Chapter 11

Mining E-Mail Messages: Uncovering Interaction Patterns and Processes Using E-Mail Logs

Wil M.P. van der Aalst
Eindhoven University of Technology, The Netherlands

Andriy Nikolov
Knowledge Media Institute, The Open University, UK

ABSTRACT

Increasingly information systems log historic information in a systematic way. Workflow management systems, but also ERP, CRM, SCM, and B2B systems often provide a so-called “event log”, i.e., a log recording the execution of activities. Thus far, process mining has been mainly focusing on structured event logs resulting in powerful analysis techniques and tools for discovering process, control, data, organizational, and social structures from event logs. Unfortunately, many work processes are not supported by systems providing structured logs. Instead very basic tools such as text editors, spreadsheets, and e-mail are used. This paper explores the application of process mining to e-mail, i.e., unstructured or semi-structured e-mail messages are converted into event logs suitable for application of process mining tools. This paper presents the tool EMailAnalyzer, embedded in the ProM process mining framework, which analyzes and transforms e-mail messages to a format that allows for analysis using our process mining techniques. The main innovative aspect of this work is that, unlike most other work in this area, our analysis is not restricted to social network analysis. Based on e-mail logs we can also discover interaction patterns and processes.

INTRODUCTION

Buzzwords such as BAM (Business Activity Monitoring), BOM (Business Operations Management), BPI (Business Process Intelligence) illustrate the interest in closing the business process management loop (van der Aalst and van Hee, 2002; Dumas et al., 2005). This is illustrated by the Figure 1, which shows the increasing level of support for closing the so-called BPM lifecycle.

The lifecycle identifies four different phases: process design (i.e., making a workflow schema), system configuration (i.e., getting a system to support the designed process), process enactment...
Mining E-Mail Messages

Figure 1. The level of support is rising - closing the business process management (BPM) cycle

(i.e., the actual execution of the process using the system), and diagnosis (i.e., extracting knowledge from the process as it has been executed). BPM technology (e.g., workflow management systems) started with a focus on getting the system to work (i.e., the system configuration phase). Since the early nineties BPM technology matured and more emphasis was put on supporting the process design and process enactment phases in a better way. Now most vendors are trying to close the BPM lifecycle by adding diagnosis functionality.

The diagnosis phase assumes that data is collected in the enactment phase. Most information systems provide some kind of event log (also referred to as transaction log or audit trail). Typically such an event log registers the start and/or completion of activities. Every event refers to a case (i.e., process instance) and an activity (i.e., the step in the process executed), and, in most systems, also a timestamp, a performer, and some additional data. Process mining techniques (van der Aalst et al., 2003; van der Aalst et al., 2004; Agrawal et al., 1998; Cook and Wolf, 1998; Herbst, 2000; de Medeiros et al., 2003; Weijters and van der Aalst, 2003) take an event log as a starting point to extract knowledge, e.g., a model of the organization or the process. In the context of our ProM tool (van der Aalst et al., 2007) we are able to extract different types of process models (e.g., Petri nets, event-driven process chains, and instance graphs), social networks, organizational models, etc.

Existing techniques for process mining assume an event log to be in place. For many process-aware information systems (Dumas et al., 2005) this assumption is valid. For example, Workflow Management (WFM) systems, Enterprise Resource Planning (ERP) systems, Customer Relationship Management (CRM), Case Handling (CH) and Product Data Management (PDM) systems log information in some transaction log or audit trail. New legislation such as the Sarbanes-Oxley (SOX) Act (Sarbanes and Oxley, 2002) and increased emphasis on corporate governance have triggered the need for improved auditing systems (Hoffman, 2004). To audit an organization, business activities need to be monitored. As enterprises become increasingly automated, a tight coupling between auditing systems and the information systems supporting the operational