Chapter 4.4
Ontological Description and Similarity-Based Discovery of Business Process Models

Khalid Belhajjame
University of Manchester, UK

Marco Brambilla
Politecnico di Milano, Italy

ABSTRACT

Project repositories are a central asset in software development, as they preserve the knowledge gathered in past development activities. Locating relevant information in a vast project repository is problematic, because it requires manually tagging projects with accurate metadata, an activity which is time consuming and prone to errors and omissions. Just like any other artifact or web service, business processes can be stored in repositories to be shared and used by third parties, e.g., as building blocks for constructing new business processes. The success of such a paradigm depends partly on the availability of effective search tools to locate business processes that are relevant to the user purposes. A handful of researchers have investigated the problem of business process discovery using as input syntactical and structural information that describes business processes. This work explores an additional source of information encoded in the form of annotations that semantically describe business processes. Business processes can be semantically described using the so called abstract business processes. These are designated by concepts from an ontology which additionally captures their relationships. This ontology can be built in an automatic fashion from a collection of (concrete) business processes, and this work illustrates how it can be refined by domain experts and used in the discovery of business processes, with the purpose of reuse and increase in design productivity.

DOI: 10.4018/978-1-61350-456-7.ch4.4

INTRODUCTION

The last two decades showed that *business process modeling (BPM)* is the solution of choice of multiple companies and government institutions for describing and enacting their internal and external work procedures. Once modelled, business processes can be made available either publicly or accessible to a specific community to share the know-how between institutions and promote the reuse of existing business processes, e.g., as building blocks for constructing new business processes. The success of such a paradigm depends partly on the availability of a means by which users can locate business processes that are relevant for their purposes.

Another important trend in software development is towards reuse of artifacts and sharing of knowledge. Software models and code repositories play a central role in software development companies, as they accumulate the knowledge and best practices evolved by skilled developers over years. Besides serving the current needs of project development, they have also an archival value that can be of extreme importance in fostering reuse and the sharing of high quality design patterns. In several cases (e.g., the open source community), software project repositories have overcome the boundaries of individual organizations and have assumed a social role in the diffusion of coding and design solutions. State-of-the-practice project repositories mostly support source code or documentation search (Bajracharya, Ossher, & Lopes, 2009; Frakes & Nejmeh, 1987; Holmes & Murphy, 2005). Source code search engines (e.g., Google code, Snipplr, Koders) are helpful if the abstraction level at which development occurs is the implementation code. However, searching project repositories at the source code level clashes with the use of high level models like business process models as the principal artifact to express solutions and design patterns. Therefore, the question arises of what tools to use to leverage the knowledge implicitly stored in repositories of models, to make them play the same role in disseminating modeling best practices and foster design with reuse.

Approaches to model-driven repository search have been recently explored in the fields UML design (Chen, Madhavan, & Halevy, 2009, Llorens, Fuentes, & Morato, 2004) and business processes. A handful of researchers have investigated the problem of business process reuse based on process similarity and discovery of processes based on search upon repositories. Beeri, Eyal, Kamenkovich, and Milo (2006) proposed a visual query language for discovering business processes modelled using BPEL. Goderis, Li, and Goble (2006) developed a framework for discovering workflows using similarity metrics that consider the activities composing the workflows and their relationships.

The above solutions to business process discovery use as input the workflows that model the business, namely the activities that constitute the business processes and their dependencies in term of control flow. Yet, a workflow is not a complete description of the business processes. In this paper, we argue that a more effective discovery of business processes can be achieved if they are semantically described. Specifically, we show how such information can be encoded within an ontology that can be used for:

- *Abstracting discovery queries:* The user is able to formulate his/her queries in terms of the tasks (semantics) fulfilled by the desired business processes.
- *Exploiting relationships between business processes:* Business processes are interdependent. These dependencies can be explicitly described in the ontology in the form of binary relationships (Figure 1) that can be used, amongst other things, for increasing the recall of discovery queries.

Our approach also takes care of automating the classification of business processes within