EMPIRICAL VALIDATION ONKNOWLEDGE PACKAGING SUPPORTING KNOWLEDGE TRANSFER

Pasquale Ardimento, Teresa Baldassarre, Marta Cimitile and Giuseppe Visaggio
Department of Informatics, University of Bari, Via Orabona 4, Bari, Italy
RCost Bari

Keywords: Knowledge Base, Knowledge Packaging, Open Innovation.

Abstract: Transfer of research results, as well as technological innovation, within an enterprise is a key success factor. The introduction of research results aims to improve efficacy and effectiveness of the production processes with respect to business goals, and also to better adapt the products to the market needs. Nevertheless, it is often difficult to transfer research results in production systems because it is necessary, among others, that knowledge be explicit and understandable by stakeholders. Such transfer is demanding, as so many researchers have been studying alternative ways to classic approaches such as books and papers that favour knowledge acquisition on behalf of users. In this context, we propose the concept of Knowledge Package (KP) with a specific structure as alternative. We have carried out an experiment which compared the efficacy of the proposed approach with the classic ones, along with the comprehensibility of the information enclosed in a KP rather than in a set of Papers. The experiment has pointed out that knowledge packages are more effective than traditional ones, for knowledge transfer.

1 INTRODUCTION

In the Software Processes, knowledge is a critical factor, both because software development (production and maintenance) is man-centred and because the products of the development process are destined to be used by humans to improve their abilities in all applicative domains. For this reason, the necessary knowledge involves two types of problems:

- transferability and reusability. Knowledge hidden in processes and products is not even readable by its authors in that it is spread out and confused in many of the process or product components (Foray, 2006) (Myers, 1996). So, until knowledge is transferable or reusable, it cannot be considered as part of an organization’s assets (Foray, 2006);
- knowledge exploitation. Research produces knowledge that should be transferred to production processes as innovation in order to be valuable. Consequently, domain knowledge must be enriched by technical and economical knowledge that allows to identify the best approach for introducing new knowledge in processes together with the resources, risks and mitigation actions (Reifer, 2003).

The first problem requires formalizing knowledge so that it is comprehensible and reusable by others that are not the author of the knowledge. The second problem requires experience packaging able to guide the user in applying the knowledge in a context.

Given these premises, this paper describes an approach for Knowledge Packaging and Representation and reports preliminary results of a first experimentation of the approach.

In our proposed approach we have formalized a KP and we have defined some packages that are stored in a Knowledge Base (Schneider, 2001) (Malone, 2003) (Basili, 1992) (Schneider, 2003). The KPs are obtained by using paper and other resources available on the Web and by giving them a predefined structure in order to facilitate stakeholders in the comprehension and the acquisition of the knowledge that they contain.

We have conducted a preliminary validation through a controlled experiment with the aim to answer to the following Research Questions (RQ):

RQ1. Is proposed knowledge description approach more efficacious than traditional ones?

Ardimento P., Baldassarre T., Cimitile M. and Visaggio G. (2007).
EMPIRICAL VALIDATION ONKNOWLEDGE PACKAGING SUPPORTING KNOWLEDGE TRANSFER.
In *Proceedings of the Second International Conference on Software and Data Technologies - PL/DPS/KE/WsMUSE*, pages 212-218

RQ2. Is the proposed knowledge description approach more comprehensible than traditional ones?

In the first question we introduce the concept of Knowledge Description Efficacy considered as the rapidity in which a usable knowledge chunk can be selected without support of the knowledge author.

In the second question we introduce the concept of Knowledge Description Comprehensibility as the capability of the adopted Knowledge Description criteria to transfer the selected knowledge in a complete and correct way.

In this work we named “traditional approach” the approach based on the use of papers, book or, in general, not structured text for knowledge transfer and acquisition.

The rest of the paper is organized as follows: related works are described in section 2; section 3 illustrates the proposed approach for knowledge representation, section 4 illustrates the measurement model used; results of the study and lessons learned are presented in section 5; finally in section 6 conclusions are drawn.

2 RELATED WORKS

The problems related to knowledge transfer and valorization are investigated in industrial and academic contexts and sometimes it’s not possible to distinguish the two because there is a convergence between industry and academia. Some companies have established internal organizations whose task is to acquire new knowledge (Halvorsen, 2004) (Hastbacka, 2004) to face knowledge transfer needs. For example, Shell Chemical has organized some groups with the aim at finding knowledge from outside sources, Hewlett Packard is commercializing not only its own ideas, but also innovations from other entities (Halvorsen, 2004), Philips Research is participating to consortiums that direct one to one collaboration with innovative organizations (Hastbacka, 2004).

There are also many studies that are focused on the use of Internet together with its Search Engines for knowledge diffusion and transfer. But in this direction our analysis shows that INTERNET, does not offer appropriate technologies for searching knowledge that is produced and published by a research organization nor by an enterprise, which is reusable in innovation projects by other research organizations or enterprises (Scoville, 1996) (Leighton, 1997) (Ding, 1996) (Leighton, 1996) (Chu, 1996) (Clarke, 1997). A validation of this statement is proposed in (Ardimento, 2007). The most accredited reason for this limitation is that usually general queries produce a large amount of documents and that there is not a natural language interface of the search engine. The latter technology improves the search precision although it does not overcome the problems described above.

There are also many approaches based on the use of specialized search engines in the way to find search results related to a specific application domain (Kitchenham, 2004).

Another approach to knowledge search and transfer is based on the use of ontology (Zhang, 2004) (Mingxia, 2005). This approach is actually object of many studies which currently lack tools for creation and management. Much attention is being focused on these issues but the available experimental evidence is not yet sufficient for large-scale use.

In this work we proposed an alternative approach to knowledge transfer based on concepts of knowledge packaging and knowledge base. The problem of knowledge packaging for better use is being studied by many research centres and companies. The current knowledge bases in literature, sometimes have a semantically limited scope. This is the case of the IESE base (Althoff, 2001), that collects lessons learned or mathematical prediction models or results of controlled experiments. In other cases the scope is wider but the knowledge is too general and therefore not very usable. This applies to the MIT knowledge base (Malone, 2003), that describes business processes but only at one or two levels of abstraction. There are probably other knowledge bases that cover wider fields with greater operational detail but we do not know much about them because they are private knowledge bases, for example the Daimler-Benz Base (Schneider, 2001).

3 PROPOSED APPROACH

Our approach focuses on a knowledge base, named Prometheus (Serlab, 2006), whose contents make it easier to achieve knowledge transfer among research centres; between research centres and production processes; among production processes. The knowledge base must be public to allow one or more interested communities to develop around it and exchange knowledge (Ardimento, 2006). The knowledge that is stored in the knowledge base must be formalized as KP. A KP is any cluster of knowledge, sufficiently familiar that it can be remembered rather than derived.

3.1 Knowledge Package Structure

The proposed KP includes all the elements shown in Figure 1. A user can access one of the package components and then navigate along all the components of the same package according to her/his training or education needs. Search inside the package starting from any of its components is facilitated by the component's Metadata.

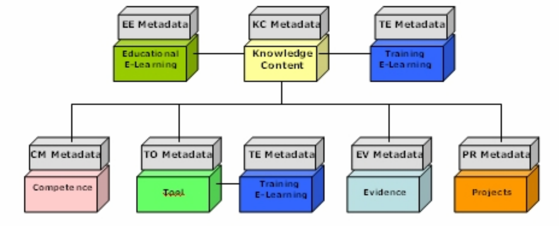

Figure 1: Diagram of a Knowledge/Experience package.

It can be seen in the figure that the Knowledge Content component (KC) is the central one. It contains the knowledge package expressed in text form, with figures, graphs, formulas and whatever else may help to understand the content. The content is organized as a tree. Starting from the root (level 0) navigation to the lower levels (level 1, level 2, …) is possible through links. The higher the level of a node the lower the abstraction of the content which focuses more and more on operative elements. The root and each intermediate node contain the reasoned index of the underlying components. The content consists of the following: research results for reference, analysis of how far the results on which the innovation should be built can be integrated into the system; analysis of the methods for transferring them into the business processes; details on the indicators listed in the metadata of the KC inherent to the specific package, analyzing and generalizing the experimental data evinced from the evidence and associated projects; analysis of the results of any applications of the package in one or more projects, demonstrating the success of the application or any improvements required, made or in course; details on how to acquire the package.

When knowledge of some concepts is a prerequisite for understanding the content of a node, the package points to an Educational E-learning course (EE). Instead, if use of a demonstrational prototype is required to become operative, the same package will point to a Training E-learning course (TE).

To integrate the knowledge package with the skills, KC refers to a list of resources possessing the necessary knowledge, collected in the CoMpetence component (CM).

When a package also has support tools, rather than merely demonstration prototypes, KC links the user to the available tool. For the sake of clarity, we point out that this is the case when the knowledge package has become an industrial practice, so that the demonstration prototypes included in the archetype they derived from have become industrial tools. The tools are collected in the Tool component (TO). Each tool available is associated to an educational course, again of a flexible nature, in the use of the correlated TE course.

A knowledge package is generally based on conjectures, hypotheses and principles. As they mature, their contents must all become principle-based. The transformation of a statement from conjecture through hypothesis to principle must be based on experimentation showing evidence of its validity. The experimentation, details of its execution and relative results, are collected in the Evidence component (EV), pointed to by the knowledge package.

Finally, a mature knowledge package is used in one or more projects, by one or more firms. At this stage the details describing the project and all the measurements made during its execution that express the efficacy of use of the package are collected in the Projects component (PR) associated with the package.

3.1.1 Metadata

As shown in Figure 1, each component in the knowledge package has its own metadata structure. For all the components, these allow rapid selection of the relative elements in the knowledge base (Ardimento, 2006), Figure 2.

Figure 2: Diagram of a Knowledge package.

To facilitate the research a set of selection classifiers and a set of descriptors summarizing the contents are used. The classifiers include: the key words and the problems the package is intended to solve. The summary descriptors include: a brief summary of the content and a history of the essential events occurring during the life cycle of the package, giving the reader an idea of how it has been applied, improved, and how mature it is. The history may also include information telling the reader that the content of all or some parts of the package are currently undergoing improvements.

The package also provides the following indicators: skills required to acquire it, prerequisite conditions for correct working of the package, acquisition plans describing how to acquire the package and estimating the resources required for each activity. To assess the benefits of acquisition, they contain a list of: the economic impact generated by application of the package; the impact on the value chain, describing the impact acquisition would have on the value of all the processes in the production cycle; the value for the stakeholders in the firm that might be interested in acquiring the innovation. There are also indicators estimating the costs and risks. Thus, all these indicators allow a firm to answer the following questions: what specific changes need to be made? What would the benefits of these changes be? What costs and risks would be involved? How can successful acquisition be measured?

3.2 Experiment Planning

3.2.1 Research Goals

The following Research Goals (RG) have been defined:

RG1:
Analyze knowledge extraction using an Knowledge Package (KP)
With the aim of evaluating it
With respect to efficacy (compared to knowledge extracted from papers)
From the view point of the knowledge user
In the context of a controlled experiment on a knowledge package tool called Prometheus.

RG2:
Analyze knowledge extraction using an Knowledge Package (KP)
With the aim of evaluating it
With respect to comprehensibility (compared to knowledge extracted from papers)
From the view point of the knowledge user
In the context of a controlled experiment on a knowledge package tool called Prometheus.

3.2.2 Variable Selection

The dependent variables of the study are Efficacy and Comprehensibility. Efficacy indicates to what point the Knowledge Representation criteria is effective (in terms of effort spent) for extracting knowledge and answering a specific set of questions. Comprehensibility indicates to what point the resources described in Prometheus or in Papers are easy to understand and to abstract in order to answer a set of questions.

The independent variables are the two treatments: the problems examined with KP and with Papers in literature. Two different types of problems were investigated: Balanced Scorecard and Reengineering Process.

For each problem a set of 4 questions have been defined. This has been considered an appropriate number that balances the need for a sufficient amount of data without having to count on an excessive amount of effort and risk to bore a tire experimental subjects.

3.2.3 Selection of Experimental Subjects

The experimental subjects involved in the experimentation are first year students of a graduate course in Informatics with background experience on collaborations with industrial case studies as result of project work carried out during their courses.

A total of 82 students have been divided in two groups (GROUP A and GROUP B) with random assignment to each one. Each group was asked to answer questions assigned using, alternatively KP or Papers extracted from literature.

All of the students have previous knowledge on the topic concerning Balanced Scorecard because it is part of their course curricula. While, they have no previous knowledge on the Reengineering Process topic.

It is important to note that the selected set of experimental subjects, even if variegate, is not completely representative of the population of all software stakeholders such as managers, end users and so on. As consequence, at this first stage, it is not possible to generalize the results of the empirical investigation. Rather, results represent a first important step towards this direction.

3.2.4 Experiment Operation

The experiment was organized in two experimental runs, RUN1 and RUN2, one per day in two consecutive days. During each run we changed the content of the KP/papers and the content of the questions used to extract information from the source. Moreover, in RUN1, the KP/papers content, along with the questions for extracting information, related to Balanced Scorecard (Becker, 1999) (Grembergen, 2000) (Abran, 2000) (Mair, 2002); and in RUN2 they referred to Reengineering (Bianchi, 2000) (Bianchi, 2001) (Bianchi 2003). Within a RUN, each group was assigned to either one of KP or Paper.

At the beginning of each run, each experimental subject received a complete set of instrumentation. It contained the papers in digital version or KP according to the treatment and group. The KP is accessible through Prometheus. The students examined the material and answered the questions reporting them on the data form. The start and end time were recorded by the researchers when handing in and collecting the forms.

Comprehensibility was evaluated according to the number of errors made, while the effort is reported on the data form.

4 MEASUREMENT MODEL

The introduced metrics are collected as Prometheus and Paper metrics. The metrics described have been collected on both types of knowledge extraction treatments.

According to the Efficacy Factor the introduced metric is:

Effort (EF): The amount of effort, measured in person/hrs, spent by each subject for carrying out their task and answer the questions:

$$EF = t' - t$$

- t: Time when packages/papers and forms are given to an experimental subject.
- t’: Time when an experimental subject hands in the data form complete with answers.

Another factor is Comprehensibility. It is measured as the average of points Pij attributed for answering the i-th question of the j-th experimental subject. All answers were evaluated according to the interval scale reported in table 1.

Table 1: Details of comprehensibility quality factor.

Evaluation of Question	P_{ij} score
Wrong Answer: the j-th subject gave a wrong answer to the i-th question.	0
Lacking Answer: the question was not answered by the j-th subject	2
Incomplete Answer: the j-th subject gave a partially correct answer to the i-th question	4
Complete Answer: the i-th question has received a correct answer by the j-th subject	6

The researchers, as domain experts involved in the investigation, corrected all the answers to the questions given by the experimental subjects.

5 EXPERIMENTAL RESULTS

5.1 Efficacy

In RUN1, the subject performances, as shown in figure 3 are closer. The mean values are respectively 0.0643 for PROMETHEUS and 0.0657 for PAPERS. Also, the dispersion of the results is very high for both knowledge representation methods. It seems as if the performances are independent from the technique used. Our explanation is that the experimental subjects were familiar with the topic (Balanced Scorecard) and so they used their previous experience and knowledge to answer the questions rather than strictly relate on the technique assigned (KP or Papers).

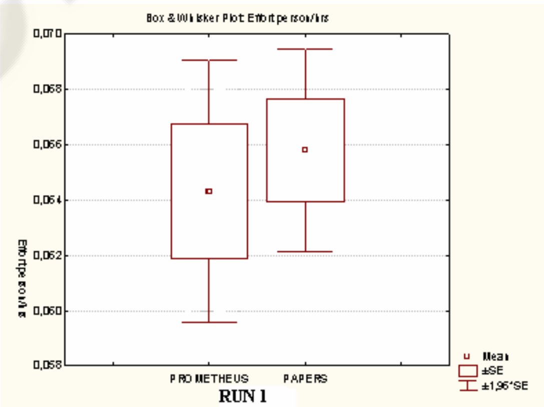

Figure 3: Effort in Prometheus and Papers during RUN 1.

Figure 4 illustrates the average effort in person/hrs spent by the experimental subjects in RUN2. It can be seen that there is less dispersion in the results for both knowledge representation techniques. Also, it can be seen how subjects using Papers spent, on average, a larger amount of time for answering the questions. This suggests that the structure of the packages promotes a more appropriate search of the knowledge contents for answering a question.

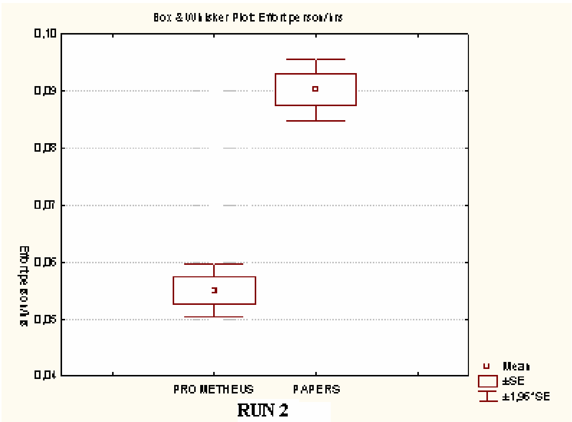

Figure 4: Effort in Prometheus and Papers during RUN 2.

5.2 Comprehensibility

In RUN1, figure 5 shows the trend of comprehensibility with respect to the questions, which appears to be analogous in both representation methods. This confirms our assumption that subjects have most likely used their previous knowledge on the topic to answer the questions within RUN1. In each case, comprehensibility with Prometheus is always better than with Papers.

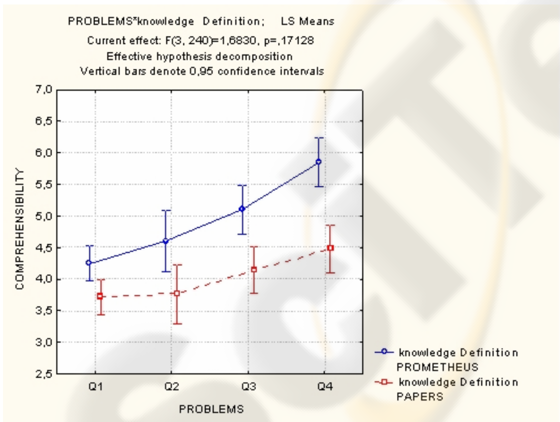

Figure 5: Comprehensibility in Prometheus and Papers for problem during RUN1.

Figure 6 shows the interaction effect between the factors Problems*Knowledge Representation with respect to comprehensibility in RUN2. The graph points out that overall comprehensibility is better when Prometheus is used.

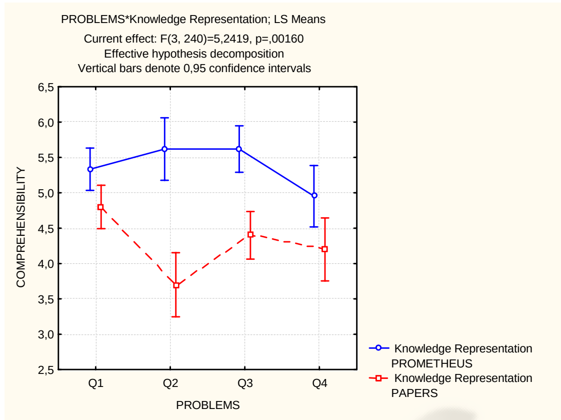

Figure 6: Comprehensibility in Prometheus and Papers for problem during RUN2.

6 CONCLUSIONS AND FUTURE WORKS

This paper proposes an approach based on the concept of Knowledge Package for knowledge transferring as alternative way to the traditional ones.

The proposed approach was implemented through a knowledge base called PROMETHEUS.

To validate the approach an empirical investigation was conducted. The experiment was carried out with university students attending first year and consisted of a comparison between proposed approach and traditional approach in terms of Efficacy and Comprehensibility.

The collected results provide some lessons learned about structure of an Knowledge Package in Prometheus. In fact the proposed approach with respect to the traditional ones:

- requires less effort for extracting information searched;
- represents explicit knowledge in a more comprehensible form.

According to our opinion and the feedback provided by students the discovered differences could be related to the use of metadata and to the multi-level structure of package.

It is clear that, in order to generalize the validity of the lessons learned proposed in this work, many replications, statistical validation and further studies, extended to other contexts, are needed. Finally it is necessary to replicate the study on a set of experimental subjects that may be even more representative of the population than the ones involved in this first empirical investigation.

REFERENCES

Foray, D., 2006. L'economia della conoscenza. *Il Mulino*.

Myers, P., 1996. Knowledge Management and Organizational Design: An Introduction, Knowledge Management and Organizational Design. *Butterworth-Heinemann, Newton, MA.*

D. J. Reifer, D., J., Nov/Dec 2003. Is the Software Engineering State of the Practice Getting Closer to the State of the Art?. *IEEE Software, vol. 20, no. 6, pp. 78-83*.

Schneider, K., Schwinn, T., 2001. Maturing Experience Base Concepts at DaimlerChrysler. *Software Process Improvement and Practice 6*.

Malone, T., W., Crowston, K., Herman, G., A., 2003. Organizing Business Knowledge-The MIT Process Handbook. *MIT Press Cambridge*.

Basili, V., R., Caldiera, G., McGarry, F., Pajerski, R., Page, G., Waligora, S., May 1992. The Software Engineering Laboratory - an Operational Software Experience Factory. *Proceedings of the International Conference on Software Engineering.*

Schneider, K., Hunnius, J., V., 2003. Effective Experience Repositories for Software Engineering. *25th International Conference on Software Engineering , ICSE*.

Halvorsen, P., K., 2004. Adapting to changes in the (National) Research Infrastructure. *Hewlett Packard Development Company, L.P*.

Hastbacka, M., A., December 2004. Open Innovation: What's mine it's mine. What if yours could be mine too. Technology Management Journal.

Scoville, R., January 1996. Special Report: Find it on the Net!. *PC World*, retrieved by http://www.pcworld.com/reprints/lycos.htm

Leighton, H., Srivastava, J., 1997. Precision among WWW search services (search engines): AltaVista, Excite, HotBot, Infoseek and Lycos. Retrieved from http://www.winona.edu/library/.

Ding, W., Marchionini, G., October 1996. A comparative *study of the Web search service performance. Proceedings of the ASIS Annual Conference.*

Leighton, H., June 1996. Performance of four WWW index services, Lycos, Infoseek, Webcrawler and WWW Worm. Retrieved from http://www.winona.edu/library/.

Chu, H., Rosenthal, M., October 1996. Search engines for the World Wide Web: a comparative study and evaluation methodology. *Proceedings of the ASIS Annual Conference*.

Clarke, S., Willett, P., 1997. Estimating the recall performance of search engines. *ASLIB Proceedings, 49 (7), 184-189*.

Ardimento, P., Caivano, D., Cimitile, M., Visaggio, 2007. G. Empirical Investigation of the efficacy and efficiency of tools for transferring Software Engineering Knowledge. *JIKM Editor World Scientific Publishing (submitted).*

Kitchenham, B., 2004. Procedures for Performing Systematic Reviews. *Technical Report TR/SE-0401-ISSN*.

Mingxia, G., Chunnian, L., Furong, C., 2005. An Ontology Search Based on Semantic Analysis. *Proceedings of the Third International Conference on Information Technology and Applications*.

Althoff, K.D., Decker, B., Hartkopf, S., Jedlitschka, A., Nick, M., Rech, J., July 2001. Experience Management: The Fraunhofer IESE Experience Factory. Proceedings Industrial Conference Data Mining, Institute for Computer Vision and applied Computer Sciences, Leipzig, Germany In P. Perner (ed.).

Ardimento, P., Cimitile, M., Visaggio, G., 2006. Knowledge Management integrated with e-Learning in Open Innovation. *Journal of e-Learning and Knowledge Society, Vol. 2, n.3, Erickson Edition*.

Becker, A. S., Bostelman, M., L., May/June 1999. Aligning Strategic and Project Measurement Systems. *IEEE Software*.

Grembergen, W., V., 2000. The Balanced Scorecard and IT Governance. *Information Systems Control Journal, Volume 2*.

Abran, A., Buglione, L., 2000. Balanced scorecards and GQM :What are the differences?. *FESMA-AEMES Software Measurement Conference*.

Mair, S., November/December 2002. A Balanced Scorecard for a Small Software Group. *IEEE Software, Volume 19, Issue 6.*

Bianchi, A., Caivano, D., Visaggio, G., 2000. Method and Process for Iterative Reengineering Data in a Legacy System. Proc. *Working Conf. on Reverse Engineering, pp. 86-96*.

Bianchi, A., Caivano, D., Marengo, V., Visaggio, G., 2001. Iterative Reengineering of Legacy Functions. *Proceedings of IEEE International Conference Software Maintenance, pp. 632-641*.

Bianchi, A., Caivano, D., Marengo, V., Visaggio, G., March 2003. Iterative Reengineering of Legacy Systems. *Software Engineering. IEEE Transactions Volume: 29, Issue: 3: 225- 241.*

SERLAB (Software Engineering Research LABoratory), 2006. Retrieved from http://193.204.187.180:8080/frame/controller_accessi