Method of Decision-Making Logic Discovery in the Business Process Textual Data

Nina Rizun[1], Aleksandra Revina[2](✉), and Vera Meister[3]

[1] Gdansk University of Technology, 80-233 Gdansk, Poland
nrizun@zie.pg.gda.pl

[2] Technical University of Berlin, 10623 Berlin, Germany
revina@tu-berlin.de

[3] Brandenburg University of Applied Sciences, 14770 Brandenburg an der Havel, Germany
vera.meister@th-brandenburg.de

Abstract. Growing amount of complexity and enterprise data creates a need for novel business process (BP) analysis methods to assess the process optimization opportunities. This paper proposes a method of BP analysis while extracting the knowledge about Decision-Making Logic (DML) in a form of taxonomy. In this taxonomy, researchers consider the routine, semi-cognitive and cognitive DML levels as functions of BP conceptual aspects of Resources, Techniques, Capacities, and Choices. Preliminary testing and evaluation of developed method using data set of entry ticket texts from the IT Helpdesk domain showed promising results in the identification and classification of the BP Decision-Making Logic.

Keywords: Business process management · Decision-making · Robotic Process Automation · Natural Language Processing · Text Mining

1 Introduction

A strong market-driven digitization trend opens many opportunities for organizations, such as cost savings and performance increase achieved by the process automation, but at the same time puts some barriers, such as how to assess enterprise processes for an optimal utilization of digitization opportunities offered on the market and in the open source communities. One of these increasingly discussed process automation technologies is Robotic Process Automation (RPA) [1]. RPA software can automatically take over the execution of routine repetitive tasks of a human employee. Today, it is also possible to augment simple RPA with so-called “cognitive” functions based on scalable Natural Language and Image Processing technologies equipped with Machine Learning capabilities. Here, marketing and consulting specialists use the term Intelligent Process Automation (IPA) [2]. However, independently from the terminology, key challenges for an organization remain the following: (1) how to identify process activities that are suitable for automation and (2) how to identify the achievable degree or form of automation (for example, RPA vs. IPA).

W. Abramowicz and R. Corchuelo (Eds.): BIS 2019, LNBIP 353, pp. 70–84, 2019.
https://doi.org/10.1007/978-3-030-20485-3_6

In this paper, motivated by the above-mentioned challenges, the authors set the research objective to develop a novel method of extracting knowledge regarding the decision-making nature of processes in a form of Decision-Making Logic (DML) taxonomy from the process textual data. Under DML, the researchers understand "cognition" level of a decision-making process, i.e. perceived processing complexity of tasks to be performed within the process by a process worker and related task automation possibility in the context of existing rules, available information for task execution and automation costs [3–5]. To the extent of the authors' knowledge, there is no other approach combining the same methodological and technological setup, thus, no other approach reveals the same merits as the method presented in the paper. The researchers look into the IT ticket texts considering diverse conceptual aspects to discover the DML, i.e. parts of speech organized as per ***R****esources* (nouns), ***T****echniques* (verbs and verbal nouns), ***C****apacities* (adjectives), and ***C****hoices* (adverbs) (RTCC). With the help of the RTCC framework, the authors can correctly capture the various aspects of the DML levels hidden in the entry ticket texts.

The rest of the paper is organized as follows. Section 2 introduces the related work. Section 3 presents the method for the DML identification, discusses its major steps and preliminary evaluation. Finally, Sect. 4 concludes the paper and outlines the future work.

2 Related Work

The authors suggest to structure the related work section into (1) the work related to knowledge extraction in general and the BP textual data specifically, (2) the sources necessary for understanding the method and applied technologies, i.e. Natural Language Processing (NLP) and Text Mining, and (3) subjects closely related to the present research, i.e. decision-making and taxonomies in the process context.

2.1 Approaches of Knowledge Extraction

The topic related to the extraction and representation of knowledge from texts is rather varied in the sense of approaches and especially results interpretation and formalization. Thus, in semantic technologies, widespread RDF (Resource Description Framework) approach [6] describes the data about (web) resources as subject-predicate-object triples [7]. Here, the subject must be an entity, whereas the object may also be a textually named literal. Approaches such as [8] use logical-linguistic models that consider the knowledge as sentences in the form of subject-predicate-object triplets. Another group of scientists deeply working on the knowledge extraction from the BP textual data highlight two research application areas [9]: (1) analysis of natural language inside process models [10]; (2) techniques that analyze the natural language captured in textual process descriptions [11]. One of the most relevant research publications in the context of the current paper deals with the identification of RPA candidate tasks in the BP textual descriptions [9]. The scientists suggest a three-step Machine Learning-based approach to automatically detect the degree of automation of tasks (manual, user, automatic) described in the textual process descriptions. However, the formalization approach of the extracted knowledge in the routine-cognitive classification context is missing. The authors of the present paper while building up upon the findings of [3, 9] suggest a novel approach of knowledge extraction, interpretation and formalization with the DML taxonomy of the BP activities.

2.2 Technologies Applied for Knowledge Extraction

In the paper, the main source of knowledge are the IT ticket texts coming either per email or directly entered in the ticketing system of the case study. Thus, the technologies for knowledge extraction are related to NLP and Text Mining. One of these knowledge types connected with semantic aspects hidden in BP texts is Latent Semantic Relations (LSR), which can be found both inside the documents and between them and are used to identify the context of the analyzed document and to classify a group of documents based on their semantic proximity. Specifically, a mathematical model of a text collection describing the words or documents is associated with a family of probability distributions on a variety of topics [12]. The aim of the LSR analysis is to extract "semantic structure" from the collection of information flows and automatically expand them into the underlying topic. A variety of approaches, such as discriminative Latent Semantic Analysis (LSA) [13] or probabilistic Latent Dirichlet Allocation (LDA) [12], evolved in this field with the time. While using the mentioned state-of-the-art technologies, the researchers aim to experiment on improving their quality and finding the ways to eliminate the limitations.

2.3 Decision-Making and Taxonomies in the Process Context

There is a number of research studies devoted to the decision-making processes in the business context. One group addresses the Theory of Decision-Making in general and related open questions [14–16], second group of studies discusses the challenges and opportunities of decision-making in an enterprise context, such as context-aware group decision-making [17], the criteria and approaches in decision-making support [18], or text mining-based extraction of decision elements out of project meeting transcripts [19]. Nonetheless, to the best of the authors' knowledge, the approach of the DML discovery suggested in the present paper is not researched so far.

Decision Mining in the context of Process Mining represents another relevant research direction. Process Mining [20] is a technique that aims to extract facts out of the event log. Hereby, the analysis of the latter can provide important knowledge that can help organizations to improve their decision-making processes. Specifically, Decision Mining aims at the detection of data dependencies that affect the event-based routing within a process model [21]. Regardless of this fact, if compared to the proposed research, Decision Mining primarily focuses on the analysis of event logs generated by the machines and not on the natural language texts generated by human workers within the process.

Being a widespread method of knowledge structuring and management [22], taxonomies find also their application in the decision-making related classifications [23, 24], nonetheless taxonomies applied on the levels of DML extracted from BP texts are not researched so far.

To sum up, these studies do represent a significant contribution to the development of science. However, first, the approach of the DML discovery in the unstructured BP textual data is not well represented, and, second, taxonomies applied to the DML levels extracted from BP texts are not researched so far. Furthermore, the present research makes a valuable contribution with its focus on the RPA bringing it into the new DML context. As [9] fairly state, automation in BPM is not a recent development. Research on RPA, not to mention the most recent IPA technology, by contrast, is still scarce [1].

3 Method for the Identification of the Decision-Making Logic

The work is based on the Design Science Research guidelines by Hevner et al. [25] and uses common methods of Information Systems research, such as case study, computer experiments, interviews, and observations.

The business need for novel process analysis methods in the context of emerging technologies has been identified while observing the R&D achievements (see Related Work Section), growth of the market offerings, increasing enterprise process complexity and costs (see Introduction Section). The envisioned artifact, i.e. method for process analysis, aims at classifying the BPs into simple (routine) and complex (cognitive) ones from the perspective of decision-making complexity of the worker responsible for the BP processing. At present research stage, the method uses unstructured textual data triggering the process and generated in a natural language by a process participant. This data can be received via different communication channels – email, chat, phone call, or directly entered in a specific task management system in a natural language form. In future work, the routine BPs can be automated with mentioned RPA technologies, and for the cognitive BPs – diverse levels of support for process workers can be proposed. In the paper, the artifact is evaluated using the case study qualitative survey approach. In this regard, the following research questions, which are proved in the evaluation, are raised:

RQ1: What measurements can enable the identification of the BP “cognition” level based on the unstructured BP textual data triggering the process?
RQ2: Does the semantic knowledge extracted from the unstructured BP textual data using the proposed method provide valuable information to identify the specific DML level?

The researchers use the methodological triangulation approach, which is based on multiple data sources in an investigation to produce rich and well-developed understanding for the research artifact [26]. Currently, the researchers implemented three major phases: (1) literature review, recent research and market observations to conceptualize the DML taxonomy; (2) DML taxonomy vocabulary population and experimental set-up based on the case study BP texts; (3) evaluation of the approaches (1) and (2) via qualitative survey with the case study process workers.

In the subsections below, the mentioned phases as well as detailed research artifact description (see Fig. 1) are presented.

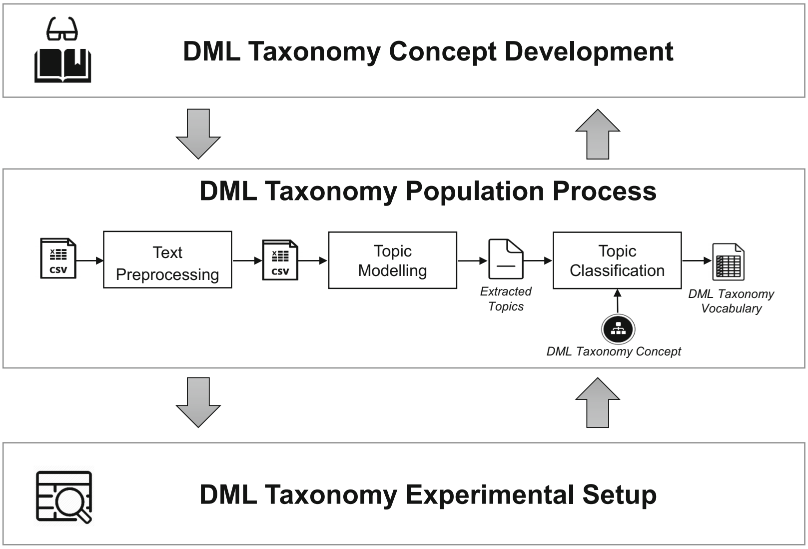

Fig. 1. Steps of method for the DML identification

3.1 DML Taxonomy Concept Development

In the first phase of the methodological triangulation, aiming to address the RQ1, i.e. measurements that could enable the identification of the BP "cognition" level, the researchers developed an understanding of the DML levels and using the systematic literature analysis enhanced with recent research and market observations distinguished three classes – *routine, semi-cognitive* and *cognitive*. Herewith, the following definitions have been accepted: (1) *routine* DML level activities or tasks are those expressible in rules so that they are easily programmable and can be performed by computers at economically feasible costs [4, 5]; (2) *semi-cognitive* DML level activities or tasks are those where no exact rule set exists and there is a clear need of information acquisition and evaluation [3, 4]. Here, computer technology cannot substitute but increases the productivity of employees [27] by partial task processing; (3) *cognitive* DML level activities or tasks are the most complex ones where not only information acquisition and evaluation is required, but also complex problem solving [4]. Computers can offer only a minimal support.

In order to measure the DML levels, a taxonomy-based approach is suggested based on the following principles: (1) consideration of a sentence as a tuple of parts of speech compound of nouns, verbs, verbal nouns, adjectives, and adverbs; (2) assumptions that DML effects of a BP in the form of *routine, semi-cognitive* and *cognitive* can be largely understood as functions of their (BP) conceptual aspects ***R****esources*, ***T****echniques*, ***C****apacities*, and ***C****hoices* (referred by the authors as RTCC semantic tagging framework). Hereby, ***R****esources* (nouns) indicate the specificity of business process task items affected by the decision-making activity; ***T****echniques* (verbs and verbal nouns) represent knowledge and information transformation activities by which the decision-making process affects existing resources; ***C****apacities* (adjectives) describe situational specificity of decision-making techniques or resources; and finally, ***Choices*** (adverbs) determine the selection of the required set of decision-making techniques or resources in the course of DML.

Furthermore, using a systematic literature approach and previous work [15, 16], the researchers drafted a set of indicators, or contextual variables, based on which the classification into the three DML levels according to the RTCC framework takes place (see Table 1).

In the subsection below, while populating the DML taxonomy with the contextual values of variables, the authors aim to develop a set of domain attributes and characteristics for each of the DML levels.

3.2 DML Taxonomy Population Process

In the second phase of the methodological triangulation, the case study is introduced. The data set in a form of entry ticket texts comes from the IT Change Management (ITIL Framework[1]) ticket-processing department of a big enterprise with more than 200,000 employees worldwide. The tickets can be opened in the system using an incoming email text from the customer or directly by a professional process worker who already understands what needs to be done having received the request via different channel. While developing the DML taxonomy vocabulary, the researchers extracted the topics with descriptive key words out of the case study data set, entry ticket description texts, using mentioned LD/SA approach [28, 29]. The vocabulary (see Table 1) was developed based on the available data set processed and converted into a CSV-formatted text corpus with more than 1,000,000 documents (text entries) of English, German and English-German ticket texts created in the period of 2015–2018. After removing duplicates and selecting English texts, the final case study data sample comprised 28,157 entries.

The architecture of the DML taxonomy population process (DML_TPP) is visualized on Fig. 1 as a part of the method for the DML identification. The unstructured use case BP textual data in natural language, entry ticket texts, served as an input. As a result, the researchers obtained domain dependent attributes in the form of descriptive key words extracted from the textual data (separate parts of speech – nouns, verbs, verbal nouns, adjectives, adverbs) and classified them with the help of the contextual variables, grouped based on the conceptual aspects of RTCC structure, into *routine*, *semi-cognitive* and *cognitive* DML levels.

The first step of DML_TPP is to pre-process, parse the text entries and build the document term matrices for the separate parts of speech. It is an important step in order to be able to perform topic modeling in the corpus and afterwards populate the DML taxonomy. The result of this step is cleaned textual process data with separate parts of speech. The second step of DML_TPP is to create topics with descriptive key words over the complete preprocessed data set. In particular, the created document term matrices for each part of speech (nouns, verbs, verbal nouns, adjectives, adverbs) are processed using the combination of LD/SA topic modeling methods [28, 29]. In the

[1] IT Infrastructure Library Framework, www.axelos.com/best-practice-solutions/itil.

Table 1. DML taxonomy vocabulary with exemplary key words

Contextual variables	Decision-making logic levels		
	Routine	*Semi-cognitive*	*Cognitive*
	Conceptual aspects		
	Resources		
	22%	8%	2%
Problem processing level	user, task, user request, interface, tool	team, leader, project, colleague, production	management, CAB, measure, server farm
Indeterminacy	time, application, product, name, ID	description, environment, requirement, solution, problem	risk
Information	server, database, file, location, dataset	requestor, case, rule, outage, power-supply	impact, approval
	Techniques		
	16%	6%	2%
Experience	send, note, deploy, document, decommission	check, assign, increase, create, modify	approve, delegate, define
Action alternative	follow, start, stop, monitor, run	implement, deploy, require, classify, process	propose
Effort	cancel, delete, activate, finish, mount	perform, support, plan, verify, migrate	freeze
	Capacities		
	12%	9%	5%
Specificity	additional, attached, online, virtual, same	separate, specific, technical, minor, successful	major, high, big, small, strong
Decisions formulation	new, old, preinstalled, fixed, ready	available, necessary, important, significant, successful	possible, desired, related, different, multiple
Predictability	actual, full, current, valid, same	temporary, normal, previous, similar, standard	random, randomized, expected
	Choices		
	11%	5%	2%
Precision	automatically, manually, internally, instead, there	normally, well, shortly, enough, recently	approximately, properly
Time	current, still, now, often, daily	newly, immediately, later, urgently	soon
Ambiguity	consequently, completely, never, simultaneously, accordingly	successfully, however, usually, temporarily, previously	randomly, likely, maybe

third step of DML_TPP, the extracted topics with descriptive key words are classified based on the contextual variables into the suggested DML levels of *routine, semi-cognitive* and *cognitive*. Here, the involvement of the process workers being familiar with the context is essential for the right key words classification. In Table 1, the exemplary extracted key words classified according to the three DML levels based on the RTCC framework are presented. The majority of the identified key words belong to the ***R**esources* counting up to 71 key words in the *routine* DML level. Due to the size limits of the paper, the researchers provided up to five exemplary key words per each RTCC element and DML level. As the key word relative distribution based on the total count of 324 (100%) vocabulary key words shows (see also Table 1), the majority of 61% belongs to the class *routine*, 28% to the *semi-cognitive* and very few 11% to the *cognitive*, what can be explained by: (1) the specificity of the data set domain, IT ticket processing, and (2) the fact that the entry texts of routine tickets contain a substantial level of details explaining every single step to be done and not ambiguous generic words that can imply a lot of action options.

The mentioned experiments of data preprocessing and topic modeling were performed using Python 3.4.3. The vocabulary population process is to be executed manually involving the process workers familiar with the context, what is essential for the right key words classification. At this research stage, the researchers performed the population and related classification based on the available process documentation, findings obtained within the workshop with process workers and qualitative survey based findings (see Sect. 3.4). To sum up, based on the developed DML levels taxonomy vocabulary, the authors suggest the following interpretation (characteristics of the DML levels of the data set):

- *Routine* DML level is characterized by prevailingly ***R**esources* indicating the specificity of BP task items followed by ***T**echniques* of knowledge and information transformation. This characteristic can be interpreted as an intensive and clear naming of the process resources, i.e. exact names of servers, databases, configuration items (CIs) related to the ticket.
- Compared with the routine DML level, *semi-cognitive* DML level is increasingly described with ***C**apacities* describing situational specificity of decision-making techniques and ***T**echniques* themselves. The accent shifts from ***R**esources* to the description of the situational specificity of the ***T**echniques,* what can be explained by the absence of the simple and exact rule set to be followed while processing the ticket.
- Compared with routine and semi-cognitive levels, *cognitive* DML level is not well represented in the taxonomy vocabulary. The rare findings, however, show the growing “cognition” within the cognitive DML level key words relative distribution, i.e. increased relative numbers of ***C**apacities* describing situational specificity and ***C**hoices* describing the selection of the ***R**esources* or ***T**echniques*. In this case, the process workers need to act based on their “gut” feeling performing mental simulations of the situation, acquiring and evaluating further information via various channels. All these complicates further ticket processing.

3.3 DML Taxonomy Experimental Setup

In the experimental setup step of the method, the developed DML taxonomy vocabulary was tested (using Python 3.4.3) on the final case study data sample with 28,157 entries (processed and English language based) to analyze the specific distributions of the DML vocabulary words in the tickets (all parts of speech). As shown in Table 2, the following analytical findings can be derived: (1) there are only 20.23% of *pure routine, cognitive* or *semi-cognitive* tickets in the data sample with the clear majority of *pure routine* tickets 18.87%, only 1.30% of *pure semi-cognitive* and scarcely 0.07% of *pure cognitive*[2]; (2) no key words were identified in the group of tickets 2.82%. These tickets are characterized by the 100% presence of unique names of process ***R****esources.* Therefore, the authors classify them as *pure routine* ones (see the interpretation of *routine* DML level above); (3) the prevailing amount of tickets in the data sample 76.95% is the mixed one. In this group, the majority of tickets 33.31% contains >=50% of *routine* key words, <=50% of *semi-cognitive* and 0% of *cognitive*. The second representative subgroup with 19.00% comprises >50% of *routine* key words, <=25% of *semi-cognitive* and <=25% of *cognitive*. Thus, also in the mixed group, the *routine* amount of key words prevails. The obtained analytical findings provide an understanding about the "cognition" level of the tasks the process workers are dealing with in the case study department. The practical value of these findings will be investigated in the future work, i.e. Recommender System approach: (1) *pure routine* tickets can be processed automatically while implementing existing templates; (2) mixed and *pure semi-cognitive* groups of tickets should be studied in more detail, and the recommendation with drop-down lists of possible choices can be provided to the process workers; (3) processing of *pure cognitive tickets* or tickets with relative high amount of the *cognitive* key words (mixed group) should be supported by the provision of the history of similar tickets.

3.4 DML Taxonomy Evaluation

In order to evaluate the approaches presented in Sects. 3.1, 3.2 and 3.3, in the third phase of the methodological triangulation, the researchers developed a qualitative survey in a form of a questionnaire. The survey was conducted in the period of January-February 2019 with the process workers responsible for the ticket processing in the case study IT Change Management department. The sample comprised 13 process managers. The respondents were asked to critically evaluate and provide their own (practical) view on: (1) the definitions of the three DML levels suggested by the researchers (Sect. 3.1); (2) the developed DML taxonomy vocabulary (Sect. 3.2); (3) the ticket classification examples (see Table 3) according to the three DML levels performed by the researchers using sentence-by-sentence approach based on the randomly selected tickets.

As one can conclude from Table 3, the underlying decision-making processes of the process worker in the exemplary anonymized ticket are predicted to be 77%

[2] The relative distributions were calculated based on the presence of the DML taxonomy vocabulary key words in a ticket and not on the overall count of words in a ticket.

Table 2. DML taxonomy vocabulary key word distribution in tickets

Routine key words in a ticket, %	Semi-cognitive key words in a ticket, %	Cognitive key words in a ticket, %	Number of tickets with such a distribution, %
100	0	0	18.87
0	100	0	1.30
0	0	100	0.07
0	0	0	2.82
>=50	<=50	0	33.31
>50	<=25	<=25	19.00
The rest			24.64

Table 3. Anonymized ticket classification example

Ticket (sentences)	Values	Contextual semantic aspect	Word DML level	Ticket DML summary
Please stop-start XYZ databases mentioned below: XYZ1, XYZ2, XYZ3, XYZ4	stop	verb (technique)	routine	77% routine 15% semi-cognitive 8% cognitive
	start	verb (technique)	routine	
	database	noun (resource)	routine	
server: xyzxyz	server	noun (resource)	routine	
Please check mentioned databases if were stop-start successfully, and if applications after start running properly	check	verb (technique)	semi-cognitive	
	database	noun (resource)	routine	
	stop	verb (technique)	routine	
	start	verb (technique)	routine	
	successfully	adverb (choice)	semi-cognitive	
	application	noun (resource)	routine	
	start	verb (technique)	routine	
	run	verb (technique)	routine	
	properly	adverb (choice)	cognitive	

routine, 15% *semi-cognitive* and only 8% *cognitive* (the second characteristic subgroup of the mixed group with 19.00%, see Table 2). Such ***Techniques*** as “stop”, “start”, “run” in the case study context are based on the simple decision-making processes implying a straightforward action (routine DML) while the activity or task “check” demands a certain amount of experience on the level of direct habits to be performed correctly (semi-cognitive DML). Such process ***Resources*** as “database”, “server”, “application” have exact names indicating high accuracy, certainty and complete information for the process worker (routine DML). However, such ***Choices*** determining the selection of the required set of ***Techniques*** as “successfully” (semi-cognitive DML) or “properly” (cognitive DML) have an implicit meaning and in the majority of the cases are based on the experience-gained “gut” feelings.

The results of the qualitative survey and their implications for the current research are presented below:

(1) the accepted DML levels definitions were complemented with the following new contextual aspects (see Summary Table 4). Thus, for example, the DML *routine* level can be characterized from the process worker perspective with the time, frequency, effort, and impact aspects, while the researchers first considered the theoretical perspective of rules, information, and automation based on the literature, recent research and market observations.

(2) the researchers enhanced the DML taxonomy vocabulary with context-based key words received from the respondents, such as: *routine* – “firewall”, “user request”, “rundown”, “decommission” and *cognitive* – “big measures”, “server farm”, “freeze”. As a result, the DML taxonomy vocabulary has been specified with the contextual key words.

(3) the majority of respondents agreed with the provided examples of ticket classification (sentence-by-sentence approach based on the identified descriptive key words) performed by the researchers using the DML taxonomy vocabulary.

The evaluation-based findings mentioned above enabled the researchers to answer the research questions posed at the beginning of the Sect. 3. Hence, while approaching the RQ1 in Sects. 3.1 and 3.2, the researchers suggested (and experimentally tested in Sect. 3.3) a set of measurements for the BP “cognition” level identification based on the unstructured BP textual data triggering the process. These measurements are the DML levels definitions, RTCC structure and the DML taxonomy vocabulary. In the evaluation phase, the proposed set of measurements was specified with the context-based definitions and key words provided by process workers. While approaching the RQ2, the researchers provided ticket classification examples for the evaluation by the process workers. Prevailingly positive evaluation results showed the plausibility of the method (RQ2).

Table 4. Summary of DML levels definitions

Theoretical definitions	Context-based definitions
Routine	
Rules: simple **Information:** complete **Automation:** easily programmable at economically feasible	**Time:** less than 5 min **Frequency:** daily occurred work **Effort:** few mouse clicks **Impact:** no impact
Semi-cognitive	
Rules: no exact rule set **Information:** need for information acquisition and evaluation **Automation:** partial task processing to increase the productivity of employees	**Number of tasks:** many **Number of CIs:** many **Impact:** with clear impact
Cognitive	
Rules: complex **Information:** arguable information demanding complex problem solving **Automation:** minimal possible	**Challenging** **Multi-solution** Thinking of **what?, where?, how?**

4 Conclusion and Future Work

In this paper, the researchers presented a method of extracting knowledge about the Decision-Making Logic ("cognition") of business processes in a form of DML taxonomy from the unstructured textual data. In contrast to the most of existing approaches in process and text analysis, the proposed method is based on the novel combination of methodological and technological approaches, offering new merits for process analysis in the automation context. Following the methodological triangulation approach, the researchers present the method in three main phases: (1) literature review, recent research and market observations to conceptualize the DML taxonomy; (2) DML taxonomy vocabulary population and experimental set-up based on the case study BP texts, IT entry ticket texts; (3) evaluation of the approaches (1) and (2) via qualitative survey with the case study process workers.

The main contribution of the paper is finding answers to the research questions suggested by the authors. The results of the evaluation phase provided additional (contextual) information on the DML definitions and taxonomy vocabulary. Furthermore, the evaluation demonstrated that the method in general and DML taxonomy vocabulary in particular are able to deliver plausible results in the classification of the entry ticket texts according to *routine, semi-cognitive* and *cognitive* DML levels. The discovered topics with descriptive key words appeared to be coherent and informative for DML taxonomy vocabulary building.

However, in the presented research, the authors performed an in-depth analysis of unstructured textual data using only one criterion – specific semantics of an entry ticket text. The researchers will include additional criteria to analyze the mentioned unstructured texts, for example length of the texts and their stylistic characteristics.

Furthermore, such event log data as time stamps, number of tasks and CIs per ticket, responsible groups will be also included into analysis to verify the semantically based DML level findings.

Besides, the researchers will address the demonstration of the research practical value. The proposed DML method can be formalized and automated with the support of an OMG framework for a rule-based decision modeling DMN [30]. The researchers plan to formalize the case study process using BPMN [31] for modeling the overall process with strict procedures while preferring CMMN [32] in case of wide range of free plannable activities depending on the situational context. Hereby, the experimental model of the case study process (Recommender System approach mentioned in Sect. 3.3) will be built using DMN extension to measure the entry ticket DML level as an attempt of partial automation. To enable a functioning prototype of such a multi-level Recommender System, such technologies as robust graph databases, ML algorithms for exact match search, RPA to automate front- and back-end rule-based tasks will be studied and evaluated in detail.

References

1. Bourgouin, A., Leshob, A., Renard, L.: Towards a process analysis approach to adopt robotic process automation. In: 15th International Conference on e-Business Engineering, Xi'an, pp. 46–53. IEEE (2018)
2. Lowes, P., Cannata, F.R.S., Chitre, S., Barkham, J.: The Business Leader's Guide to Robotic and Intelligent Automation. Automate This (2016). Deloitte: https://www2.deloitte.com/content/dam/Deloitte/nl/Documents/operations/deloitte-nl-operations-the-business-leaders-guide-to-robotic-and-intelligent-automation-automate-this-report.pdf. Accessed 1 Jan 2019
3. Koorn, J.J., Leopold, H., Reijers, H.A.: A task framework for predicting the effects of automation. In: Twenty-Sixth European Conference on Information Systems Proceedings, Portsmouth (2018)
4. Autor, D.H., Levy, F., Murnane, R.J.: The skill content of recent technological change: an empirical exploration. Q. J. Econ. **118**, 1279–1333 (2003)
5. Levy, F., Murnane, R.: With what skills are computers a complement? Am. Econ. Rev. **86**(2), 258–262 (1996)
6. Jetschni, J., Meister, V.G.: Schema engineering for enterprise knowledge graphs - a reflecting survey and case study. In: Conference on Intelligent Computing and Information Systems Proceedings, Cairo, pp. 271–277. IEEE (2017)
7. RDF Scheme. https://www.w3.org/TR/rdf-schema/. Accessed 1 Jan 2019
8. Khairova, N.F., Petrasova, S., Gautam, A.P.S.: The logical-linguistic model of fact extraction from English texts. In: Dregvaite, G., Damasevicius, R. (eds.) ICIST 2016. CCIS, vol. 639, pp. 625–635. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-46254-7_51
9. Leopold, H., van der Aa, H., Reijers, H.A.: Identifying candidate tasks for robotic process automation in textual process descriptions. In: Gulden, J., Reinhartz-Berger, I., Schmidt, R., Guerreiro, S., Guédria, W., Bera, P. (eds.) BPMDS/EMMSAD -2018. LNBIP, vol. 318, pp. 67–81. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-91704-7_5

10. Leopold, H., Mendling, J.: Automatic derivation of service candidates from business process model repositories. In: Abramowicz, W., Kriksciuniene, D., Sakalauskas, V. (eds.) BIS 2012. LNBIP, vol. 117, pp. 84–95. Springer, Heidelberg (2012). https://doi.org/10.1007/978-3-642-30359-3_8
11. Leopold, H., van der Aa, H., Pittke, F., Raffel, M., Mendling, J., Reijers, H.A.: Searching textual and model-based process descriptions based on a unified data format. Softw. Syst. Model. **18**(2), 1179–1194 (2017)
12. Blei, D.M.: Probabilistic topic models. Commun. ACM **55**(4), 77–84 (2012)
13. Dumais, S.T., Furnas, G.W., Landauer, T.K.: Using latent semantic analysis to improve access to textual information. In: Conference on Human Factors in Computing Proceedings, pp. 281–285. ACM, New York (1988)
14. Xinghua, L., Zhixin, M.: Research on the problem of the intuitive decision making. In: The 2nd International Conference on E-Education, E-Business and E-Technology Proceedings, Beijing, pp. 192–196. ACM (2018)
15. Rizun, N., Shmelova, T.: Decision-making models of the human-operator as an element of the socio-technical systems. In: Batko, R., Szopa, A. (eds.) Strategic Imperatives and Core Competencies in the Era of Robotics and Artificial Intelligence, pp. 167–204. IGI Global, Hershey (2017)
16. Rizun, N., Taranenko, Y.: Simulation models of human decision-making processes. Manage. Dyn. Knowl. Econ. **2**(5), 241–264 (2014)
17. De Maio, C., Fenza, G., Loia, V., Orciuoli, F., Herrera-Viedma, E.: A framework for context-aware heterogeneous group decision-making in business processes. Knowl.-Based Syst. **102**, 39–50 (2016)
18. Pérez-Álvareza, J.M., Maté, A., Gómez-López, M.T., Trujillo, J.: Tactical business-process-decision support based on KPIs monitoring and validation. Comput. Ind. **102**, 23–39 (2018)
19. Chibelushi, C., Thelwall, M.: Text mining for meeting transcript analysis to extract key decision elements. In: International MultiConference of Engineers and Computer Scientists Proceedings, IMECS, Hong Kong (2009)
20. van der Aalst, W.: Process Mining. Springer, Heidelberg (2016). https://doi.org/10.1007/978-3-662-49851-4
21. Rozinat, A., van der Aalst, W.: Decision mining in ProM. In: Dustdar, S., Fiadeiro, J.L., Sheth, A.P. (eds.) BPM 2006. LNCS, vol. 4102, pp. 420–425. Springer, Heidelberg (2006). https://doi.org/10.1007/11841760_33
22. Guttmann, C.: Towards a taxonomy of decision making problems in multi-agent systems. In: Braubach, L., van der Hoek, W., Petta, P., Pokahr, A. (eds.) MATES 2009. LNCS (LNAI), vol. 5774, pp. 195–201. Springer, Heidelberg (2009). https://doi.org/10.1007/978-3-642-04143-3_19
23. Scherpereel, C.M.: Decision orders: a decision taxonomy. Manag. Decis. **44**(1), 123–136 (2006)
24. Gibcus, P., Vermeulen, P.A., de Jong, J.P.: Strategic decision making in small firms: a taxonomy of small business owners. Entrepreneurship Small Bus. **7**(1), 74–91 (2009)
25. Hevner, A.R., March, S.T., Park, J., Ram, S.: Design science in information systems research. MIS Q. **78**(1), 75–105 (2004)
26. Patton, M.Q.: Enhancing the quality and credibility of qualitative analysis. Health Serv. Res. **34**(5), 1189–1208 (1999)
27. Spitz-Oener, A.: Technical change, job tasks, and rising educational demands: looking outside the wage structure. J. Labor Econ. **24**(2), 235–270 (2006)

28. Rizun, N., Taranenko, Y., Waloszek, W.: The algorithm of modelling and analysis of latent semantic relations: linear algebra vs. probabilistic topic models. In: Różewski, P., Lange, C. (eds.) KESW 2017. CCIS, vol. 786, pp. 53–68. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-69548-8_5
29. Rizun, N., Taranenko, Y., Waloszek, W.: Improving the accuracy in sentiment classification in the light of modelling the latent semantic relations. Information **9**(12), 53–68 (2018)
30. OMG: Decision Model and Notation (DMN) – Version 1.1. Object Management Group, Needham (2016)
31. OMG: Business Process Model and Notation (BPMN) – Version 2.0.2. Object Management Group, Needham (2013)
32. OMG: Case Management Model and Notation (CMMN) – Version 1.1. Object Management Group, Needham (2016)