Towards a Framework for Context Awareness Based on Textual Process Data

Aleksandra Revina[1[0000-0002-8405-0018]], Nina Rizun[2[0000-0002-4343-9713]], and Ünal Aksu[3[0000-0002-0807-902X]]

[1] Technical University of Berlin, Straße des 17. Juni 135, 10623 Berlin, Germany
Technische Hochschule Brandenburg, Magdeburger Str. 50, 14770 Brandenburg an der Havel, Germany
revina@th-brandenburg.de

[2] Gdansk University of Technology, G. Narutowicza 11/12, 80-233 Gdansk, Poland
nina.rizun@pg.edu.pl

[3] Utrecht University, Princetonplein 5, 3584 CC Utrecht, The Netherlands
u.aksu@uu.nl

Abstract. Context awareness is critical for the successful execution of processes. In the abundance of business process management (BPM) research, frameworks exclusively devoted to extracting context from textual process data are scarce. With the deluge of textual data and its increasing value for organizations, it becomes essential to employ relevant text analytics techniques to increase the awareness of process workers, which is important for process execution. The present paper addresses this demand by developing a framework for context awareness based on process executions-related textual data using a well-established layered BPM context model. This framework combines and maps various text analytics techniques to the layers of the context model, aiming to increase the context awareness of process workers and facilitate informed decision-making. The framework is applied in an IT ticket processing case study. The findings show that contextual information obtained using our framework enriches the awareness of process workers regarding the process instance urgency, complexity, and upcoming tasks and assists in making decisions in terms of these aspects.

Keywords: Context Awareness, Textual Data, Text Analytics, Business Process Management.

1 Introduction

Business processes are one of the key assets of organizations through which they provide products and services to their customers. To lead to an outcome that is valuable for customers, activities, events, and decisions are instrumented within business processes by process workers [1]. Moreover, how such process elements will be instrumented is influenced by the context of the processes, i.e., the environmental properties [2]. Hence, business processes are bound to their context. In fact, many Business Process Management (BPM) projects are known to fail as they base on standard guidelines and schemes neglecting the context [3]. In empirical research, context is frequently viewed as an outside threat that needs to be controlled or removed [4]. There are multiple attempts in information systems research to develop solutions to control the context, such as promoting generalizability [7], increasing causality, and improving robustness [8]. However, one of the most profound and prominent theoretical findings in information systems research is that context matters [4].

In line with the importance of context in business processes, to be able to execute processes appropriately, workers need to comprehend their context. More specifically, decision-making of humans while performing processes requires awareness and processing of all possible information about the process context. However, humans are constrained in their ability to effectively comprehend and process large amounts of data. As the majority of the information about process context is textual data and unstructured, how to facilitate contextualization of such data is considered rather challenging [5, 6].

BPM literature introduces a number of methods to deal with context. For example, in the recent work [9], building on a special type of directed graph, the authors present an approach incorporating the contextual information in the analysis and visualization of the process execution data, i.e., event log. Other research focuses on the development of holistic frameworks [10], ontologies [11], taxonomies, and specific event log- [12] and business process model-based [6] methods. Despite the abundance of these approaches, the potential of textual data for context awareness and decision-making support of process workers in process execution has not been comprehensively explored [13]. Human workers are known for their limited ability to process large amounts of data [14]. Hence, their awareness of certain process-relevant information, such as customer-related or expected process complexity, naturally supports efficient process execution. Additionally, most studies do not address the practical implementation of context awareness approaches and focus on conceptual work [9]. However, providing or using publicly available tools and techniques as well as specifying the implementation details are important aspects of the work reproducibility and value for both research and practice.

In this paper, we propose a framework for context awareness, explicitly focusing on textual data *related to process executions*. In particular, we focus on linguistic features of textual data (such as syntactic structure, meaning, style, text parts, word choice, and order) that can provide process context and assist process workers in executing processes in a proper way. We take an established BPM context model as a basis [2] and adopt it for leveraging textual data on context awareness. More specifically, we enhance the model, particularly its external, internal, and immediate context layers, with common text analytics techniques. This way, we aim to "practically implement" or operationalize the model. This forms our theoretical and methodological contribution. Practical contributions are demonstrated in an IT ticket processing case study. The awareness regarding the process instance *urgency*, *complexity*, as well as *expected tasks* can help process workers to *prioritize* their work, *facilitate task assignment* and *resource allocation* in the short run. In the long run, it may be beneficial for successful and fast process executions increasing the satisfaction of multiple stakeholders such as managers, process workers, and customers. Moreover, as we use a well-established comprehensive context model and common text analytics techniques, we believe that the framework can be applied in several real-life settings and domains with justified effort.

The remainder of the paper is as follows. Section 2 presents a review of related literature on context awareness in BPM. In Section 3, the research methodology we used in the study is explained. Section 4 describes how we conceptualize the framework for context awareness based on textual process data. Then, we provide the case study-based insights, evaluate the framework, and discuss the findings and limitations in Section 5. Finally, in Section 6, we conclude our paper with a summary and present ideas for future work.

2 Related work

The context of processes and information about such context received much attention in BPM [15, 16]. At the same time, detecting and incorporating contextual factors in processes is considered rather challenging [17]. These factors may be encountered at different levels, and their number and range can be diverse [2]. Moreover, they can be close to the process itself, for example, a minimal time required to execute a process or far beyond, like country import regulations. Therefore, developing context-aware solutions requires a profound knowledge of internal and external factors impacting the processes [18, 19]. In this regard, various context awareness approaches in BPM are suggested for detecting contextual factors. We analyzed state-of-the-art studies about context awareness in BPM aligned with our goal. With this, we aim to identify a comprehensive BPM context model serving as a basis for our framework. We elaborate on the studies[1] that are relevant to our framework development.

The research on contextual BPM is still in its early stages [15] and demands more context awareness inclusion in BPM method design and exploration [20]. However, we could identify both rather extensive, like [15], and very specific, such as [12], approaches to context awareness. In recent context awareness studies, various topics are covered, for example, process modeling [21], decision-making [22], process mining [23], IoT [6], and cloud computing [24].

At the same time, automated analyses of the language and text characteristics are valuable in various settings [25]. Text generated as communication via emails, chats, social media, and documents can naturally imply rich information on different contextual factors. Hereby, big data analytics in general [26] and text analytics in particular [27] have become popular techniques to extract contextual information from large amounts of textual data. In BPM, we observe that context awareness approaches inherently consider textual data. However, this consideration is prevailingly limited to textual data employed in process models [28], event logs [12], and ontologies [6]. Moreover, no studies suggest the linking of various semantic aspects inherent in textual data to context types. For example, one of the latest studies on BPM context [20] proposes linking various BPM contexts to appropriate BPM methods leaving the contextual aspects and textual data focus out of scope.

[1] The overview of the studies can be found on the Github page.

Hence, our study aims to address these shortcomings while developing the framework. To base our framework development, we searched for the groundwork in the related literature, i.e., a BPM context model that could incorporate varying multilayered knowledge inherent in textual data [29]. In this regard, the context model by [2] is one of the earliest efforts in BPM context awareness, providing a foundation for future, more focused research [30, 31]. It represents a comprehensive universal taxonomy as an initial reference for process contextualization in organizations. Unlike other approaches, the model is characterized by a layered structure, each layer (immediate, internal, external, environmental) having a broader coverage. For example, whereas the immediate layer covers those aspects directly influencing process execution, like data and resources, the environmental layer addresses the aspects outside the business network of an organization, such as national policies [2]. These layers are intertwined so that the innermost layer gathers data from all the outer layers. Such a structure provides the necessary flexibility and semantics to integrate various knowledge types extracted from textual data.

Nonetheless, the selected BPM context model [2] reveals certain limitations. For example, it provides only a high-level construct that lacks a strategy on how to practically implement the model in the sense of necessary data, analysis methods, and techniques, i.e., make it operational for an organization. Further, the demonstration of how a process worker can benefit from the model is also missing. The next section describes the methodology used to address these limitations while developing the framework.

3 Methodology

Taking Design Science Research Methodology (DSRM) as a basis [32], this study aims to build a framework for context awareness based on textual data. According to the first phase of DSRM, we define the problem we focus on as follows: (1) process workers must be aware of the process context to be able to carry out the business process properly, (2) textual data inherently contains such contextual information, (3) however, workers are constrained in their ability to comprehend textual data, especially in large amounts.

In the second DSRM phase, we set the objectives of the envisioned framework. Relying on solid prior research in this context [2] and dealing with its limitations, the envisioned framework aims to analyze available textual data relevant for process execution using various text analytics methods. With this, process-related insights can be extracted at the immediate, internal, and external contextual layers [2] and provided to process workers in a comprehensive manner.

In the third design phase, we employ conceptual modeling. It serves the purpose to abstractly represent specific aspects of a domain with the help of graphical representations [33]. In this paper, we use conceptual modeling to develop our framework based on the context of a selected organization. As the context itself may imply an endless amount of information, the decision was made to select an organization interested in such a solution and involve its experts in the framework development. Hence, the framework has been conceptualized in an industrial IT Service Management (ITSM) Change Management (CHM) [34] process from a large international telecom provider.

The CHM department of the company started a project to get insights into its CHM operations using the textual data of IT tickets issued to perform changes in IT products and services offered by the company. To achieve data-driven service delivery and provide real-time decision-making support for the workers, the CHM department declared an interest in creating context awareness solutions using the IT ticket textual descriptions. As this setting is highly relevant to our framework, we conceptualize it in the CHM department. Hereby, we worked with 12 domain experts (CHM managers and coordinators) who handle IT tickets daily and are well-versed in this process. In a one-day workshop, the experts provided us with the necessary knowledge on the problems related to IT ticket processing. Moreover, they explained that information contained in textual data which is important for and related to the process context.

In the fourth DSRM phase, we use the case study to evaluate our framework and demonstrate its applicability. Specifically, we use the real-world IT ticket textual data and related process goals to showcase the value of the framework for process worker decision-making support.

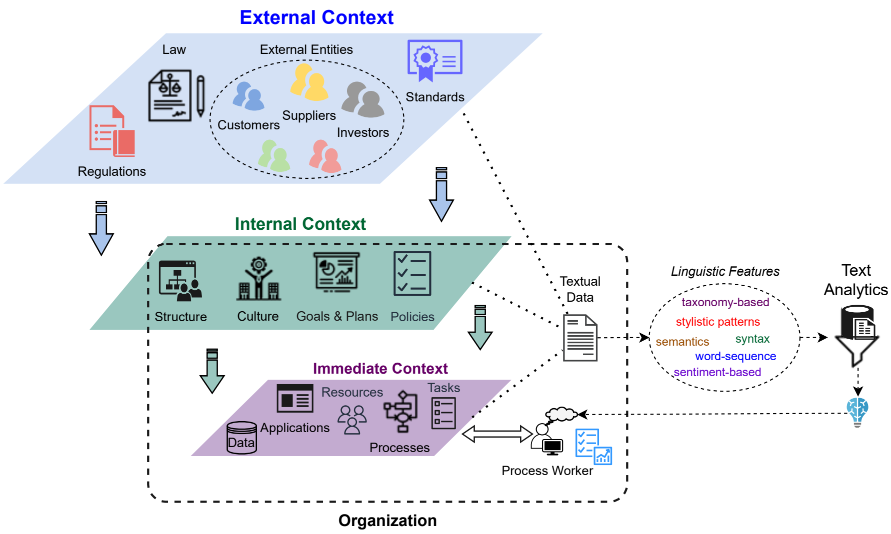

Fig. 1. Design of a framework for context awareness

4 Conceptual framework development

As presented in the related work section, Rosemann et al. introduce four layers of the model serving as a basis for our framework: immediate, internal, external, and environmental [2]. We take the first three context layers and enrich them with text analytics and process worker support in decision-making. We exclude the environmental layer, i.e., society, nature, and technological developments, as being far from process operations. According to [2], this layer addresses weather, strikes, policies, and work norms. Although this information can be considered in our framework, the impactful cases involving such information are rare compared to others. Hence, we propose considering the environmental layer as a part of our future work.

The framework design reveals the exemplary elements in each layer based on [2] and enriching elements, i.e., text analytics, linguistic features, and process worker (see Fig. 1). Below, we elaborate on the conceptual framework layer by layer, starting with the external context layer. While doing that, we base the elaborations on our previous work [13, 35, 36]. The main reason is that, in this work, we showed the textual data potential in decision-making and awareness of process workers.

As shown in Fig. 1, the external layer includes concepts such as law, regulations, standards, and external entities, i.e., customers, suppliers, and investors. According to the experts from the CHM department of the organization where the framework is conceptualized, the most significant external layer information is related to the customers, i.e., the authors of the tickets. Further, at the internal layer addressing the organization or department-specific information, such as goals, plans, and policies, ticket processing implicit guidelines would play an important role. Finally, the immediate layer deals with the process execution and involves applications, resources, tasks, and process instances. While the experts consider all this information important, we focus primarily on the tasks necessary to process an IT ticket due to the organizational privacy policy.

4.1 Framework layers

External context. As presented in Fig. 1, the external context represents the outer layer of our framework comprising regulations, laws, standards, and external entities. Referring to a definition of a process [1], the ultimate goal of any process is bringing value to its customers. Knowledge about customers, essential for building sustainable businesses [37], naturally belongs to the external context layer and has also been identified as the most relevant for the organization by the experts. Hereby, sentiment analysis allows extracting subjective, i.e., customer-related, information such as attitude, opinion, and emotions from the text [38]. This can help process workers better understand the customer's pain points and react accordingly while executing the processes. Hence, we suggest sentiment analysis as an operational and easily applicable approach to obtain external information regarding customers.

In brief, at the external context layer, with the help of sentiment analysis, the process worker receives important latent information regarding the author of the IT ticket, i.e., the customer anxiety level. This knowledge extends the awareness with the customer context element. Hence, the worker can assess the urgency and significance of the request, which facilitates the planning. For example, if the number of urgent tickets gets high, a delegation of work or involvement of other colleagues can occur.

Internal context. The internal context denotes the middle layer of the framework. It includes information about the internal organizational environment affecting process execution, such as a company or department structure, goals, and work plan (see Fig. 1). In BPM, knowledge, i.e., awareness of the organization- or department-specific rules, can be regarded as the next level of decision-making support closer to direct process operations. The extraction of this information is known to be achieved with the help of text analytics related to organizational knowledge management like lexicons, thesauri, taxonomies, and ontologies [39]. Such knowledge representation can not only help experienced workers to manage the processes efficiently and faster but also enable novice workers to execute any process instance. Hence, we propose using taxonomies to include organizational knowledge in the decision-making support of process workers.

In short, at the internal context layer, the process worker gets an understanding of the process cognition in the organizational unit context based on the mutually agreed meaning of keywords formalized in the form of taxonomy. Then, using this knowledge enriched by personal, contextual experience, the worker can estimate the complexity of the effort and time needed to execute the process and its tasks.

Immediate context. The immediate context constitutes the innermost layer of our framework. According to [2] and Fig. 1, the immediate context includes those aspects directly assisting in process execution, for example, information on required data, organizational resources, activities, IT, and applications. In its meaning, the immediate context is somewhat similar to the internal context. The distinction is its proximity to the direct process execution. Hence, we draw on the same considerations of the internal context in our choice of the text analytics technique. Hereby, we highlight the difference related to the thematic specificity of the text analytics commonly applied for knowledge management, i.e., lexicons, thesauri, and taxonomies. In the immediate context, this thematic specificity should reflect tasks of processes that should be realized in the taxonomic approach.

Thus, at the immediate context layer, the process worker gets an understanding of those tasks and their content based on the keywords grouped in the form of taxonomy. Using this knowledge enriched by personal, contextual experience, the worker can plan specific tasks and resources needed to execute them.

4.2 Decision-making support enabled by the framework

After getting a meaningful representation of the three context layers, we elaborate on the solution for decision-making support of a process worker: (1) assessment of the process instance based on the textual data related to the process execution in terms of (i) the customer-related information (sentiment, external layer) and (ii) unit-specific knowledge (taxonomies, internal and immediate layers) related to the processing of the similar process instances, (2) based on (1), recommendation of possible actions. Together with the experts, we searched for a simple technique enabling us to integrate all the knowledge from three context layers within one text analytics approach. Due to their common usage and implementation simplicity, association rules represent an appropriate technique to enable such integration.

To sum up, based on the information collected in all the layers with the help of association rules, the process worker gets (i) an immediate comprehension of the current situation, i.e., process instance urgency, complexity in terms of time and resources, and (ii) a recommendation of tasks relevant for the process execution. Using this recommendation, the worker is able to adequately assess the situation and plan the implementation of the ticket.

5 Case study

We used both ticket and corresponding task textual descriptions from the same organization to evaluate and demonstrate the applicability of the framework in a case study. After cleaning and preprocessing, the final dataset is comprised of 4623 entries. The time of the dataset covers prevailingly the first half of 2019. In the case study, experiments were performed using Python 3.6, and the association rules were implemented in R, using the arules package. As further important textual data sources, we used ITIL handbooks and process descriptions existing in the company. These data have been used to develop case-study-specific vocabularies and taxonomies necessary in each context layer.

For the purpose of illustration, we provide an anonymized IT ticket example from our case study. A CHM worker receives the following customer request via email: "*Dear colleagues, we need a service pack installation on production SAP HANA. Online installation is possible. Kind regards, XXX*". This request is entered into the IT ticketing system, as a rule, in its original form. However, it might be slightly modified or extended by the CHM workers. Afterward, while planning the ticket execution, the CHM worker breaks down the ticket into separate tasks. In the example, the following three tasks were planned: "*execute SAP HANA service pack installation on YYY*", "*4EP*", "*QA task, QA task will include healthcheck validation by ZZZ, check of logs, check of application*". Below, we explain the application of the proposed framework layer by layer and illustrate it using the example.

5.1 External layer

In line with the consideration of the context at the external layer (see Section 4.1), we deploy a lexicon-based context-specific Business Sentiment (BS) to measure the "emotional" component, or level of anxiety, implied by the customer in the request description. As a rule, standard lexicons do not function well in domain-specific applications [40]. This prompted us to create a domain-specific BS lexicon utilizing the state-of-the-art VADER [41] and Latent Dirichlet Allocation (LDA) algorithm [42]. For details, we refer to [36] and the Github page. Using VADER, we, *first*, identify the emotionally loaded keywords and expressions and create the BS lexicon. Afterward, we enrich the BS lexicon with the keywords obtained from the LDA implementation. To do so, two sources are used: (1) IT ticket texts and (2) CHM process descriptions from the ITIL handbook. Each keyword is associated with a positive, negative, or neutral sentiment. Keywords with valence scores greater than 0 are considered positive, whereas those with less than 0 are marked as negative. Other keywords are denoted as having a neutral sentiment. *Second*, we compute the normalized total score of BS keywords with the pre-assigned valence and unique significance markers (syntactic and semantic intensifiers) for each ticket text in the dataset. *Third*, using the CHM workers' feedback, threshold rules are developed. *Fourth*, based on the normalized score and threshold rules, the BS is measured as the customer anxiety level for each ticket on the qualitative scale of normal, moderate, and severe anxiety [45].

In the motivating example, we observe the BS lexicon keywords with neutral valence (*dear, kind regards*) and no syntactic and semantic intensifiers, i.e., normal anxiety.

5.2 Internal layer

Based on the contextual information considerations at the internal layer in Section 4.1, we apply the taxonomic approach. Its realization is accomplished in two steps. *First*, we build a hierarchical taxonomy to determine a process cognition level. Thus, we call it Decision-Making Logic (DML) taxonomy. With this, we aim to discover the decision-making character of activities inherent in processes. We differentiate between the following three DML levels: routine, semi-cognitive, and cognitive. The most significant keywords in the IT ticket texts are identified using LDA. Combined with the CHM workers' feedback, we organize these keywords into one of the three DML levels. For details of DML taxonomy building, we refer to [35] and the Github page. Accordingly, the internal context is realized by assigning the keywords to the DML levels based on the knowledge intensity and process complexity with which these keywords are contextually associated [43].

In our motivating example, the keyword *online* indicates a routine, i.e., simple, activity as no service outage needs to be planned for this ticket. The keyword *colleagues* refers to the semi-cognitive level as one needs to have a contextual understanding and know the history of similar requests to determine who will process the ticket. *Second*, based on the total number of detected keywords in the IT ticket text, we calculate the relative occurrence of keywords and determine the DML level using the context-specific threshold rules defined by the case study employees. Our motivating example reveals four routine *online, install, pack, need* (4/7=0.57) and three semi-cognitive (3/7=0.43) *colleague, service, production* keywords. Thus, the overall DML level of such a ticket is routine.

5.3 Immediate layer

According to the conceptualization of the immediate layer in Section 4.1, to get a meaningful representation of task sets, we develop a typology of ticket types and task types and subtypes in the form of a hierarchical taxonomy [44]. Under ticket and task type, we infer an actual activity that the customer requests. In the motivating example, the type of ticket is *install*. *First*, to build a ticket and task types taxonomy, we extract topics from the ticket and task descriptions using LDA. Afterward, we fine-tune the topics with the involved CHM workers. In the case of tasks, we used the IT ticketing system manual, where three types of tasks are distinguished: (1) projected service outage (PSO), (2) implementation of the ticket itself, and (3) quality assurance (QA). The PSO task means the service disconnection for performing requested changes. The implementation task directly references activities such as installation, update, and migration. QA is necessary for ensuring the service level quality, for example, the four-eyes principle (4EP). As a result, ticket types and task types (including subtypes) are organized into one hierarchical taxonomy. *Second*, we label tickets and tasks in our dataset with the types and subtypes using the developed taxonomy.For details of typology building, we refer to the Github page.

In the motivating example, the ticket type *installation* reveals one implementation task (subtype *installation*), two QA tasks (four-eyes principle, i.e., *4EP*, and *healthcheck* subtypes), and no PSO task since the installation can be executed online.

5.4 Decision-making support enabled by the framework

As suggested in Section 4.1 and shown in Table 1, we enhance the association rules with the information obtained at the internal and external layers, i.e., DML/process cognition and BS/customer anxiety values. In particular, the association rules are applied to determine possible sets of task types and subtypes (rule body, i.e., consequent) based on the ticket type (rule head, i.e., condition). For details of the implementation and more examples, we refer to the Github page.

In Table 1, using our motivating example, row 1 represents association rules based on ticket types and task types and subtypes. Our ticket of type *installation* is predicted to have three tasks with the support of 7%: *implementation* (subtype *installation*) and two *QA* tasks (subtype *4EP* and *healthcheck*). It means that such a pattern occurs in 7% of the cases, i.e., rows in the dataset. The confidence of 55% shows how frequently the rule head *{Ticket: Installation}* appears among all other rows containing the rule body *{Implementation: Installation, QA: 4EP, QA: Healthcheck}*. The rows below indicate the contextual enhancements of the same ticket, i.e., ticket type (rule head) is enhanced with DML/process cognition and BS/customer anxiety values. Performing this experiment, we could make the following main observations: (1) on average, when adding the contextual information, the support and confidence values increase, (2) using the support and confidence values, most frequent patterns can be identified, (3) the influence of specific contextual information in the rule head, i.e., DML/process cognition and BS/customer anxiety, on the task sets in the rule body can be identified. Based on this information, the process worker can assess the urgency of the customer request (ticket) and how to better deal with/contact that customer, estimate the ticket complexity, and get a recommendation of task sets needed to process the ticket.

Table 1. Association rules based on motivating example

Nr.	Association rule	Supp.[2] (%)	Conf.[2] (%)
1	{Ticket: Installation} => {Implementation: Installation, QA: 4EP, QA: Healthcheck}	7	55
2	{Ticket: Installation (routine)} => {Implementation: Installation, QA: 4EP, QA: Healthcheck}	13	88
3	{Ticket: Installation (routine, normal)} => {Implementation: Installation, QA: 4EP, QA: Healthcheck}	12	89
4	{Ticket: Installation (routine, moderate)} => {Implementation: Installation, QA: 4EP, QA: Healthcheck}	10	83
5	{Ticket: Installation (semi-cognitive)} => {Implementation: Installation, QA: 4EP, QA: Healthcheck, QA: Backup}	10	71
6	{Ticket: Installation (semi-cognitive, normal)} => {Implementation: Installation, QA: 4EP, QA: Healthcheck, QA: Backup}	9	70
7	{Ticket: Installation (semi-cognitive, moderate)} => {Implementation: Installation, QA: 4EP, QA: Healthcheck, QA: Backup}	11	72
8	{Ticket: Installation (cognitive, normal)} => {Implementation: Installation, QA: 4EP, QA: Healthcheck, QA: Backup, QA: Test}	8	66

[2] supp: Support, conf: Confidence

In Table 2, we summarize all the steps necessary to apply our framework, from data collection and preprocessing to developing decision-making support. Hereby, we describe inputs, processing, and outputs in each step.

Table 2. Framework for context awareness based on textual data. Overview of steps

1. Data collection and preprocessing
Input: textual data serving as input to a process (i.e., containing requests), tools: standard NLP processing software, e.g., Python and NLTK library *Processing*: 1) special preprocessing (retaining capitalization, exclamation and question marks, specific symbols), 2) standard preprocessing (removal of numbers, special symbols, punctuation, converting to lowercase, stemming), *Output*: files 1) and 2) with preprocessed textual data
2. External context knowledge extraction
Input: file 1), BS lexicon, threshold rules for BS level assignment, tools: Python, NLTK *Processing:* identification of BS keywords and their valence, intensifiers, calculation of the normalized total score, BS level assignment *Output:* file 3) with BS total scores, normalized total scores, assigned BS level for each textual entry
3. Internal context knowledge extraction
Input: file 2), DML taxonomy, threshold rules for DML levels assignment, tools: Python, NLTK *Processing*: identification of DML keywords, calculation of the relative occurrence of the keywords of each category, DML level assignment *Output:* file 4) with DML keywords and assigned DML for each textual entry
4. Immediate context knowledge extraction
Input: file 2) enriched with textual data related to process execution (i.e., containing tasks or activities), task typology taxonomy, tools: Python, NLTK *Processing:* identification of ticket types, task types and subtypes based on the principle of the maximum relative distribution *Output:* file 5) with ticket type, task type and subtype keywords/expressions and their number, assigned ticket type, task type and subtype for each task text
5. Decision-making support
Input: files 3), 4), 5), tools: R (arules package) *Processing:* application of association rules *Output*: file 6) with the association rules (see Table 2)

Below, we provide details on our evaluation and findings as well as discuss threats to validity and limitations.

5.5 Evaluation and discussion

When performing the evaluation, we set out to test how far the association rules can predict a potentially relevant task set for an incoming ticket. The prediction considers the following: contextual information describing the ticket content, i.e., ticket type, DML/process cognition, and BS/customer anxiety. While assessing the prediction quality, we rely on two indicators: (1) support and confidence values and (2) proprietary rule-based algorithm based on [45] involving the following steps:

- Generation of a training and test dataset (70%:30%) to assess the prediction quality.
- Generation of prediction rules by transforming the association rules obtained at the mining stage from the training dataset into *condition* - *consequent* pairs. *Condition* is the head of the association rule in one of the three formats: (1) ticket type, (2) ticket type and DML/process cognition, (3) ticket type, DML/process

cognition, BS/customer anxiety. *Consequent* is the rule body, i.e., a set of tasks associated with a given *condition*.

- Forming a rule-based algorithm for the assessment of prediction quality consisting of the following. (a) We look for the *condition* matching the *condition* in the test dataset (in one of the three formats) to perform prediction and use the top-3 *consequents* as predictions of a task set ordered by the support score. (b) We evaluate the quality of our prediction by the number of attempts to find an exact or partial match between the top-3 *consequents* and the corresponding task sets from the test dataset.
- Finally, we evaluate the *consequents* and compare the results for three different types of *conditions*. Hereby, we determine how often the next task set can be correctly predicted in the first three prediction attempts (top-3). We suppose that the three suggestions are a reasonable number to display to a process worker as a recommendation.

Thus, we correctly predicted the task set, i.e., rule body or *consequent*, based on the ticket type *condition* in 43% of cases using one attempt. With increasing attempts, this number has grown to 50% (three attempts). Ticket type and DML/process cognition in the rule head (*condition*) demonstrated an evident increase in prediction quality, i.e., 59% of cases (one attempt) and 65% (three attempts). Finally, adding the BS/customer anxiety contextual information in the rule head showed no substantial influence: 57% (one attempt) and 63% (three attempts).

The described two-fold evaluation of the prediction quality allows the process worker to choose the best option from the top-3 recommended task sets. The enrichment of the rule head (*condition*) with the DML/process cognition contextual information positively influences the support and confidence values, in contrast to BS/customer anxiety. Similarly, DML/process cognition apparently impacts the task sets from the content viewpoint, i.e., task types and subtypes. The higher the DML/process cognition, the higher the amount of QA and PSO tasks. On the contrary, the BS/customer anxiety has a minimal impact on the task sets. However, this information can be used by the process workers to make a correct prioritization in the process execution, i.e., IT ticket processing.

Due to the nature of context information [46], our study reveals several threats to validity and limitations, which we have either partially addressed or plan as a part of future work:

- *Context information has a range of temporal characteristics*. Currently, a process worker gets a recommendation in the form of a task set and needs to adjust the correct execution order based on the individual experience. Using the task execution time information, we plan to perform task mining at the immediate context layer to provide recommendations on the task execution sequence.
- *Context information is imperfect*. It might contain issues in reflecting reality. In our study, the decision support has a recommendation character and consists of a set of possible options. It is transparent regarding the deployed text analytics and recommended task sets.
- *Context information is highly interrelated*. We have already managed to identify several relationships between the DML/process cognition context information in the rule head and certain task types in the rule body (or their absence in the

case of BS/customer anxiety). While including the recommendation on the task execution order, we plan to examine possible relationships in this regard.

A further limitation is related to the textual process data itself. As one can conclude from the framework conceptualization section, two important requirements on the data are the following: (1) these contain a request from a customer, and (2) textual data records on the tasks or activities necessary to process the requests are available. The ITSM case study has been selected as it fulfills both requirements. Although ITSM has become rather popular, especially in large organizations, we suggest other cases where our framework can be applied. In particular, there are many cases where textual data used as an input to a process impacts its execution. In healthcare, for example, patients' complaints arriving per email or over a form on the website usually determine the required diagnostics and, hence, activities necessary to address the reported health problem. Similarly, in public administration, citizens' requests directly influence the required activities to handle the request. In other words, our framework can be applied in those settings where a process execution relies on textual data input.

Finally, we would like to note another limitation related to the knowledge extracted as linguistic features, i.e., DML/process cognition, BS/customer anxiety, and ticket types, task types and subtypes. Same as context and textual data, linguistic features can also potentially bring cognitive overload and subconsciously misinform process workers. To address the first limitation of potential cognitive overload, as a part of future work, we aim to develop a dashboard with two focus points: (1) color-coded representation of DML/process cognition and BS/ customer anxiety and (2) recommendation of the tasks or activities and their order to solve the incoming request. In the case of the second limitation regarding possible errors in our recommendations, we highlight that our approach and implemented text analytics techniques enable the transparency of those keywords and phrases, which might lead to the misinformation of the process worker. Further, our approach has a recommendation character, i.e., process workers can decide to rely on their own experience or other sources of information.

6 Conclusion and future work

Under conditions of increased remote communication, the amounts and importance of textual data in organizations cannot be underestimated. As humans are limited in processing large amounts of data, decision-making support solutions extracting important context-related aspects are in demand. In this study, we aimed to develop a framework for context awareness based on textual process data. Specifically, we extended a prominent BPM context model where we bring forward the human aspect in BPM through the textual data perspective. Hereby, the framework, our main theoretical and methodological contribution, emphasizes the value of non-technical artifacts in BPM.

We developed the framework in the IT ticket processing case study, which allowed us to demonstrate its applicability and practical value. In particular, we highlight the context layers, linguistics features, and text analytics to extract the latter. As a result, process worker context awareness is significantly improved, enabling specific decision-making support on predicting process cognition and customer anxiety level, i.e., complexity and urgency, expressed in the textual request. Further, the process worker is supported with a list of recommended task sets for handling each incoming IT ticket.

For the recommendation, the framework obtains additional contextual information from the textual data contained in IT tickets. Our findings showed that considering the contextual information enabled by our framework increases the quality of decision-making support. As a part of future work, from the performance viewpoint, we aim to enhance our framework with machine learning approaches to predict process cognition and customer anxiety as well as develop a dashboard providing recommendations. From the context viewpoint, we will consider cases to enrich the framework with the environmental layer. Further, we will search for different settings like healthcare or public administration to test our framework.

References

1. Dumas, M., La Rosa, M., Mendling, J., Reijers, H.A.: Fundamentals of Business Process Management. Springer (2013).
2. Rosemann, M., Recker, J., Flender, C.: Contextualization of business processes. International Journal of Business Process Integration and Management. 3, 47–60 (2008).
3. vom Brocke, J., Schmiedel, T., Recker, J., Trkman, P., Mertens, W., Viaene, S.: Ten principles of good business process management. Business Process Management Journal. 20, 530–548 (2014).
4. Avgerou, C.: Contextual explanation: alternative approaches and persistent challenges. MIS Quarterly. 43, 977–1006 (2019).
5. Sundermann, C.V., de Pádua, R., Tonon, V.R., Domingues, M.A., Rezende, S.O.: A Context-Aware Recommender Method Based on Text Mining. In: EPIA Conference on Artificial Intelligence. pp. 385–396. Springer (2019).
6. Song, R., Vanthienen, J., Cui, W., Wang, Y., Huang, L.: Context-aware BPM using IoT-integrated context ontologies and IoT-enhanced decision models. In: IEEE Conference on Business Informatics. pp. 541–550. IEEE (2019).
7. Whetten, D.A.: An Examination of the Interface between Context and Theory Applied to the Study of Chinese Organizations. Management and Organization Review. 5, 29–55 (2009).
8. Johns, G.: The Essential Impact of Context on Organizational Behavior. Academy of Management Review. 31, 386–408 (2006).
9. Pentland, B.T., Recker, J., Wolf, J.R., Wyner, G.: Bringing Context Inside Process Research with Digital Trace Data. Journal of the Association for Information Systems. 21, 5 (2020).
10. Müller, O., Junglas, I., Debortoli, S., vom Brocke, J.: Using Text Analytics to Derive Customer Service Management Benefits from Unstructured Data. MIS Quarterly Executive. 15, 243–258 (2016).
11. Hoang, H.H., Jung, J.J.: An Ontological Framework for Context-Aware Collaborative Business Process Formulation. Comput. Informatics. 33, 553–569 (2014).
12. Hompes, B.F.A., Buijs, J.C.A.M., van der Aalst, W.M.P.: A generic framework for context-aware process performance analysis. In: OTM Confederated International Conferences "On the Move to Meaningful Internet Systems." pp. 300–317. Springer Verlag (2016).
13. Rizun, N., Revina, A., Meister, V.G.: Assessing business process complexity based on textual data: Evidence from ITIL IT ticket processing. Business Process Management Journal. 27, (2021).
14. Paas, F., Sweller, J., Paas, F., Sweller, J.: An Evolutionary Upgrade of Cognitive Load Theory: Using the Human Motor System and Collaboration to Support the Learning of Complex Cognitive Tasks. Educational Psychology Review. 24, 27–45 (2011).
15. vom Brocke, J., Zelt, S., Schmiedel, T.: On the role of context in business process management. International Journal of Information Management. 36, 486–495 (2016).
16. Zelt, S., Recker, J., Schmiedel, T., vom Brocke, J.: A theory of contingent business process

management. Business Process Management Journal. 25, 1291–1316 (2019).
17. Weber, M., Grisold, T., vom Brocke, J., Kamm, M.: Context-Aware Business Process Modeling: Empirical Insights from a Project with a Globally Operating Company. In: European Conference on Information Systems. AIS, Marrakesh, Morocco (2021).
18. Rosemann, M., Recker, J.: Context-aware Process Design: Exploring the Extrinsic Drivers for Process Flexibility. In: Workshop on Business Process Modeling, Development, and Support at CAiSE. pp. 149–158. CEUR, Luxembourg (2006).
19. Zelt, S., Recker, J., Schmiedel, T., Brocke, J. vom: Development and validation of an instrument to measure and manage organizational process variety. PLOS ONE. 13, e0206198 (2018).
20. vom Brocke, J., Baier, M.-S., Schmiedel, T., Stelzl, K., Röglinger, M., Wehking, C.: Context-Aware Business Process Management. Business & Information Systems Engineering. 63, 533–550 (2021).
21. Boukadi, K., Chaabane, A., Vincent, L.: Context-Aware Business Processes Modelling: Concepts, Issues and Framework. In: IFAC Proceedings Volumes. pp. 1376–1381. Elsevier (2009).
22. Enrique, H.V., De Maio, C., Fenza, G., Loia, V., Orciuoli, F.: A context-aware fuzzy linguistic consensus model supporting innovation processes. In: IEEE International Conference on Fuzzy Systems. pp. 1685–1692. IEEE (2016).
23. Mounira, Z., Mahmoud, B.: Context-aware process mining framework for business process flexibility. In: International Conference on Information Integration and Web-Based Applications and Services. pp. 421–426 (2010).
24. Hidri, W., M'tir, R.H., Ben Saoud, N.B., Ghedira-Guegan, C.: A Meta-model for context-aware adaptive Business Process as a Service in collaborative cloud environment. Procedia computer science. 164, 177–186 (2019).
25. Graesser, A.C., McNamara, D.S., Kulikowich, J.M.: Coh-Metrix: Providing Multilevel Analyses of Text Characteristics. Educational Researcher. 40, 223–234 (2011).
26. Dinh, L.T.N., Karmakar, G., Kamruzzaman, J.: A survey on context awareness in big data analytics for business applications. Knowledge and Information Systems. 62, 3387–3415 (2020).
27. Purnomo, F., Heryadi, Y., Gaol, F.L., Ricky, M.Y.: Smart city's context awareness using social media. In: International Conference on ICT for Smart Society. pp. 119–123. IEEE (2016).
28. Cartelli, V., Di Modica, G., Tomarchio, O.: A cost-centric model for context-aware simulations of business processes. In: International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management. pp. 303–314. SciTePress (2015).
29. Daelemans, W.: Explanation in Computational Stylometry. In: International Conference on Intelligent Text Processing and Computational Linguistics. pp. 451–462. Springer, Samos (2013).
30. Anastassiu, M., Santoro, F.M., Recker, J., Rosemann, M.: The quest for organizational flexibility: Driving changes in business processes through the identification of relevant context. Business Process Management Journal. 22, 763–790 (2016).
31. Ploesser, K., Recker, J., Rosemann, M.: Building a methodology for context-aware business processes: insights from an exploratory case study. In: European Conference on Information Systems IT to Empower. pp. 1–12. University of Pretoria, South Africa (2010).
32. Peffers, K., Tuunanen, T., Rothenberger, M.A., Chatterjee, S.: A design science research methodology for information systems research. Journal of Management Information Systems. 24, 45–77 (2007).
33. Wand, Y., Weber, R.: Research commentary: Information systems and conceptual modeling - A research agenda. Information Systems Research. 13, 363–376 (2002).
34. Axelos: ITIL® Service Transition. TSO, London (2011).

35. Rizun, N., Revina, A., Meister, V.: Method of Decision-Making Logic Discovery in the Business Process Textual Data. In: International Conference on Business Information Systems. pp. 70–84. Springer (2019).
36. Rizun, N., Revina, A.: Business Sentiment Analysis. Concept and Method for Perceived Anticipated Effort Identification. In: Information Systems Development: Information Systems Beyond 2020. pp. 1–12. AIS eLibrary (2019).
37. Grossnickle, J., Raskin, O.: The Handbook of Online Marketing Research: Knowing Your Customer Using the Net. McGraw-Hill Education (2000).
38. Beigi, G., Hu, X., Maciejewski, R., Liu, H.: An overview of sentiment analysis in social media and its applications in disaster relief. In: Studies in Computational Intelligence. pp. 313–340. Springer (2016).
39. Medelyan, O., Witten, I.H., Divoli, A., Broekstra, J.: Automatic construction of lexicons, taxonomies, ontologies, and other knowledge structures. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery. 3, 257–279 (2013).
40. Hammer, H., Yazidi, A., Bai, A., Engelstad, P.: Building Domain Specific Sentiment Lexicons Combining Information from Many Sentiment Lexicons and a Domain Specific Corpus. In: IFIP Advances in Information and Communication Technology. pp. 205–216. Springer (2015).
41. Hutto, C.J., Gilbert, E.: VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text. In: International Conference on Weblogs and Social Media (2014).
42. Blei, D.: Probabilistic Topic Models. Communications of the ACM. 55, 77–84 (2012).
43. Eppler, M.J., Seifried, P., Röpnack, A.: Improving Knowledge Intensive Processes through an Enterprise Knowledge Medium. In: Conference on Managing Organizational Knowledge for Strategic Advantage: The Key Role of Information Technology and Personnel. pp. 222–230. Gabler (1999).
44. Rizun, N., Revina, A., Meister, V.G.: Analyzing content of tasks in Business Process Management. Blending task execution and organization perspectives. Computers in Industry. 130, 103463 (2021).
45. Wright, A.P., Wright, A.T., McCoy, A.B., Sittig, D.F.: The use of sequential pattern mining to predict next prescribed medications. Journal of Biomedical Informatics. 53, 73–80 (2015).
46. Henricksen, K., Indulska, J., Rakotonirainy, A.: Modeling context information in pervasive computing systems. In: International Conference on Pervasive Computing. pp. 167–180. Springer (2002).