Available online at www.sciencedirect.com
ScienceDirect
Simulation Modelling Practice and Theory 15 (2007) 1068–1088
SIMULATION MODELLING PRACTICE AND THEORY
www.elsevier.com/locate/simpat

On design, verification, and dynamic modification of the problem-based scientific workflow model

Liu Xiping [a,b], Dou Wanchun [a,*], Chen Jinjun [c], Fan Shaokun [a], S.C. Cheung [d], Cai Shijie [a]

[a] State Key Laboratory for Novel Software Technology, Department of Computer Science and Technology, Nanjing University, Nanjing, China
[b] College of Computer, Nanjing University of Posts and Telecommunications, Nanjing, China
[c] Faculty of Information and Communication Technology, Swinburne University of Technology, Melbourne, Australia
[d] Department of Computer Science and Engineering, Hong Kong University of Science and Technology, Hong Kong

Received 30 September 2006; received in revised form 25 June 2007; accepted 26 June 2007
Available online 1 July 2007

Abstract

A science process is a process to solve complex scientific problems which usually have no mature solving methods. Science processes if modeled in workflow forms, i.e. scientific workflows, can be managed more effectively and performed more automatically. However, most current workflow models seldom take account of specific characteristics of science processes and are not very suitable for modeling scientific workflows. Therefore, a new workflow model named problem-based scientific workflow model (PBSWM) is proposed in this paper to accommodate those specific characteristics. Corresponding soundness verification and dynamic modification are discussed accordingly based on the new modelling method. This paper makes three main contributions: (1) three new constructs are proposed for special logic semantics in science processes; (2) verification is deployed with the consideration from both data-specific perspective and control-specific perspective; and (3) a set of rules are provided to automatically infer passive modifications caused by other modifications.

Keywords: Scientific workflow; Modelling; Verification; Dynamic modification; Context data

1. Introduction

With the fast development of computer technologies, very large-scale complex science processes which cannot be deployed in the past can now be explored. A science process can be viewed as a process to solve a scientific problem with goals such as knowledge discovery, knowledge innovation, and so on. A large-scale complex science process needs to handle very complicated logics in scientific problems. Besides, in many cases,

* Corresponding author. Tel./fax: +86 25 8368 6399.
E-mail addresses: xpliunju@hotmail.com (X. Liu), douwc@nju.edu.cn (W. Dou).

1569-190X/$ - see front matter © 2007 Elsevier B.V. All rights reserved.
doi:10.1016/j.simpat.2007.06.003

such a process involves a great number of scientists from different domains as well as distributed resources. To some extent, deploying such a science process is a huge engineering which requires automatic and reasonable management to the whole process. Fortunately, workflow technologies [24] provide useful ideas and paradigms for the automation of science processes.

Workflow was originally emerged for the automation of a business process in whole or part, where a business process means a kind of process in the domain of business organizational structure and policy. The definition of a workflow consists of a coordinated set of activities that are connected in order to achieve a common goal. These activities can be organized in various routing forms such as sequential, parallel, and so on. Many workflow instances may be created and executed based on a workflow definition, and this process is deployed automatically under the control of workflow engines [14].

Many ideas and concepts from workflow technologies can be adapted to science processes, such as the idea of organizing a process as a workflow; the idea of using conditions to control automatic transitions between activities; the concept of building blocks, e.g. and-join, or-split, used for the representation of activity routings; and so on. More specifically, a science process can be modeled as a workflow before it is deployed. Such a workflow definition organizes the whole complex science process as a coordinated set of activities with data relationships and control relationships. The workflow should be verified according to certain rules or strategies. Then workflow engines control the deployment of the science process automatically based on the verified workflow. In this way, the science process can be deployed effectively and efficiently through a workflow form. Consequently, the scientific workflow is proposed to represent science processes in workflow forms, which can facilitate the automatic and effective performance of science processes.

Though the scientific workflow roots in the traditional workflow i.e. workflow for automation of business processes, it is unsuitable to apply workflow models designed for business processes to the definitions of scientific workflows directly. This is because compared with business processes, science processes hold many characteristics which cannot be satisfied by current workflow models. Science processes differ a lot from business processes, where a science process is a tentative solving process without foregone stable solving schemes while a business process is a comparatively fixed procedure with a specific behavior description for each logic step. The detailed differences between business processes and science processes are analyzed as follows.

First, science processes are more data-centric and knowledge-intensive [18], and require more powerful computing ability and mental thinking behaviors. In contrast, business processes pay more attention to control specifications than data specifications. Most activities seem like fixed mechanical operations.

Second, activities in science processes perhaps are indeterminate at the definition stage because mature solving schemes usually do not exist. Therefore, different from business processes, science processes generally cannot be formalized as a stable series of ordered activities with specific behavior and fixed input or output specifications.

Third, science processes are often highly creative, innovative and dynamic. The predefined scientific workflow models are often modified a lot at run-time execution stage, while business processes are comparatively stable.

Fourth, a scientific workflow model is often modified at run-time execution stage because of indeterminacy and high dynamics. Hence, a scientific workflow model is high likely to be performed only one time while a business workflow model is usually executed repeatedly and has many instances.

Fifth, participants of a science process, i.e. scientists, play an important role in the whole science process and maybe are the key factor to decide whether a scientific problem can be solved successfully. On the contrary, what participants do in business processes is comparatively simple interactive behaviors.

Due to essential characteristics of science processes analyzed above, science processes are hardly organized in traditional workflow forms which consist of definite and concrete descriptions on each step, behavior, or task. Here, these traditional workflow forms mean Petri Nets based models [1,9,15,16], UML based models [5,8,11], and so on. It is necessary to design a new workflow model which can really accommodate characteristics and laws of science processes. Unfortunately, although more and more researchers have realized the significance of applying workflow technologies to the deployment of large-scale complex science processes, most of their efforts have been put on the development of special scientific workflow management systems for science processes, such as Pegasus [10], Kepler [4], and GridNexus [7]. Up to now, design of scientific workflow models fully satisfying characteristics of science processes has not attracted enough attention.