Workflow Reduction for Reachable-path Rediscovery in Workflow Mining

Kwang-Hoon Kim[1] and Clarence A. Ellis[2]

[1] Collaboration Technology Research Lab.
Department of Computer Science
KYONGGI UNIVERSITY
San 94-6 Yiuidong Youngtongku Suwonsi Kyonggido, South Korea, 442-760
kwang@kyonggi.ac.kr

[2] Collaboration Technology Research Group
Department of Computer Science
UNIVERSITY OF COLORADO AT BOULDER
Campus Box 0430, Boulder, Colorado, USA, 80309-0430
skip@cs.colorado.edu

Summary. This paper[3] newly defines a workflow reduction mechanism that formally and automatically reduces an original workflow process to a minimal set of activities, which is called minimal-workflow model in this paper. It also describes about the implications of the minimal-workflow model on workflow mining that is a newly emerging research issue for rediscovering and reengineering workflow models from workflow logs containing workflow enactment and audit information gathered being executed on workflow engine. In principle, the minimal-workflow model is reduced from the original workflow process by analyzing dependencies among its activities. Its main purpose is to minimize discrepancies between the modeled workflow process and the enacted workflow process as it is actually being executed. That is, we can get a complete set of activity firing sequences (all reachable-paths from the start to the end activity on a workflow process) on buildtime. Besides, we can discover from workflow logs that which path out of all reachable paths a workcase (instance of workflow process) has actually followed through on runtime. These are very important information gain acquiring the runtime statistical significance and knowledge for redesigning and reengineering the workflow process. The minimal-workflow model presented in this paper is used to be a decision tree induction technique for mining and discovering a reachable-path of workcase from workflow logs. In a consequence, workflow mining methodologies and systems are rapidly growing and coping with a wide diversity of domains in terms of their applications and working environments. So, the literature needs various, advanced, and specialized workflow mining techniques and architectures that are used for finally feed-backing their analysis results to the redesign and reengineering phase of the existing workflow and business

[3] This research was supported by the University Research Program (Grant No. C1-2003-03-URP-0005) from the Institute of Information Technology Assessment, Ministry of Information and Communications, Republic of Korea.

process models. We strongly believe that this work might be one of those impeccable attempts and pioneering contributions for improving and advancing the workflow mining technology.

1 Introduction

In recent, workflow (business process) and its related technologies have been constantly deployed and so gradually hot-issued in the IT arena. This atmosphere booming workflows and business processes modeling and reengineering is becoming a catalyst for triggering emergence of the concept of workflow mining that collects data at runtime in order to support workflow design and analysis for redesigning and reengineering workflows and business processes. Especially real workflow models going with e-Commerce, ERP (Enterprise Resource Planning), and CRM (Customer Relationship Management) are getting larger and more complex in their behavioral structures. These large-scaling movements trigger another big changes in workflow administration (the responsibility of redesign) and monitoring (the responsibility of rediscovery) functionality that has been featured and embedded in the workflow build-time and run-time functionality of the traditional workflow systems. The more a workflow system's architecture is distributed, the more its administrative functions ought to be extended and play important roles for improving the integrity of the workflow process models and systems. At the same time, the more transactional applications are involved in workflow models, the more workflow monitoring features are closely related with statistic-oriented workflow monitoring information. Meanwhile, in the traditional workflow systems, the workflow monitoring features generate status-oriented workflow monitoring information.

In other words, there have been prevalent research and development trends in the workflow literature - workflow mining techniques and systems that collect runtime data into workflow logs, and filter out information and knowledge from the workflow logs gathered by the administration and monitoring features of workflow management system. The workflow mining techniques and systems tend to have completely distributed architectures to support very large-scale workflow applications based upon object-oriented and internet-based infrastructures. Their key targets have been transformed from the passive, centralized, human-oriented, and small/medium scale workflow process models and systems to the active (object), distributed (architecture), system-oriented (transaction), and large-scale (application) workflow process models and systems. So, in order for WfMSs to slot in workflow mining features, it is necessary for their administration and monitoring features to be extended to gathering and analyzing statistical or workload status information of the workflow architectural components dispersed on the distributed environment, and performing feedback the analyzed results to the redesigning and reengineering phase of workflow process models. The advanced workflow management