Available online at www.sciencedirect.com

ScienceDirect

Computers in Industry 59 (2008) 420–437

COMPUTERS IN INDUSTRY

www.elsevier.com/locate/compind

Evaluating workflow process designs using cohesion and coupling metrics

Irene Vanderfeesten *, Hajo A. Reijers, Wil M.P. van der Aalst

Technische Universiteit Eindhoven, Department of Technology Management, PO Box 513, NL-5600 MB Eindhoven, The Netherlands

Received 9 January 2007; received in revised form 5 December 2007; accepted 13 December 2007
Available online 4 March 2008

Abstract

Building on the similarities between software programs and workflow processes, this paper proposes a heuristic that offers guidance for the creation and evaluation of process designs in administrative settings. Designers can use this heuristic to select from several alternatives the process design that is strongly cohesive and weakly coupled. It is argued that such a design will result in fewer errors during information exchanges and in more understandable activity descriptions. The paper includes an application of the heuristic in an industrial workflow setting, which supports its feasibility and practical value. The paper also presents the freely available *CoCoFlow* tool that implements the heuristic and its associated metrics.

Keywords: Business process modelling; Workflow management; Quality metrics

1. Introduction

In 2006, one of the world's largest asset managing companies redesigned its business process to produce the annual reports for its investment funds. On the basis of a meticulous analysis of the essential information processing function of the process, a new process design was developed using the method of Product-Based Workflow Design (PBWD) [45]. The new design reshuffles the order of the original steps and exploits the opportunities of an automated, XML-based message exchange between investment managers, accountants, print shop, etc. It is anticipated that this new way of working will cut the cycle time of the process by half. Not only does this imply a better service to the company's customers, e.g. corporate investors and pension funds, but also it facilitates compliance with deadlines from the financial authorities.

This case illustrates that the combination of *process redesign* and the *application of IT* are viable today, even though Business Process Reengineering (BPR) was introduced over a decade ago [25]. Particularly in administrative settings, where the focus is on information processing (e.g. the evaluation of a damage claim, the issuing of a building permit, or the handling of a mortgage application), various redesign opportunities exist [39]. For example, by using electronic documents and workflow technology it is relatively easy to change the routing of a file along the various decision-making steps in a process. When doing this properly, the average execution cost of the overall process can be minimized [3].

However, coping with such design freedom may also be problematic. The main issue that we focus on in this paper is the proper size of the individual *activities* (or *tasks*) in a process. This design choice is known as the issue of *process granularity* [18]. Badly chosen sizes of activities in a process may negatively affect its performance when being executed or enlarge the maintenance burden of the process model in case of updates. Small activities, on the one hand, may increase the number of hand-offs between activities leading to an increase of errors [42,49]. Large activities, on the other hand, may become unworkable to be executed well by humans [18,42].

This paper addresses the problem of activity design in the domain of information-intensive processes, typically found in the service industry. This issue is particularly relevant within the setting of BPR projects [26] where also the application is being considered of workflow management technology [5]. In the remainder of this paper, we shall refer to such processes as *workflow processes* or simply as *workflows*. We will present *cohesion* and *coupling metrics* that focus on the content of activities, i.e. their *operations*. By using the proposed set of metrics, it can be quantitatively expressed to what extent operations "belong" to each other within one activity or, in

* Corresponding author. Tel.: +31 40 247 4366; fax: +31 40 243 2612.
E-mail addresses: i.t.p.vanderfeesten@tue.nl (I. Vanderfeesten), h.a.reijers@tue.nl (H.A. Reijers), w.m.p.v.d.aalst@tue.nl (W.M.P. van der Aalst).

0166-3615/$ – see front matter
doi:10.1016/j.compind.2007.12.007

other words, how *cohesive* such an activity is. In addition, it is important to measure to what extent the various activities are dependent on each other or, in other words, how much they are *coupled*. The inspiration for the proposed metrics comes from software engineering, where an old design aphorism is to strive for *strong cohesion, and loose coupling*.

We build on work by Selby and Basili [50] and Xenos et al. [59], who defined coupling and cohesion metrics for software programs. The proposed solution based on comparable metrics fits conceptually with the PBWD method, as applied in the case we mentioned in the opening. PBWD is a BPR method which is adopted and actively applied by one of the Big Four consultancy firms [45]. For a more detailed application of the method in the financial services, see ref. [41].

In this paper, findings from our earlier work [43,46] are integrated and extended in several directions. A first version of the workflow cohesion metric was introduced in ref. [43], but it lacked some facilities. For example, it could not handle conditional alternatives to achieve the same output (XOR-construct), which is a construct very common in workflow processes [6]. Secondly, we introduced in ref. [46] an additional coupling metric, but the integrated design heuristic did not yet consider resource issues that in real life impose constraints on good designs. Thirdly, in our experiments we noticed that the manual application of the metrics can be time-consuming and may easily lead to human errors. Therefore, this paper introduces the freely available *CoCoFlow* tool that can be used for the automatic computation of the metrics. Finally, we applied the integrated design heuristic to the case of a Dutch governmental agency to demonstrate its feasibility and practical value.

The structure of this paper is as follows. In Section 2, we introduce a motivating example that clarifies the goal of our research. Section 3 describes the cohesion and coupling metrics and the heuristic for selecting a favorable design among various alternatives. The fourth section contains an application in an industrial setting, which is followed by a description of the *CoCoFlow* tool in Section 5. In Section 6, we elaborate on related work in the fields of software engineering and workflow to position our contribution. The paper ends with a discussion and prospects for future work.

2. Motivating example

To motivate the application of cohesion and coupling metrics on the one hand and to introduce some relevant concepts on the other, we present in this section a workflow process that deals with requests for governmental student grants in the Netherlands. The presented process is a simplified version of the actual procedure as implemented by the Informatie Beheer Groep[1](IBG) under the authority of the Dutch Ministry of Education, Culture and Science. We will first present the essential information processing function that the workflow process must perform. Then, we will illustrate the design dilemma that we focus on in this paper, by representing three alternative process designs.

2.1. An information processing perspective

As in many other workflow processes, the essential output of the IBG workflow process can be clearly distinguished. In this case, it is the calculation of the government grant that any new student can apply for. The proper amount of money to be granted must be determined by taking several factors into account: the student's background, the income of the student's parents, the student's living situation, the kind of study grant the student applied for, etc. Each of these factors in turn may rely on other factors.

We shall refer to the various relevant factors as *information elements*. In the activities of a workflow process the values of information elements are used to produce new values for other information elements. In the end, the essential output is produced as a set of information element values.[2]

An activity in a workflow process will now be considered as a number of *operations* on information elements. Each operation has one or more *input* information elements and one or more *output* information elements. An operation is an atomic processing step, so it cannot have ‘‘half-assembly’’ products, i.e. intermediate information element values. A workflow activity, on the contrary, is composed of one or more operations. Therefore, it can have ‘‘half-assembly’’ products. The problem we raised in Section 1 can now be reformulated as *the proper clustering into activities of operations on information elements*. In Fig. 1, the complete *information element structure* of the IBG process can be found. At the top, information element 42 can be distinguished, which represents the total grant amount that is decided to be assigned to the applicant. (Note that a complete description of the individual information elements is provided in Appendix A (see Table A.1).) The information element structure expresses that a value for information element 42 can be determined on the basis of values for information elements 39, 40, and 41, which respectively represent the granted amount of supplementary scholarship, basic scholarship, and the loan component. This relation between information elements 39, 40, 41, and 42 is expressed by arrows leading from the input elements, ‘‘knotted’’ together, to the output element. The operation to be performed on the input elements in this case is to simply add their values.[3]

Considering the information element structure for this case again (Fig. 1), it is interesting to note that a value for information element 42 can also be computed in a different way. This different computation is based on the value of information element 27 alone, which represents the outcome of the decision whether the respective student may receive a government grant at all. If this value is determined to be ‘‘no’’ – which is the case if the student is older than thirty or does not

[1] http://www.ib-groep.nl.

[2] Typically, one value is of prime interest.

[3] Note that the content of operations is not expressed in this figure.