FUZZY MODEL OF SOFTWARE IMPLEMENTATION BASED ON ONTOLOGY

Przemysław PLECKA, Krzysztof BZDYRA

Abstract: IT project concerning a selected class of ERP while related to dimensioning (cost estimation) of the project within the scope of any selected properties (metrics), e.g. changes in the cost of software is considered. According to specific customer’s requirements 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. A method, which gives more accurate results at the stage of trade talks is a method based on the ontology of implementation costs. This paper proposes a 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 effectively negotiate a contract and increase the chances of a successful completion of the project.

Keywords: software cost estimation, implementation of IT projects, fuzzy ontology, modelling, methods for software evaluation.

1. Introduction

The methods that help to estimate software production costs are known and described in the literature. Concerning the results of research on the costs [1] and the effectiveness of estimation methods according to algorithmic methods, such as COCOMO [2], or Function Points Analysis [3]. Few publications concern non-algorithmic methods (e.g. Estimation by Analogy) [4]. These publications do not restrict software development works but they refer to production, development and modification together. In the course of the research the effectiveness of dimensioning methods arising from the practice with the projects was verified [5]. On the basis of the implementation of three ERP systems in production companies significant errors in the results of estimates were noted. They were due to improper selection of methods to the information (requirements) obtained from the customer. In the next step 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 [6]. However, this solution preferred algorithmic methods (Function Point Analysis, COSMIC Full Function Points, 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 a basis of recording knowledge on the information system and the customer's requirements [7]. The use of the BPMN standard (Business Process Model and Notation) 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 if the
recording of model processes with the metrics done in BPEL (*Business Process Execution Language*) or BPML (*Business Process Modeling Language*). The disadvantage of this solution is a limited scope of the representation of data structures (entities). Unsatisfactory results of previous studies have led to further search for a model, for methods that will give more accurate results. Specification of an ERP system which is the subject of an IT project, understood as a limited scope of knowledge about the company can be formalized in a number of ways, one of them being an ontology [8], which is a recognized and increasingly common way of formalizing knowledge. Using an ontology to implement IT systems can be found in the literature increasingly often [9][10][11]. Its use for the software valuation implies that it should contain information about the computer system as such, i.e. data structures and processes implemented by the system and the related information about properties. In this context ontology allows both recording IS classes with the relationship between them (static model), and the record of the relationship between the classes (dynamic model). Because of the completeness and consistency of information, the record of requirements in the ontology is more comprehensive than in other formal models. No satisfactory proposals for dimensioning IT projects in the early stages in the literature and unsatisfactory results of previous studies justify the search for another model of the project, based on which the design methods will give more accurate results.

Section 2 describes previous research that led to the development of the method based on the ontology of implementation costs, verification its usefulness and which guided towards further exploration on the use of fuzzy attributes, classes, instances, and relationships. The next Section 3 defines the problem of fuzzification of ontology attributes and reduces it to a specific class of ERP systems. Section 4 provides brief overview of knowledge about the use of fuzzy ontologies. While Section 5 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. Software Implementation Cost Ontology Driven Approach

Previous studies [12] show that the basis for a more accurate dimensioning is a model containing a complete set of specific requirements for changes in IS. The changes requirements analysis is part of the pre-implementation analysis and its results are known only after this stage. In order to obtain adequate information earlier, a novel solution consisting in the use of ontologies, as a form of representation of requirements change was proposed. Consequence of this is a new method of dimensioning based on the metrics included in the ontology of implementation costs (*SICO - Software Implementation Cost Ontology*) [13]. The IT system ontology is prepared once and then gradually supplemented with experience from previous development works. Ontological cost model allows for the representation of both data structures and processes. 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 with metrics relating to the costs, for example: *cost of training, cost of parameterisation, 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 a 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 analysed (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 illustrative description of the steps implemented in the course of estimating costs with the SICO method are available in [14].

For illustration let us consider the stage of negotiation of the contract, when 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, e.g. 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 a supplier loses, if it is too high, he will not be competitive with other suppliers and will lose a 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 solution is the proposed, the modified SICO method using fuzzy 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 than 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. Such crispy stated, deterministic cases seem to quite unrealistic. Following from the practice real-life situations have to take into account data inaccuracy and uncertainty. The consequence of fuzzification the attributes of the component ontology needs to be re-determined:

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

3. Problem definition

The IT project concern a selected class of ERP (Enterprise Resource Planning) system. In particular, attention is drawn to the issues related to dimensioning (cost estimation) of the project within the scope of any selected properties (metrics), e.g. changes in the cost of software. Starting an IT project the manager (managers) would like to know the features of the system at the end of the project (after changes). The already known methods of evaluation, such as: Individual Expert Assessment, Evaluation Expert Group, and Counting, Calculation and Assessment, which can be used at the initial stage of the project generate results that are significantly different from the actual values. The errors include the use of IT
project models, containing incomplete information about the subject of estimation. The valuation methods, such as Function Point Analysis, COSMIC, or COCOMO, used in later stages of the project life utilize different data models. Their results are accurate, but they are less important when making project decisions. Besides, the cost of obtaining data to the model is much larger than in other methods.

All in all, in the process of IT project management it is essential to use such models and methods that allow dimensioning many metrics in the early stages of project planning and are not inferior in terms of accuracy and cost of use than other known methods. The need to take into account of the uncertain (fuzzy) values of decision variables implies the demand for fuzzy ontological models. The above reasons justify the proposed research hypothesis: Fuzzy ontological valuation models of an IT project allow the development of the entailed valuation methods resulting in more efficient i.e. more accurate results obtained at a lower estimation costs.

It is assumed that an ERP-class information system (IS) with a specific set of functionalities and the experience from previous projects therein are known. Experience is defined as a set of properties of individual elements of IS, e.g. cost, deployment time, etc. The result of cost estimation is expected as early as at the initial stage of planning the project. A formal model of an IT project that allows the formulation of specific conditions is sought. Their compliance guarantees that estimation methods based on this model will be more accurate, faster (available for early use) and less expensive than the already known methods.

The question is whether the use of fuzzy attributes in the ontology of implementation costs - fuzzy SICO gives not worse estimation results than using "sharp" attribute values. During the evaluation the results of proposed method are compared with the results obtained using the method of "crisp" - SICO and other well-known methods. The problem is limited to a specific class of medium-sized companies and their respective ERP systems.

4. Fuzzy Ontology Driven Knowledge Models

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 was described by C. Carsson [19]. In contrast, P. Alexopoulos proposed a method to convert an "crisp" 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. Orlowski and A. Czarniecki for evaluating software [21] and the use of ontologies for modelling requirements [22].

5. Fuzzy SICO Method

In the most popular and widely used method for estimating the costs by an expert or a group of experts [23] 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 [13]. 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 a 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 implementation 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.

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 \( x \) 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 a form of partial membership, whose degree is determined by the number of the range \([0, 1]\).

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 a superior attribute, which is based on the concept of the values of subordinate attributes or instances, will be a summation according to Zadeh [24]:

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

Examples of two cases of the summation function of the triangle shape is shown in Figure 1 and 2.

![Figure 1. Aggregation of parameterization costs for two warehouse documents](image)
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 Figure 3 and 4. 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 Figure 5.

At the next stage, having prepared a reference ontology, the supplier runs an analytical sessions with the client and makes changes in the ontology based on customer’s requirements. The following cases of changes in the reference ontology are possible:

- the addition of classes, instances or relationships,
- removal of classes, instances or relationships.

Changes in classes, instances or relations 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 Figure 6.
In the case of added relationships such an operation is not always possible. If the added relationships combines classes \( A \) and \( B \), then:

- if class \( A \) and class \( B \) do not have any "old" relationships, the expert assigns the costs of attributes using *Estimation by analogy* [4, 23],
- 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 result 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 \textit{cost of removing} are selected and aggregated. The second list contains the objects which were added to the ontology as a result of customer’s requirements. From this list the values of attributes e.g.: \textit{the cost of the addition}, \textit{the cost of parameterisation}, \textit{the cost of training} are selected and aggregated. Aggregation is done according to the following formula:

$$\mu_{A+B}(x) = \max\{\mu_A(y), \mu_B(z) | x = y+z\}.$$ \hfill (3)

An example of aggregation of membership function is presented in Figure 7.

The values of the attributes of \textit{the cost of the addition}, \textit{the cost of parameterisation}, \textit{the cost of training} define supplier’s additional costs of implementation arising from specific (non-standard) customer’s requirements.

![Figure 7. Example of aggregation of two membership functions - the cost of parameterisation](image)

6. Illustrative example

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 in ship-building 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 to the production and specifies that the allocated budget for this change in the amount of 9000 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 which can be used are the \textit{Individual Assessment by the Expert or a Group of Experts}. The supplier estimates the value of the work by the \textit{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 \textit{Function Point Analysis} and after the adjustment the value of 32 FP (Function Point) 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 7360 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 2 classes associated with classes \textit{Purchase documents} and \textit{Sales documents}, two instances (\textit{Certificate_doc} and \textit{Certificate_img}) and 12 relationships between the added and the existing classes. Classes and instances inherit the metric \textit{cost of adding} from the superior objects, while the added relationships the 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 \textit{cost of adding} is 6400 PLN.
After pre-implementation analysis another 9 relationship are added what 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 $M_1 = (4600 \text{ PLN}, 6800 \text{ PLN}, 7800 \text{ PLN}, 9800 \text{ PLN})$. This means that the most likely cost of implementation is between 6800 and 7800 PLN. In extremely disadvantageous situation it does not exceed 9800 PLN, but is not lower than 4600 PLN. Among the many strategies of defuzzification [25] only the strategy of 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 7800 PLN. After pre-implementation analysis the supplier adds another 9 relationships to the ontology and the estimate increases to $M_2 = (4600 \text{ PLN}, 7800 \text{ PLN}, 9000 \text{ PLN}, 10800 \text{ PLN})$. This means that the most likely cost of implementation is between 7800 and 9000 PLN. In extremely disadvantageous situation it does not exceed 10800 PLN and it is not lower than 4600 PLN. On the basis of LOM the final cost of the implementation is 9000 PLN. The supplier consciously undertakes the work considering the risk of exceeding the cost up to 10800 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 8600 PLN. A comparison of the estimation errors of each method is shown in Table 1.

<table>
<thead>
<tr>
<th>Cost estimation method</th>
<th>The stage of trade talks</th>
<th>The stage of pre-implementation analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Estimation of a group of experts</td>
<td>35%</td>
<td>-</td>
</tr>
<tr>
<td>Function Point Analysis</td>
<td>-</td>
<td>14%</td>
</tr>
<tr>
<td>SICO method</td>
<td>26%</td>
<td>8%</td>
</tr>
<tr>
<td>Fuzzy SICO method</td>
<td>9%</td>
<td>5%</td>
</tr>
</tbody>
</table>

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 results of the approach presented, which fit within the scope of software engineering will bring new knowledge in the area of modelling processes of computer-aided development and implementation of IT projects, in particular concerning the ontological fuzzy models employed therein. The confirmation of the research hypothesis opens the possibilities for the development of new valuation methods based on fuzzy ontologies meeting the requirements of appropriately defined and relevant conditions. The proposed fuzzy ontological model of IT project represents a new perception of properties changes in IT systems, in particular allowing the observation of the dynamics of changes in IS. The
implementation of the above mentioned options will significantly contribute to the
development of methods of valuing projects. The proposed method of estimating the
implementation costs that uses fuzzy attributes in fuzzy SICO ontology gives suppliers
greater benefits than previously known methods. 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 defuzzification strategies. The
known strategies are not fully satisfactory for the suppliers in the process of determining
the value of IS implementation.

Bibliography

3. Arnuphaptairong T.: Early Stage Software Effort Estimation Using Function Point
Analysys Empirical Evidence, at International MultiConference of Engineers and
Computer Science, Hong Kong, 2013.
5. Plecka P.: Selected Methods of Cost Estimation of ERP Systems' Modyfications,
Zarządzanie Przedsiębiorstwem, No. 4/2013, pp. 27-34.
Software Implementation, Applied Computer Science, Lublin University of
7. Plecka P., Bzdyra K.: Identyfikacja wymagań użytkownika ERP w procesie wyceny
kosztów wdrożenia, Information Systems in Management, Vol. XVIII, Warsaw
University of Life Science, 2013.
Raport KSL 92-71, Knowledge Systems Laboratory, Computer Science Dept.
Stanford University, California, 1993.
First Ontology, Stanford Knowledge Systems Laboratory Technical Report KSL-01-
05, Standford, California, 2001.
10. Czarnecki A., Sitek T.: Ontologies vs. Rules — Comparison of Methods of
Knowledge Representation Based on the Example of IT Services Management,
Information Systems Architecture and Technology: Intelligent Information Systems,
Knowledge Discovery, Big Data and High Performance Computing, Wroclaw
11. Czarnecki A., Orlowski C.: Możliwości zastosowania ontologii do oceny technologii
informatycznych, Komputerowo zintegrowane zarządzanie. Vol. 2, Polskie
Towarzystwo Zarządzania Produkcyj, Opole, 2007, pp. 143-152.
12. Plecka p., Bzdyra K.: Usefulness of software valuation method at initial stage of ERP
implementation., Foundations of Management, Warsaw University of Technology,
Vol. 05, No. 03 (11), 2013, pp. 33-48.

Mgr inż. Przemysław PLECKA
Dr inż. Krzysztof BZDYRA
Wydział Elektroniki i Informatyki
Politechnika Koszalińska
75-453 Koszalin, ul. Śniadeckich 2
tel.: (94) 343 34 85, (94) 3478-706, fax (94) 343 34 79
e-mail: przemek.plecka@gmail.com
krzysztof.bzdyra@tu.koszalin.pl