

# Process ontology development using natural language processing: a multiple case study

Ozge Gurbuz

*Informatics Institute, Middle East Technical University, Ankara, Turkey and  
Huawei Turkey R&D Center, Istanbul, Turkey*

Fethi Rabhi

*School of Computer Science and Engineering, University of New South Wales,  
Sydney, Australia, and*

Onur Demirors

*Izmir Yuksek Teknoloji Enstitusu, Urla, Turkey and  
School of Computer Science and Engineering, University of New South Wales,  
Sydney, Australia*

## Abstract

**Purpose** – Integrating ontologies with process modeling has gained increasing attention in recent years since it enhances data representations and makes it easier to query, store and reuse knowledge at the semantic level. The authors focused on a process and ontology integration approach by extracting the activities, roles and other concepts related to the process models from organizational sources using natural language processing techniques. As part of this study, a process ontology population (PrOnPo) methodology and tool is developed, which uses natural language parsers for extracting and interpreting the sentences and populating an event-driven process chain ontology in a fully automated or semi-automated (user assisted) manner. The purpose of this paper is to present applications of PrOnPo tool in different domains.

**Design/methodology/approach** – A multiple case study is conducted by selecting five different domains with different types of guidelines. Process ontologies are developed using the PrOnPo tool in a semi-automated and fully automated fashion and manually. The resulting ontologies are compared and evaluated in terms of time-effort and recall-precision metrics.

**Findings** – From five different domains, the results give an average of 70 percent recall and 80 percent precision for fully automated usage of the PrOnPo tool, showing that it is applicable and generalizable. In terms of efficiency, the effort spent for process ontology development is decreased from 250 person-minutes to 57 person-minutes (semi-automated).

**Originality/value** – The PrOnPo tool is the first one to automatically generate integrated process ontologies and process models from guidelines written in natural language.

**Keywords** Process ontology, Ontology development, Business process modelling, Natural language processing

**Paper type** Research paper

## 1. Introduction

An ontology is defined as a “formal specification of a shared conceptualization.” Ontologies improve communication between people, enhance the reuse of domain knowledge, make domain assumptions explicit and enable differentiation between domain and operational knowledge (Gruber, 1993). Business process models, on the other hand, are sequential representations of activities associated with specific business goals and performed in a variety of styles and are used for a variety of purposes, such as establishing execution consistency, optimization, automation, measurement and certification (Aysolmaz *et al.*, 2018; Dumas *et al.*, 2013; Kaya and Demirors, 2011; Tarhan and Demirors, 2012; Turetken and Demirors, 2011).

Ontological approaches on business process modeling concepts are used for automating business process model matching by semantic alignment (Fengel, 2014), enhancing understandability of process models and easing the process model navigation



---

through the repository by semantic annotations (Mturi and Johannesson, 2013) and increasing richness of semantics for process mining by using domain ontologies (Jareevongpiboon and Janecek, 2013). Moreover, process ontologies are developed by integrating ontologies with business process models to interlink the process information with business dictionaries, improve the explicit representation of process information and to ensure the consistency between external descriptions (Haller *et al.*, 2008). Process ontologies can incorporate reusable and comprehensive knowledge about an organization's processes, activities, roles, application systems, business rules, process interfaces, inputs/outputs and the relationships between them.

Several ontology development methodologies have been proposed for building, reusing and merging domain ontologies manually (Garcia *et al.*, 2010; Suárez-Figueroa *et al.*, 2012; Uschold and Gruninger, 1996). Such domain ontologies are composed of concepts and their relations belonging to a domain of a particular context. Therefore, domain ontologies do not necessarily contain information regarding the process entities of the organization. Several methods have been proposed for integrating domain ontologies with process models to build process ontologies (Alkhaldi *et al.*, 2015; Cherfi *et al.*, 2013; Dimitrov *et al.*, 2007; Fan *et al.*, 2016; Francescomarino *et al.*, 2014; Thomas and Fellmann, 2009). While these approaches proved to be useful, they have an inherent limitation; the formal representation of resulting processes is limited to the knowledge that exists in the underlying domain ontologies. Another approach is transforming process models to process ontologies (De Cesare *et al.*, 2014; Coskuncay *et al.*, 2017; Eisenbarth, 2013; Haller *et al.*, 2008; Höfferer, 2007; Leopold *et al.*, 2015). These studies, in most cases, included an automation tool for transformation and increased the efficiency. However, the process ontologies only cover the knowledge represented in the process models. Hence, both approaches result in incomplete process ontologies and therefore make formal specification of the process information difficult.

In this paper, a different approach is advocated, process ontologies are developed from organizational guidelines. The proposed process ontology methodology is called process ontology population (PrOnPo). It uses natural language parsers as its underlying technology for process ontology development. Previous work, which involved manual construction of process ontologies (Gurbuz and Demirors, 2017a), has identified the requirements of process ontology development from organizational guidelines. The preliminary study has also exposed the requirements for an efficient tool to automatically populate a process ontology from textual guidelines using natural language parser and led to implementation of process model generation tool called ProModGen. Overall, the results of these exploratory studies showed the potential of efficiency gain from the automated process ontology development and the process model creation from the process ontology. In continuation with these efforts, the methodology and the tool is refined and applied in diverse domains. This paper presents PrOnPo, its associated tools and implementation of the approach in five cases. The cases are selected from multiple domains and from processes and guidelines with different characteristics. Process ontologies are developed from these five cases and their accuracy and coverage are measured with precision and recall metrics. Based on the results, the applicability and generalizability for the PrOnPo tool is discussed on different domains and guidelines.

The paper is structured as follows. Section 2 presents the PrOnPo methodology and the associated tool. Section 3 presents the multiple cases study. Section 4 discusses the results and the related work. Section 5 concludes the paper.

## 2. The PrOnPo methodology and tool

This section describes the proposed methodology, which is based on the acquisition of process information from organizational guidelines, the evaluation of the extracted

information and the conceptualization of the extracted knowledge in order to build the process ontology. It also includes a description of the tools that support the methodology.

2.1 The PrOnPo methodology

The methodology focuses on analyzing and interpreting organizational guidelines written in natural languages. Organizational guidelines can be in the form of regulations, directives or procedures that define the set of actions and processes with business rules to be followed within or by entities interacting with the organization. These documents are known to be a primary unstructured source (written in natural language) for extracting process information. Organizational guidelines consist of sentences containing activities and their relations with associated entities, i.e., the role which performs the activity, the information carrier used as input or output, the time when the activity is performed or triggered, the associated event and the business rule which should be considered when performing the activity.

The methodology relies on a predefined conceptual model which consists of classes, individuals, data and object properties. In general, a business process model consists of process elements which have different forms. In PrOnPo ontology, “process” is a super class with “process element” as a sub-class. The concepts “function,” “event,” “information carrier,” “role,” “location,” “application,” “business rule” and “technical term” are sub-classes of “process element” as shown in Figure 1 (Coskuncay *et al.*, 2017). These are the concepts, in other words elements, belonging to event-driven process chain (EPC) modeling notation. A process starts with an event and this event triggers the function. Roles or locations support the execution of the function whereas application automates the execution of the function. A function may have input information carriers or create output information carrier. Technical terms may be related in a function and business rules may rule the execution of the function. A function may also have a sub-process. Consequently, when the execution of the function finishes, an event will be created. All these concepts are process elements and each process element belongs to a process. From ontology point of view, these process elements can be regarded as ontology classes and the relations can be expressed by object properties.

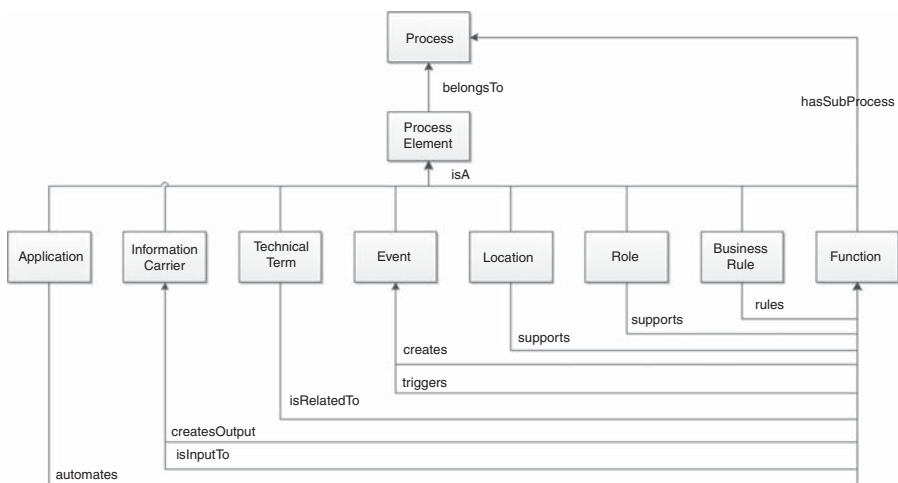


Figure 1. Meta-model for EPC elements and relations

Source: Coskuncay *et al.* (2017)

The aim of the methodology is to populate predefined upper level process ontology classes by extracting ontology individuals from the text. For example, in the sentence of “location plant engineer administers the written contractor safety program,” “location plant engineer” is the individual/instance of the “role” class, “administers” is the individual/instance of the “function” class and “contractor safety program” is the individual/instance of the “information carrier” class. According to the predefined relations (object properties) in Figure 1, “location plant engineer” supports “administers” function, in which “contractor safety program” is input-to. A sample expected output of process ontology in RDF/OWL turtle format is given below:

```
<Location_Plant_Engineer> a <processontology.com/class/role> ,
<processontology.com/property/supports> <administers> .

<administers> a <processontology.com/class/function> .
<the_written_Contractor_Safety_Program>
a <processontology.com/class/informationCarrier> ;
<processontology.com/property/isInputTo> <administers> .
```

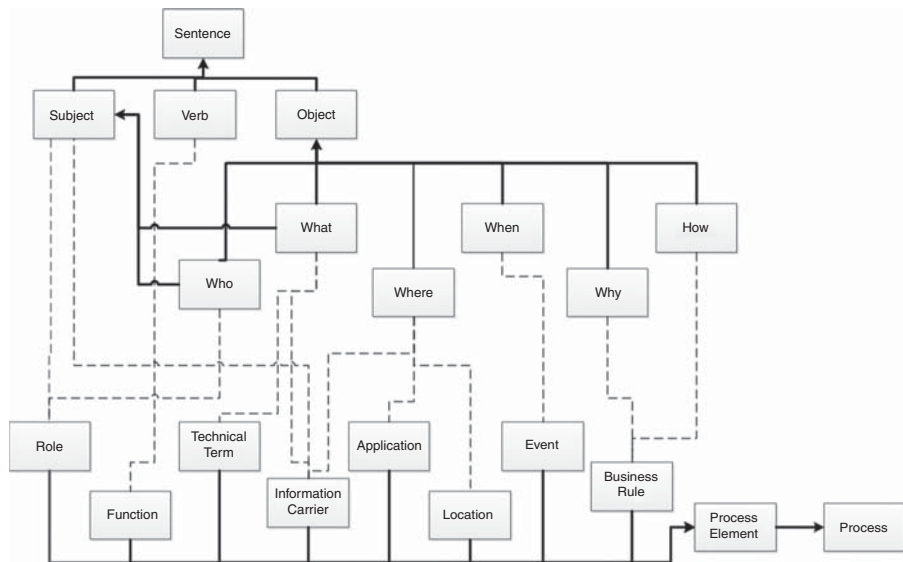
The PrOnPo ontology population is performed in three phases:

- (1) Parsing phase: sentences are extracted and each sentence is structured in a subject–verb–object form.
- (2) Acquisition phase: objects of the sentences are partitioned into what, where, why, when, who and how (5W1H) questions’ answers. Each of these answers and subject–verb are considered as process entities. This phase uses patterns created based on the part-of-speech (POS) tags. The natural language parser marks each word in the sentence with a tag to define whether it is a noun, verb, adjective or adverb, etc., based on its context which are called POS tags. The patterns used for partitioning the sentences using the tags created by the parser (using the Penn TreeBank ([www.surdeanu.info/mihai/teaching/ista555-fall13/readings/PennTreebankConstituents.html](http://www.surdeanu.info/mihai/teaching/ista555-fall13/readings/PennTreebankConstituents.html))) are given in Table I. The relationship “→” in Table I indicates parent→child relation. For example, if the child of verb phrase (VP) tag is noun phrase (NP), then it is placed under “what,” if the child of the root is prepositional phrase (PP) then it is placed under “when.” The rules for each question are separated by commas.
- (3) Conceptualization phase: the process entities extracted in the previous step are conceptualized as ontology individuals under the predefined upper layer process ontology as mentioned earlier and shown in Figure 1. To do so, a meta-model was created to match process entities with process elements as given in Figure 2 (Gurbuz and Demirors, 2017b). This meta-model shows correspondences of process elements in the sentence. For instance, the role element is the subject of the sentence

What	VP → NP, VP → ADJP, S → NP
Where	SBAR → NP, VP → SBAR
Why	PP → IN, VP → TO
When	Root → PP
Who	VP → PP, PP → IN (for, from, by), PP → TO
How	Root → ADVP, PP → VBG, VP → VBG, S → VP → VBG

**Notes:** VP, verb phrase; NP, noun Phrase; ADJP, adjective phrase; S, simple declarative clause; SBAR, clause introduced by a (possibly empty) subordinating conjunction; PP, prepositional phrase; IN, preposition or subordinating conjunction; TO, to; ADVP, adverb phrase; VBG, verb, gerund or present participle

**Table I.**  
Template for  
partitioning the  
object with tags  
and abbreviations



**Figure 2.**  
Meta-model for  
matching the entities  
of the sentence with  
process entities

and answer to who partition, function element is the verb of the sentence, technical term and information carrier is answer to what partition and event element is answer to when partition, etc. For matching subjects, verbs and answers of 5W1H questions are automatically transformed to process ontology individuals and matched with the process ontology, as shown in Figure 2.

### 2.2 The PrOnPo tool

The PrOnPo tool (<http://doi.org/10.5281/zenodo.1439383> (accessed October 14, 2018)) supports the process ontology extraction phases described earlier. It is implemented in Java 8 in an Eclipse programming environment and takes the guidelines in PDF/DOC format as input and creates the output in RDF/OWL and CSV formats. It uses the Stanford Parser, the most common parser (Manning *et al.*, 2014) for extracting sentences and building a tree structure. With the help of POS tags, each sentence is structured into subject–verb–object form. The tool utilizes the Apache Jena (<https://jena.apache.org/> (accessed October 14, 2018)) libraries for the conceptualization phase.

Before the conceptualization phase, the user has the option to review the results of the early phases. Therefore, the tool can be utilized in a semi-automated or fully automated form. In the semi-automated mode, the user can inspect the CSV file for correcting misplaced newly created partitions, when required. In the fully automated mode, the user can skip the evaluation of the process entities created in the acquisition phase and directly conceptualize them into process individuals.

The output of the PrOnPo tool, in which process instances are extracted from the guidelines, can also be used for process modeling. A process model consists of roles, functions, information carriers, business rules, applications, technical terms and events which are already extracted by the PrOnPo tool. Even though the developed process ontology may not fulfill all the requirements of a desired process model, it can provide a skeleton for reflecting on what is written in the document parsed. Process model generation (ProModGen) view plugin (Gurbuz and Demirors, 2017b) was implemented to transform the process ontology instances to process model elements which are stored on a PROMPTUM

Ontology Server as ontology resources (Coskuncay *et al.*, 2016). This transformation not only leads to time and effort savings for process modeling but also helps to precisely reflect on what is written in the guidelines on process models.

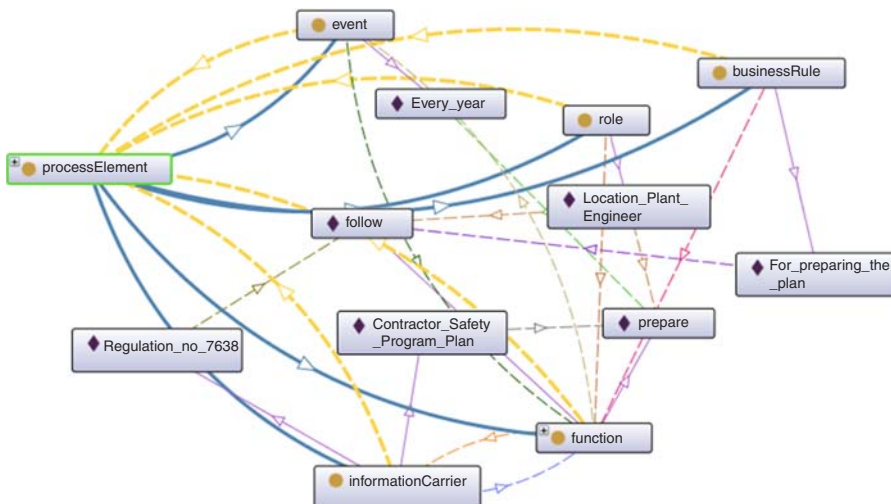
Let us consider two sample sentences given below from a manufacturing process case, which is one of the five selected cases:

“Every year, Location Plant Engineer prepares the Contractor Safety Program Plan. For preparing the plan, Location Plant Engineer follows the rules written in Regulation no. 7638.”

These two sentences are first parsed with Stanford Parser into 5W1H questions as given in Table II. Next, these partitions are converted to process individuals under their related process element classes. Figure 3 shows the output of the PrOnPo tool in which the process elements from the given sentences are formally specified under the process element classes. Business rule, role, function and information carrier are ontology classes and connected to process element class with dashed and bold line indicating subclass relation. The ontology classes are shown with a (yellow) circle node in their labels. These classes' individuals are represented with a (purple) diamond node in their labels and are connected to their classes with a straight (blue) line (such as location plant engineer role, contractor safety program plan-information carrier, for preparing the plan-business rule). The dashed lines between individuals show the relations, such as location plant engineer support-follow- and prepare- functions and contractor safety program plan is an input-to prepare- function.

Role – subject	Function – verb	Information carrier – what	Location – where	Event – when	Business rule – why – how
Location plant engineer	Prepare	Contractor safety program plan		Every year	
Location plant engineer	Follow	Regulation no. 7638			For preparing the plan

**Table II.**  
Motivating example for partitioning the sentences



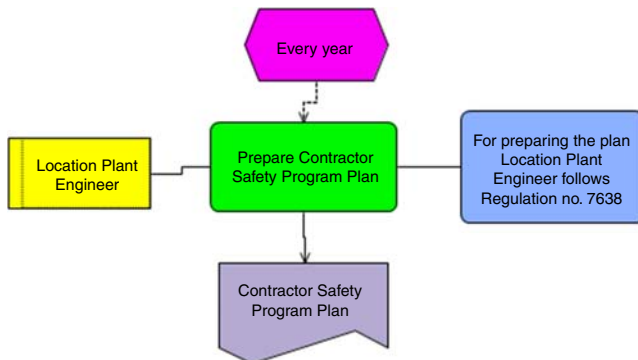
**Figure 3.**  
Excerpt from process ontology

This output process ontology is then uploaded to ProModGen for process modeling. The process ontology is converted back to table format as shown in Table III for the sake of the user to view and select process instances. In this phase, a user can select the rows which he/she thinks should be in the process model, and create process elements which are connected with each other. In other words, ProModGen guides reviewing the document by creating a list of essential constructs, and interactively records the decisions on what should be included in the model and creates a process model based on these decisions.

The automation of an otherwise knowledge intensive process has potential to save time and effort. It also enables a systematic mapping that might eliminate some of the modeling errors. In the example, the first sentence defines an activity of a location plant engineer with an information carrier and event and the second sentence defines a business rule for location plant engineer to follow for executing the activity defined in the previous sentence. In practice, a modeler may easily pass on the second sentence and only reflect the first sentence in a model. However, as PrOnPo generates the complete list of process ontology instances and they can be previewed by ProModGen, the modeler would not miss the opportunity to use them. Using PrOnPo, the modeler can choose the first row on Table III and directly transform each instance to corresponding process element. Second, modeler can choose the second row to create a business rule element as a whole sentence and afterwards connect the business rule element to the function element manually. In brief, these process instances are already created and shown to the user. It is relatively trivial to use them in the process model. An excerpt from the process model which reflects these two sample sentences for the motivating example is shown in Figure 4. This excerpt only shows how two sample sentences are converted to process elements. It is not a complete or proper EPC model of course, but however, an initial motivation for the process ontology to process model transformation. At this point, the user is free to edit and update the model.

**Table III.**  
Sample process  
ontology instances  
preview on ProModGen

Role	Function	Information carrier	Business rule	Event
Location plant engineer	Prepare	Contractor safety program plan		Every year
Location plant engineer	Follow	Regulation no. 7638	For preparing the plan	



**Figure 4.**  
Excerpt from the  
process model for the  
motivating example

### 3. Multiple case study

#### 3.1 *Research questions and validation criteria*

A multiple case study is performed in order to observe the usability of the PrOnPo methodology and its tool in different domains and different types of guidelines. In addition, we aim to determine the efficiency and coverage of model generation using ProModGen. In relation with these goals we have determined two research questions:

*RQ1.* How well does the PrOnPo methodology adapts to different domains with different types of organizational guidelines?

Validation method for *RQ1*: the PrOnPo tool will be used to develop process ontologies from organizational guidelines selected from five different domains with two treatments: fully automated and semi-automated. The results of both treatments will be compared with the manually developed process ontology using precision and recall metrics:

*RQ2.* How does the creation of process models from process instances affect the efficiency and coverage?

Validation method for *RQ2*: the process models generated automatically by ProModGen will be compared with existing process models in terms of efficiency and coverage. The efficiency will be evaluated based on the time and effort required for process modeling. The coverage will be evaluated by comparing the number of process elements in both models. The time and effort required for semi-automated process ontology development will also be compared with those required for manual development.

#### 3.2 *Multiple case study design*

*3.2.1 Case selection.* The case selection criteria is based on eight characteristics: type, organization type, focus domain, focus process, number of pages, number of process model, real-life context and process complexity. The cases that were selected according to these criteria are listed in Table IV. We grouped the type of organizations into government, private and university. We categorized the government and private organization's processes into the most common processes: service and finance. For these organizations, the most common processes were in the domain of human resources. For the specialized processes, we selected the manufacturing domain to generalize the application of the PrOnPo methodology. For a scalable evaluation of the developed process ontology, the number of pages of the selected guideline were restricted to ten pages. We looked for guidelines that include information for at least one process model. The selected cases must reflect the real-life context and must be ecologically valid, in other words should be used in real life. In addition, we selected guidelines which are written in English, preferably by a native speaker.

*3.2.2 Process ontology development.* During the case studies, the process ontologies are developed using three different methods: manual, semi-automated (user guided) and fully automated. For semi-automated and fully automated ontology development, the PrOnPo tool is used. For the manual process ontology development, the following steps are followed:

- (1) analyze the document to identify the processes, business rules, terms and irrelevant information it contains;
- (2) identify activity, role, input/output sources, information systems, business rules or terms related to the selected domain;
- (3) construct the relation of the extracted process elements;
- (4) match the process elements with the related process ontology classes;



Case characteristics	Case 1	Case 2	Multiple cases Case 3	Case 4	Case 5
Name	The banking and financial institutions regulation <sup>a</sup>	Family reunification policy <sup>b</sup>	Contractor safety procedure <sup>c</sup>	Leave policy <sup>d</sup>	Registration of graduate students <sup>e</sup>
Short name	Banking Regulation	Visa Policy	Manufacturing Procedure	HR Policy	University Official website
Type					University
Organization type	Government	Government	Private/ specialized processes	Private/ common processes	University
Focus domain	Finance	Services	Manufacturing	Human resources	University
Focus process	Application and decision making for bank license	Family reunification visa approval process	Approval process, orientation process, training process	Application and monitoring of annual leave process	Registration process
Number of pages	4	10	8	6	2
Number of process model	1	1	3	1	1
Real-life context	Average	Average	Average	Average	Average
Problem complexity	Average	Average	Simple	Average	Simple
<b>Notes:</b>	<sup>a</sup> <a href="http://www.bot.go.tz/BankingSupervision/documents/New%20Docs/The%20Banking%20and%20Financial%20Institutions%20(Licensing)%20Regulations,%202014.pdf">www.bot.go.tz/BankingSupervision/documents/New%20Docs/The%20Banking%20and%20Financial%20Institutions%20(Licensing)%20Regulations,%202014.pdf</a> (accessed March 21, 2017); <sup>b</sup> <a href="http://www.inis.gov.ie/en/INIS/Family%20Reunification%20Policy%20Document.pdf/Files/Family%20Reunification%20Policy%20Document.pdf">www.inis.gov.ie/en/INIS/Family%20Reunification%20Policy%20Document.pdf/Files/Family%20Reunification%20Policy%20Document.pdf</a> (accessed April 11, 2017); <sup>c</sup> <a href="http://www.schneider-electric.us/documents/company/supplier-resource-center/P-P_Contractor_Safety.doc">www.schneider-electric.us/documents/company/supplier-resource-center/P-P_Contractor_Safety.doc</a> (accessed September 22, 2016); <sup>d</sup> <a href="http://www.nice.org.uk/media/default/about/who-we-are/policies-and-procedures/leave-policy.pdf">www.nice.org.uk/media/default/about/who-we-are/policies-and-procedures/leave-policy.pdf</a> (accessed March 9, 2017); <sup>e</sup> <a href="http://oidb.metu.edu.tr/en/registration-graduate-students-0">http://oidb.metu.edu.tr/en/registration-graduate-students-0</a> (accessed April 18, 2017)				

**Table IV.**  
Summary of the  
selected cases for  
the study

- (5) build the process ontology;
- (6) evaluate the process ontology according to the competency questions; and
- (7) record the ontology individuals that are required.

This manually developed process ontology was used as the baseline to evaluate the process ontologies developed by PrOnPO. The manual process ontology development results are recorded using an Excel sheet. The RDF output of the developed process ontology from the tool is stored in Protégé 4.3 (<https://protege.stanford.edu/> (accessed September 29, 2018)).

The semi-automated and fully automated process ontology individuals are compared using recall and precision metrics (Brewster *et al.*, 2004) with the manually developed process ontology individuals. These metrics are utilized in the information retrieval process to define the percentage of the correct and relevant retrievals. For each case, time and effort for each task is recorded.

**3.2.3 Process model generation.** We have utilized existing process models for two cases studies. The registration process from university domain had already been modeled as part of a course assignment in a BPMN Business Process Modeling Notation (BPMN) editor. The students were given the registration process guideline and asked to model the process on a web-based editor in BPMN notation. In addition, contractor approval, orientation and training processes from manufacturing domain were already modeled by a different analyst using the guideline on UPROM (Aysolmaz and Demirörs, 2015).

---

Consequently, we compared the prior manually established models with the process models generated using the ProModGen tool using the process instances extracted during process ontology development for these two case studies.

### 3.3 Mitigation of threats to validity

We performed the following actions to prevent threats that could affect the internal, construct and external validity (Wohlin *et al.*, 2012). For this case study, one potential threat is the comparison of the process models which are created by the same modeler. In order to mitigate this threat, manual and assisted process models are created by different analysts. Additionally, we assumed that the lengths of the texts in the selected guidelines could also affect the results. Thus, we limited the length of the documents, in order to prevent any bias that could result from the exhaustion of the analysts during the process ontology development and evaluation. On the other hand, in order to avoid another potential threat regarding the length selection, we increased the number of selected cases, thus preventing a mono-operation bias.

A potential external validity threat is related with the restricted domain of application. We mitigated this threat by conducting a multiple case study by collecting different types of guidelines from different domains.

We explicitly defined and used objective measures to mitigate threats to construct validity. This study focuses on the process ontology development from the selected guidelines. The focus of the research questions is the extent to which the proposed technique is successful in developing a process ontology from selected domains, and measures the applicability in terms of precision and recall metrics.

However, this study's focus is to analyze the adaptability and generalizability of the presented PrOnPo methodology. Therefore, this study does not cover the examining of the overall benefit to business experts nor conducting interviews with the process owners in terms of ease of use.

### 3.4 Results and analysis regarding RQ1

The results of the analysis are given in three dimensions: guideline, ontology individuals and time. The metrics collected for the comparison are given as a true-false matrix and precision/recall results. The results are summarized in Table V. The sentences presented in the guidelines are grouped into four categories of irrelevant sentences, terms, business rules and processes. Table V also contains the results of the analysis on the PrOnPo tool regarding the parsed and corrupted sentences. "The number of sentences" row shows the actual number of sentences in the guidelines. "The parsed sentences" row shows the number of sentences which could be parsed by the parser. The "Missing, Parsed and Corrupted sentences" row shows the number of sentences which were incorrectly parsed and the number of sentences which could not be parsed by the parser, in other words the forming of subject-verb-object was unsuccessful.

The numbers of the actual ontology individuals identified in the manual process ontology development is given in the first row of Table VI. The next rows give the number of extracted, replaced, missing and irrelevant individuals, respectively. Using these numbers, the true-false metrics are created. The correctly retrieved individuals are considered as true positives and the retrieved but irrelevant individuals are considered as false positives. The missing individuals which were not retrieved but are relevant are considered as false negatives. The true-false matrix summary given in Tables VII and VIII depicts the semi-automated and fully automated process ontology development, respectively. Regarding the fully automated matrix, the individuals which should be modified by the user are considered either as false positives since they are in the wrong place or as false negatives since they are relevant but not retrieved.

**Table V.**  
Summary of the  
selected guidelines

	Manufacturing	Visa	University	HR	Banking
<i>Number of sentences</i>					
Total	153	37	40	29	36
Irrelevant	7	1	0	2	0
Terms	28	0	0	0	0
BusinessRules	71	21	27	15	17
Process	47	15	13	12	19
<i>Parsed sentences</i>					
Total	129	27	39	26	32
Irrelevant	5	0	0	2	0
Terms	27	0	0	0	0
BusinessRules	58	12	26	14	15
Process	39	15	13	10	17
<i>Missing, parsed and corrupted sentences</i>					
Total	26	11	2	4	10
Irrelevant	2	1	0	1	0
Terms	1	0	0	0	0
BusinessRules	15	7	2	1	5
Process	9	3	0	2	5

**Table VI.**  
Summary of  
the ontology  
individual analysis

	Manufacturing	Visa	University	HR	Banking
Number of actual individuals	486	134	164	126	168
Number of individuals extracted	448	100	157	114	136
Number of individuals replaced by user	52	17	28	24	32
Number of missing individuals	63	34	7	15	32
Number of irrelevant individuals	25	0	0	3	0

In the next step, the precision and recall metrics are calculated according to the formulas given below:

$$\text{True positive}/(\text{true positive} + \text{false positive}) = \text{Precision.}$$

$$\text{True positive}/(\text{true positive} + \text{false negative}) = \text{Recall.}$$

The summary of the results is given in Table IX. These metrics are calculated separately for the semi- and fully automated process ontology development according to Tables VII and VIII. The recall metric calculates the correctly extracted individuals over all the relevant individuals whereas the precision metric calculates the correctly extracted individuals over all the retrieved individuals. According to the precision and recall metrics, it is obvious that the precision and recall metrics for semi-automation are higher than full automation. These results provide an insight concerning the effectiveness of the user guide during the process ontology development. The success of the PrOnPo tool for retrieving the relevant individuals (recall) ranged from 75 to 96 percent with an average of 85 percent. Discarding the missing individuals, the success of the PrOnPo tool for retrieving relevant individuals over all the retrieved individuals (precision) ranged from 93 to 100 percent with an average of 98 percent. These results show that the PrOnPo tool and methodology are applicable for different domains and guideline types.

	Manufacturing		Human resources		University		Visa		Banking	
	Relevant	Irrelevant	Relevant	Irrelevant	Relevant	Irrelevant	Relevant	Irrelevant	Relevant	Irrelevant
Semi Retrieved	486	FP = 25	126	FP = 3	164	FP = 0	134	FP = 0	168	FP = 0
Not retrieved	TP = 423 FN = 63		TP = 111 FN = 15		TP = 157 FN = 7		TP = 100 FN = 34		TP = 136 FN = 32	

**Notes:** TP, true positive; FP, false positive; FN, false negative

**Table VII.**  
True-false matrix  
summary for  
semi-automated

**Table VIII.**  
True-false matrix  
summary for  
fully automated

	Manufacturing		Human resources		University		Visa		Banking	
	Relevant	Irrelevant	Relevant	Irrelevant	Relevant	Irrelevant	Relevant	Irrelevant	Relevant	Irrelevant
Full Retrieved	486		126		164		134		168	
Not retrieved	TP = 371 FN = 63+52	FP = 25+52	TP = 87 FN = 15+24	FP = 3+25	TP = 129 FN = 7+28	FP = 28	TP = 83 FN = 34+17	FP = 17	TP = 104 FN = 32+32	FP = 32
<b>Notes:</b> TP, true positive; FP, false positive; FN, false negative										

Regarding the precision and recall metrics for different domains and guideline types, we calculated the  $F$ -measure (the harmonic means of precision and recall metrics) to make a comparison between the different cases, as shown in Table X.

Regulation and policies which belong to the government case consist of more complex structured sentences. We observed that process ontology development for banking and visa have the lowest  $F$ -measure compared to the other categories. University has the highest  $F$ -measure which also has the least complex structure. In other words, the website consists of to-the-point process sentences. Likewise, the private companies' specialized procedure also consists of less complex sentences which exactly define the processes. The  $F$ -measure for specialized processes (manufacturing) is high. We can state that, the performance of the PrOnPo tool regarding the extraction of the relevant and correct process information is mostly dependent on the structure and complexity of the sentences.

### 3.5 Results and analysis regarding RQ2

The summary of the time and effort required to perform the studies for the manual and semi-automated process ontology development is given in Table XI. The time and effort data are collected in terms of person-minutes. The time and effort for semi-automated approach includes the process ontology development time of the PrOnPo tool and the replacement of the individuals. The total effort for manual development was 250 person-minutes whereas for the semi-automated development, it was 57 person-minutes.

Case	Precision TP/(TP + FP)		Recall TP/(TP + FN)	
<i>Manufacturing</i>				
Semi	423/448	94%	423/486	87%
Full	371/448	83%	371/486	76%
<i>Visa</i>				
Semi	100/100	100%	100/134	75%
Full	83/100	83%	83/134	62%
<i>University</i>				
Semi	157/157	100%	157/164	96%
Full	129/157	82%	129/164	79%
<i>Human resources</i>				
Semi	111/114	97%	111/126	88%
Full	87/114	76%	87/126	69%
<i>Banking</i>				
Semi	136/136	100%	136/168	81%
Full	104/136	76%	104/168	62%

**Table IX.**  
Summary of precision  
and recall metrics

Case		$F$ -measure
Specialized processes	Manufacturing	0.79
Common process	HR	0.72
University process website	University	0.80
Government regulations	Banking	0.68
Government regulations	Visa	0.71

**Note:**  $F$ -measure =  $2 \times \text{precision} \times \text{recall} / (\text{precision} + \text{recall})$

**Table X.**  
 $F$ -measure summary

These results show that process ontology development with the PrOnPo methodology and the tool is significantly more efficient than manual process ontology development.

The summary of the comparison of the process models created by ProModGen to the existing process created before the case study is shown in Table XII. Existing process models were created from scratch using guidelines by different modelers and the process models created by ProModGen were created from the process instances which were the output of PrOnPo. *M* stands for manual creation (existing models) and *P* stands for the ProModGen models. The required time and effort is recorded in person-minutes.

In terms of time and effort, the process models created by ProModGen was significantly better than the manual process modeling time and effort. In terms of coverage, the results required further analysis. We analyzed the process models according to their information coverage by analyzing the function elements. We first prepared a list of activities (activity checklist) that the process models should include. Then, for each process model, we compared this activity checklist with their function elements to see if the process model mentions the activity. Afterwards, we checked if the function elements had their supporting process elements, such as role, information carrier and business rule elements. The results are depicted in Table XII.

The process information of contractor safety procedure guidelines is mainly composed of high-level activities, thus high-level descriptive process models are generated. The time comparison results of the manufacturing processes reveal that the process model generation was more effective when the process model is complex. The required time is decreased from 38 to 10 person-minutes for the approval process, which is the most complex process model, from 23 to 14 person-minutes for the orientation process and from 8 to 6 person-minutes for the training process. For approval process, all the functions are the same, but the generated model has an additional function which is not stated in the existing model. The generated model has more business rules which refers to the documents used in the functions. For orientation process, although both models cover the same information based on the

**Table XI.**  
Summary of the effort analysis

Type	Manufacturing	Visa	University	HR	Banking
<i>Time</i>					
Manual	120 p-m	30 p-m	30 p-m	30 p-m	40 p-m
Semi	15 p-m	12 p-m	10 p-m	10 p-m	10 p-m

**Table XII.**  
Process model coverage and time

Focus domain	Manufacturing						University	
	Approve contractor		Give orientation		Train contractor		Registration process	
Process models	<i>M</i>	<i>P</i>	<i>M</i>	<i>P</i>	<i>M</i>	<i>P</i>	<i>M</i>	<i>P</i>
Treatments	eEPC	eEPC	eEPC	eEPC	eEPC	eEPC	BPMPN	eEPC
Notation	38 p-m	10 p-m	23 p-m	14 p-m	8 p-m	6 p-m	1 person 4 h	1 person 23 min
No. of functions	5	7	8	6	2	3	30	9
No. of events	4	5	5	6	2	2	X	3
No. of business rules	1	10	0	7	0	4	9	19
No. of information carriers	9	3	5	X	2	X	14	5
No. of roles	4	8	9	5	3	3	2	11
No. of connector	2	3	2	3	0	X	12	1
No. of application	X	X	X	X	X	X	7	X
Total	25	36	29	27	9	12	77	48

activity checklist, the generated model is more descriptive while the existing model is more general. The generated model again has more business rules. For training process, the generated model has one or more functions and hence includes the information that is not available in prior manually established models. For registration process, since the notation of the process models are different, we compared the information coverage of the models by comparing activities with functions. The prior manually established model has more detail whereas the generated model is a high-level descriptive one. The prior manually established model gives more detailed information in relation to the document collection for registration.

Overall, regarding the answer to *RQ2*, ProModGen is more efficient in terms of time and effort compared to manual development. It has created process models with a high-level description that match with activity checklist. In addition, the functions in the generated process models have more detailed supporting elements (roles, business rules and information carriers). However, as the models are always generated at the same abstraction level of the document it does not depict the details at various abstraction levels as the case might require.

#### 4. Related work and implications

In the literature, studies related to process ontologies has addressed the problem using two different approaches. The first approach consists of methods to integrate domain ontologies with process models to build process ontologies (Alkhaldi *et al.*, 2015; Bulegon Gassen *et al.*, 2017; Cherfi *et al.*, 2013; Dimitrov *et al.*, 2007; Fan *et al.*, 2016; Francescomarino *et al.*, 2014; Thomas and Fellmann, 2009). Although these approaches improve efficiency, the formal representation of processes is limited to the knowledge that already exists in domain ontologies. The second approach focus on transforming process models to process ontologies (De Cesare *et al.*, 2014; Coskuncay *et al.*, 2017; Eisenbarth, 2013; Haller *et al.*, 2008; Höfferer, 2007; Leopold *et al.*, 2015). Transforming process models to process ontologies is possible but the resulting process ontologies only cover the knowledge represented in the process models. This leads to the incompleteness of the process ontology and formal specification of the process information.

In the manner of using ontologies for process modeling, the works by Alkhaldi, Fan, Bulegon and Francescomarino (Alkhaldi *et al.*, 2015; Bulegon Gassen *et al.*, 2017; Fan *et al.*, 2016; Francescomarino *et al.*, 2014) are the most related studies to the approach we propose. They have validated the improvement in the quality of process models using domain ontology (Fan *et al.*, 2016), semantic consistency using enterprise ontology (Alkhaldi *et al.*, 2015) and semantic annotation (Francescomarino *et al.*, 2014). Francescomarino *et al.* (2014) use semantic annotations for collaborative process modeling in BPMN. Their experiment concentrates on the effect of using semantic annotations on the variables defined for collaborative process modeling and as well as the quality of process models. There is an agreement in the proposed and existing research, on preventing conflicts using semantic annotations in process modeling. Similar to Francescomarino's study, Alkhaldi *et al.* (2015) use the enterprise ontology for creating consistent process models in BPMN notation. However, they do not propose a transformation approach. In this manner, the work of Fan *et al.* (2016) is similar to ours in that they convert domain ontologies to process ontologies and semi-automatically generate a BPMN model using these process ontologies. But, however, the proposed method expects that domain ontology is already built. Lastly, Bulegon Gassen *et al.* (2017) also claims the enhancement in process model by the support of ontologies with an experiment. They note that the ontology usage enhances the usage of good labels and terminology of the process models by examining two groups, in which one uses ontology support and one does not.

Overall, our approach differs significantly from the available studies in the literature. Our approach does not rely on existing formal process models or existing ontologies.



---

We use widely available guidelines in organizations written in natural language as the sole input. Using natural language processing methods, we transform them into process ontologies and generate integrated process models. Main implication of this transformation according to the results of the case study is that what is written in the guidelines are what is represented in the process ontology and models. In other words, roles, functions, information carriers, etc., are explicitly defined in the process ontology as stated in the guidelines, thus preventing any misunderstanding of a concept based on its context and creating a shared understanding within the organization. Moreover, from the business process modeling point of view, the work effort for defining process elements from guidelines (from scratch) is reduced since they will be already defined. Second implication of this study based on the findings is that it does not require ontology knowledge for the business expert. Since the most of the work is done by the method and the sentences are transformed to the ontology file in the background, only work left to the business expert is correcting the sentences.

### 5. Conclusion and future work

In this paper, we presented the application of the PrOnPo methodology to different guideline types in a multiple case study involving five different domains. Unlike existing approaches which use existing domains ontologies or process models for creating process ontologies, PrOnPo methodology uses the organizational guidelines for developing process ontologies. Encouraged by the studies which use domain ontologies for creating process models, this study also presents ProModGen plugin which uses process ontologies for creating process models. Overall, our methodology is the first one to automatically generate integrated ontologies and process models from guidelines written in natural language. It can be used by analysts/users that are not necessarily experienced in ontology building as the proposed methodology and associated tools provide guidance for ontology development process. The integration of ontology and process models enables organizations to improve consistency among models and significantly enhances process querying and visualization alternatives (Bulegon Gassen *et al.*, 2017; Haller *et al.*, 2008).

We observed that the methodology and supporting tools can be useful in practice to model processes and process ontologies directly from the guidelines written in a natural language. To measure its effectiveness, we analyzed the precision and recall metrics for both fully automated and semi-automated development. The overall average precision (extracted the relevant process individuals over all the extracted individuals) was 98 percent in the semi-automated development and 80 percent in the fully automated development. The output process ontology of semi-automation covered 85 percent of the overall process information whereas the output of full automation covered 70 percent.

We also identified the impact of the guideline types. PrOnPo is recorded to be more effective for a specialized process guideline than for a government regulation. Based on these results, we conclude that the PrOnPo methodology is promising in terms of usability in different domains. Regarding the efficiency of the PrOnPo tool, we observed that there was a four-fold reduction in the required time and effort for process ontology development. Regarding the efficiency of the ProModGen tool, the required time and effort for process modeling reduced by more than half. Overall, the required time and effort for process ontology development and process modeling for manufacturing process is reduced three-fold.

We also identified challenges that can be addressed by future studies. We observed that the generated process ontologies could be useful for solving the natural language ambiguity and inconsistency problems within guidelines. Our methodology can be integrated with

WordNet (Fellbaum, 2012) to link process ontology classes with their synonyms. Furthermore, the developed process ontology can also be used as a knowledge base for question answering systems or enterprise search engines for exploring process information as well as relating to the domain ontologies or data analytics (Milosevic *et al.*, 2016) of the corresponding domain. Taking advantages of the Semantic Web technologies, the process ontology can be turned into a knowledge graph of processes and can be used for graph mining to provide recommendation systems based on the similarities found between different processes. Moreover, inference and reasoning capabilities of ontologies may lead to finding hidden relations between entities which would help the end user to discover process information deeply.

Moreover, we also plan to address the limitations of the research in the future, by practicing the methodology on all of the guidelines of one organization and examining the overall benefit to business experts. Additionally, conducting interviews with process owners would also enhance the validity of the methodology in terms of its ease of use and contribution to the business and therefore left for future work.

We are planning future work on a supervised machine learning approach for partitioning objects to minimize user involvement and obtain better performance. We will collect sentences from different sources and create a data set for 5W1H question answering. We will train this data set so that the system will learn the patterns for partitioning the sentence into 5W1H questions. Thus, it is believed that the algorithm will learn the patterns according to POS tags and dependency paths between words itself and create new patterns according to different sentence structures.

We also plan to extend our methodology to guide users in reusing the generated process ontologies when constructing domain ontologies and/or when performing data analytics. In addition, we plan to integrate both tools within a single framework which will provide comprehensive support for process ontology development from text analysis to model generation. This would involve an experiment to test end-to-end model generation process from natural language sources.

## References

- Alkhalidi, N., Casteleyn, S. and Gailly, F. (2015), "Enterprise-specific ontology-driven process modelling papers", in Cordeiro, J., Hammoudi, S., Maciaszek, L. and Camp, O.F.J. (Eds), *Lecture Notes in Business Information Processing*, Vol. 227, Springer, Charm, pp. 472-488.
- Aysolmaz, B. and Demirörs, O. (2015), "Unified process modeling with UPROM tool", in Nurcan, S.P.E. (Ed.), *Information Systems Engineering in Complex Environments, CAiSE 2014, Lecture Notes in Business Information Processing*, Vol. 204, Springer, Charm, pp. 250-266.
- Aysolmaz, B., Leopold, H., Reijers, H.A. and Demirörs, O. (2018), "A semi-automated approach for generating natural language requirements documents based on business process models", *Information and Software Technology*, Vol. 93, pp. 14-29.
- Brewster, C., Alani, H., Dasmahapatra, S. and Wilks, Y. (2004), "Data driven ontology evaluation", *Fourth International Conference on Language Resources and Evaluation*, pp. 641-644.
- Bulegon Gassen, J., Mendling, J., Bouzeghoub, A., Thom, L.H., Palazzo, M. and de Oliveira, J. (2017), "An experiment on an ontology-based support approach for process modeling", *Information and Software Technology*, Vol. 83, pp. 94-115.
- Cherfi, S.S.-S., Ayad, S. and Comyn-Wattiau, I. (2013), "Aligning business process models and domain knowledge: a meta-modeling approach", in Morzy, T., Härder, T. and Wrembel, R. (Eds), *BT - Advances in Databases and Information Systems*, Springer, Berlin and Heidelberg, pp. 45-56.
- Coskuncay, A., Gurbuz, O. and Demirors, O. (2017), "Transformation from business process models to process ontology: a case study", *MCIS 2017 Proceedings*, p. 40.

- Coskuncay, A., Gurbuz, O., Demirors, O. and Ekinci, E.E. (2016), "PROMPTUM toolset: tool support for integrated ontologies and process models", in Dumas, M. and Fantinato, M. (Eds), *Business Process Management Workshops, BPM 2016*, Lecture Notes in Business Information Processing, Vol. 281, Springer, Rio de Janeiro, pp. 93-105.
- De Cesare, S., Juric, D. and Lycett, M. (2014), "Toward the automation of business process ontology generation", *Proceedings – 16th IEEE Conference on Business Informatics*, Vol. 1, pp. 70-77.
- Dimitrov, M., Simov, A., Stein, S. and Konstantinov, M. (2007), "A BPMO based semantic business process modelling environment", *Proceedings of the Workshop on Semantic Business Process and Product Lifecycle Management*, Vol. 251, pp. 1-4.
- Dumas, M., La Rosa, M., Mendling, J. and Reijers, H.A. (2013), *Fundamentals of Business Process Management*, Vol. 1, Springer, Heidelberg.
- Eisenbarth, T. (2013), *Semantic Process Models: Transformation, Adaption, Resource Consideration*, University of Augsburg, Augsburg.
- Fan, S., Hua, Z., Storey, V.C. and Zhao, L. (2016), "A process ontology based approach to easing semantic ambiguity in business process modeling", *Data & Knowledge Engineering*, Vol. 102, pp. 57-77.
- Fellbaum, C. (2012), "WordNet", in Chappelle, C. (Ed.), *The Encyclopedia of Applied Linguistics*, doi: 10.1002/9781405198431.wbeal1285.
- Fengel, J. (2014), "Semantic technologies for aligning heterogeneous business process models", *Business Process Management Journal*, Vol. 20 No. 4, pp. 549-570.
- Francescomarino, C.D., Rospocher, M., Ghidini, C. and Valerio, A. (2014), "The role of semantic annotations in business process modelling", *IEEE 18th International Enterprise Distributed Object Computing Conference*, pp. 181-189.
- Garcia, A., O'Neill, K., Garcia, L.J., Lord, P., Stevens, R., Corcho, O. and Gibson, F. (2010), "Developing ontologies within decentralised settings", *Semantic E-Science*, Vol. 10, pp. 99-139.
- Gruber, T.R. (1993), "A translation approach to portable ontology specifications", *Knowledge Acquisition*, Vol. 5 No. 2, pp. 199-220.
- Gurbuz, O. and Demirors, O. (2017a), "A comparison of process ontology discovery from organizational guidelines in two different languages", *Proceedings – 43rd Euromicro Conference on Software Engineering and Advanced Applications*, pp. 363-370.
- Gurbuz, O. and Demirors, O. (2017b), "From organizational guidelines to business process models: exploratory case for an ontology based methodology", *IEEE 19th Conference on Business Informatics*, Vol. 1, pp. 320-329.
- Haller, A., Marmolowski, M., Oren, E. and Gaaloul, W. (2008), "A process ontology for business intelligence", *Digital Enterprise Research Institute, Technical Report*, April 1.
- Höfferer, P. (2007), "Achieving business process model interoperability using metamodels and ontologies", *ECIS 2007 Proceedings*, pp. 1620-1631.
- Jareevongpiboon, W. and Janecek, P. (2013), "Ontological approach to enhance results of business process mining and analysis", *Business Process Management Journal*, Vol. 19 No. 3, pp. 459-476.
- Kaya, M. and Demirors, O. (2011), "E-cosmic: a business process model based functional size estimation approach", *37th EUROMICRO Conference on Software Engineering and Advanced Applications*, pp. 404-410.
- Leopold, H., Meilicke, C., Fellmann, M., Pittke, F., Stuckenschmidt, H. and Mendling, J. (2015), "Towards the automated annotation of process models", *27th International Conference Advanced Information Systems Engineering*, pp. 401-416.
- Manning, C.D., Bauer, J., Finkel, J., Bethard, S.J., Surdeanu, M. and McClosky, D. (2014), "The Stanford CoreNLP natural language processing toolkit", *Proceedings of 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, pp. 55-60.
- Milosevic, Z., Chen, W., Berry, A. and Rabhi, F.A. (2016), "An open architecture for event-based analytics", *International Journal of Data Science and Analytics*, Vol. 2 Nos 1-2, pp. 13-27.

- 
- Mturi, E. and Johannesson, P. (2013), "A context-based process semantic annotation model for a process model repository", *Business Process Management Journal*, Vol. 19 No. 3, pp. 404-430.
- Suárez-Figueroa, M.C., Gómez-Pérez, A., Motta, E. and Gangemi, A. (2012), "Ontology engineering in a networked world", *Springer Science & Business Media*, p. 444.
- Tarhan, A. and Demirors, O. (2012), "Apply quantitative management now", *IEEE Software*, Vol. 29 No. 3, pp. 77-85.
- Thomas, O. and Fellmann, M. (2009), "Semantic process modeling – design and implementation of an ontology-based representation of business processes", *Business & Information Systems Engineering*, Vol. 1 No. 6, pp. 438-451.
- Turetken, O. and Demirors, O. (2011), "Plural: a decentralized business process modeling method", *Information & Management*, Vol. 48 No. 6, pp. 235-247.
- Uschold, M. and Gruninger, M. (1996), "Ontologies: principles, methods and applications", *The Knowledge Engineering Review*, Vol. 11 No. 2, pp. 93-136.
- Wohlin, C., Runeson, P., Höst, M., Ohlsson, M.C., Regnell, B. and Wesslén, A. (2012), *Experimentation in Software Engineering*, Springer Science & Business Media.

**Corresponding author**

Ozge Gurbuz can be contacted at: [gurbuzozge@gmail.com](mailto:gurbuzozge@gmail.com)