Designing like a Pro: The automated composition of workflow activities

Han van der Aa [a,*], Hajo A. Reijers [a,b], Irene Vanderfeesten [c]

[a] Department of Computer Sciences, Vrije Universiteit Amsterdam, Faculty of Sciences, De Boelelaan 1081, 1081 HV Amsterdam, The Netherlands
[b] Department of Mathematics and Computer Science, Eindhoven University of Technology, PO Box 513, 5600 MB Eindhoven, The Netherlands
[c] School of Industrial Engineering, Eindhoven University of Technology, PO Box 513, 5600 MB Eindhoven, The Netherlands

ABSTRACT

To design a workflow process that is efficient, meaningful, and understandable it is important to properly consider how to compose the activities it will consist of. In this paper, guidelines are presented for this exact purpose. These focus on the elementary data-processing steps that are at the core of a workflow process. The guidelines help to determine the relative importance of these data-processing steps as well as their relatedness, such that activities can be composed in a fully automated manner. We implemented this approach in freely available software. A thorough evaluation that incorporates real-life workflow designs indicates that the use of these guidelines leads to activities that closely resemble those designed by experienced modelers. As such, the proposed guidelines provide a proper and automated alternative to what is otherwise a complex and time-consuming task.

© 2015 Elsevier B.V. All rights reserved.

1. Introduction

Data-centric business processes pursue the production of an informational product. These processes, which are here referred to as *workflows*, are often found in the service domain. Typical informational products one can think of are, e.g. a mortgage contract, a decision on an insurance claim, or a personalized commercial offer. The structure of a workflow is to a large extent determined by the data-flow underlying the process. The data-flow comprises numerous elementary data processing steps, which all contribute to the ultimate computation of a desired end result. When designing a workflow, these elementary data processing steps are grouped into *activities*. Activities should be constructed in such a way that each of these represents a *logical piece of work* within a process [1]. An activity may or may not comprise multiple elementary processing steps. Creating a large activity, which is composed of many data processing steps, may often be preferable over including the elementary processing steps as separate activities. For example, the activity of calculating a mortgage amount may consist of entering the current interest rate, choosing the discount rate negotiated by the customer, and calculating the amortized amount of debt. Individually, each of these three steps may appear insignificant and overly fine-grained, while an activity that combines all three denotes an important and recognizable part of a workflow. In this paper, our attention is with the grouping of elementary data processing steps into activities. We will refer to this act as *activity composition*.

Properly carrying out activity composition is important, because the size and contents of activities affect three aspects of a process design: execution efficiency, experienced meaningfulness, and model understandability. First, activities can influence the *execution efficiency* of a process. Namely, activities that have a proper size, i.e. are of the right granularity, provide a balance between an increased number of work hand-overs that results from many small activities, against reduced flexibility caused by too many large activities [2,3]. Furthermore, incorrectly composed activities can result in the redundant or unnecessarily delayed execution of tasks. Second, activity composition affects the *experienced meaningfulness* of activities for those employees that execute these [4]. This aspect is associated with job satisfaction, motivation, and work performance [5,6]. As a result, more meaningful tasks can result in increased productivity. Finally, the design of activities also affects the *understandability* of process models. By grouping elementary steps into larger activities, the number of process model elements is reduced, which results in an increased understandability of the process [7]. Furthermore, activity designs that emphasize the important process steps, as well as properly depict the general process flow, can provide useful

* Corresponding author. Tel.: +31 20 59 87788.
  E-mail address: j.h.vander.aa@vu.nl (H. van der Aa).

insights into the most important aspects of a process. Constructing such a quick process overview has been found to be a highly demanded use case for Business Process Model Abstraction [8].

Due to the impact that activity design can have on the quality of a workflow design, proper activity composition is a highly desirable task. Despite this importance, there exists a lack of support for this task [9]. Earlier work proposed metrics that objectively evaluate the quality of activity designs on the basis of job design insights [3,4], but these can only be retrospectively applied on activities already composed. Without support, activity composition requires expertise and case knowledge. Expertise is required, because one must be familiar with, for example, concepts of task and process design to ensure that activities are of the proper size, represent meaningful steps, and do not negatively affect execution efficiency. A modeler must furthermore be well-acquainted with the elementary information processing steps of a particular workflow. Due to the potential number of steps and possible complexity of their inter-relations, acquiring and applying the required case knowledge when designing activities can become a time-consuming task.

In this paper we set out to limit the expertise and domain knowledge that is required to properly compose workflow activities, whilst simultaneously reducing the time-intensiveness of this task. We achieve this by the provision of composition guidelines, that can be used to objectively construct *well-designed activities*. These guidelines, which extend a preliminary set introduced in [10], exploit the structural data-flow relations in a workflow. To make our ideas operational, we here capture such relations in a Product Data Model (PDM). However, the guidelines can be easily transferred to comparable data-flow specifications, such as the data flow matrices of [11]. The PDM that we build on stems from Product Based Workflow Design (PBWD), a methodology for the radical redesign of workflows [12]. We propose that data-flow relations, captured in the structure of a PDM, can be used to determine the *semantic relatedness* and *relative importance* of elementary information processing steps. These notions form the basis for the proposed composition guidelines, which result in activities that (i) are *meaningful* to workflow users executing them, (ii) form the basis for *understandable* process models, and (iii) do not result in *redundant* or *unnecessarily delayed* execution of tasks. Since the guidelines only consider the structural properties of a PDM, they can be applied without specific domain knowledge and can therefore be fully automated. This is demonstrated through an example implementation, which is freely available and can be used to automatically generate activities for any PDM. To illustrate the usefulness of our approach, we performed a thorough evaluation: We compare activity designs that have been automatically generated according to the propositions with designs that have been manually composed by experienced modelers.

The remainder of this paper is structured as follows. Section 2 introduces the PBWD methodology as well as a running example. The example illustrates the act of activity composition and its impact on workflow design. Section 3 describes the composition guidelines and illustrates how these can be applied in an automated fashion. Next, Section 4 shows the application of the proposed guidelines on a real-world business process. In Section 5, the performance of the guidelines is evaluated by comparing automatically generated activity designs with designs composed by experienced modelers. Section 6 discusses limitations of the current approach and directions for future research. Section 7 considers related work, after we conclude this paper with Section 8.

2. Activity composition

The composition guidelines are presented in this paper in the context of Product Based Workflow Design. This method is described in Section 2.1. Before the guidelines are proposed, we first provide a running example (Section 2.2) and demonstrate the impact that activity composition can have on the quality of process designs (Section 2.3).

2.1. Product Based Workflow Design

Product Based Workflow Design (PBWD) is a business process (re)design method [13]. It is one of the data-centric process (re)design methods that have emerged over the past decade to counter the overwhelming number of activity-oriented process modeling languages. Other well-known data-centric methods are *large process structures* by Müller et al. [14] and *artifact centric* process models introduced by Nigam and Caswell [15].

The PBWD method takes the informational product that is produced in the business process as the central concept. This product is built from several information processing steps, similar to how a Bill-of-Material (BoM) in [16] in manufacturing describes the composition of a physical product from its materials. The informational end product and its decomposition into sub data elements and input data elements is modeled in an hierarchical model, which is called the Product Data Model (PDM) [9]. This model describes the data elements and their relationships.

Fig. 1 contains a small and simple example of a PDM. It describes the calculation of the maximum amount of mortgage a bank is willing to provide to a client. The figure shows that there a three ways to compute the maximum mortgage amount (element *A* in Fig. 1). This amount is based on either (i) a previous mortgage offer (*E*), (ii) on the registration in the central credit register (*H*), or (iii) on the combination of the percentage of interest (*B*), the term of the mortgage (*C*), and the annual budget to be spent on the mortgage (*D*). The annual budget (*D*) is determined from the gross
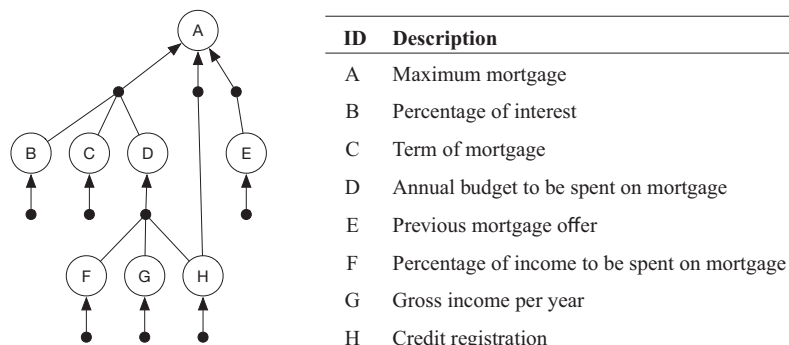

ID	Description
A	Maximum mortgage
B	Percentage of interest
C	Term of mortgage
D	Annual budget to be spent on mortgage
E	Previous mortgage offer
F	Percentage of income to be spent on mortgage
G	Gross income per year
H	Credit registration

Fig. 1. Small example of a Product Data Model describing the computation of a mortgage.