Wirtschaftsinformatik 2022 Proceedings | Track 19: Business Process Management

Jan 17th, 12:00 AM

Towards a Business Process Complexity Analysis Framework Based on Textual Data and Event Logs

Aleksandra Revina
Technische Hochschule Brandenburg, Germany, oleksandra.revina@gmail.com

Ünal Aksu
Utrecht University, The Netherlands, u.aksu@uu.nl

Towards a Business Process Complexity Analysis Framework Based on Textual Data and Event Logs

Aleksandra Revina[1] and Ünal Aksu[2]

[1]Brandenburg University of Applied Sciences, Faculty of Economics,
Brandenburg an der Havel, Germany
revina@th-brandenburg.de
[2] Utrecht University, Department of Information and Computing Sciences, Utrecht,
The Netherlands
u.aksu@uu.nl

Abstract. Being an established discipline, Business Process Management (BPM) confronts various challenges related to digitization and rapid penetration of technologies into business processes (BPs). As a result, both generated and used data, such as textual data and event logs, grow exponentially, complicating the decision-making. Event logs are typically used to analyze BPs from several perspectives, including complexity. Recent approaches to BP complexity analyses focus on BP models and event logs, limiting the consideration of textual data. Hence, we propose a BP complexity analysis framework combining textual data and event logs. The framework has been conceptualized based on the IT Service Management (ITSM) case study of an international telecom provider and further developed in the IT department of an academic institution. The latter has also been used to investigate the value of the framework. Our preliminary findings show that the framework can enable comprehensive process redesign and improvements.

Keywords: Business process analysis, complexity, event log, textual data

1 Introduction

Business processes (BPs) are one of the most valuable assets of any organization. They have a significant influence on the quality of products and services, customer satisfaction, and, as a result, the sustainability of organizations [1]. BPs define activities, roles, and responsibilities, framing the work of every individual and technology in the organization. Not by chance, they are often called arterial systems of organizations, as any BP malfunction or interruption might paralyze business operations and even the whole process ecosystem [1].

To build efficient BPs and maintain them, a rich set of tools, techniques, methods, and methodologies shaping the Business Process Management (BPM) discipline have been established over the past two decades [1]. To achieve evidence-based redesign and improvement, BP models and event logs, i.e., BP execution data, are commonly considered as a starting point for analyzing BPs. In this regard, current BP complexity

17th International Conference on Wirtschaftsinformatik,
February 2022, Nürnberg, Germany

analysis approaches prevailingly focus on BP models and event logs [2]. For example, process mining (PM) is widely used to extract various process insights, including insights about complexity, via analyzing event logs [3]. Nevertheless, the analysis of textual data is limited to BP descriptions and models [4].

Though difficult to assess, the unstructured text makes up more than 80% of data in companies [5], which also brings up new opportunities for BPM [6]. In fact, textual data is one of the important inputs influencing BP execution. For example, due to the recent pandemics, many organizations, such as the public sector or medical, are increasingly offering various online interfaces to remotely collect their customers' requests, for example, requests related to obtaining a visa or describing a health problem. The clarity and completeness of these requests strongly influence the speed and quality of request processing. Similarly, in IT Service Management (ITSM), the processing of the customer requests, including their urgency, criticality, roles, and teams, highly depends on the textual descriptions. Thus, textual data as an input to a BP impacts decisions, steps, their flow, and resource involvement, i.e., BP complexity. In our previous work [7, 8], we have investigated and proved the potential of textual data for BP complexity prediction. As also observed in this work, no other studies use textual data and event logs in one framework to analyze BP complexity.

This paper proposes a framework (see Figure 1) combining textual data and event logs for BP analysis. Our framework is conceptualized based on the industrial ITSM Change Management (CHM) process from a prominent international telecom provider. In [8–12], we mainly focused on textual data to develop a set of linguistic features suitable for BP complexity prediction. Further, using another ITSM case study from an academic institution, we extended the textual data perspective by event log complexity analysis combining both in one framework. Though addressing the same area (ITSM), the two case studies have been performed in various processes from organizations in entirely unrelated domains, i.e., telecommunication and education. Hence, the characteristics of textual data, such as terminology, style, length, and process characteristics, strongly differ. With this, we aim to achieve the generalizability of the framework.

In the framework, first, the analysis of textual data serving as an input to an IT ticket handling process is performed. Hereby, a set of linguistic features is extracted. Using these features, IT tickets are classified according to a complexity label on the agreed-upon complexity scale. This way, textual data-based (TD) complexity is predicted. Second, the event logs and relevant complexity metrics are used to calculate event log-based (EL) complexity. Finally, correlation analysis of two complexities followed by the analysis of the significant differences in correlations is performed. Based on the results, the recommendations and insights are derived.

The remainder of the paper is structured as follows. Section 2 provides an overview of the related work and highlights unsolved issues. Section 3 discusses the research methods used in the study, whereas Section 4 introduces the framework. We present and discuss the findings in Section 5 and conclude in Section 6.

2 Related Work

Due to the research artifact nature, our study handles such topics as Natural Language Processing (NLP) application in BPM, BP complexity, and textual complexity, which are subsequently reviewed in the subsections below. Additionally, we highlight the unsolved issues in each of the subsections.

2.1 Natural Language Processing in Business Process Management

As more than 80% of enterprise data is estimated unstructured, its investigation offers scholars and practitioners an attractive research and application field [13]. For decades, it has been considered a challenging task [13]. Compared to structured data, analyzing textual data is more complicated, ambiguous, and, hence, time-consuming. Further, its analysis has been limited to the missing ability of computers to understand the human language [14].

At present, the NLP technologies have matured considerably, opening new and broader opportunities in many disciplines. Recent BPM research has also demonstrated that NLP can potentially assist a wide range of BPM tasks in a scalable manner [4]. However, these tasks mainly focus on analyzing BP descriptions and texts related to BP models, i.e., supporting BP discovery and modeling activities [15].

In particular, [4] note the following NLP-related research areas in BPM: identifying the similarity [16], matching [17] and merging of BP models [18], textual-based [19] and semantic [20] search of BP models, lexical ambiguity [21], the transformation of textual descriptions into BP models [22] and vice versa [23]. Additionally, other studies focus on addressing the problem of different abstraction levels in BP models [24], multiple languages and semantic quality check [25], compliance check [26], and auto-completion of BP descriptions [27].

Despite such an abundance in the related literature, there are many unsolved challenges in applying NLP in BPM, such as semantic enhancements and domain or organization-specific adaptations of NLP solutions [4]. In fact, a more rigorous interaction between these two areas has the opportunity to improve the BPM toolset considerably.

2.2 Business Process Complexity

BP complexity has always been a popular research direction in BPM. Being a natural BP characteristic, complexity is usually related to uncertainty, high variability, low analyzability of BPs, negatively influencing the performance [28]. The complexity studies in BPM originate from software complexity. Thus, getting inspiration from software engineering, cognitive science, and graph theory, [29] discuss how analogous metrics can be applied to BP models. Likewise, [30] analyze software complexity metrics and extend the latter to suggest metrics for measuring BP model complexity. [31] evaluate existing theoretical thresholds for complexity metrics with real-life BP models and classify the models accordingly. Further, [32] investigate BP models complexity for implementing the Robotic Process Automation software. A row of studies has been devoted to managing and reducing the complexity of BP models [33, 34] and performing systematic reviews [35].

Apart from that, PM offers another useful toolset to analyze BP complexity using event logs. For example, [36] presents a metric to calculate the log-based complexity of workflow patterns. [37] analyze the design and applicability of event log complexity metrics to facilitate the effectiveness of the process discovery in case management. [38] measure data-center workflows complexity using PM. [39] suggest pre-processing event logs to reduce the discovered BP model complexity. Finally, a recent study reviews state-of-the-art event log complexity metrics to determine the relationship between the event log and the resulting BP model [2].

To conclude, the major focus on measuring complexity in BPM lies in BP models and event logs, whereas textual data consideration is limited to BP discovery and modeling support. This observation has been confirmed in our previous work on complexity management in organizations [7] and on assessing BP complexity based on textual data [8]. The latter study has also allowed us to demonstrate a noticeable impact of textual data on BP complexity.

2.3 Textual Complexity

Naturally taking its origin in linguistics, textual complexity is closely related to text comprehension, understandability, and readability [40]. Despite inherent differences between readability and understandability metrics [41], both aim to measure those aspects, making some texts easier to read than others [42]. In general, readability is defined as the ease of understanding or comprehending words and sentences caused by the writing style [43, 44]. Hence, the readability metrics commonly consider a number of words, sentences, letters, vowels, syllables, and monosyllabic vs. polysyllabic words [40]. Another approach more focused on text understandability is the Cloze test, which is also a popular exercise used in teaching a foreign language. Hereby, certain words or phrases are removed from the text, and the participants are asked to fill in the missing parts [45].

These metrics and textual complexity analyses are widely applied not only in Pedagogy [46, 47] and Linguistics [48, 49] but also in Finance [50], Legal [51, 52], and Politics [53]. Accordingly, complex texts have been associated with the attempt to manipulate certain contents. For example, by purposely adopting a textual complexity, underperforming organizations tend to conceal the content of their corporate narratives, i.e., the so-called obfuscation hypothesis [54]. Similarly, complex contract texts can impede a sufficient understanding of the contractual clauses for the reader (for example, buyer), which can be beneficial for the other party (for example, seller) [51].

However, the complexity that can be derived from text is not only limited to the text readability, i.e., writing style, but also should address the complexity of the subject matter. In an organizational and BPM context, it could imply how far this complexity perceived by the reader affects further decision-making and the choice of the related BP activities. In our previous work [7], while performing a systematic literature review (SLR), we observed that such approaches are missing in the related literature.

3 Research Method

Motivated by the shortcomings mentioned in the introduction and related work sections, we aim to develop a framework to analyze BP complexity considering both EL and TD perspectives. We extend the latter with our elaborations guided by the organizational context specificity.

To accomplish our research goal, we use various research methods: literature review, text analytics- and PM-based data analysis performed in the two case studies, including interviews and expert knowledge.

Specifically, we conducted a literature review to achieve several objectives. In [7], we followed an SLR process to identify the types of complexities in organizations and their shortcomings in terms of BP and textual complexities. We used the selected works supported by a backward and forward search to conceptualize EL complexity [2] and TD complexity [8]. To include the decision-making component into the latter and develop an understanding of the BP complexity concept in decision-making, we performed additional literature reviews in [10, 11, 55]. The obtained knowledge helped us to extend common textual complexity interpretations and metrics with decision-making component and suggest linguistic features for predicting TD complexity.

The linguistic features design was performed in the industrial ITSM CHM case study process of an international telecom provider. CHM deals with the processing of so-called Requests for Change (RfCs), IT tickets issued to perform changes in IT products and services of organizations [56]. Such requests either enter via e-mail or are directly recorded by the CHM workers in the IT ticketing system in a textual form, for example, based on the phone call with the customer. The requests range from the simple addition of a user to complex software upgrades. To support a data-driven IT ticket assignment and processing, the CHM department indicated its interest in investigating the complexity of incoming textual requests. Hence, we used the works selected in the SLR and, in total, 32782 IT ticket textual descriptions registered in a period of 2015-2019 to conceptualize and evaluate a set of linguistic features. As a result, two vocabularies (Decision-Making Logic taxonomy[1] and Business Sentiment lexicon[2]), including their development process, and the three groups of linguistic features, namely taxonomy-based [10], sentiment-based [11], and stylistic patterns-based[3] [9], were suggested. Then, in our further work [12], the best-performing features were selected. The team of 13 CHM department workers was involved in refining the features, defining IT ticket complexity, labeling IT tickets, and investigating the feasibility of using event logs for complexity analysis (EL complexity).

Whereas the focus of the first case study was TD complexity conceptualization (see [8] for details), the second case study of an ITSM IT ticket process of an academic institution aimed at investigating the potential of the TD and EL complexities used together in one framework. It is planned to be achieved by analyzing (i) correlations between TD and EL complexities and (ii) significant differences analysis in

[1] See our Github project page for taxonomy and further details
[2] See our Github project page for lexicon and further details
[3] See our Github project page for stylistic patterns and further details

correlations. Based on these analyses, recommendations and insights for process improvement will be derived. The correlation analysis part has already been completed and will be presented in Section 4. The significant differences analysis and deriving recommendations and insights based on the two analyses are currently work-in-progress. Selected preliminary results will be mentioned in Section 5.

4 Conceptual Framework

In this paper, we propose a framework for BP analysis from a complexity perspective in which we combine textual data and event logs. The inputs that the framework takes and its phases are shown in Figure 1. The framework uses three inputs: a complexity scale, textual data, and event log. In the first phase, these inputs are employed, and TD and EL complexities are calculated. Then, in the second phase, the relation between these complexities is analyzed. Based on the analysis, recommendations and insights are derived. These two phases are elaborated in the following subsections.

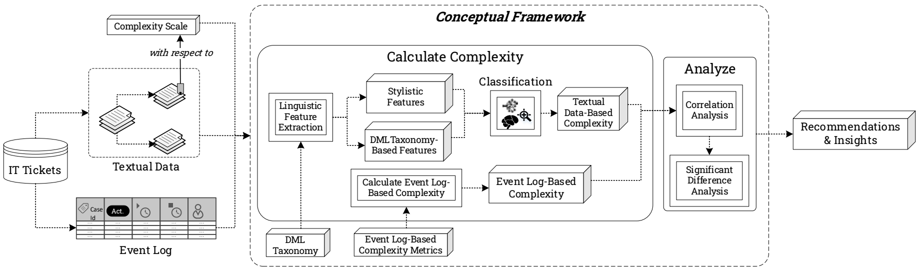

Figure 1. Business Process Complexity Analysis Framework

4.1 Phase 1: Calculate Complexity

The first phase of the framework is devoted to the calculation of TD and EL complexities. Textual data used and generated in BPs are often either unlabeled or contain few labeled points, as labeling is a time-consuming and costly endeavor for organizations. Hence, in our framework, a set of linguistic features are extracted from the given labeled textual data and transferred to a classification pipeline for calculating TD complexity. In particular, we extract the features that are investigated in [8] and determined as the most significant for prediction in [12], namely taxonomy-based and stylistic features. To extract these features, we focus on the following parts-of-speech (PoS) distribution: nouns, verbs, adjectives, and adverbs. The main reason is that they transmit essential information about business context, decision-making, and style in textual data. We use all words as the search space for extracting stylistic features. For taxonomy-based features, we extract PoS distributions based on a given Decision-Making Logic (DML) taxonomy developed with experts. A DML taxonomy consists of the most important words extracted from a given text and their DML levels, each denoting complexity of the subject matter for making a decision. The stylistic features denoting the easiness to understand the subject matter constitute an extensible part of the framework. For more details on the feature sets, we refer to [12].

Within the classification pipeline, semi-supervised learning techniques are employed for enriching unlabeled textual data. Thus, the problem that the framework tackles for calculating TD complexity is reduced to a typical prediction problem. Accordingly, training and test data sets are created from the combined unlabeled and labeled textual data. Then, commonly used prediction modeling techniques are trained, and the best-performing prediction model is selected. This model is run on the unlabeled textual data, and TD complexity for IT tickets is calculated.

As presented in Figure 1, an event log and a set of EL complexity metrics [2] serve as inputs in addition to the mentioned complexity scale. These metrics are applied to the given event log, a value for each metric is computed. Afterward, the computed values are mapped to the points in the complexity scale. With a majority voting, an EL complexity for each IT ticket is determined.

4.2 Phase 2: Analyze

In the second phase of the framework, we focus on how TD and EL complexities are related. To find out the connection between TD and EL complexities, we perform a correlation analysis. With this, we aim to identify in what way TD complexity affects EL complexity. Mainly, strong-positive, strong-negative, and no correlations between these two complexities can be used as indicators to mitigate BP complexity in organizations. For example, the value of textual data in predicting the execution complexity of BPs can be further investigated. In doing so, we go beyond the use of textual data in PM, which is typically limited to BP descriptions or labels in BP models. Moreover, such correlations can help organizations to discover what data type and attributes should be in focus when identifying root causes for BP complexity. Further, we perform a significant difference analysis [57] on the event log to identify the reasons for changes in correlations, i.e., root causes. For this, we use typical IT ticket attributes, for example, category and channel. Based on the two analyses, recommendations and insights are provided to organizations, motivating them to redesign and improve the processes.

5 Findings and Discussion

The IT department, hereafter Org-IT, of an academic institution in the Netherlands seeks ways to reduce the complexity of service requests. All requests within the institution are sent to Org-IT via various *channels*: phone, e-mail, IT ticketing system, and online chat. To handle incoming requests, Org-IT performs a Service Request Management (SRM) process. The primary input to that process is textual data coming through those channels. Hence, the case study setting is highly relevant for applying our framework. For the case study, we worked together with five experts in Org-IT who coordinate ticket resolution teams and have substantial knowledge of tickets. They provided us the required inputs for the framework, namely textual data, event log of tickets, and a complexity scale consisting of *low*, *medium*, and *high* values. The textual data contain 4982 tickets handled between Jan 2019 and May 2021. Among these tickets, 134 are randomly selected and labeled based on the defined three-point complexity scale by the same experts. The given event log consists of 37K events carried out while processing these tickets. To check the changes in two complexities over time, the provided textual data and event log are split into two parts: *before* (Jan 2019-Feb 2020) and *during* (Mar 2020-May 2021) the Covid-19 pandemic.

Phase 1: Calculate Complexity. TD complexity is calculated by means of classification. To do so, we enriched the unlabeled textual data using the commonly used semi-supervised learning techniques. Specifically, Pseudo-Labeling, Self-Training, Label Spreading, and Label Propagation from the Python scikit-learn library are employed. Then, training and test data sets are created from the combined labeled and enriched unlabeled textual data. The split is done using the 80 to 20 % rule of thumb. Common prediction modeling techniques such as tree-based algorithms, Logistic Regression, and Support Vector Machines are trained in ten-fold cross-validation. The best-performing prediction model is selected and run on the unlabeled data. Thus, TD complexity is calculated. Similarly, using a set of EL complexity metrics [2] and the given event log, we computed EL complexity for the tickets.

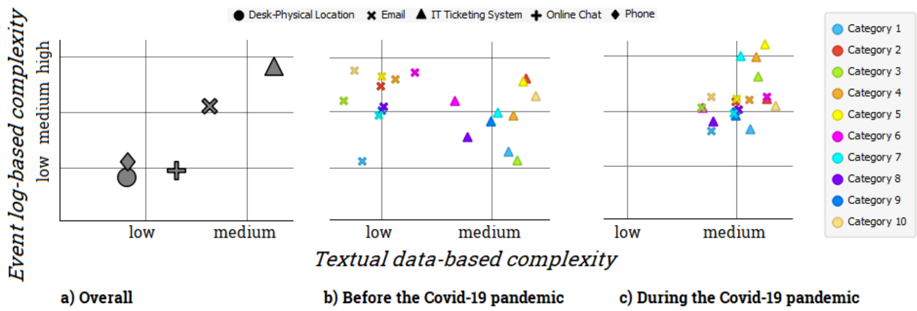

Figure 2. Correlation analysis of TD and EL complexities

Phase 2: Analyze. We analyze the correlation between TD and EL complexities considering the *channels* tickets are sent. We chose Spearman's correlation [23] for correlation analysis for two reasons. First, the complexity scale contains categorical and ranked points. Second, the distribution of data is not close to a normal distribution. We obtained 0.96 as the correlation value, indicating that the two complexities have a strong positive correlation (see Figure 2). Notably, both complexities increase when the *channel* is less interactive. In other words, less verbal communication while collecting the required information raises complexity. Our analysis addresses three important challenges of IT ticket processing in Org-IT: ticket *categorization*, work *assignment*, and *prioritization*. IT tickets are handled by resolution teams based on a *category* identified using textual data describing tickets. Accordingly, we aim to check the effects of TD complexity on the *categorization* activities in the event log. Based on the figure fragments b and c in Figure 2 (categories are enumerated due to privacy restrictions) showing the correlation between the two complexities for the two time periods, we made the following main observations:

O1: As presented in the figures, the TD complexity of tickets coming via the *e-mail channel* increased from low to medium *during* the Covid-19 pandemic. To investigate the reasons, we looked into the EL complexity of these tickets. Although it remained the same, within the significant difference analysis, we identified that the median duration for handling these tickets was doubled. It indicates that textual data might influence the execution time, the fact requiring further study.

O2: We observed that in the three ticket *categories* (Category 3, 4, and 5), the EL complexity of tickets registered via the *IT ticketing system channel* increases to high despite that the TD complexity does not change. To identify the reasons, we analyzed the significant differences in the event logs of these tickets. We found that the *assigned resolution teams* were changed frequently. The interviewed experts interpreted this as a "ping-pong" behavior between teams. It often happens when textual data is not clear and comprehensive to identify the resolution team correctly. Hence, end-users need to be contacted to obtain the required information.

O3: Further, we analyzed the tickets coming via less interactive *channels*: e-mail and IT ticketing system. We checked the significant differences in the event logs of the tickets that have low and medium TD complexities. In both before and during the pandemic, we detected that changing the priority of one and the same ticket is more frequent in the case of medium TD complexity. We infer that TD complexity can negatively influence accurate *prioritization*.

Based on the strong positive correlations shown in Figure 2, we conclude that TD complexity affects EL complexity. Moreover, we observed that increasing TD complexity could cause longer execution time (O1) and less accurate ticket categorization (O2) and prioritization (O3).

To sum up, in the beginning, we have suggested that TD complexity can potentially influence the EL, i.e., actual BP execution, complexity, which has been confirmed in our observations. Hence, the organizations relying on textual data as an input to their processes can benefit from these data to predict BP execution complexity. In case of scarce no correlations, changes over time can be used to search for, analyze, and address root causes. For example, if TD complexity is not changing, then root causes, such as bottlenecks, rework, and "ping-pong" behavior, should be analyzed in the EL complexity. This way, organizations can capture the BP redesign opportunities more comprehensively. Suppose EL and TD complexities are high. In that case, organizations can work on improving both (i) TD complexity by real-time text suggestions and chatbots for customers to better collect the input information and (ii) EL complexity by automating or reordering specific BP steps.

6 Conclusion

In this paper, we presented a conceptual framework to analyze BP complexity in ITSM. We focused on identifying the relation between textual data and event logs in terms of complexity. Linguistic features are used to calculate the TD complexity of IT tickets.

For EL complexity, the framework employs EL complexity metrics. The relation between these two complexities is studied by means of correlation analysis. Early results of the significant difference analysis aimed to identify the reasons for the changes in EL complexity are presented. In future work, we will consider other techniques to better capture the overlap between two complexities, such as clustering. In addition, we will focus on other processes in ITSM and incorporate conversations in BP executions into textual data analysis. Another future work direction we would like to pursue is comparing our findings with the ongoing process mining initiative in the case study organization.

References

1. Dumas, M., La Rosa, M., Mendling, J., Reijers, H.A.: Fundamentals of Business Process Management. Springer (2013).
2. Augusto, A., Mendling, J., Vidgof, M., Wurm, B.: The Connection between Process Complexity of Event Sequences and Models discovered by Process Mining. Submitted to Information Sciences. (2021).
3. van der Aalst, W., van der Aalst, W.: Data Science in Action. In: Process Mining. pp. 3–23. Springer Berlin Heidelberg (2016). https://doi.org/10.1007/978-3-662-49851-4_1.
4. van der Aa, H., Carmona, J., Leopold, H., Mendling, J., Padró, L.: Challenges and Opportunities of Applying Natural Language Processing in Business Process Management. In: The 27th International Conference on Computational Linguistics. pp. 2791–2801. Association for Computational Linguistics, Santa Fe (2018).
5. Kobayashi, V.B., Mol, S.T., Berkers, H.A., Kismihók, G., Den Hartog, D.N.: Text Mining in Organizational Research. Organizational research methods. 21, 733–765 (2018). https://doi.org/10.1177/1094428117722619.
6. Leopold, H.: Natural Language in Business Process Models. Springer, Cham (2013). https://doi.org/10.1007/978-3-319-04175-9.
7. Revina, A., Aksu, Ü., Meister, V.G.: Method to Address Complexity in Organizations Based on a Comprehensive Overview. Information 2021, Vol. 12, Page 423. 12, 423 (2021). https://doi.org/10.3390/INFO12100423.
8. 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. ahead-of-print, (2021). https://doi.org/10.1108/BPMJ-04-2021-0217.
9. Rizun, N., Revina, A., Meister, V.: Discovery of Stylistic Patterns in Business Process Textual Descriptions: IT Ticket Case. In: Soliman, K.S. (ed.) Innovation Management and Education Excellence through Vision 2020, International Business Information Management Association (IBIMA-2019). pp. 1–10. , Granada (2019).
10. Rizun, N., Revina, A., Meister, V.: Method of Decision-Making Logic Discovery in the Business Process Textual Data. In: Abramowicz, W. and Corchuelo, R. (eds.) Proceedings of 22nd International Conference on Business Information Systems. pp. 70–84. Springer (2019).
11. Rizun, N., Revina, A.: Business Sentiment Analysis. Concept and Method for Perceived

Anticipated Effort Identification. In: Siarheyeva, A., Barry, C., Lang, M., Linger, H., and Schneider, C. (eds.) Information Systems Development: Information Systems Beyond 2020 (ISD2019 Proceedings). pp. 1–12. AIS eLibrary (2019).
12. Revina, A., Buza, K., Meister, V.G.: IT Ticket Classification: The Simpler, the Better. IEEE Access. 8, 193380–193395 (2020). https://doi.org/10.1109/ACCESS.2020.3032840.
13. Muller, O., Junglas, I., Debortoli, S., Brocke, J. vom: Using Text Analytics to Derive Customer Service Management Benefits from Unstructured Data. MIS Quarterly Executive. 15, (2016).
14. Jurafsky, D., Martin, J.H.: Speech and Language Processing. Pearson (2009).
15. De Almeida Bordignon, A.C., Dani, V.S., Thom, L.H., Fantinato, M., Silva, T.S., Ferreira, R.C.B.: Natural language processing in business process identification and modeling: A systematic literature review. ACM International Conference Proceeding Series. 191–198 (2018). https://doi.org/10.1145/3229345.3229373.
16. Dijkman, R., Dumas, M., Van Dongen, B., Krik, R., Mendling, J.: Similarity of business process models: Metrics and evaluation. Information Systems. 36, 498–516 (2011). https://doi.org/10.1016/j.is.2010.09.006.
17. Weidlich, M., Dijkman, R., Mendling, J.: The ICoP framework: Identification of correspondences between process models. In: Pernici, B. (ed.) Advanced Information Systems Engineering. CAiSE 2010. Lecture Notes in Computer Science. pp. 483–498. Springer, Berlin (2010). https://doi.org/10.1007/978-3-642-13094-6_37.
18. La Rosa, M., Dumas, M., Uba, R., Dijkman, R.: Business process model merging: An approach to business process consolidation. ACM Transactions on Software Engineering and Methodology. 22, 1–42 (2013). https://doi.org/10.1145/2430545.2430547.
19. 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. Software and Systems Modeling. 18, 1179–1194 (2019).
20. Thomas, O., Fellmann, M.: Semantic Process Modeling – Design and Implementation of an Ontology-based Representation of Business Processes. Business & Information Systems Engineering. 1, 438–451 (2011).
21. Pittke, F., Leopold, H., Mendling, J.: Automatic Detection and Resolution of Lexical Ambiguity in Process Models. IEEE Transactions on Software Engineering. 41, 526–544 (2015). https://doi.org/10.1109/TSE.2015.2396895.
22. Friedrich, F., Mendling, J., Puhlmann, F.: Process model generation from natural language text. In: Mouratidis, H. and Rolland, C. (eds.) Advanced Information Systems Engineering. CAiSE 2011. Lecture Notes in Computer Science. pp. 482–496. Springer, Berlin (2011).
23. Leopold, H., Mendling, J., Polyvyanyy, A.: Supporting process model validation through natural language generation. IEEE Transactions on Software Engineering. 40, 818–840 (2014).
24. Leopold, H., Pittke, F., Mendling, J.: Towards measuring process model granularity via natural language analysis. In: Business Process Management Workshops - BPM 2013 International Workshops. pp. 417–429. , Beijing, China (2013).
25. Leopold, H., Eid-Sabbagh, R.H., Mendling, J., Azevedo, L.G., Baião, F.A.: Detection

of naming convention violations in process models for different languages. Decision Support Systems. 56, 310–325 (2013). https://doi.org/10.1016/j.dss.2013.06.014.
26. van der Aa, H., Leopold, H., Reijers, H.A.: Comparing textual descriptions to process models – The automatic detection of inconsistencies. Information Systems. 64, 447–460 (2017). https://doi.org/10.1016/j.is.2016.07.010.
27. Hornung, T., Koschmider, A., Oberweis, A.: 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 (2007).
28. Schäfermeyer, M., Rosenkranz, C., Holten, R.: The Impact of Business Process Complexity on Business Process Standardization. Business & Information Systems Engineering. 4, 261–270 (2012). https://doi.org/10.1007/s12599-012-0224-6.
29. Cardoso, J., Mendling, J., Neumann, G., Reijers, H.: A Discourse on Complexity of Process Models. In: Business Process Management Workshops. pp. 117–128. Springer, Vienna (2006). https://doi.org/10.1007/11837862_13.
30. Gruhn, V., Laue, R.: Approaches for business process model complexity metrics. Technologies for Business Information Systems. 13–24 (2007). https://doi.org/10.1007/1-4020-5634-6_2.
31. Boomsma, R.: An evaluation of thresholds for business process model metrics , (2017).
32. Qvick, T.: Understanding business process complexity for robotic process automation, (2019).
33. La Rosa, M., ter Hofstede, A.H.M., Wohed, P., Reijers, H.A., Mendling, J., van der Aalst, W.M.P.: Managing Process Model Complexity via Concrete Syntax Modifications. IEEE Transactions on Industrial Informatics. 7, 255–265 (2011). https://doi.org/10.1109/TII.2011.2124467.
34. La Rosa, M., Wohed, P., Mendling, J., ter Hofstede, A.H.M., Reijers, H.A., van der Aalst, W.M.P.: Managing Process Model Complexity Via Abstract Syntax Modifications. IEEE Transactions on Industrial Informatics. 7, 614–629 (2011). https://doi.org/10.1109/TII.2011.2166795.
35. Polančič, G., Cegnar, B.: Complexity metrics for process models – A systematic literature review. Computer Standards and Interfaces. 51, 104–117 (2017). https://doi.org/10.1016/j.csi.2016.12.003.
36. Cardoso, J.: Business Process Quality Metrics: Log-Based Complexity of Workflow Patterns. In: On the Move to Meaningful Internet Systems 2007: CoopIS, DOA, ODBASE, GADA, and IS. pp. 427–434. Springer , Vilamoura (2007). https://doi.org/10.1007/978-3-540-76848-7_30.
37. Benner-Wickner, M., Book, M., Bruckmann, T., Gruhn, V.: Examining Case Management Demand Using Event Log Complexity Metrics. In: 2014 IEEE 18th International Enterprise Distributed Object Computing Conference Workshops and Demonstrations. pp. 108–115. IEEE, Ulm (2014). https://doi.org/10.1109/EDOCW.2014.25.
38. Fernández-Cerero, D., Varela-Vaca, Á.J., Fernández-Montes, A., Gómez-López, M.T., Alvárez-Bermejo, J.A.: Measuring data-centre workflows complexity through process mining: the Google cluster case. The Journal of Supercomputing 2019 76:4. 76, 2449–2478 (2019). https://doi.org/10.1007/S11227-019-02996-2.
39. Richetti, P.H.P., Baião, F.A., Santoro, F.M.: Declarative Process Mining: Reducing

Discovered Models Complexity by Pre-Processing Event Logs. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). 8659 LNCS, 400–407 (2014). https://doi.org/10.1007/978-3-319-10172-9_28.

40. Gervasi, V., Ambriola, V.: Quantitative assessment of textual complexity. In: Barbaresi, L.M. (ed.) Complexity in Language and Text. pp. 197–228. PLUS (2003).
41. Smith, M., Taffler, R.: Readability and Understandability: Different Measures of the Textual Complexity of Accounting Narrative. Accounting, Auditing & Accountability Journal. 5, 84–98 (1992). https://doi.org/10.1108/09513579210019549.
42. DuBay, W.H.: The Principles of Readability. ERIC Clearinghouse (2004).
43. Hargis, G., Carey, M., Hernandez, A.K., Hughes, P., Longo, D., Rouiller, S., Wilde, E.: Developing Quality Technical Information: A Handbook for Writers and Editors. IBM Press, Indianapolis, Indiana, USA (2004).
44. Klare, G.R.: The measurement of readability. Iowa State University Press, Ames (1963).
45. Taylor, W.L.: “Cloze Procedure”: A New Tool for Measuring Readability: Journalism Quarterly. 30, 415–433 (1953). https://doi.org/10.1177/107769905303000401.
46. Bunch, G.C., Walqui, A., Pearson, P.D.: Complex Text and New Common Standards in the United States: Pedagogical Implications for English Learners. TESOL Quarterly. 48, 533–559 (2014). https://doi.org/10.1002/TESQ.175.
47. Schade-Eckert, L.: Protecting Pedagogical Choice: Theory, Graphic Novels, and Textual Complexity. Language Arts Journal of Michigan. 29, 40–43 (2013). https://doi.org/10.9707/2168-149X.1984.
48. Biber, D.: On the complexity of discourse complexity: A multidimensional analysis. Discourse Processes. 15, 133–163 (1992). https://doi.org/10.1080/01638539209544806.
49. Loukina, A., Yoon, S.-Y., Sakano, J., Wei, Y., Sheehan, K.M.: Textual complexity as a predictor of difficulty of listening items in language proficiency tests. In: Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers. pp. 3245–3253. , Osaka (2016).
50. Hrasky, S., Mason, C., Wills, D.: The textual complexity of annual report narratives: A comparison of high- and low-performance companies. New Zealand Journal of Applied Business Research. 7, 31–45 (2009).
51. Rameezdeen, R., Rodrigo, A.: Textual complexity of standard conditions used in the construction industry. Australasian Journal of Construction Economics and Building. 13, 1–12 (2013). https://doi.org/10.5130/ajceb.v13i1.3046.
52. Richards, G., Fisher, R., Van Staden, C.: Readability and Thematic Manipulation in Corporate Communications: A Multi-Disclosure Investigation. In: Business and Law: Conference Contributions. University of Canterbury. Department of Accounting and Information Systems (2015).
53. Benoit, K., Munger, K., Spirling, A.: Measuring and Explaining Political Sophistication through Textual Complexity. American Journal of Political Science. 63, 491–508 (2019). https://doi.org/10.1111/AJPS.12423.
54. Smith, M., Jamil, A., Chik Johari, Y., Ahmar Ahmad, S.: The chairman’s statement in Malaysian companies: A test of the obfuscation hypothesis. Asian Review of Accounting. 14, 49–65 (2006).

https://doi.org/10.1108/13217340610729464/FULL/PDF.

55. Rizun, N., Shmelova, T.: Decision-Making models of the human-operator as an element of the socio-technical systems. In: Strategic Imperatives and Core Competencies in the Era of Robotics and Artificial Intelligence (2016). https://doi.org/10.4018/978-1-5225-1656-9.ch009.
56. Axelos: ITIL® Service Transition. TSO, London (2011).
57. Bolt, A., De Leoni, M., Van Der Aalst, W.M.P.: A Visual Approach to Spot Statistically-Significant Differences in Event Logs Based on Process Metrics. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). 9694, 151–166 (2016). https://doi.org/10.1007/978-3-319-39696-5_10.