ABSTRACT

Software Process Intelligence (SPI) is an emerging and evolving discipline involving mining and analysis of software processes. This is modeled on the lines of Business Process Intelligence (BPI), but with the focus on software processes and its applicability in software systems. Process mining consists of mining event log and process trace data for the purpose of process discovery (run-time process model), process verification or compliance checking (comparison between design-time and run-time process model), process enhancement and recommendation. Software Process Mining or Intelligence is a new and emerging discipline which falls at the intersection of Software Process & Mining, and Software & Process Mining. Software Process Mining is integral to discovering and verifying the processes in a software system.

Software Process Mining is a three word phrase which can be viewed from two perspectives: Software + Process Mining and Software Process + Mining. Software development and evolution involves usage of several workflow management and information systems and tools such as Issue Tracking Systems (ITS), Version Control Systems (VCS), Peer Code Review Systems (PCR) and Continuous Integration Tools (CIT). Such information systems log data consisting of events, activities, time-stamp, user or actor and context specific data. For example, a popular open-source ITS such as Bugzilla logs the submission time of a bug report, user id and name of the bug reporter, title and description of the bug report in addition to several other information. Any change made to the bug report during its life-cycle is logged. Mining this information could provide more insight into the processes involved and also help in predictive analysis. Such events or trace data generated by information systems used during software construction (as part of the software development process) contains valuable information which can be mined for gaining useful insights and actionable information.

The tutorial will be of half-day (3-4 hours) duration and the level will be intermediate. The target audiences for the tutorial will be industry practitioners and researchers in the area of data mining, process mining, software analytics and mining software repositories. The pre-requisites for the tutorial is basic background in data mining and software engineering. The tutorial will be divided into 3 parts and cover topics such as: fundamentals of process mining, familiarity with open-source process mining framework ProM, basics of Business Process Modeling Notation (BPMN), overview of mining software repositories and software analytics, understanding of common software repositories and archives, important applications of mining software repositories, basics of process mining software repositories, techniques and applications.
KEYWORDS
Software Process Intelligence, Mining Software Repositories, Software Engineering, Business Process Intelligence, Process Mining, Software Archives, Software Processes, Software Analytics, Software Intelligence

DURATION Half-Day (3-4 Hours)

LEVEL Intermediate

TARGET AUDIENCES
Industry practitioners and researchers, Academics in the area of Software Analytics, Software Intelligence, Mining Software Repositories and Process Mining

PRE-REQUISITITES
Basic background in Software Engineering and Data Mining

LEARNING OUTCOME
- Fundamentals of process mining
- Familiarity with open-source process mining framework ProM
- Basics of Business Process Modeling Notation (BPMN) and Petri-Nets
- Overview of mining software repositories and software analytics
- Understanding of common software repositories and archives
- Important applications of mining software repositories
- Basics of software process intelligence and process mining software repositories
- Techniques and applications of software process intelligence
- Research directions in the area of software process intelligence

TUTORIAL OUTLINE AND DESCRIPTION
Table 1, 2 and 3 displays the tutorial outline. The tutorial is divided into 3 parts. Part 1 is on Process Mining (refer to Table 1). Part 2 and 3 are on Mining Software Repositories and Process Mining Software Repositories respectively (refer to Table 2 and 3). Table 1, 2 and 3 shows the sub-topics and the duration for each of the topic. In Part 1 of the tutorial, we will give an introduction to process modeling and cover basics of process modeling, business process modeling notation, process discovery, conformance checking and mining organizational perspective. We will discuss the goals and purpose of a process model. We will introduce the audiences to BPMN which is a graphical representation for specifying business processes in a business process model. Process conformance checking is an important topic and will be covered in the initial 30-45 minutes. Process conformance checking is about comparing the design time process model with run-time process model (derived from event logs) and measuring the extent of conformance. In addition to introducing the audiences to control flow perspective, we may talk about organizational perspective (optional topic) depending on the available time [5][6].
**Table 1: Part 1 - Process Mining**

<table>
<thead>
<tr>
<th>Part 1</th>
<th>Process Mining</th>
</tr>
</thead>
</table>
| 1 Introduction to Process Mining | Process Modeling  
Business Process Modeling Notation (BPMN)  
Process Discovery  
Conformance Checking and Verification  
Mining Organizational Perspective | 45 Minutes |
| 2 ProM Demo | Installation and Overview  
Throughput time of cases  
Paths and their time  
Service time of tasks  
Time spent between tasks  
Case execution | 30 Minutes |
| 3 DISCO Demo | Automated process discovery  
Demo Statistics and charts  
Case explorer | 25 Minutes |

ProM (http://www.processmining.org/) is an extensible framework that supports a wide variety of process mining techniques in the form of plug-ins. It is platform independent as it is implemented in Java, and can be downloaded free of charge. We will conduct a demo and show how to get started with ProM. We will cover ProM user interface and explain the basic features of the user interface by explaining the different objects in this interface. We will then show how to use ProM to answer some of the common questions that managers have about processes in organizations. The tutorial will cover how users can inspect and clean (or pre-process) an event log in ProM. The tutorial will also cover how to mine the control flow perspective of process models and how to mine information regarding certain aspects of cases. DISCO (http://fluxicon.com/disco/) is a commercial process mining tool. Depending on the available time, we will demo automated process discovery, process map animation and computing detailed statistics from event logs using DISCO.

**Table 2: Part 2 - Mining Software Repositories**

<table>
<thead>
<tr>
<th>Part 2</th>
<th>Mining Software Repositories</th>
</tr>
</thead>
</table>
| 1 Software Repositories and Archives | Version Control Systems  
Issue Tracking Systems  
Peer Code Review System  
Mail Archives  
Source Code Repositories | 30 Minutes |
| 2 Overview of Mining Software Repositories | Duplicate Bug Report Detection  
Fault Localization  
Effort and Contribution Estimation  
Automatic Triaging  
Defect Proneness | 30 Minutes |

Mining Software Repositories (MSR) and Software Analytics is an emerging field and has attracted several researcher's attention. We will give an overview of various software repositories such as Version...
Control Systems (VCS), Issue Tracking System (ITS), Peer Code Review Systems (PCR), Mail Archives and Source Code Repositories. Then we will cover some the important applications of MSR such as Duplicate Bug Report Detection, Fault Localization, Effort and Contribution Estimation, Automatic Triaging, Code Clone Detection and detecting Defect Prone areas in the code [1][7]. The last section (Part 3) consists of software process intelligence and process mining software repositories (intersection of process mining and mining software repositories). We will cover topics such as: software development processes, technical challenges in process mining software repositories and some of the applications of process mining software repositories [2][3][4]. In the end we will talk about the research and future research directions in software process intelligence.

Table 3: Part 3 - Software Process Intelligence

<table>
<thead>
<tr>
<th>Part 3</th>
<th>Software Process Intelligence</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Software Process Intelligence</td>
<td>1. Software Development Processes Technical Challenges in Software Process Intelligence</td>
</tr>
<tr>
<td></td>
<td>2. Applications of Software Process Intelligence</td>
</tr>
<tr>
<td></td>
<td>3. Industrial Case Studies</td>
</tr>
<tr>
<td></td>
<td>4. Research and Future Directions</td>
</tr>
</tbody>
</table>

60 Minutes

BRIEF BIO OF PRESENTERS

Ashish Sureka is a Faculty Member at Indraprastha Institute of Information Technology, Delhi (IIIT-D). His current research interests are in the area of Mining Software Repositories, Software Analytics, and Social Media Analytics. He graduated with an MS and PhD degree in Computer Science from North Carolina State University (NCSU) in May 2002 and May 2005 respectively. He has worked at IBM Research Labs in USA, Siemens Research Lab (India) and was a Senior Research Associate at the R&D Unit of Infosys Technologies Limited before joining IIIT-D in July 2009. He has received research grants from Department of Information Technology (DIT, Government of India), Confederation of Indian Industry (CII) and Department of Science and Technology (DST, Government of India). He has published several research papers in international conferences and journals, graduated several PhD and MTech students, organized workshops co-located with conferences, and received best paper awards. He was selected for ACM India Eminent Speaker Program. He holds seven granted US patents.

Atul Kumar is a Senior Researcher at Siemens Corporate Research and Technology. Before that he was at ABB Corporate Research, India. He has worked at IBM Research, Microsoft and Accenture Technology Labs. His research interest are in the areas of Distributed Systems, Software Engineering, Internet Technologies and Data Engineering. He has co-organized workshops and special sessions related Software Engineering and Cloud Computing at various conferences ICSE, ISEC, i-Society etc. He is serving as tutorials co-chair at ICIS 2014. Atul holds a Master’s degree and a PhD in Computer Science from IIT Kanpur. Atul is a senior member of both IEEE and ACM.

Shrinath Gupta is a Technical Expert at Siemens Corporate Research and Technology. His research interests include Code, Design, Architecture Assessment and Analysis, Process and Defect Analysis, and Software Analytics. He has published papers in top conferences like ICSE, ICPC, and ISEC. Shrinath holds a Master’s degree in Computer Science and Engineering from IIT Bombay, India.


