View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by MAnnheim DOCument Server

Challenges and Opportunities of Applying Natural Language Processing in Business Process Management

Han van der Aa
Humboldt University
Berlin, Germany
vanderah@hu-berlin.de

Josep Carmona
Univ. Politècnica de Catalunya
Barcelona, Spain
jcarmona@cs.upc.edu

Henrik Leopold
Vrije Univ. Amsterdam
Amsterdam, The Netherlands
h.leopold@vu.nl

Jan Mendling
Vienna Univ. of Economics and Business
Vienna, Austria
jan.mendling@wu.ac.at

Lluís Padró
Univ. Politècnica de Catalunya
Barcelona, Spain
padro@cs.upc.edu

Abstract

The Business Process Management (BPM) field focuses in the coordination of labor so that organizational processes are smoothly executed in a way that products and services are properly delivered. At the same time, NLP has reached a maturity level that enables its widespread application in many contexts, thanks to publicly available frameworks. In this position paper, we show how NLP has potential in raising the benefits of BPM practices at different levels. Instead of being exhaustive, we show selected key challenges were a successful application of NLP techniques would facilitate the automation of particular tasks that nowadays require a significant effort to accomplish. Finally, we report on applications that consider both the process perspective and its enhancement through NLP.

1 Introduction

In the last decade, the maturity achieved by Natural Language Processing (NLP) technologies, together with the explosion of *big data* and *deep learning* techniques, have turned the spotlight to the possibilities offered by NLP approaches for a variety of novel applications. Many of these applications are situated in a business setting where documents and textual data is extensively used to manage production, logistics, accounting, and procurement, to give just a few examples. These application areas have in common that they relate to various business processes that are executed inside a company and beyond.

Organizing business processes in an efficient and effective manner is the overarching objective of business process management. Classically, business process management has been many concerned with the quantitative analysis of key performance dimensions such as time, cost, quality, and flexibility (Dumas et al., 2013) without taking the automatic processing of textual data too much into account. Recent research highlights the potential of NLP-based analysis techniques (Leopold, 2013; Mendling et al., 2015; Mendling et al., 2017) to support many business process management tasks in a scalable fashion.

In this paper, we describe the application of NLP techniques to BPM, where the focus rests on the processes that an organization must carry out on its daily activities, and on how are they modeled, updated, optimized, and shared with the relevant stakeholders. We believe that the NLP community has much to offer to BPM, as well as many interesting challenges to address, and we aim to display the most relevant in this paper.

The rest of the paper is structured as follows. Section 2 describes the background of our research. Section 3 provides an outline how NLP could inform business process management tasks in the future. Section 4 highlights the potential impact of NLP-supported business process management in various domains. Finally, Section 5 concludes the paper.

2 Background

In this section, we describe the essential ideas of business process management and its major design and analysis artifacts, which are process representations and event logs. Likewise, we provide a current

This work is licenced under a Creative Commons Attribution 4.0 International Licence. Licence details: http://creativecommons.org/licenses/by/4.0/

Proceedings of the 27th International Conference on Computational Linguistics, pages 2791–2801
Santa Fe, New Mexico, USA, August 20-26, 2018.

perspective on NLP and how it can be oriented towards improving BPM in a general setting.

2.1 Requirements of Business Process Management

Business process management is concerned with various management activities that are related to business processes (Dumas et al., 2013). In line with Mendling et al. (2017), we describe three levels of business process management as illustrated in Figure 1. The top level called multi process management is concerned with the identification of the major processes of an organization and the prioritization of these. The middle level is concerned with the management of a single process along the classical BPM lifecycle (Dumas et al., 2013) including the steps of discovery, analysis, redesign, implementation and controlling. The level of process instance management deals with planning the tasks of a process, executing them, monitoring them, and potentially adapting the instance if required. All the three layers make use of business process models and event data.

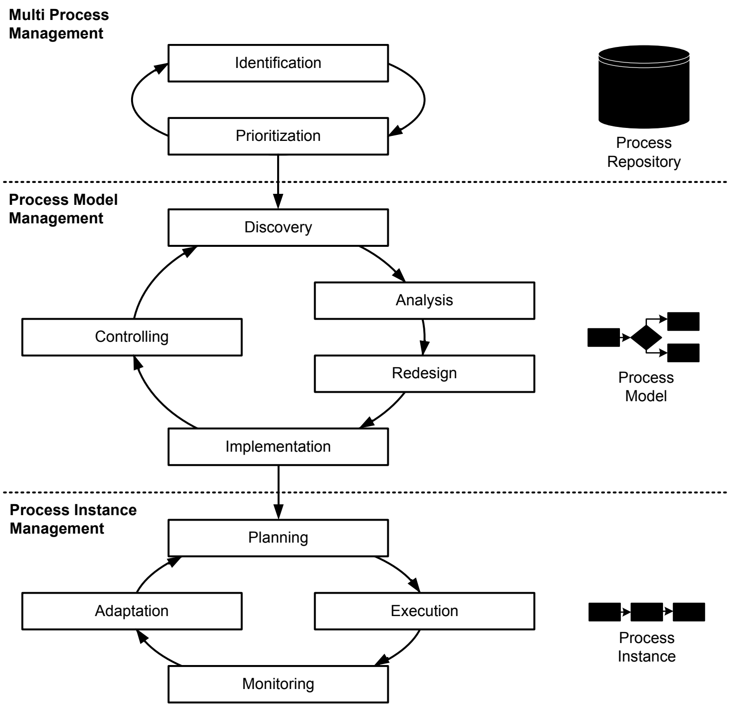

Figure 1: Three Levels of Business Process Management (taken from (Mendling et al., 2017)).

2.2 Process Representations

A variety of representation formats can be used to capture process information in informational artifacts (Wolter and Meinel, 2010), including process models (Davies et al., 2006), natural language descriptions (Phalp et al., 2007), spreadsheets (Krumnow and Decker, 2010), and checklists (Reijers et al., 2017). The representation format used to provide process information to users should be well-suited for its particular purpose, in two ways: A format should convey its informational content in a useful manner and the intended users should be able to appropriately understand the received format (van der Aa, 2018).

Representation formats emphasize different aspects of business processes. This means that the choice for a certain representation format depends on the intended focus of an informational artifact. For instance, *natural language text* can be very useful to provide process participants with detailed insights on how to perform complex tasks (Baier and Mendling, 2013). However, for a process participant who needs to be sure that all necessary steps are performed, a *checklist* might be more useful. This later format could be more suitable because it emphasizes the information that is of primary importance for that purpose. Furthermore, process models have been found to be better suited to express complex execution logic of a process in a more comprehensive manner than natural language (Mendling, 2008, p.23).

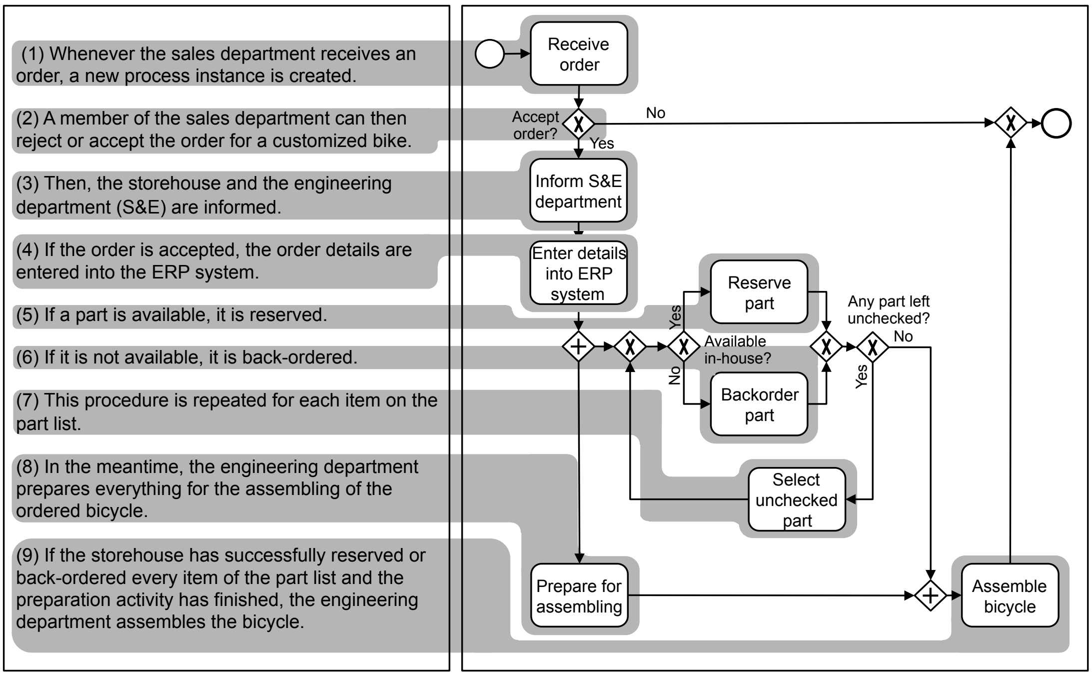

Figure 2: A textual and a model-based description of a bicycle manufacturing process.

It is also important that users of an informational artifact are able to work well with the employed representation format. The ability of users to do so can depend on their familiarity and preferences with respect to different formats. Research by Figl and Recker (2016) shows that people prefer different process representation formats depending on the application purpose and on the *cognitive style* of the user. For example, some participants were found to prefer textual descriptions over process models, whereas others preferred models over text for the same purpose. The influence of user preferences on the choice for model or text-based process representations is also recognized by Recker et al. (2012) and Chakraborty et al. (2010).

Figure 2 shows an example of two different process representations, namely a process model and a textual process description. On the left- hand side, we observe a textual description, which comprises nine sentences. On the right-hand side, a corresponding model-based description can be seen, modeled using the Business Process Model and Notation (BPMN). The model contains eight activities, which are depicted using boxes with rounded edges. The diamond shapes that contain a plus symbol indicate concurrent streams of action; the diamond shapes containing a cross represent decision points. The gray shades suggest correspondences between the sentences and the activities of the process model.

2.3 Event Data for the Footprints of Process Executions

Information systems supporting processes in organizations enable the persistent storage of process executions, in large log files denoted *event logs*. Event logs can be seen as a tabular representation of all the necessary events for a process execution (denoted as *case*) to be accomplished. Table 1 provides an excerpt of an event log for the bicycle manufacturing process used in this paper. Three important elements (columns in the table) that identify an event (row) in an event log are: i) the case ID (to identify the process instance or case executed), ii) the timestamp (to identify when the event was executed), and iii) the activity name (to identify the task performed). Extra information in form of additional columns may exist, to provide contextual information to the execution of events. For instance, in the table three cases appear: case 1 denotes an order that was rejected, case 2 denotes an order that was completely executed till the bicycle assembly, and case 3 is ongoing.

In contrast to the process representations described in the previous section, event logs describe the reality, i.e., the recorded events witnessing all process instances stored in the information system. We will refer to event logs in this paper in particular parts, to highlight the importance to relate observed and

Event	Case ID	Activity	Timestamp	Dept.	Add. Data
1	1	`Receive Order`	10-04-2015 9:08am	Sales	Reject
2	2	`Receive Order`	10-04-2015 10:03am	Sales	Accept
3	2	`Inform S&E dept.`	10-04-2015 10:05am	Sales	–
4	2	`Enter details ERP`	10-04-2015 10:09am	Sales	–
5	2	`Prepare for assemb.`	10-04-2015 10:10am	Eng.	–
6	2	`Reserve part`	10-04-2015 10:12am	Storeh.	Wheel
7	2	`Sel uncheck. part`	10-04-2015 11:12am	Storeh.	Breaks
8	2	`Backorder part`	10-04-2015 12:06am	Storeh.	Breaks
9	3	`Receive Order`	10-04-2015 13:18am	Sales	Accept
10	2	`Assemble bycicle`	11-04-2015 10:03am	Eng.	–

Table 1: Part of an event log for the bicycle manufacturing process.

modeled behavior.

2.4 NLP Capabilities and Frameworks

There are several existing NLP technologies that can be more or less straightforwardly applied to BPM. As illustrated in Fig. 2, a textual description of a bussiness process is a text where several actors are mentioned, and their actions and interactions described. Thus, out-of-the-box analyzers can be used to structure the content of the text. Tools such as Stanford Core (Manning et al., 2014), FreeLing (Padró and Stanilovsky, 2012), Apache OpenNLP[1], or NLTK (Bird et al., 2009) can be used to extract predicates and involved actors via SRL, perform WSD to identify domain objects that may be mentioned using different words, solve coreferences to decide which of the mentioned actors correspond to the same entity, decide which order the described actions must follow or whether there are choices or loops, etc.

However, existing analysis pipelines are still below optimal performance, specially in the more semantic tasks (WSD, coreference, SRL, temporal relation extraction, etc.) and on an unrestricted domain (process descriptions may report organization processes in a huge range of sectors –health, education, banking, manufacturing, etc).

So, several challenges for NLP researchers can be found in the application to BPM:

- Improvement of the performance of individual analyzers, specially at the semantic/pragmatic level (e.g. in Figure 2, the mentions *the sales department* in sentence (1), and *a member of the sales department* in sentence (2) are actually referring to the same actor in the process, but a coreference system would consider them two different entities).
- Domain adaptation methods to tune generic NLP processors to deal with process descriptions in a specific organization or sector. This may require the creation/acquisition of tailored ontologies that help specifying with the right terms important parts of the process and relations among these relevant domain concepts.
- Definition of new tasks, such as the detection of exclusivity, parallelism/concurrency, decision points, or iteration of tasks described in the text (e.g. the phrase *this procedure* or the marker *in the meantime* in sentences (7) and (8) in Figure 2 are ambiguous with respect their antecedent, and each interpretation leads to different formal models).
- Use of world knowledge to improve the results (often some steps or relations among tasks are omitted because the reader is assumed to be able to understand it using common sense or domain knowledge, such as that a payment happens always after an invoice is emitted, or that a bicycle can not be assembled until all required parts have been obtained).

[1] `https://opennlp.apache.org`

- Information extraction from event logs. Among others, interesting directions is the elicitation through NLP techniques of process steps for loosely specified processes, projecting event logs for events at a given granularity level, etc.

3 Expanding BPM Capabilities through NLP

In this section, we provide interesting research directions based on incorporating/extending NLP capabilities in the three levels described in Fig. 1.

3.1 Multi-Process Management

Multi process management is concerned with the identification of the major business processes of a company and the prioritization of these processes. The overall landscape of business processes is often represented as a process architecture, which is stored in a business process repository. Large multinational corporations often maintain several thousand process models in such repositories.

Prior research on multi process management has focused on repository management, building on techniques for determining process similarity (Dijkman et al., 2011), automatically matching business process models (Weidlich et al., 2010), and other querying techniques (Leopold et al., 2017). Several concepts have been proposed on top including automatic refactoring (Weber et al., 2011) including harmonization of terminology (Pittke et al., 2015), automatic service derivation (Leopold et al., 2015), semantic search (Thomas and Fellmann, 2009) and merging business process models (Rosa et al., 2013).

Several challenges remain in this area including the following (Mendling et al., 2015). First, the availability of a business process repository bears the potential to discover an overarching formal ontology that captures the full spectrum of operations of a company. Second, content captured in the repository might be automatically categorized, for example with respect to documentation standards such as ISO:9001. Such tasks require the application and adaptation of existing NLP techniques in this specific context.

3.2 Process Management

Several important challenges need to be tackled so that tasks in the process model management stage of Fig. 1 such as discovery, analysis, redesigning and controlling of processes are excelled at organizations. In this section we focus on providing examples of key enablers for tackling these challenges.

The discovery of processes concerns the identification of the main processes in an organization, and the corresponding elicitation into a (graphical) notation such as BPMN, which facilitates the communication between the stakeholders involved in a process. In practice, this phase is often implemented using workshop meetings, requiring quite significant efforts to materialize a process model from the conversations arising in these meetings. An alternative is the use of process mining techniques, which enable, for instance, the automatic discovery of process models from event logs (van der Aalst, 2016).

NLP techniques can be applied at very different granularities to boost the discovery of precise process model descriptions of a process. Among the available techniques, we highlight the most disruptive ones:

- *Transform textual descriptions into process models*: The creation of process models consumes up to 60% of the time spent on process management projects. This is a paradox, because there are often extensive textual process documentations available in organizations. Therefore, automatically transforming textual process descriptions to process models represents a particularly attractive use case. Several techniques have been developed for this purpose (Friedrich et al., 2011). These use tailored NLP techniques in order to identify actions, e.g. "*The sales department receives an order*", and their inter-relations, e.g. "*If the order is accepted, the order details are entered into the system*. These extracted components represent the foundation for the generation of a process model from a text. However, a number of challenges still remain for this endeavor. For instance, techniques must be able to identify sentences that provide contextual information, rather than describe process steps. Furthermore, the inherent ambiguity of natural language can lead to different interpretations regarding the process that is described (Van der Aa et al., 2016).

- *Translate process models*: Sometimes the same process (e.g., the admission process to enroll in a university) is defined multiple times (e.g., several universities in several countries, or a worldwide company that applies a given process across several countries). In these contexts, it is crucial to have multilingual support for translating a given process model description (in any form) into a target language. For the case of process model graphical descriptions (e.g., BPMN), translation of individual model elements may be sufficient to obtain a translated description with a certain quality, since the structure of the process is retained. In contrast, translating textual descriptions of a process may become more challenging, due to the particularities each language has in describing certain constructs.

- *Text Annotation and Inference*: In the scope of the two previous use cases (from text to process models, process model translation), the difficulty of generating the output would be significantly alleviated if the text was correctly annotated, reducing the noise introduced by automatic NLP tools. One can envision that, in the narrow scope of the description of a process one could define a limited set of template annotations that address particular perspectives (control-flow, roles, data, among others) which can help a user to (partially) annotate a textual description of a process. Nowadays, there exist advanced tools for text annotation such as Brat (Stenetorp et al., 2012). From annotated textual descriptions of a process model (e.g., control-flow annotations establishing relations between two sentences), inferences can elicit new relations that can be used to have a more precise and refined descriptions (e.g., transitive closure of the relations). The same annotations could be used as training data to improve or adapt automatic languages analyzers tuned to this particular tasks.

The analysis phase is oriented towards finding weakness of current process model candidates. This phase can also be significantly improved by a tailored used of NLP techniques in particular situations. We describe here some examples of challenges to face in order to materialize such improvements:

- *Verify semantic correctness and completeness*: The semantic quality of process models is crucial to the proper understanding of business processes. A number of techniques aim to verify this quality in an automated manner. These techniques achieve this, for example, by detecting and/or correcting inconsistent use of terminology (Koschmider and Blanchard, 2007) or violations of labeling conventions (Becker et al., 2009; Leopold et al., 2013a; Van der Vos et al., 1997). Others aim to improve modeling quality by detecting common modeling errors (Gruhn and Laue, 2011) or ambiguously labeled activities (Pittke et al., 2015; Pittke et al., 2014).

- *Calculate consistency between process model and text*: As mentioned before, keeping different process descriptions helps improving the knowledge about processes across an organization. However, as processes evolve continuously, it is necessary to detect inconsistencies between process descriptions in order to ensure the expectations for process outcomes are the same for every actor (van der Aa et al., 2017). The challenge here is to find correspondences between sentences in the text and elements in the process model, and warn in case important parts are not mapped. This requires a respective NLP analysis in both process descriptions and computation of similarities while considering the different discourse level, ambiguities, anaphora/coreference in the text, among others. Fig. 2 shows an example of a possible mapping between the textual and the model description of the process.

In the redesign phase, issues detected in the previous phase are amended by a refactoring of the process model, so that the to-be model is produced. NLP-based techniques can also be oriented towards this goal:

- *Fine-tuning of semantic abstraction levels*: Larger sets of process models, so-called, process model collections, are typically hierarchically organized. This means, that process models on higher level provide a rather course-grained view of a process, while process model on a lower level provide a more detailed view. One of the major challenges in this context is to make sure that process models on the same level in the hierarchy also have a comparable level of abstraction. First works that addressed this problem have used rather simple linguistic measures, such as the specificity of individual words, to determine whether two process models provide a comparable level of abstraction (Leopold et al., 2013b). However, a solution that assesses the abstraction level based on the conveyed semantics is still missing.

- *Process description semantic auto-completion*: Process descriptions may be semantically incomplete. Ideally, a process description should be semantically complete before it is used. There are different situations where a process description could be auto-completed. For instance, imagine that the second sentence in the process description of Fig. 2 is: *A member of the sales department must check the order for a customized bike and decide its acceptance*. Semantically, this sentence suggests that there is also the possibility to reject the order of a bike. However, it is not explicit in the textual description what to do if the result of the check is negative. Warning about this incompleteness of the textual description may help improving it, by reducing its ambiguity. A similar situation may occur in the graphical description in the right, i.e., if the arc labeled *No* (and the target gateway) are missing in the process model description in Fig. 2. Some techniques are available for this task (Hornung et al., 2007), but there are still several challenges, mainly on the automation of the problem.

In the controlling phase, apart from the traditional monitoring techniques that focus on checking the performance and conformance requirements, Natural Language Generation techniques can be applied to keep different descriptions of the process available:

- *Transform process model to textual descriptions*: Not all stakeholders are able to understand a process model descriptions like the one shown at the right of Fig. 2. However, virtually everyone can understand the textual description shown at the left. Recent studies on process understandability advocate for the use of several descriptions in order to boost the understanding (Ottensooser et al., 2012). Hence, generating textual descriptions of processes that complement formal ones helps ensuring different stakeholders will have the same expectations for a certain process (Leopold et al., 2014), and NLG is a good fit for this task. There are two main challenges associated with this task. First, we need to properly infer which words from the short process model labels refer to verbs and which words refer to nouns, and due to the shortness of the labels and the lack of a proper sentences structure (e.g., consider the label *Order reservation*) this is a non-trivial task which may require domain knowledge. Second, parallel behavior as well as choices from the process model have to be communicated using a sequence of sentences, without compromising the ability of the reader to comprehend the process semantics.

3.3 Instance Management

Managing the execution of a single process instance (e.g., a particular order for manufacturing a bicycle in the process of Fig. 2, e.g., case 2 of Table 1) is the primary objective of the bottom level of BPM.

There exists a significant amount of research from the last years about *conversational systems* (Mott et al., 2004), ultimately implemented as conversational bots or chatbots. Now let us assume that conversational systems are trained to support the execution of business processes. The ultimate goal would be to allow stakeholders of the process to navigate through it by means of querying a dialog system. For instance, for the process described in Fig. 2, a new worker may have doubts on what to do after a new order is received. A tailored conversation system may come into rescue, by first describing her what are the formal requirements for an order to be accepted. Then, when these requirements have been evaluated by the worker, it will communicate back to the dialogue system the outcome of the evaluation (accepted or not), and the dialogue system will provide the next step correspondingly (in case of the order being accepted, to inform the engineering department; in case of the order being rejected, finish the process). The challenge here is how to build useful conversational systems when a description of the process is available.

In general, BPM tailored conversational systems may incorporate important NLP features like semantic understanding, context resolution, NLG, among others.

4 Applications

4.1 Education

The possibility of transforming a formal process description (e.g., a BPMN like the one on the right in Fig. 2) into a text (Leopold et al., 2014), or vice-versa, as well as the ability to align a textual and a formal description of a process (Sànchez-Ferreres et al., 2017) opens the door to a range of educational applications for modeling students (e.g. Computer Science or Business students that need to learn to formalize a process): For instance, the model created by a student can be automatically compared to the text given as statement, not only for automatic grading, but also to provide feedback regarding missing/redundant tasks or inconsistent paths in the control flow. Also, an initial model automatically generated from text can be given to the student to complete or correct.

4.2 Troubleshooting

An important application domain for ensuring the correct execution of a process in a organization is through tailored dialogue systems. An example of this is chatbot-aided troubleshooting (Acomb et al., 2007), where artificial agents are used to complement human operators in contact centers. So far, knowledge representation in such chatbots comprises several components like natural language understanding and/or generation, together with a planner that encompasses the possible reactions to provide (Thorne, 2017). The incorporation of process descriptions may help into attaining a more precise dialogue system, so that the planning of the dialogue is done under the constraints of the process, and that the agent can better assist the human in following the appropriate steps.

4.3 Regulations

Non-compliance represents a risk for many organizations. According to a recent study by Thomson Reuters, non-compliance may even represent a possible cause of bankruptcy, also for the so-called "behemoths" in the financial sector (English and Hammond, 2014). Recognizing the risk that is associated with non-compliance, organizations in a wide range of domains are stepping up their spending in order to ensure their compliance with laws, regulations, and procedures. In this context, automated compliance checking techniques that consider the process model and the event log play a crucial role thanks to their ability to automatically identify compliance violations (Accorsi and Stocker, 2012; Van der Aalst et al., 2010). For this reason, numerous approaches have been developed to perform this task (Van der Aalst et al., 2012; Weidlich et al., 2011; Awad et al., 2008). While the majority of such techniques require the allowed process behavior to be specified in formal models, such as process models or business rules, recent advances of NLP in the context of BPM overcome this restriction. In particular, a technique (van der Aa et al., 2018) has been developed that can perform compliance checks on the basis of textual process documentation. This technique employs probabilistic methods in order to deal with the inherent ambiguity of natural language, which can cause uncertainty about the truly allowed process behavior specified in a text.

4.4 Healthcare

Electronic health records (ERHs) contain detailed information of patients. They can and have been used for monitoring adherence to clinical guidelines. There has been several studies on how using these ERHs can lead to improving the management of patients in the healthcare domain. NLP tooling related to ERH processing can also provide a significant step towards improving the efficiency in the treatment of certain diseases (Garvin et al., 2018).

Both clinical guidelines and ERHs can be the source for elicitating formal process descriptions, using a combination of the techniques listed in the Sect. 3.2. Moreover, recent techniques for comparing process models with event logs stored in Healthcare Information Systems (HIS) have proven to be successful in improving the analysis of patients (Mans et al., 2015). We therefore envision the connection of both disciplines: on the one hand, NLP-based techniques to formalize into models ERHs or clinical guidelines, that can be then used to accurately analyze patients with the event data stored in a HIS.

5 Conclusions

In this paper, we have highlighted research directions and prospective applications for NLP in the area of business process management. A more intensive exchange between the two fields has the potential to significantly enhance the tool set of business process management and to fruitfully provide both practical challenges and industrial application scenarios to the NLP community.

Acknowledgements

This work supported by the Spanish Ministerio de Economía y Competitividad, under projects TIN2017-86727-C2-1-R and TIN2016-77820-C3-3-R.

References

Rafael Accorsi and Thomas Stocker. 2012. On the exploitation of process mining for security audits: the conformance checking case. In *Proceedings of the 27th Annual ACM Symposium on Applied Computing*, pages 1709–1716. ACM.

Kate Acomb, Jonathan Bloom, Krishna Dayanidhi, Phillip Hunter, Peter Krogh, Esther Levin, and Roberto Pieraccini. 2007. Technical support dialog systems: Issues, problems, and solutions. In *Proceedings of the Workshop on Bridging the Gap: Academic and Industrial Research in Dialog Technologies*, NAACL-HLT-Dialog '07, pages 25–31, Stroudsburg, PA, USA. Association for Computational Linguistics.

Ahmed Awad, Gero Decker, and Mathias Weske. 2008. Efficient compliance checking using bpmn-q and temporal logic. In *International Conference on Business Process Management*, pages 326–341. Springer.

Thomas Baier and Jan Mendling. 2013. Bridging abstraction layers in process mining by automated matching of events and activities. In *International Conference on Business Process Management*, pages 17–32. Springer.

Jörg Becker, Patrick Delfmann, Sebastian Herwig, L. Lis, and Armin Stein. 2009. Formalizing linguistic conventions for conceptual models. In *Conceptual Modeling - ER 2009*, LNCS, pages 70–83. Springer Berlin Heidelberg.

Steven Bird, Edward Loper, and Klein Ewan. 2009. *Natural Language Processing with Python*. O'Reilly Media Inc.

Suranjan Chakraborty, Saonee Sarker, and Suprateek Sarker. 2010. An exploration into the process of requirements elicitation: A grounded approach. *J. AIS*, 11(4).

Islay Davies, Peter Green, Michael Rosemann, Marta Indulska, and Stan Gallo. 2006. How do practitioners use conceptual modeling in practice? *Data & Knowledge Engineering*, 58(3):358–380.

Remco M. Dijkman, Marlon Dumas, Boudewijn F. van Dongen, Reina Käärik, and Jan Mendling. 2011. Similarity of business process models: Metrics and evaluation. *Information Systems*, 36(2):498–516.

Marlon Dumas, Marcello La Rosa, Jan Mendling, and Hajo A. Reijers. 2013. *Fundamentals of Business Process Management*. Springer Berlin Heidelberg, Berlin, Heidelberg.

Stacey English and Susannah Hammond. 2014. *The rising costs of non-compliance: from the end of a career to the end of a firm*. Thomson Reuters.

Kathrin Figl and Jan Recker. 2016. Exploring cognitive style and task-specific preferences for process representations. *Requirements Engineering*, 21(1):63–85.

Fabian Friedrich, Jan Mendling, and Frank Puhlmann. 2011. Process model generation from natural language text. In Haralambos Mouratidis and Colette Rolland, editors, *Advanced Information Systems Engineering*, pages 482–496, Berlin, Heidelberg. Springer Berlin Heidelberg.

Hornung Jennifer Garvin, Youngjun Kim, Temple Glenn Gobbel, E. Michael Matheny, Andrew Redd, E. Bruce Bray, Paul Heidenreich, Dan Bolton, Julia Heavirland, Natalie Kelly, Ruth Reeves, Megha Kalsy, Kane Mary Goldstein, and M. Stephane Meystre. 2018. Automating quality measures for heart failure using natural language processing: A descriptive study in the department of veterans affairs. *JMIR Med Inform*, 6(1):e5, Jan.

V. Gruhn and R. Laue. 2011. Detecting Common Errors in Event-Driven Process Chains by Label Analysis. *Enterprise Modelling and Information Systems Architectures*, 6(1):3–15.

Thomas Hornung, Agnes Koschmider, and Andreas Oberweis. 2007. Rule-based autocompletion of business process models. In *CAiSE'07 Forum, Proceedings of the CAiSE'07 Forum at the 19th International Conference on Advanced Information Systems Engineering, Trondheim, Norway, 11-15 June 2007*.

Agnes Koschmider and Emmanuel Blanchard. 2007. User assistance for business process model decomposition. In *Proceedings of the 1st IEEE International Conference on Research Challenges in Information Science*, pages 445–454.

Stefan Krumnow and Gero Decker. 2010. A concept for spreadsheet-based process modeling. In *International Workshop on Business Process Modeling Notation*, pages 63–77. Springer.

Henrik Leopold, Rami-Habib Eid-Sabbagh, Jan Mendling, Leonardo Guerreiro Azevedo, and Fernanda Araujo Baião. 2013a. Detection of naming convention violations in process models for different languages. *Decision Support Systems*, 56:310–325.

Henrik Leopold, Fabian Pittke, and Jan Mendling. 2013b. Towards measuring process model granularity via natural language analysis. In *Business Process Management Workshops - BPM 2013 International Workshops, Beijing, China, August 26, 2013, Revised Papers*, pages 417–429.

Henrik Leopold, Jan Mendling, and Artem Polyvyanyy. 2014. Supporting process model validation through natural language generation. *IEEE Trans. Software Eng.*, 40(8):818–840.

Henrik Leopold, Fabian Pittke, and Jan Mendling. 2015. Automatic service derivation from business process model repositories via semantic technology. *Journal of Systems and Software*, 108:134–147.

Henrik Leopold, Han van der Aa, Fabian Pittke, Manuel Raffel, Jan Mendling, and Hajo A Reijers. 2017. Searching textual and model-based process descriptions based on a unified data format. *Software & Systems Modeling*, pages 1–16.

Henrik Leopold. 2013. *Natural language in business process models*. Ph.D. thesis, Springer.

Christopher D. Manning, Mihai Surdeanu, John Bauer, Jenny Finkel, Steven J. Bethard, and David McClosky. 2014. The Stanford CoreNLP natural language processing toolkit. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, pages 55–60.

Ronny Mans, Wil M. P. van der Aalst, and Rob J. B. Vanwersch. 2015. *Process Mining in Healthcare - Evaluating and Exploiting Operational Healthcare Processes*. Springer Briefs in Business Process Management. Springer.

Jan Mendling, Henrik Leopold, and Fabian Pittke. 2015. 25 challenges of semantic process modeling. *International Journal of Information Systems and Software Engineering for Big Companies*, 1(1):78–94.

Jan Mendling, Bart Baesens, Abraham Bernstein, and Michael Fellmann. 2017. Challenges of smart business process management: An introduction to the special issue. *Decision Support Systems*, 100:1–5.

Jan Mendling. 2008. *Metrics for process models: empirical foundations of verification, error prediction, and guidelines for correctness*, volume 6. Springer Science & Business Media.

Bradford W. Mott, James C. Lester, and Karl Branting. 2004. Conversational agents. In *The Practical Handbook of Internet Computing.*

Avner Ottensooser, Alan Fekete, Hajo A. Reijers, Jan Mendling, and Con Menictas. 2012. Making sense of business process descriptions: An experimental comparison of graphical and textual notations. *Journal of Systems and Software*, 85(3):596–606.

Lluís Padró and Evgeny Stanilovsky. 2012. Freeling 3.0: Towards wider multilinguality. In *Proceedings of the Language Resources and Evaluation Conference (LREC 2012)*, Istanbul, Turkey, May. ELRA.

Keith Thomas Phalp, Jonathan Vincent, and Karl Cox. 2007. Improving the quality of use case descriptions: empirical assessment of writing guidelines. *Software Quality Journal*, 15(4):383–399.

Fabian Pittke, Henrik Leopold, and Jan Mendling. 2014. When language meets language: Anti patterns resulting from mixing natural and modeling language. In *Business Process Management Workshops*, pages 118–129. Springer.

Fabian Pittke, Henrik Leopold, and Jan Mendling. 2015. Automatic detection and resolution of lexical ambiguity in process models. *IEEE Transactions on Software Engineering*, 41(6):526–544.

Jan Recker, Norizan Safrudin, and Michael Rosemann. 2012. How novices design business processes. *Information Systems*, 37(6):557–573.

Hajo A Reijers, Henrik Leopold, and Jan Recker. 2017. Towards a science of checklists. In *Proceedings of the 50th Hawaii International Conference on System Sciences*.

Marcello La Rosa, Marlon Dumas, Reina Uba, and Remco M. Dijkman. 2013. Business process model merging: An approach to business process consolidation. *ACM Transactions on Software Engineering and Methodology*, 22(2):11:1–11:42.

Josep Sànchez-Ferreres, Josep Carmona, and Lluís Padró. 2017. Aligning textual and graphical descriptions of processes through ILP techniques. In *Advanced Information Systems Engineering - 29th International Conference, CAiSE 2017, Essen, Germany, June 12-16, 2017, Proceedings*, pages 413–427.

Pontus Stenetorp, Sampo Pyysalo, Goran Topić, Tomoko Ohta, Sophia Ananiadou, and Jun'ichi Tsujii. 2012. brat: a web-based tool for NLP-assisted text annotation. In *Proceedings of the Demonstrations Session at EACL 2012*, Avignon, France, April. Association for Computational Linguistics.

Oliver Thomas and Michael Fellmann. 2009. Semantic process modeling - design and implementation of an ontology-based representation of business processes. *Business & Information Systems Engineering*, 1(6):438–451.

Camilo Thorne. 2017. Chatbots for troubleshooting: A survey. *Language and Linguistics Compass*, 11(10).

Han Van der Aa, Henrik Leopold, and Hajo A Reijers. 2016. Dealing with behavioral ambiguity in textual process descriptions. In *International Conference on Business Process Management*, pages 271–288. Springer.

Han van der Aa, Henrik Leopold, and Hajo A. Reijers. 2017. Comparing textual descriptions to process models - the automatic detection of inconsistencies. *Inf. Syst.*, 64:447–460.

Han van der Aa, Henrik Leopold, and Hajo A Reijers. 2018. Checking process compliance against natural language specifications using behavioral spaces. *Information Systems*.

Han van der Aa. 2018. *Comparing and Aligning Process Representations*. Ph.D. thesis, Springer.

Wil M P Van der Aalst, Kees M van Hee, Jan Martijn van Werf, and Marc Verdonk. 2010. Auditing 2.0: using process mining to support tomorrow's auditor. *Computer*, 43(3):90–93.

Wil M P Van der Aalst, Arya Adriansyah, and Boudewijn van Dongen. 2012. Replaying history on process models for conformance checking and performance analysis. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 2(2):182–192.

Wil M. P. van der Aalst. 2016. *Process Mining: Data Science in Action*. Springer.

Bram Van der Vos, Jon Atle Gulla, and Reind van de Riet. 1997. Verification of conceptual models based on linguistic knowledge. *Data & Knowledge Engineering*, 21(2):147 – 163.

Barbara Weber, Manfred Reichert, Jan Mendling, and Hajo A. Reijers. 2011. Refactoring large process model repositories. *Computers in Industry*, 62(5):467–486.

Matthias Weidlich, Remco M. Dijkman, and Jan Mendling. 2010. The icop framework: Identification of correspondences between process models. In Barbara Pernici, editor, *Advanced Information Systems Engineering, 22nd International Conference, CAiSE 2010, Hammamet, Tunisia, June 7-9, 2010. Proceedings*, volume 6051 of *Lecture Notes in Computer Science*, pages 483–498. Springer.

Matthias Weidlich, Jan Mendling, and Mathias Weske. 2011. Efficient consistency measurement based on behavioral profiles of process models. *IEEE Transactions on Software Engineering*, 37(3):410–429.

Christian Wolter and Christoph Meinel. 2010. An approach to capture authorisation requirements in business processes. *Requirements engineering*, 15(4):359–373.