COLLECTION ANALYSIS METRICS APPLICABLE TO PROJECTS-BASES

PhD. Ion IVAN, Bucharest, RO, ionivan@ase.ro
MSc. Andrei SANDU, Bucharest, RO, andreisandu@yahoo.com

Abstract: *Projects-base concept is defined and the projects-base management system is presented. Next, quality characteristics applicable to projects-bases are defined. There are associated indicators for estimating the quality level of each indicator. The operations applicable to projects-bases are the prerequisites for defining the vocabulary, the frequency of using specific words and the domain specific thesaurus. By using these tools it is possible to later analyze the dynamics of projects, by comparing the specifics of the text of the project plan with the text of the description of ongoing project phases. This leads to a primary level of knowledge management directly applicable in project management profession.*

Key words: *projects-base, metrics, quality, orthogonality.*

1. Base Concepts

In order to understand what projects-base stands for we will start by looking at what a project is, how they can be structured and we will then present the major quality characteristics which could be taken into account when evaluating projects for obtaining financing.

The most spread definition of a project is that of a temporary endeavor undertaken in order to create a unique, product, service or result. No matter their type and structure, projects can be grouped based on different criteria. A grouping can be called a program, a portfolio or simply a collection, based on the ownership of the projects contained, based on the project goals and on the project types. We a call a portfolio of projects a set of projects, not necessarily inter-dependent, which are performed by a project-oriented company at a certain point in time. If the projects are inter-dependent and, at a strategically they are perceived as an ongoing long term correlated effort, then we are talking about a program. On the other hand, a collection of projects is a set of projects which can be run by different project-oriented companies.

In the case of projects collection $\{P_1, P_2, \ldots P_n\}$ which must be evaluated in a unitary manner in order for the projects to be accepted for obtaining financing, there comes the idea of building up a projects-base. The projects-base contains the project proposals which will be submitted to the evaluation process.

2. Projects-base Management

In order to successfully manage a projects-base operations have to be defined [1]. The operations have to be implemented via an application, in order to assure a human error-free environment. The most important operations are as follows:

- *insertion*, which means the process of populating the projects-base;
- *modification / replacement*, which can occur in the following situation of existence of some errors found out as a result of the insertion operation, when some additions/modifications are required for the text of the project or when a project needs to be replaced by another one;
- *retrieving projects based on vocabulary*, which means listing only the projects that contain the words from a certain search vocabulary;
- *evaluation and projects selection for being financed*, which means that a predefined list of evaluation criteria is applied – the title has to include a predefined set of key words, the value of the project team cohesion coefficient must be greater or equal to a minimum imposed value, the sum requested has to be within the allocated financing budget;
- *checking projects for imposed limits and structure integrity*, which means controlling the adherence to an imposed structure for each project, by measuring each qualitative or quantitative property, considered to be useful for projects evaluation; the value resulted from measurement is put against a specific evaluation grid and the minimum acceptability level.

The operations collection is open and allows the creation of new operations which could be applicable to large collections of projects, like:

- creating project categories, based on complexity level;
- extraction of projects which comprise a certain particularity;
- automatic valuation of projects and delimitation of affiliation to a certain typology, based on predefined criteria;
- automatic archiving of projects which have been completed and which are not referenced within a certain timeframe;

– recalculation of trust thresholds in relation to quality criteria, which is based on information from projects-base sample.

In order to implement operations which are based on numbers of projects, the software components which are added to the existing system need a resource allocation process which is based on neuronal networks. Therefore it is needed that existing software modules are also gradually modified in order to comply with this approach.

3. Quality Metrics

The most important quality metrics to implement on projects-bases are complexity, orthogonality, correctitude and completeness.

Regarding project *complexity*, there are 3 types: estimated, planned and actual.

Estimated complexity is based mostly on expertise gathered from of similar past projects. Planned complexity is a refinement of the estimated complexity, as some corrections are applied in order to adapt to the distinct project context. Actual complexity is measured after the project has been implemented.

Project complexity is the universal measuring unit for all the projects. Complexity represents a characteristic, which is common to all the projects. Projects differentiate through the complexity degree. Some projects are simple. Others are complex. The resources asked by a project vary very much, depending on its complexity. Thus, a project could be realized in a few weeks or it can last more months; a project may need from one team with few people to teams with hundreds.

One approach to measuring complexity is that of taking into account the variety of existing resource types and their quantities.

$$C = \sum_{i=1}^{m} q_i * \log_2 q_i \text{ , where}$$

C = project complexity
m = total number of distinct resources used by the project
q_i = quantity of resource i

Orthogonality means determining the degree to which 2 text entities are different, considering their presentation means and content.

Two projects will be considered orthogonal if their texts only have in common the domain specific technical terms, whereas the others words differ in terms of frequency and positioning within the sentences. It is more important that projects texts are orthogonal and not completly different, because totally different texts could mean that the 2 projects do not belong to the same domain.

A very simple implementation of such an indicator for orthogonality would be:

$$ORTO = \frac{NW_{both}}{NW_{max}} \text{ , where}$$

NR_{both} = total number of words, which are part of the project domain standard terminology, which are found in both texts analyzed
NR_{max} = the highest of the total number of domain specific words of the 2 projects.
Therefore, for this indicator a value of 1 means the 2 projects are cloned one after the other, whereas a value of 0 means that the 2 projects are orthogonal.

Correctness as applicable to projects means that the project text is accepted as being in accordance with the basis requirements for the domain which is being financed. Correctness envisages to the naming of processes, technologies and operations, the usage of appropriate concepts, models and the usage of proper variables. Therefore, by respecting the industry standards and code of professional ethics all that previously existed has to be correctly cited and referenced in the bibliography. Correctness also refers to the logic behind activities planning, resource consumption and estimating. If project valuation procedure is transparent, correctness means that auto valuation differs very little from the official valuation.

In case of partial valuation [6], it is needed to determine the weight of correct criteria from the whole set of criteria.

The relative correctness indicator, RCI, is computed as follows:

$$RCI = \frac{\sum_{i=1}^{n} R_{ia}}{\sum_{i=1}^{n} R_{iMax}} * w \text{ , where}$$

R_{ia} = actual result of valuation, based on criterion i

R_{iMax} = maximum result of valuation, based on criterion i

w = weight of correct criteria from the whole set of criteria.

Completeness is the project quality characteristic which shows the degree by which, when turning project plan into action, the desired project outcomes are obtained.

There are 3 matrixes to be considered as relevant in project completeness valuation:

- A-C matrix, which shows all resources used for performing each activity;
- A-TM matrix, which shows the allocation of project team members on activities
- A-A matrix, which shows the dependencies between the activities.

One indicator of completeness, CMPL, has the following form:

$$CMPL = \frac{3 - \frac{\Delta_1 + \Delta_2 + \Delta_3}{n}}{3}, \text{ where}$$

n = number of project planned activities;

Δ_1 = number of activities which have at least 1 difference between the planed and the executed A-C matrix;

Δ_2 = similar to Δ_1, except it applies for A-TM matrix;

Δ_3 = similar to Δ_1, except it applies for A-A matrix.

The level of completeness is to be considered 1 if supplementary costs are not present. The more the completeness level is closer to 0, the more it is clear that the project responsible for project delivery has not taken into account some activities, has badly estimated durations or has not properly managed the deliverables dependencies.

4. Statistical Analysis on Vocabularies

The projects-base contains a large number of projects, which are part of one or many financing initiative. Financing programs are based on a core set of key words.

The key words allow the identification of domain specific thesaurus. The projects-base also contains thesaurus collections, implemented by using vocabularies attached to domains such as technical, economical, administrative, social and artistic. We will consider VD1, VD2 ... VDk as being the vocabularies of each domain and for each project Pi there will a pair (Cij, Fij), j=1,2,...,Ni, where Ni=maximum number of words comprised within project Pi, whereas Cij is word j from project Pi and Fij is the frequency of appearance of word Cij in project Pi.

By analyzing the list of key word of project Pi against domain specific vocabularies VD1, VD2, ..., VDk, it will be concluded to which domain does project Pi belong to.

The statistical analysis will take into account the frequency of each word, the initial ranking of the words in the project vocabulary and the final ranking of the words after sorting by frequency. For each domain there should be a ranking of the words from the thesaurus, according to their expected usage in a project belonging to that domain.

An aggregated indicator is computed, in order to determine is the vocabulary for that domain is properly used.

$$VOC = \frac{TPWD}{\sum_{i=1}^{TPWD} (|rankP_i - rankD_i| + 1)_i}, \text{ where}$$

TPWD = total number of project distinct words that are part of the domain thesaurus
rankPi = project word i rank after sort by frequency
rankDi = domain rank of word i

Therefore, if VOC indicator is closer to 0, then the project vocabulary is not adequately used. Otherwise, if VOC is 1 or very close to this value, then the project team is using the domain terminology in an adequate manner and the project should be valuated higher and increase its chances of obtaining financing.

By creating domain thesaurus and maintaining it on a regular basis, projects valuation could gain in precision and in speed. The challenges to this are terminology advancement and lack of sufficient human power. One step in implementing the process is automating it, via the usage of a software tool for projects comparison, whereas the second step would be implementing knowledge management in the sense of creating an initial domain specific knowledge base and then letting the system evolve and learn by itself.

5. Conclusions

Projects-bases need to be implemented by using a software application which has unbiased valuation criteria.

In the valuation process, quality characteristics form a system and, based on project type, there appears a difference in the importance given to each specific indicator - i.e. if there are financial restrictions the most important indicators will be considered completeness and correctness, whereas if the project is innovative, complexity, consistency and solution feasibility are taken into account.

Moreover the valuation should also take into account the project vocabulary and its matching against the relevant domain thesaurus. Over the years, by valuating more and more projects, the system learns new terminology and is able to adapt and include or exclude words from the vocabulary and also change their ranking.

The valuation of projects and the implementation auditing are dependant upon the common strategy of the financer and of the beneficiary.

Bibliography

[1] Ion IVAN, Andrei SANDU - Projects-bases management software, 20th IPMA World Congress on Project Management, 15 - 17 October, 2006, Shanghai, China;

[2] Ion IVAN, Cătălin BOJA – Construction of Aggregated Indicators for Evaluating the Quality of Software Metrics, Revista de Studii şi Cercetări de Calcul Economic şi Cibernetică Economică, vol. 40, nr. 3, 2006, ISSN 0585-7511, pg. 59-68;

[3] Andrei SANDU - Security of Projects-bases, Proceedings of 8th International Conference on Informatics in Economy, INFORMATICS IN KNOWLEDGE SOCIETY, 17-18 May, 2007, Bucharest, ASE Printing House, p. 974-983, ISBN 978-973-594-921-1;

[4] Roland GAREIS – Happy Projects!, 2nd Edition, ASE Printing House, Bucharest, 2006, ISBN 978-973-594-848-1;

[5] Constanta BODEA - Knowledge management into modern university, ASE Printing House, 2007, ISBN 9789735949532;

[6] Nicolae Iulian ENESCU – Models for Evaluating the Correctitude of Distributed IT Systems, ASE Bucharest, 2008, PhD thesis;

[7] M. AUBRY, B.HOBBS, D. THUILLIER – A new framework for understanding organizational project management through the PMO, International Journal of Project Management, Vol. 25, Issue 4, May 2007, ISSN 0263-7863, p. 328-336.