Ontological approach to enhance results of business process mining and analysis

Wirat Jareevongpiboon and Paul Janecek
Department of Computer Science, Asian Institute of Technology, Pathumthani, Thailand

Received 21 June 2012
Accepted 10 October 2012

Abstract

Purpose – The purpose of this paper is to propose a solution to the problem of a lack of machine processable semantics in business process management.

Design/methodology/approach – The paper introduces a methodology that combines domain and company-specific ontologies and databases to obtain multiple levels of abstraction for process mining and analysis. The authors valuated this approach with a real case study from the apparel domain, using a prototype system and techniques developed in the Process Mining Framework (ProM). The results of this approach are compared with similar research.

Findings – Semantically enriching process execution data can successfully raise analysis from the syntactic to the semantic level, and enable multiple perspectives of analysis on business processes. Combining this approach with complementary research in semantic business process management (SBPM) can provide results comparable to multidimensional analysis in data warehouse and on line analytical processing (OLAP) technologies.

Originality/value – The approach and prototype described in this paper improve the richness of semantics available for open-source process mining and analysis tools like ProM, and the richness and detail of the resulting analysis.

Keywords Semantics, Process analysis, Business process, Process mining and analysis, Semantic process mining and analysis, Semantic business process management, Ontological approach, Multi-perspective process analysis, Multidimensional analysis, Semantic enhancement, Semantic annotation log, Ontology-database mapping, Ontology layers

Paper type Research paper

1. Introduction

Business process management (BPM) is an approach for managing the lifecycle of business processes from a business perspective. Current BPM suffers from a limitation in automation due to a lack of machine processable semantics to smooth the gaps between business and IT functions (Hepp *et al.*, 2005; Pedrinaci *et al.*, 2008b). To address this problem, two major efforts have been proposed in recent years. One is the attempt to provide semantics to support process analysis in data warehouse and OLAP technologies. With this approach, the semantic information is correlated to process executions stored in a data warehouse, and semantic inquiries in the form of templates are used to query in the process space. An example of this effort is the development of an integrated BPI tool suite (Grigori *et al.*, 2004). The other is the ontological approach pursuing to the direction of SBPM which combines BPM with semantic web and semantic web services (SWS) technologies (Hepp *et al.*, 2005). Most solutions in SBPM were developed in the context of the semantic utilized for process management within and between enterprises (SUPER) project. In this project, a stack

Business Process Management Journal
Vol. 19 No. 3, 2013
pp. 459-476

1463-7154
DOI 10.1108/14637151311319905

of ontologies (Pedrinaci *et al.*, 2008b) was defined to be a core component in providing semantics to the business process space and machine reasoning is used to support the correlation tasks between business and IT (Pedrinaci and Domingue, 2007). Examples of the work in this project are described in the literature section. Our work focuses on the necessity of adding semantics to business processes in order to implement the capability of SBPM, particularly in the process mining and analysis phase.

Semantic process mining and analysis is an active area of research, and several related algorithms and tools have been developed in recent years. The process execution data, usually in the form of event logs, are targeted for enrichment with knowledge as they are the primary input for process mining and analysis and they are generally recorded by information systems. Most researches have claimed that the process data elements in logs are label-based and lower in abstraction. To enhance the interpretation of results, the data elements in the event log are enriched with their concepts from domain ontologies. This would allow semantic mining and analysis algorithms to extend the analysis results with the higher levels of abstraction defined in the ontologies. The use of concepts instead of labels also makes it possible to aggregate or group data elements having the same concept. As a result, the process model discovered from enhanced logs tends to become a generic model. This generalization of the process model is favorable to unreadable spaghetti-like models, or when the business analyst wants to measure key performance figures in grouped units. However, only using the semantics provided by the domain ontology may limit the applicability of the analysis at the company specific level, even though the company itself is the primary source of semantics for processes. To extend analysis down to the company level as well as up to the domain level, we propose linking semantics from the domain ontology, company ontology and databases.

The contribution of this work is an approach to improve the availability of rich semantics in process mining and analysis. We introduce a methodology to combine domain and company specific ontologies and databases to obtain multiple levels of abstraction for mining and analysis. We show that multidimensional results can be achieved when process execution data are enriched with the proper level of semantics. To evaluate this approach, we developed a prototype system providing semantics to execution data that works with open-source process mining and analysis tools like ProM (van Dongen *et al.*, 2005).

The rest of this paper is organized as follows. Section 2 reviews related research mostly in SBPM. Section 3 describes our approach and its steps. Section 4 demonstrates the enhancement of outcomes from process mining and analysis via a real case study. Section 5 evaluates the approach and discusses related issues. Section 6 concludes this paper.

2. Related work

This section organizes work related to SBPM into six categories. The first category discusses the approaches that apply ontology at different phases of BPM in general. The second emphasizes on research related to the process mining and analysis phase specifically. The third presents research in the data mining area that uses ontology to facilitate data preparation. The fourth lists papers involving ontology-database mapping techniques. The fifth reviews papers introducing techniques we use for process mining and analysis. And the last describes other approaches to add semantic capability to data warehouses.

SBPM was first introduced in Hepp *et al.* (2005), and ontologies were used to provide machine-processable semantics in business processes (Hepp and Roman, 2007). SBPM was further developed in the SUPER project (Pedrinaci *et al.*, 2008b), and the ontology stack was described as an approach to increase the automation of tasks within the overall lifecycle of BPM. Several ontologies were created as extensions, such as the process mining ontology, the event ontology (Pedrinaci and Domingue, 2007), and the metric ontology (Pedrinaci and Domingue, 2009) as part of core ontology for business process and analysis (COBRA) (Pedrinaci *et al.*, 2008a). These ontologies define core terminologies of BPM which are used by machines for task automation. Our work reuses the EVO to provide semantics to event data as these concepts are common to any domain of business.

Several research works focus on the preparation of process execution data for semantic process mining and analysis. An extension to the MXML log file format, named SA-MXML, was introduced in Alves de Medeiros *et al.* (2007) that supported semantic annotation by linking terms to concepts in ontologies. To date, the only supporting tool to generate SA-MXML files is a plug-in to ProMImport (Günther and van der Aalst, 2006) version 7.0, but it only serves log files from process-aware information systems (PAIS) under SUPER. In Verbeek *et al.* (2011), a new format of log file named eXtensible Event Stream (XES) was introduced to solve problems with the semantics of additional attributes and the nomenclature used for different concepts. However, most of its supporting algorithms are still under development.

The use of ontologies to facilitate data preparation is also found in the data mining area. Domain ontologies were used to categorize attributes in the preparation step of a medical case study to obtain more meaningful results (Kuo *et al.*, 2007). A similar approach was proposed in Zeman *et al.* (2010) to use domain ontology to store attribute categories. The authors suggested assigning higher level semantics to individual values of data to avoid opaque mining results. These works are similar to ours in that we also use attribute categorization in ontologies to provide semantics to data attributes to improve the results from decision point analysis.

The semantic annotation to process execution data sometimes requires the mapping of concepts to instances stored in a database. In Stojanovic *et al.* (2002), the authors suggest a data migration approach based on mapping the given relational schema into an already existing ontology structure using reverse engineering. To manage the size of these ontologies, Konstantinou *et al.* (2006) suggested keeping instances in the database and only storing links to the instances in the ontologies. Our approach combines both approaches by storing instances in the database generally, and loading instances into the ontology temporarily when performing reasoning tasks.

The following describes some other works related to techniques for process mining and analysis. Ontologies, model references and a reasoner are described as three essential building blocks for semantic process mining and analysis tools in Alves de Medeiros *et al.* (2008a). The "Alpha algorithm" is presented in van der Aalst *et al.* (2004) to extract a process model from an event log. This algorithm supports both semantic and non semantic process data. Our work used it to mine event logs for result comparison. The "decision miner" in Rozinat and van der Aalst (2006) is a decision point analysis technique performed on a discovered process model to detect data dependencies that affect the routing of a case. Although it does not support semantic analysis, our work showed how its results can be improved by supplying semantic input. The technique described in Alves de Medeiros *et al.* (2008b) analyzes basic performance of a process only from the view of task and originator while ours adds a business object perspective. Some other semantic techniques developed in ProM to support SA-MXML are also discussed and demonstrated in Alves de Medeiros *et al.* (2007, 2008a) and Alves de Medeiros and van der Aalst (2009).

Attempts to extend semantic analysis capabilities to data warehouse are described in Casati and Shan (2002) and Grigori *et al.* (2004). The BPI tool (Casati and Shan, 2002) is an extended version of the HPPM Intelligent Process Data warehouse (Grigori *et al.*, 2004) that allows business analysts to perform multidimensional analysis and classify process instances via three concepts: behaviors, taxonomy and process regions. OntoDSS (Sell *et al.*, 2005) uses data warehouse and semantic web technology based on IRS-III to integrate business semantic into analytical tools, but it emphasizes the semantics of data items for better explanations rather than the semantics of processes in general.

Although the described approaches above specifically addressed the same problem and proposed different solutions for BPM, none of them can solely fulfill the entire requirement of SBPM. The extension, continuation, and new coming approaches of this research area are still motivating.

3. Enriching process execution data with semantics from ontology

In the SBPM approach, semantic annotation starts at the process modeling phase of BPM to capture all facts in an organization, such as software, hardware and business logic, and represent them in a set of ontologies. In a real situation, companies having existing business processes may find it difficult to redefine their process related components semantically. To avoid dramatic changes, semantic enhancement on process mining and analysis, a process post execution phase, is a preferable alternative. Process mining and analysis (Figure 1) are performed on past execution data, recorded by most IT systems, to assess the quality of business operations. When these data are gathered and transformed into a unified format, they become our target for semantic enrichment.

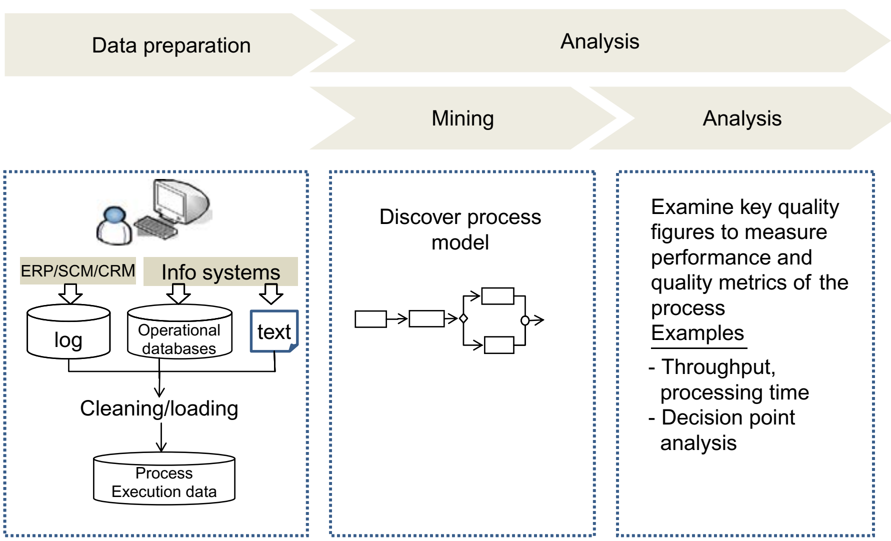


Figure 1.
Typical process mining and analysis steps

3.1 Examining process execution data

Companies usually record what happens during a process in several forms, i.e. databases, data warehouse, text files, log files, etc. These forms contain similar contents about execution of process instances. Each process instance consists of several audit trail entries, each of which includes data about task, event, the originator who performed the task, timestamp, and additional data attributes related to the task. These contents are logged as string labels, and therefore mining and analysis techniques are not able to reason over them. The semantics of labels, concerning some specific domains, are typically only in the head of the business analyst. To enable semantic analyses, the semantics of the labels should be incorporated at the data preparation step via the use of ontologies.

3.2 Defining ontologies

Ontologies can be defined at various abstract levels, such as an upper level enterprise ontology, or a domain specific telecommunication ontology. A higher level ontology containing general concepts can be reused by many lower level ontologies. For process mining and analysis, our work focuses on the use of domain and company level ontologies that produce analysis results that are closer to the actual operation of the company. Based on common information found in process execution data, five domain ontologies describe data elements necessary for multi-perspective analysis: task, originator, event, time and data attributes. We describe the contents of these ontologies in Table I.

3.3 Creating and combining ontologies

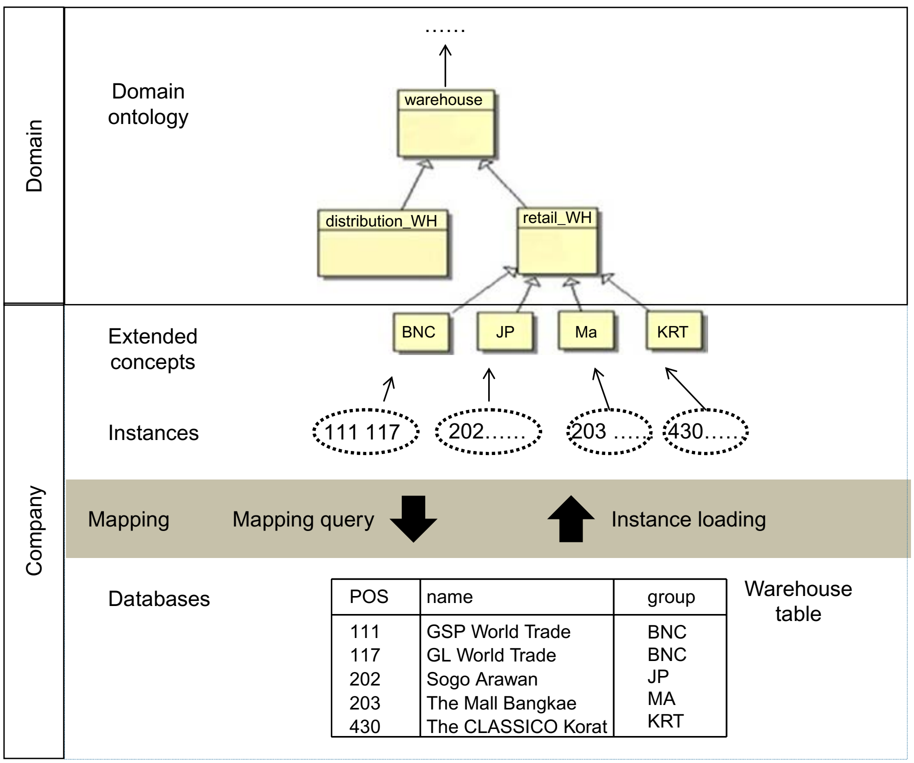
To date, various ontology engineering tools with user friendly interfaces are available to the public and most of them require little technical effort to use. For example, WSML Editor (Kerrigan, 2005), a plug-in to the Web Service Modeling Toolkit (WSMT) framework (Kerrigan and Mocan, 2008), allows users to create and edit WSML (de Bruijn *et al.*, 2006) ontologies in both text and graphical format. Business analysts can consider them to create ontologies by extending from upper ontologies or create simple ontologies for individual cases.

Data attributes are usually logged by information systems (IS) as additional details of tasks in a process. For example, the point of sale (POS) attribute in a "stock empty notification" task is recorded to identify the store that has recently sold products. To save

Ontologies	Description
Task ontology	Concepts of activities performed in the process being analyzed
Originator ontology	Concepts related to actors performing activities which can be concepts about roles, department or resources
Event ontology	Concepts of event for process execution like concepts defined in EVO from SUPER
Time ontology	Concepts related to date and time which are universal to any domain. This ontology, therefore, can be reused from other sources
Data attributes ontology	Data attributes are additional info of tasks and they may identify business object being updated or created in process. Our approach focuses on only data attributes related to business object. To emphasize its contents, we later call it business object ontology

Table I. Description of domain ontologies for process mining and analysis storage space and avoid typing errors, IS typically only record key values in the process execution log. The semantics of these values may differ from company to company in that one company may assign "POS ID" based on its location while another assigns "POS ID" consecutively with no special meaning. So, the primary semantics of data attributes should be derived from an individual company rather than from a general description. To allow multi level abstraction analysis related to data attributes, we suggest combining concepts from the domain and company level ontologies with the database. As such, data in databases becomes concepts or instances at the bottom level of the ontology tree structure. Figure 2 shows the example of a warehouse ontology having the extended concepts derived from a company's internal *retail_WH* grouping. The "POS" data stored in a warehouse table are instances of these concepts and they can be loaded when a semantic query is performed. This layer of concepts can be viewed as a semantic hierarchy for "POS" data attributes.

The extension from domain ontologies to databases is possible for all ontologies listed in Table I. This allows business analysts to add detailed knowledge related to their business as part of the analysis. For instance, the addition of role, department or category of employees to the originator ontology allows the business analyst to examine originators in various dimensions specific to the company.



Note: Concepts are represented by rectangles and instances by ellipses

Figure 2.
A warehouse ontology having concepts and instances from a database

3.4 Mapping concepts from ontologies to process execution data

The mapping step is necessary to make these semantics available to the process mining and analysis tool. We categorize our mapping into concept mapping (mapping a data element directly to its concept in the domain ontologies) and instance mapping (mapping a data element to ontology instances retrieved from the database). During the mapping phase, not all data attributes are crucial for analysis. We are only interested in data attributes having values that identify business objects being updated or created by a given task. These objects are usually persistent in a company, and their relevant info are stored in databases. For example, the data attribute named "customer id" identifying a customer of a company and its description details are stored in a customer table. To avoid confusion and emphasize the business object, we will refer to ontologies related to data attributes as business object ontologies for the rest of this paper. In most cases, a business object ontology is created to categorize the business objects of interest, so its concepts constitute a taxonomy of that object. For this reason, we view an individual business object in a database as an instance of a concept defined in a business object ontology. When concepts and instances reside in separate data sources, the linking of data sources before mapping is required so that the semantics of data attributes can be efficiently retrieved by the mapping engine.

4. Use case

We evaluate our methodology with a use case from a company in the apparel domain that owns several major brands of ready-made dresses and leatherwear for women. We focus on the "Restocking" process, which involves distributing and transferring products to/from department stores as well as the company's retail stores. If products are available in the central warehouse, they are sent daily to those stores that have recently sold products ("refill"). If products are not available in the warehouse, they are redistributed among retail stores and department stores that have product in stock ("transfer"). This process is managed by "merchandisers", the group of people responsible for promoting their individual brand, using the restocking information system (RIS), which is a non PAIS. Currently, the management team has difficulty analyzing and evaluating this process as the only information they have are periodic stock and sales reports. These reports can be generated from various perspectives, such as sales reports categorized by brands, or by product types, but they are only static and data-centric. Management cannot drill down to the process and task level to identify the source of the problem if it exists. Also, some analysis questions cannot be answered instantly such as:

- *Q1.* What are the tasks related to restocking the process?
- *Q2.* How many items were restocked and how are these numbers distributed between refill and transfer?
- *Q3.* What is the distribution of restocking among types of stores, types of products or brand of products?

Or:

- *Q4.* What are the patterns of product distribution found in the process? Do they follow the typical business rules? Are these patterns different among brands?, etc.

Our approach to addressing such problems is to use a process-centric mining and analysis tool and supplying the tool with the company specific knowledge. This allows the company to extend the analysis down to the sources where those static reports are derived and the analysis results are bound to the specifics of the company. The following sections demonstrate the steps in our approach to handle this case.

4.1 Gathering process execution data

The process related data scattered across many tables in the company's database are gathered and transformed to the format used by the process mining and analysis tool. We selected the restocking of products in three major brands that occurred in May 2009. Table II shows the example of some data.

4.2 Creating ontologies

Three ontologies having company specific concepts are emphasized: task, originator and business object. For this process, the business object is the stock of clothing categorized by clothing type, brand and storage location, each of which has its own viewpoint of interest. For instance, the clothing type may have concepts about style, fabric, color and size. To simplify our example, this study demonstrates only clothing style. We derived most general concepts from the "women's clothing" categories on the Yahoo shopping page[1] as they are similar to clothing styles used by this company.

Instance no.	Task	Originator	Eventtype	Timestamp	Data attributes
1	Stock empty notification	RIS	Complete	06/05/09	POS2 = 187, prd_id = G, prd_code = 2006040288895
1	Check from distribution WH	RIS	Complete	06/05/09	
1	Refill	RIS	Complete	06/05/09	POS1 = 777
1	Create Moving Notes	RIS	Complete	06/05/09	
2	Stock Empty Notification	RIS	Complete	06/05/09	POS2 = 117, prd_id = G, prd_code = 2006040332321
2	Check from distribution WH	RIS	Complete	06/05/09	
2	Refill	RIS	Complete	06/05/09	POS1 = 7BP
2	Create Moving Notes	RIS	Complete	06/05/09	
3	Stock Empty Notification	RIS	Complete	06/05/09	POS2 = 142, prd_id = G, prd_code = 2006040332116
3	Check from distribution WH	RIS	Complete	06/05/09	
3	Transfer	Ms. Hatairat	Complete	06/05/09	POS1 = 185
3	Create Moving Notes	RIS	Complete	06/05/09	
4	Stock Empty Notification	RIS	Complete	06/05/09	POS2 = 116, prd_id = G, prd_code = 2006040332109
4	Check from distribution WH	RIS	Complete	06/05/09	
4	Transfer	Ms. Ratirom	Start	06/05/09	

Notes: RIS – Restocking information system; POS1 – from POS; POS2 – to POS (stock out POS); prd_id – brand; prd_code – product code

Table II. The extract of process execution data in table format

For brand and location ontologies, most concepts come from the company itself because this company has its own brands and sells products in the local market. These two ontologies are created to provide some degree of semantics to raw data attributes. Parts of these ontologies are shown in Figure 3.

The central node is the name of the ontology with links to all the concepts it contains. The other nodes represent concepts, and arrows link super concepts to their sub concepts (the oval the arrow points to). For instance, in the clothing ontology, the concept “long-sleeve shirts” is a sub concept of “Shirts”, and “Shirts” is the super concept of “long-sleeve shirts”, “short-sleeve shirts”, etc.

4.3 Mapping of data elements to ontological concepts

To demonstrate and evaluate our approach, we developed a prototype system to facilitate the mapping process, add semantics to process execution data, and perform simple analysis. We selected ProM for process mining and analysis, although our approach is not restricted to any specific tool. To work with ProM, the prototype system accepts as input process execution data in the form of MXML log files (van Dongen and van der Aalst, 2005) and ontologies in WSML.

Data elements of task, originator and event are mapped directly with concepts in their corresponding ontologies using concept mapping. Figure 4 shows the example of concept mapping on a task where a business analyst selects a task listed in a log and maps it with the desired concept listed in the task ontology. For a business object, concept mapping is also applicable if only few data elements need to be mapped. For our case, there are many products restocked each day, therefore mapping them to concepts

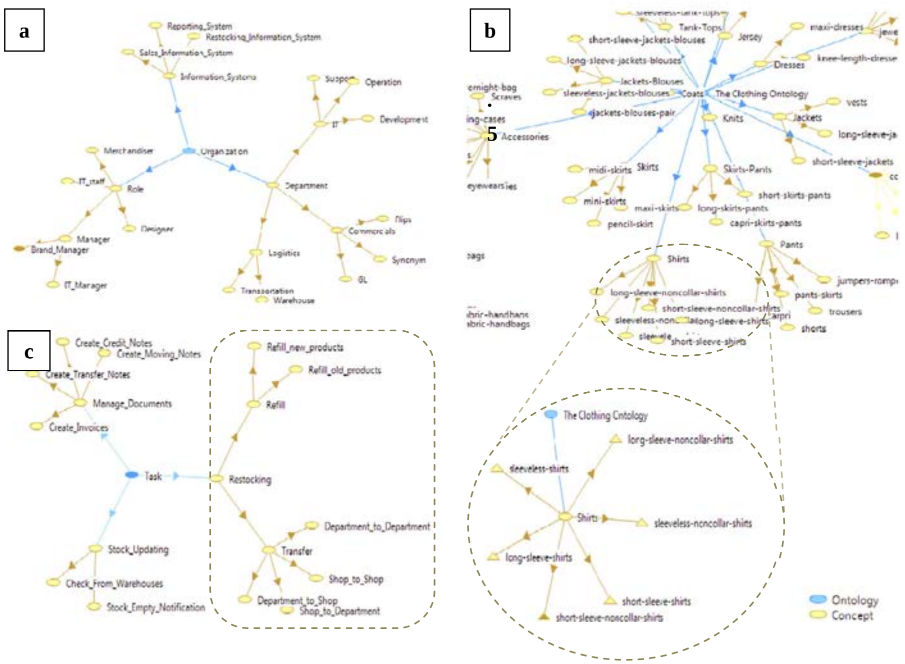


Figure 3.
(a) Organizational; (b) clothing; and (c) task ontologies shown in WSML Visualizer

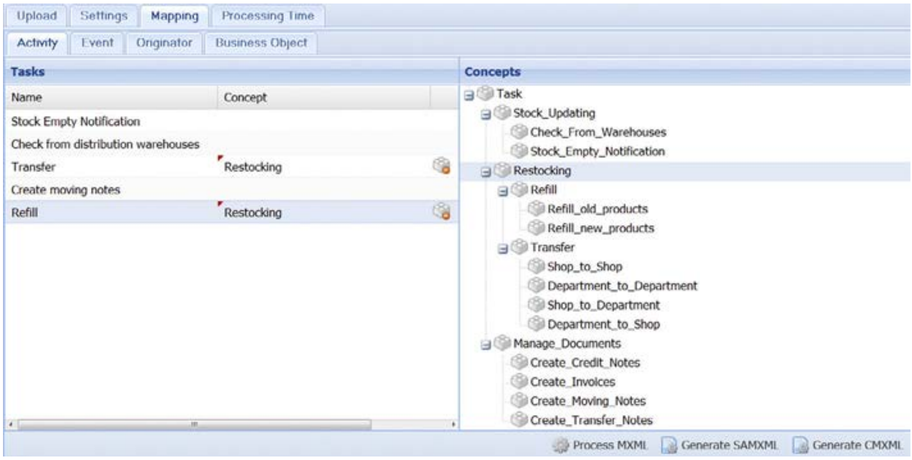


Figure 4.
Concept mapping for tasks

individually is a tedious and time-consuming task. Stock data are actually key fields in the database; we can link them as groups of instances to their corresponding concepts via SQL queries and allow the system to perform auto-mapping. Figure 5 demonstrates this process by listing all the lowest concepts of business objects (clothing) with their associated queries to load instances.

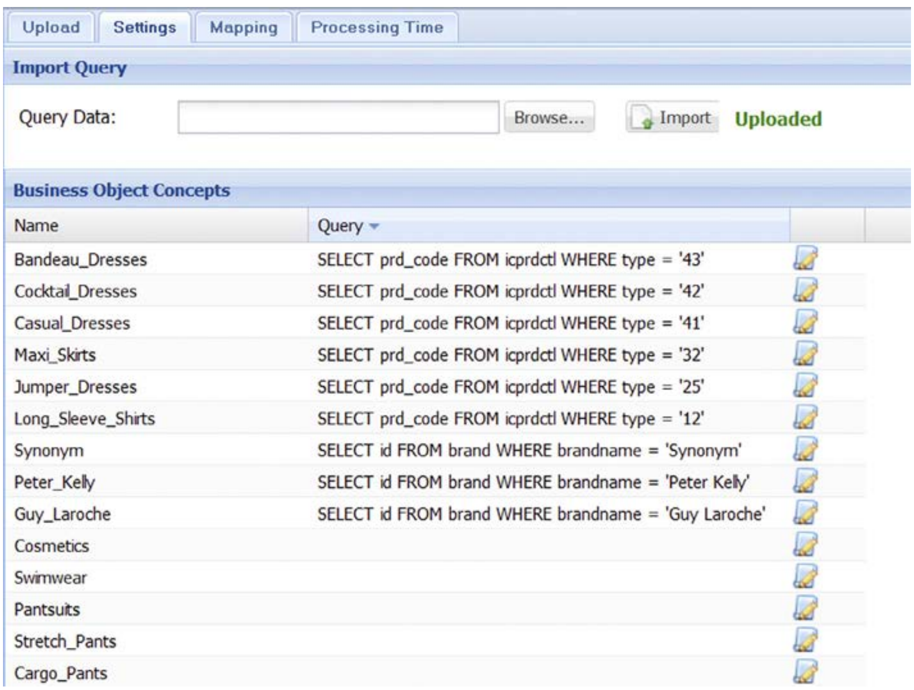


Figure 5.
Instance mapping
for clothing

In Figure 5, a query *select prd_code from icprdctl where type* = 12 is assigned to the "long-sleeve shirts" concept, illustrating that all product codes (*prd_code*) having type equal to 12 are instances of "long-sleeve shirts". The steps can be repeated for other lowest-level concepts required for analysis. The system can also import a pre-configuration file containing all desired concepts and their corresponding queries. Once the concepts and instances are linked, the system performs auto-mapping between the attributes found in the log, the related instances in the database, and the concepts in this ontology. This step results in two files: a semantically enhanced log file in the SA-MXML format, where data elements are annotated with links to their concepts in ontologies; and a log file in the C-MXML format, where data elements are replaced with their concepts. The former can be used with semantic algorithms whereas the latter can be used with non semantic algorithms.

4.4 The improvement of results in process mining and analysis

This section shows the improvement of results when applying process mining and analysis techniques in ProM to the semantically enhanced log.

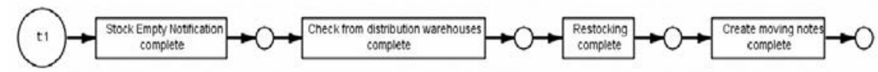
Enhanced results of process mining. The results from process mining can be enhanced in two ways. First, the discovered process model can be viewed at the concept level so that task concepts instead of task name (if mapped) are used for mining. If consecutive tasks in the discovered model have the same concepts, those tasks are combined or aggregated. The resulting model is more abstract (Figure 6(b)). Second, the discovered model can be viewed at a company specific level where tasks can be decomposed by concepts of their data attributes into sub tasks. Recall that we are only interested in data attributes that identify business objects. This view displays a discovered process model having tasks categorized by business objects. Figure 6(d) shows the "stock empty notification" task on our restocking process decomposed by clothing type.

The decomposition starts from mapping a data attribute with its concept (if any) from the business object ontology, then appending this concept to the end of the task name in the original log, so a single task name is broken up into several groups, each of which has the same data attribute concept. Considering the process execution data of Table II, the "stock empty notification" task has three additional data attributes named *POS2* (requesting store), *prd_code* (product code) and *prd_id* (brand). Focusing on the clothing type perspective, we append the concept of *prd_code* to every "stock empty notification" task in the log. After appending, the new log file with *Stock Empty Notification_*(*Long-sleeve-shirts*) tasks is generated and passed to process mining algorithms to gain task decomposition results. The process in aggregated and decomposed task views in Figure 6 allows the business analyst to be able to answer the analysis questions *Q1-Q3*.

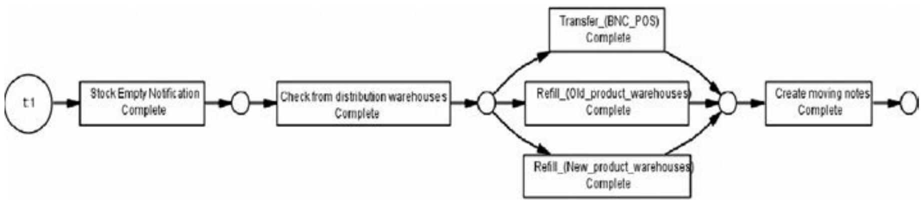
Enhanced results of decision point analysis. "Decision miner" is a label-based decision point analysis algorithm that only uses data attributes and shows how they influence the choices made in the process based on past process executions. The rule discovered in Figure 7(a) found that product code (*prd_code*) is a major influence on the type of product movement (refill or transfer). In this case, the business analyst may find it difficult to interpret the result unless he/she can remember information directly from the product code. Figure 7(b) shows the results after running "Decision miner" with the enhanced log in C-MXML format.

The new rule discovered that the brand of product (*prd_id*) determines the type of stock movement. For the first two brands ("Synonym" and "Peter Kelly"), the company

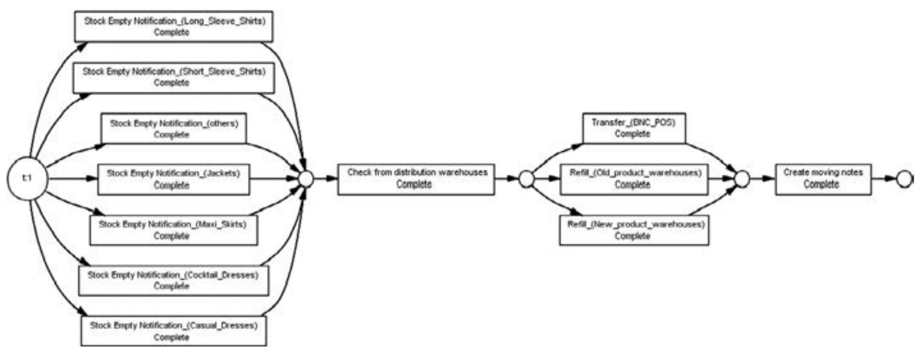
(a) Process model from non semantic log



(b) Process model when tasks are aggregated by *Restocking* concept



(c) Process model with *Transfer* and *Refill* tasks decomposed by *POS1* concept

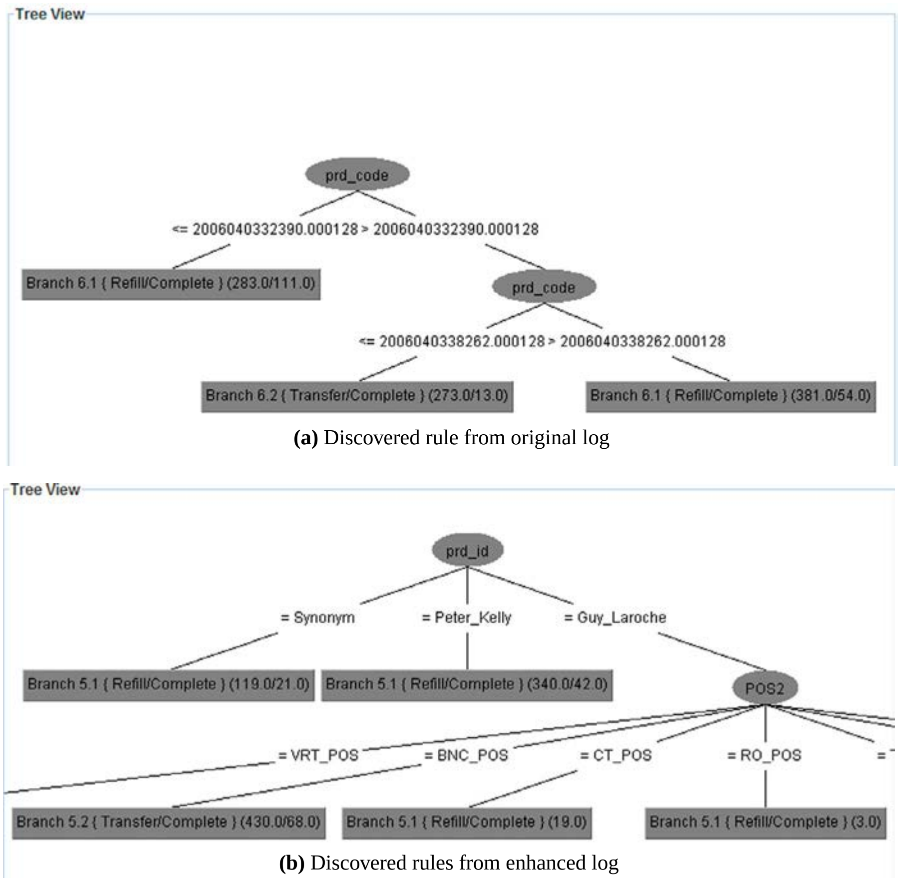


(d) Process model showing task decomposition on *prd_code* and *POS1* concepts

Figure 6.
Various options of process models from enriched log discovered by Alpha algorithm

refills products from the central warehouse to the requesting stores, while for the third brand ("Guy Laroche") they refill and transfer products depending on the type of the requesting store (*POS2*). If the requesting stores are in the BNC or KRT groups, transfer is preferred, otherwise refill is applied. The result from this example can answer the analysis question *Q4* showing the different restocking patterns for each brand, which partly involves preference of individual merchandiser.

Enhance results of measuring basic performance. One of the basic requirements for process analysis is to measure performance, which reflects how well the company manages its processes. In ProM, a semantic plug-in named "performance metrics in ontologies" computes metrics for every concept found in the task and originator ontologies, such as processing time and throughput time. However, this plug-in does not consider the enhancement from semantic data attributes. We show that users can view another dimension of metrics on business objects via semantic data attributes. Figure 8 shows processing time[2] and throughput in view of the clothing style. This concept filtering ability can help answer the questions *Q2* and *Q3*. Also, the resulting metrics,



(a) Discovered rule from original log

(b) Discovered rules from enhanced log

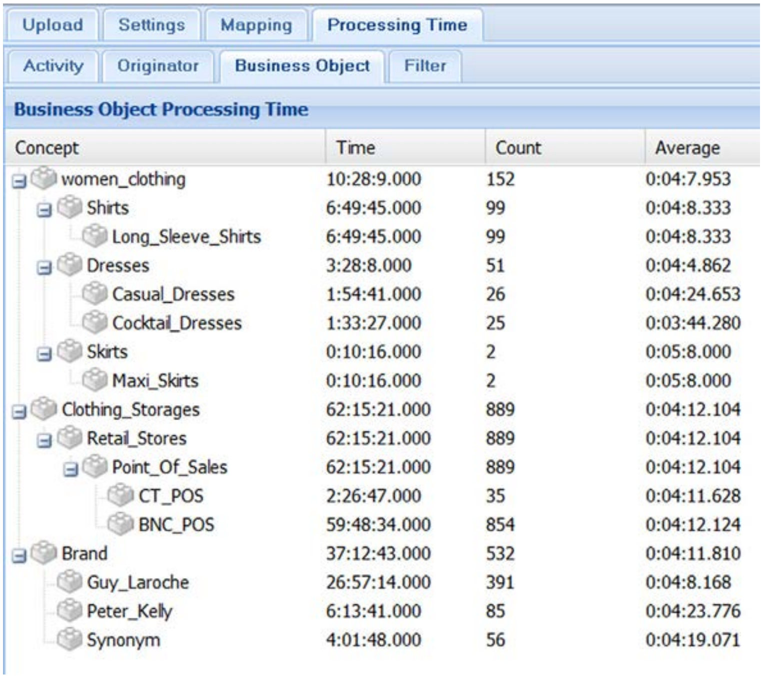
Figure 7.
The discovered rules from decision miner

when linked to originator, allow the business analyst to evaluate personal preference and personal performance.

5. Evaluation and discussion

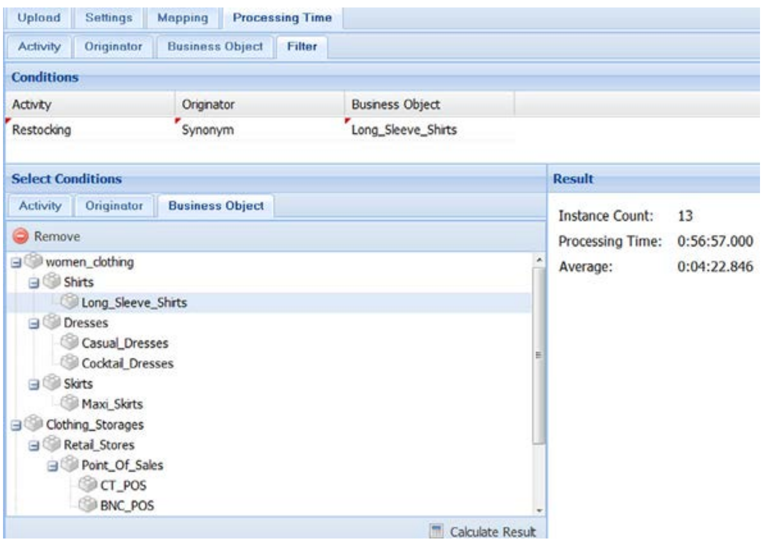
We compare our approach with other SBPM research in (Table III). In general, our approach is similar to current semantic approaches in that the analysis of business processes can be improved from string-based to concept-based. While most semantic approaches focus on semantics derived from the domain, ours adds a way to derive semantic info from the company data source. Thus, the analysis from our approach can be extended in two opposite directions: domain and company specific. Our approach also improves the business object dimension of analysis through enriching the semantics of data attributes.

Our approach is complementary to the one presented in Alves de Medeiros *et al.* (2007) in providing semantics to process execution data via ontology. However, the focus in Alves de Medeiros *et al.* (2007) is on a generalized view of process using task and



Concept	Time	Count	Average
women_clothing	10:28:9.000	152	0:04:7.953
Shirts	6:49:45.000	99	0:04:8.333
Long_Sleeve_Shirts	6:49:45.000	99	0:04:8.333
Dresses	3:28:8.000	51	0:04:4.862
Casual_Dresses	1:54:41.000	26	0:04:24.653
Cocktail_Dresses	1:33:27.000	25	0:03:44.280
Skirts	0:10:16.000	2	0:05:8.000
Maxi_Skirts	0:10:16.000	2	0:05:8.000
Clothing_Storages	62:15:21.000	889	0:04:12.104
Retail_Stores	62:15:21.000	889	0:04:12.104
Point_Of_Sales	62:15:21.000	889	0:04:12.104
CT_POS	2:26:47.000	35	0:04:11.628
BNC_POS	59:48:34.000	854	0:04:12.124
Brand	37:12:43.000	532	0:04:11.810
Guy_Laroche	26:57:14.000	391	0:04:8.168
Peter_Kelly	6:13:41.000	85	0:04:23.776
Synonym	4:01:48.000	56	0:04:19.071

(a) Processing time in the view of clothing type



(b) Processing time filtered by task, originator and business object

Figure 8.
Processing time in view of business object

Type of enhancement	The approach Non-semantic	Current (e.g. SUPER)	Ours
General			
Level of analysis	String-based	Concept-based	Concept-based
Direction of analysis		Domain	Domain and company specific
Dimensions of analysis	Task, originator or combination	Task, originator or combination	Task, originator, business object or combination
Functional			
Process mining			
Task aggregation	No	Yes	Yes
Task decomposition	No	No	Yes
Process analysis			
Decision point analysis	Label-based rules	Label-based rules	Semantic rules
Performance measurement	Task-originator dimensions	Task-originator (concepts) dimensions	Task-originator-business object (concepts) dimensions

Table III. Type of enhancement comparison of non semantic, current semantic and our approaches

originator ontologies while ours focuses on specifying a view of process using a business object (data attributes) ontology. The results from the combination of both approaches could be comparable to the results from the approach of multidimensional analysis in data warehouse and OLAP technologies, but different in focus and technical requirements. Data warehouse and OLAP are widely used as data-centric storage and analysis tools, but our approach is based on process mining and analysis which are process-centric. Although, data warehouse and OLAP technologies currently also support semantic analysis of business processes as found in the BPI tool suite mentioned in Section 2, such technologies are commercial and limit data formats. The ontology approach, on the other hand, is used in the semantic web, so its supporting tools are widely available for the public. The extendibility and reusability of ontologies also help reduce the time for research. In addition, ontologies are able to provide semantics to various formats of process execution data, and support analysis of business processes in various aspects (roll up/aggregate and drill down/decompose) with less technical effort.

Although our approach can be applied to various business processes, there are several limitations. First, our approach yields semantic enriched data in formats compatible with MXML, but the only tool that currently works with MXML is ProM. Second, this system is a prototype developed to evaluate our approach, so it may not be able to handle a high volume of data when performing instance mapping. These technical difficulties together with more techniques to handle several formats of data should be studied and improved in future work.

6. Conclusion

BPM is currently limited by a lack of machine-readable semantics to smooth the transition of work between business and IT. SBPM addresses this problem by providing semantics in the form of ontologies to many phases of BPM. In the process mining and analysis phases, ontologies are used in two ways: to increase the automation of tasks and to enhance the interpretation of results. Our research introduces an approach that combines domain ontologies, company ontologies and a company's database to form a knowledge source for process mining and analysis. We demonstrated that semantic enrichment in process execution data via these multiple layers of ontologies can raise the analysis from the syntactic to the semantic level, and enable multiple perspectives of analysis on processes. To assess our approach, we analyzed a real case study in the apparel domain using our prototype system and techniques developed in ProM. The outcome showed that without semantically enriched data, especially on data attributes, some analysis questions are difficult or impossible to answer. In conclusion, enhancing the semantics available as input for SBPM tools has the potential to greatly improve the results of process mining and analysis.

Notes

1. http://shopping.yahoo.com/browse/womens-clothing/
2. No time information is provided in the real log, so we randomly generated timestamps to calculate processing time.

References

Alves de Medeiros, A.K. and van der Aalst, W.M.P. (2009), "Process mining towards semantics", in Tharam, S.D., Elizabeth, C., Robert, M. and Katia, S. (Eds), *Advances in Web Semantics I*, Springer, Berlin, pp. 35-80.

Alves de Medeiros, A.K., van der Aalst, W.M.P. and Pedrinaci, C. (2008a), "Semantic process mining tools: core building blocks", paper presented at 16th European Conference on Information Systems in Galway, Ireland, 1953-1964.

Alves de Medeiros, A.K., van den Brand, P., van der Aalst, W.M.P., Weijters, T., Gaaloul, W. and Pedrinaci, C. (2008b), "Semantic process mining tool – final implementation", SUPER Project Deliverable 6.11.

Alves de Medeiros, A.K., Pedrinaci, C., van der Aalst, W.M.P., Domingue, J., Song, M., Rozinat, A., Norton, B. and Cabral, L. (2007), "An outlook on semantic business process mining and monitoring", in Meersman, R., Tari, Z. and Herrero, P. (Eds), *On the Move to Meaningful Internet Systems 2007: OTM 2007 Workshops*, Springer, Berlin, pp. 1244-1255.

Casati, F. and Shan, M.-C. (2002), "Semantic analysis of business process executions", in Jensen, C., Šaltenis, S., Jeffery, K., Pokorny, J., Bertino, E., Böhn, K. and Jarke, M. (Eds), *Advances in Database Technology – EDBT 2002*, Springer, London, pp. 119-135.

de Bruijn, J., Lausen, H., Polleres, A. and Fensel, D. (2006), "The web service modeling language WSML: an overview", *Proceedings of the 3rd European Conference on the Semantic Web: Research and Applications in Budva, Montenegro*, Springer, Berlin, pp. 590-604.

Grigori, D., Casati, F., Castellanos, M., Dayal, U., Sayal, M. and Shan, M.-C. (2004), "Business process intelligence", *Computers in Industry*, Vol. 53 No. 3, pp. 321-343.

Günther, C. and van der Aalst, W.M.P. (2006), "A generic import framework for process event logs", in Eder, J. and Dustdar, S. (Eds), *Business Process Management Workshops*, Springer, Berlin, pp. 81-92.

Hepp, M. and Roman, D. (2007), "An ontology framework for semantic business process management", *Proceedings of Wirtschaftsinformatik*.

Hepp, M., Leymann, F., Domingue, J., Wahler, A. and Fensel, D. (2005), "Semantic business process management: a vision towards using semantic web services for business process management", *ICEBE – IEEE Computer Society*, pp. 535-540.

Kerrigan, M. (2005), "The WSML editor plug-in to the web services modeling toolkit", *Proceedings of 2nd WSMO Implementation Workshop* (*WIW 2005*) *in Innsbruck, Austria, Citeseer*, pp. 1-9.

Kerrigan, M. and Mocan, A. (2008), "The web service modeling toolkit", *Proceedings of the 5th European Semantic Web Conference on the Semantic Web: Research and Applications in Tenerife, Canary Islands, Spain*, Springer, Berlin, pp. 812-816.

Konstantinou, N., Spanos, D., Chalas, M., Solidakis, E. and Mitrou, N. (2006), "VisAVis: an approach to an intermediate layer between ontologies and relational database contents", paper presented at International Workshop on Web Information Systems Modeling (WISM 2006), Luxembourg, June 5-9.

Kuo, Y.-T., Lonie, A., Sonenberg, L. and Paizis, K. (2007), "Domain ontology driven data mining: a medical case study", *Proceedings of the 2007 International Workshop on Domain Driven Data Mining in San Jose, California*, ACM, New York, NY, pp. 11-17.

Pedrinaci, C. and Domingue, J. (2007), "Towards an ontology for process monitoring and mining", in Hepp, M., Hinkelmann, K., Karagiannis, D., Klein, R.d. and Stojanovic, N. (Eds), *Proceedings of the Workshop on Semantic Business Process and Product Lifecycle Management* (*SBPM-2007*), available at: CEUR-WS.org.

Pedrinaci, C. and Domingue, J. (2009), "Ontology-based metrics computation for business process analysis", *Proceedings of the 4th International Workshop on Semantic Business Process Management in Heraklion, Greece*, ACM, New York, NY, pp. 43-50.

Pedrinaci, C., Domingue, J. and Alves de Medeiros, A.K. (2008a), "A core ontology for business process analysis", *Proceedings of the 5th European Semantic Web Conference on the Semantic Web: Research and Applications in Tenerife, Canary Islands, Spain*, Springer, Berlin, pp. 49-64.

Pedrinaci, C., Brelage, C., van Lessen, T., Domingue, J., Karastoyanova, D. and Leymann, F. (2008b), "Semantic business process management: scaling up the management of business processes", *2nd IEEE International Conference on Semantic Computing* (*ICSC*) *in Santa Clara, CA, USA*, Institute of Electrical and Electronics Engineers, Inc., New York, NY, pp. 546-553.

Rozinat, A. and van der Aalst, W.M.P. (2006), "Decision mining in prom", *Proceedings of the 4th International Conference on Business Process Management in Vienna, Austria*, Springer, Berlin, pp. 420-425.

Sell, D., Cabral, L., Motta, E., Domingue, J. and Pacheco, R. (2005), "Adding semantics to business intelligence", *Proceedings Sixteenth International Workshop on Database and Expert Systems Applications, August 25-26*, IEEE Computer Society, Washington, DC, pp. 543-547.

Stojanovic, N., Stojanovic, L. and Volz, R. (2002), "A reverse engineering approach for migrating data-intensive web sites to the semantic web", *Proceedings of the IFIP 17th World Computer Congress – TC12 Stream on Intelligent Information Processing*, Kluwer, B.V., Deventer, pp. 141-154.

van der Aalst, W.M.P., Weijters, A.J.M.M. and Maruster, L. (2004), "Workflow mining: discovering process models from event logs", *IEEE Trans. on Knowl. and Data Eng.*, Vol. 16 No. 9, pp. 1128-1142.

van Dongen, B.F. and van der Aalst, W.M.P. (2005), "A meta model for process mining data", paper presented at Conference on Advanced Information Systems Engineering.

van Dongen, B.F., Alves de Medeiros, A.K., Verbeek, H.M.W., Weijters, A.J.M.M. and van der Aalst, W.M.P. (2005), "The prom framework: a new era in process mining tool support", *Proceedings of the 26th International Conference on Applications and Theory of Petri Nets in Miami*, Springer, Berlin, pp. 444-454.

Verbeek, H.M.W., Buijs, J.C.A.M., Dongen, B.F. and Aalst, W.M.P. (2011), "XES, XESame, and ProM 6", in Soffer, P. and Proper, E. (Eds), *Information Systems Evolution – CAiSE Forum 2010, Hammamet, Tunisia, June 7-9*, Selected Extended Papers, Springer, Berlin, pp. 60-75.

Zeman, M., Ralbovský, M., Svátek, V. and Rauch, J. (2010), *Ontology-Driven Data Preparation for Association Mining*, available at: http://keg.vse.cz/onto-kdd-draft.pdf (accessed October 29, 2011).

About the authors

Wirat Jareevongpiboon is an Assistant Professor at the Department of Computer Science, Thammasat University, Thailand. She received her Master degree in MIS from the University of Illinois, Urbana-Champaign and is now a PhD student at the Asian Institute of Technology, Thailand. She is working on research to enhance business process mining and analysis with ontologies. Her research interests include business information systems, database, data mining and applied ontologies. Wirat Jareevongpiboon is the corresponding author and can be contacted at: Wirat.Jareevongpiboon@ait.ac.th

Dr Paul Janecek is an Adjunct Professor at the Asian Institute of Technology, Thailand, and CEO of Think Blue Data, a data analytics software company in Bangkok. He received his PhD in Computer Science from the Swiss Federal Institute of Technology, Lausanne (EPFL). His research interests include e-business, information visualization (including business intelligence) and human-computer interaction.