \in R$. In particular cases, it may possible that: $a_2 = a_3$ or $a_2 = a_4$, or $a_3 = a_4$, which has influence only on the shape of membership function.
At the first stage, the supplier builds the ontology of IS, attributes expert knowledge on various categories of costs (adding, deleting, parameterization, training) to appropriate relationships and instances. Then the information is transferred towards the superior concepts (propagation). In the method of fuzzy SICO, the operation of calculating the value of a superior attribute, which is based on the concept of the values of subordinate attributes or instances, will be a summation according to Zadeh [23]:

\[ \mu_{A \cup B}(x) = \max\{\mu_A(x), \mu_B(x)\} \]  (2)

Examples of two cases of the summation function of the triangle shape are shown in Fig. 3 and 4.

Concavity of membership function is solely due to incomplete knowledge of experts. In such cases, it is reasonable to highlight their features, as it is shown in Fig. 5 and 6.
At this stage, the supplier already has the ontology containing both a record of IS classes and concepts, as well as the relationship between the classes together and associated attributes that describe the corresponding costs. A part of fuzzy SICO ontology of the implementation cost is shown in Fig. 7.
At the next stage, having prepared a reference ontology, the supplier runs an analytical session with the client and makes changes in the ontology based on the customer’s requirements.

The following cases of changes in the reference ontology are possible:

i. the addition of classes, instances or relationships,

ii. removal of classes, instances or relationships.

Changes in classes, instances or relationships are realized by removing the object and creating a new one in its place. The added ontology components do not have attributes associated with the costs of implementation. In the process of succession, they are assigned with a set of attributes of the superior object. An example of attributes inheritance by the added class and instance is shown in Fig. 8.

In the case of added relationships, such an operation is not always possible. If the added relationships combine classes A and B, then:

i. if class A and class B do not have any “old” relationships, the expert assigns the costs of attributes using Estimation by analogy [3],

ii. if class A or class B have the “old” relationships, the relationship A-B receives all the attributes that exist in the relationship of class A or B, and its values are calculated as the sum by Zadeh.

The next process in cost estimation is the differentiation of the reference ontology and the ontology changed as a result of specific customer’s requirements. The results of differentiation are two lists of objects, such as classes, relationships and instances with attributes assigned to them. In both lists, classes and their attributes are removed as they do not affect the costs.

The first list contains the objects removed from the reference ontology. From this list, the values of the attribute cost of removing are selected and aggregated. The second list contains the objects that were added to the ontology as a result of customer’s requirements. From this list, the values of attributes, for example: the cost of addition, the cost of parameterization, and the cost of training are selected and aggregated. Aggregation is done according to the following formula:

\[
\mu_{A+B}(x) = \max \{\mu_A(y), \mu_B(z) \mid x = y+z\} \tag{3}
\]

An example of aggregation of membership function is presented in Fig. 9.
The values of the attributes of the cost of the addition, the cost of parameterization, the cost of training, and the cost of removing define the supplier’s additional costs of implementation arising from specific (non-standard) customer’s requirements.

6 Example of application

The verification of the proposed method was carried out on the example of estimating the cost of IS changes in a medium-sized manufacturing company of the shipbuilding industry. After a year of implementation of a standard version of the IS in the production area, the client reports a new requirement associated with product safety certifications introduced into production and specifies that the allocated budget for this change in the amount of 9,000 PLN. During the trade talks, the company reveals, among other things, that the raw material from group A are supplied with safety certificates to be attached to the final product sold to the customer, and a copy should be stored in electronic form in the archive.

At this stage, the only methods of estimating costs that can be used are the Individual Assessment by the Expert or a Group of Experts. The supplier estimates the value of the work by the Individual Expert Assessment at 5600 PLN, which means he can undertake the works. During the implementation analysis, the supplier specifies the requirements of the exact contents of the new document (number and types of fields), and obtains additional information. Having completed the requirements, the cost of IS changes is estimated by the supplier according to the Function Point Analysis and after the adjustment, the value of 32 Function Point (FP) is obtained.

On the basis of previous projects, the supplier determines the value of 1 FP equal to 230 PLN, which gives an estimated value of 7,360 PLN.

Then the supplier performs the valuation using the \( SICO \) method. The first estimate is made at the stage of trade talks. On the basis of the disclosed customer’s requirements, the reference ontology increases by two classes associated with classes Purchase documents and Sales documents, two instances (Certificate_doc and Certificate_img) and 12 relationships between the added and the existing classes.
Classes and instances inherit the metric cost_of_adding from the superior objects, while the added relationships value of this metric is determined by the expert. After determining the difference between the modified and reference ontology, the total value assigned to the metric cost_of_adding is 6400 PLN. After pre-implementation analysis, another 9 relationships are added, which increases the estimate up to 7900 PLN.

At the same time, the supplier carries out the valuation using the method of fuzzy SICO. Changes in classes, instances and relationships are the same as in the SICO method. After determining the difference between the modified and reference ontologies, membership function is the sum of the metrics cost of adding M1 = (4,600 PLN, 6,800 PLN, 7,800 PLN, 9,800 PLN), as shown in Fig. 10. This means that the most likely cost of implementation is between 6,800 and 7,800 PLN. In extremely disadvantageous situation, it does not exceed 9,800 PLN, but is not lower than 4,600 PLN.

Among the many strategies of de-fuzzing [24] only the strategy of the highest maximum (LOM - Last of Maximum) is useful for the supplier in the process of contract negotiations. On this basis, the supplier determines the value of the contract at 7,800 PLN. After pre-implementation analysis, the supplier adds another 9 relationships to the ontology and the estimate increases to M2 = (4,600 PLN, 7,800 PLN, 9,000 PLN, 10,800 PLN), as shown in Fig. 11. This means that the most likely cost of implementation is between 7,800 and 9,000 PLN. In extremely disadvantageous situation, it does not exceed 10,800 PLN and it is not lower than 4,600 PLN. On the basis of LOM, the final cost of the implementation is 9,000 PLN. The supplier consciously undertakes the work considering the risk of exceeding the cost up to 10,800 PLN.

The primary criterion of evaluation for each method of estimating is the error of the results. For IT projects, this error is known only after the implementation of the entire IS, when the estimates and the actual cost of the work are compared. After the implementation of this project, the analysis of cost reports shows that the value of the work delivered to the customer is 8,600 PLN. A comparison of the estimation errors of each method is shown in Table 1.

On the basis of the analyzed case, it can be noted that cost estimation, according to the method based on fuzzy attribute of fuzzy SICO ontology gives no worse results than the other known methods. In particular, it should be noted that at the stage of trade talks, the result is much better (more accurate) than the estimation by a group of experts and the SICO method.

7 Conclusions

The method of estimating the implementation costs that uses fuzzy attributes in fuzzy SICO ontology gives suppliers greater benefits than previously known methods. Its advantages should be seen not only in no worse results, but above all in the fact that supplier’s knowledge is extended by the range of the expected costs. This allows suppliers to take conscious risk of cost exceeding, if during the negotiations they are forced to reduce the value of the contract because of competitiveness.

In future studies, verification of fuzzy SICO methods will be extended with further estimation cases, in order to confirm the accuracy that is not worse than other known methods. In addition, the research will be complemented by de-fuzzing strategies. The known strategies are not fully satisfactory for the suppliers in the process of determining the value of IS implementation.
8 Bibliography

[1] Burns M. - How to select and implement an ERP System, 2005


