Utrecht University

Master Thesis

The Organizational Thermometer: Incorporating Feedback from Process Participants in Business Process Models

Author: Benjamin Kleppe (6865887)

1st supervisor:	Prof. dr. ir. H.A. Reijers
Daily supervisor:	Ing. J.M. Bakker (BPM Consult)
2nd supervisor:	Dr. I.M. Beerepoot

A thesis submitted in fulfillment of the requirements for the University Utrecht Master of Science degree in Business Informatics

July 19, 2023

Abstract

Business Process Management (BPM) stands as a key methodology for optimizing organizational operations and efficiency. However, the feedback of process participants is often not sufficiently integrated, leading to unrealized opportunities for process optimization. This research aims to bridge this gap by designing and developing an innovative tool that analyzes and visualizes process participants' feedback onto a BPMN model, fostering a bottom-up process redesign approach. The tool leverages a state-of-the-art Natural Language Processing (NLP) model, ChatGPT-3.5 Turbo, to map the feedback of process participants to the corresponding BPMN activities. The results are visualized as a heat map overlay on the BPMN model, providing a clear and concise representation of the process participants' insights.

The tool was evaluated in a real-world business process with 13 process participants and three process managers, resulting in 64 usable feedback messages. The results highlight the tool's effectiveness in collecting, processing, analyzing, and integrating the insights of process participants in the organizational knowledge-base. Moreover, the tool's ease of use, efficiency, generalizability, and operationality were evaluated. The findings indicate that the insights obtained by the tool are novel and lead to actionable improvement ideas.

The research contributes an anonymized dataset of the feedback and the corresponding BPMN model, along with the Python code of the tool for future studies. The findings demonstrate the potential of the application of AI models within the realms of BPM. Future research could further examine the application of similar models, how to further increase the scalability of the tool, and the psychological benefits of process participant feedback. Despite the limitations, this research sets a foundation for a more efficient, data-driven, and bottom-up approach to BPM.

Acknowledgements

First and foremost, I would like to express my gratitude to my supervisor, Prof. dr. ir. H.A. Reijers, for his support, guidance, and infectious enthusiasm throughout this research project. His expertise and mentoring improved both the outcome of this research project and how I would handle a project of such magnitude.

Moreover, I would like to thank my second supervisor, Dr. I.M. Beerepoot, for her constructive feedback, the interesting talks during our digital meetings, and her enthusiasm that sparked my interest in the field of BPM.

A special note goes to the partner organization, BPM Consult, for all the interesting opinions, discussions, and help during this project. I would like to thank Ing. J.M. Bakker for providing me with the opportunity to conduct my research within your organization. Moreover, I am grateful to my colleagues, whose ongoing encouragement have been a source of motivation and inspiration.

Furthermore, I would like to thank the participating organization (CorpX), all the participants, and managers involved in this study. This research would not have been possible without their input, cooperation, and decisiveness.

Last but certainly not least, I am incredibly grateful to my wife, family, and friends for your faith, encouragement, and well-timed food deliveries during late-night study sessions.

Your support made this journey more enjoyable than I had anticipated.

Contents

Chapter 1

Introduction

In today's dynamic and competitive business environment, organizations constantly strive to improve their processes to maintain efficiency, responsiveness, and adaptability. Business Process Management (BPM) is a discipline that focuses on organizing business processes efficiently, by utilizing a systematic approach to analyze, design, execute, monitor, and continuously improve business processes within an organization [14].

In the realm of BPM, process models serve as a visual and analytical representation of an organization's business processes, forming an important part of the organizational knowledge-base. Yet, while process analysis plays a crucial role in identifying areas of improvement, traditional methods often fail to capture the valuable insights of process participants because BPM has traditionally been an expert-driven field [1, 5, 10, 15, 36, 40]. As a result, organizations are unable to leverage the knowledge and expertise of process participants regarding their day-to-day business activities.

This research aims to develop a tool that leverages the power of Natural Language Processing (NLP) techniques to collect, process, and analyze feedback from process participants regarding their daily performed business activities in an efficient manner. The tool, developed following the principles of design science [49], aims to extract valuable insights from the feedback with the overarching goal to integrate the obtained insights into the organizational knowledge-base by enriching the existing process models.

The integration of these insights enables organizations to initiate process analysis and redesign in a bottom-up manner. The tool also has the potential to improve the working day of process participants by allowing them to submit their feedback about their daily executed business activities, fostering a culture of continuous improvement.

The derived insights are visualized by plotting a heat map overlay on top of the corresponding process model, in Business Process Management Notation (BPMN) format. The visualization provides a clear and intuitive representation of areas that require attention by highlighting the positive and negative sentiments towards business activities.

1.1 Problem Statement

Currently, process optimization is usually initialized using a top-down approach [15, 36, 40]. One of the key reasons for using a top-down approach is the lack of options for process participants to provide input on their daily executed business activities [15]. Furthermore, most participants do not possess the modelling skills to partake in the modelling process [15]. However, participants do possess vast amounts of practical information regarding these processes that could aid in identifying processes that need improvement [15, 36, 40].

Prilla and Nolte [33] stressed the importance of a more bottom-up, people-centric BPM as there is surprisingly little research on how to integrate stakeholders of processes properly. Their research suggested that a significant gap exists in the current understanding of how to integrate process stakeholders effectively. One of their key proposals is to provide suitable interactions for non-modelers to enable active user participation in modeling. This study aligns with their proposal, aiming to design a tool that collects textual, unstructured feedback from process participants, analyzes the data to extract insights, and integrates the obtained insights in the organizational knowledge-base by enriching the existing process models.

However, handling large amounts of unstructured, textual data introduces a considerable challenge [15, 36, 40]. The complexity of human language and the richness of its nuances, including emotional states, can pose significant challenges to conventional data analysis techniques [12]. Understanding the emotions of process participants is crucial as they can impact the efficiency of business processes and contribute to identifying improvement opportunities [2, 25, 26, 29, 43].

Herein lies the importance of Natural Language Processing (NLP). NLP is an emerging discipline that resolves around the analysis of unstructured, textual data [6]. The literature review indicates that there has been little research regarding the enrichment of BPMN models by invoking the feedback of process participants with the help of NLP techniques [7, 46, 48], despite various researchers suggesting examining this area [25, 26, 28, 29, 43]. In this research, NLP is utilized to analyze the feedback in order to extract the insights of the process participants regarding their daily executed business activities.

This research is a step into the direction of a more bottom-up, people-centric approach to process optimisation. This approach enables process participants to proactively contribute their feedback and insights to process management, potentially revealing bottlenecks and frustrations experienced by process participants. Consequently, these insights can point to improvement opportunities, enhancing the efficiency of both the process redesign phase and the process itself.

1.2 Thesis Outline

Chapter 2 presents the literature study that investigates the problem context and multiple NLP techniques. Chapter 3 presents the research plan. Chapter 4 presents the results. Chapter 5 discusses the main findings and limitations. Finally, Chapter 6 contains the conclusion.

Chapter 2

Literature Review

In this section, the scientific literature review is presented in order to provide a theoretical basis for this research. The review was conducted in two parts. First, a literature review was conducted to explore the existing work in order to position this research in the existing literature. This includes sections 2.1, 2.2, and 2.3. The literature research protocol for this part, which can be found in Appendix E, was followed to conduct the literature search in a systematic manner. Traditionally, BPM models are constructed in a top-down manner. Section 2.1 covers several papers regarding the top-down approach. Recently, however, research has focused on a more bottom-up approach that incorporates process participants into the modeling process. Social BPM is one such area of research, where social technology is involved in BPM. Section 2.2 contains related literature pertaining to this field. In this research, NLP is used to integrate stakeholder feedback at the process level. Section 2.3 covers the intersection of the fields of BPM and NLP. A second literature review was consulted in order to evaluate three candidate NLP techniques that could be used for this research. The selection of the techniques was based on expert interviews, which are presented in Chapter 3. Section 2.4 aims to provide knowledge on these techniques in order to make an informed decision on which technique should be used for this research. Appendix G presents the literature review protocol which was followed for this section. Finally, the conclusion is presented in Section 2.5.

2.1 Top-down Approach

Process models are often created top-down. Antunes et al. [4] propose a reassessment of the way business processes are modeled, emphasizing that BPM should not rely on underlying technical constraints. Instead, BPM should retain the human ability to handle ingenuity, variations, exceptions, and unique contexts. One of the flaws of the traditional, top-down approach is that the insights of the process participants, which are at the instance level, cannot be captured effectively [10]. These insights arise as process participants execute their daily business activities. This leads to a gap between the model and reality [1, 5, 10, 15]. Furthermore, traditional BPM fails to capture the improvement ideas of process participants, which is commonly referred to as *lost innovation* [15, 36, 40]. Prilla and Nolte [33] infer that BPM is mainly driven by experts, such as analysts, consultants, and managers, who have specialized knowledge and skills in designing, implementing, and managing business processes. Furthermore, the authors emphasize the importance of a more user-centric approach because process participants possess a great deal of knowledge about the business activities they encounter on a daily basis. The non-expert use of process models in practice and the notation of processes were analyzed in several experiments. The authors concluded, among other things, that further research should be conducted on appropriate interactions for non-experts in order to encapsulate their knowledge about processes.

An example of a proposed interaction for process participants in modeling is Processpedia [42]. The authors suggested using the Wikipedia model to engage stakeholders, since traditional top-down approaches do not consider the potential of process participants' tacit knowledge. In this approach, the process experts, business experts, developers, and process participants are encouraged to share their knowledge related to the process models. The results indicate that the proposed method is fast and flexible for feedback. However, the results imply that the organizational structure itself might need to be restructured to adapt the proposed method.

2.2 Social BPM

Limitations in the traditional BPM approach have triggered research that is inspired by other emerging trends. The combination of social software and BPM was first introduced in a Business Process Management and Social Software workshop (BPMS2'08), as part of the International Conference on Business Process Management in Milan [15]. Social software refers to software that allows users to communicate and share data (e.g., WhatsApp and e-mail). Social BPM combines Business Process Management with social networking applications by enabling more stakeholders to participate in the BPM cycle [9, 10, 15, 30, 40].

2.2.1 Principles

The main goal of social BPM is to improve business processes (e.g., accelerate decision-making, exchanging knowledge as well as information). Erol et al. [15] enumerated the main causes of both the model-reality gap and lost innovation. The first is the information pass-through threshold, meaning that in traditional BPM, passing on improvement ideas requires too much effort from the process owner and process participants. Examples include difficult participation, or restrictive change management. The other cause is the lack of information fusion, which implies that not all stakeholders are involved in process modeling. Examples include the use of formal modeling tools or a top-driven management approach.

Social BPM attempts to solve these problems by using multiple underlying principles. These include self-organization, egalitarianism, collective intelligence, and social production [30, 40]. First, self-organization implies that social BPM should not be regulatory. Moreover, the allocation of redesign processes should be done in a bottom-up manner. Second, egalitarianism refers to the equal rights and inclusion of all stakeholders. Third, collective intelligence assumes that the collective wisdom of multiple stakeholders is superior in creating process solutions compared to individual process experts. Finally, social production refers to the use of social software with the intention of bringing more stakeholders into the modeling process.

2.2.2 Challenges

Applying the five principles of social BPM introduces several challenges. Pflanzl and Vossen [30, 31] conducted two literature studies concerning the most important human-oriented challenges of social BPM. The authors examined the human related challenges that arise when the principles of social BPM are applied. By investigating related literature, the authors located seven main human-oriented challenges [30].

- Engaging external stakeholders As stakeholders usually have their own interests and motivations for participation, selecting the right stakeholders at the right time can be challenging.

- Motivating participation It is essential that all participants are motivated to share their meaningful contributions because the first step to a successful social BPM project is achieving a critical mass of participants.

- Training novice modelers Experts argue that novice modelers lack the competencies needed for modeling. Therefore, contributions from novice modelers are often not included even if they are motivated to participate.

- Software selection Providing novice modelers with the appropriate modeling software and languages can be challenging.

- Model quality It can become quite complex to ensure model quality when non-expert modelers are involved.

- Handling the information The involvement of many stakeholders can lead to an information overload. Techniques to counteract this range from annotating contributions, reviewing contributions for validity, and providing social information.

- Integrating semantics Due to differences in background, and thus differences in terminology, integrating the semantics can be hard. With respect to our research, it is especially interesting to look at involving external stakeholders and motivating participation.

To overcome these challenges, it is important to select key individuals who promote the use of social BPM to others and to encourage people to participate not just once, but continuously share their feedback [31]. Therefore, during this research project, key individuals will be selected for this task.

Vugec et al. [48] conducted a literature review regarding the implementation and practical use of social BPM. Their findings suggest that many research is conducted in the field of BPM. However, surprisingly little research is done in the field of social BPM. Furthermore, there is limited research regarding the software implementation and knowledge management possibilities of social BPM. Finally, the authors noticed that there is a lack of research in applying social BPM in professional organizations. This research has the goal of utilizing social software to integrate the insights of process participants into the organizational knowledge-base of a professional organization.

Bazan and Estevez [7] expand these findings by examining the contemporary state-of-the-art approaches, tools, and challenges of social BPM. Their findings illustrate that social software has a strong impact on how businesses conduct their operations, and that there is a lack of effective social BPM solutions for businesses. Moreover, there has been insufficient research on improving business processes using tacit and unstructured knowledge from the process participants. The three main challenges are structuring the knowledge from process participants' informal interactions, using this knowledge to improve processes, and developing automated tools for the previous challenges. In this research project, NLP techniques are utilized to automatically structure the unstructured, tacit knowledge of process participants with the overarching goal to enrich process to improve processes.

2.3 BPM and Natural Language Processing

Considerable research has been done on the process identification, discovery, and analysis phases of the BPM life cycle regarding the intersection of BPM and NLP [8]. For our research, however, it is important to focus on the process analysis and redesign phase as we explore how insights of process participants can be included in existing BPM models. Therefore, the following related work concentrates on the enrichment of these models with the help of NLP tools.

2.3.1 Model Enrichment

An intensive exchange between the fields of BPM and NLP has the potential to enrich BPM models [46]. Mustansir et al. [29] conducted research with respect to enriching BPM models with the help of NLP. The authors state that business process redesign is mainly carried out by specialized process analysts. Feedback of process participants is not incorporated in the redesign process since it takes too much time and effort to process all the feedback suggestions manually. By doing so, however, a lot of valuable contributions are lost. Therefore, the authors developed a method in order to incorporate the feedback suggestions of process participants. First of all, the feedback was classified into three categories: non-suggestion feedback, non-redesign suggestion feedback, and redesign suggestion feedback. Furthermore, three annotation guidelines are presented, consisting of three classification levels to automatically classify the feedback into these three categories. The first classification level distinguishes non-suggestive and suggestive feedback. The second level identifies redesign suggestions. Lastly, the third level identifies target process model elements to which the feedback relates to. Multiple experiments illustrated that it is possible to achieve high F1-scores with respect to the first two levels. However, the automatic identification of target business process elements did not yield great results as this task is still challenging. Sentiment analysis would be a good direction for further research, according to the authors.

2.3.2 Text-Model Alignment

The task of automatically mapping feedback onto business process model elements is called text-model alignment. Recently, Ahmed and Shahzad [2] examined the process of augmenting business process models by automatically mapping feedback of process participants onto these models. This is a challenging task as process participants use informal language to express themselves, while business process models are comprised of business terminology. In order to address this challenge, the authors designed mapping guidelines. These guidelines classify the feedback into three categories: syntactic, semantic, and business semantic correspondence. First of all, feedback should be classified as syntactic correspondence if the process element label is present in the feedback. Second, the feedback should be classified as semantic correspondence when the feedback contains words that are similar to the process element label. Third, feedback should be classified as business semantic correspondence if the words have a complex lexicon relationship with the process element label. The authors conducted 2880 experiments by a cross comparison by testing various word vectors, data balancing methods, and machine learning algorithms on six datasets. The results indicate that state-of-the-art machine learning models are capable of delivering adequate results when mapping feedback onto business process models. Therefore, future research should perform sentiment analysis on the feedback, and map it onto the process models in order to develop a perception of user feedback related to a process model.

2.3.3 Sentiment Analysis on Business Processes

As mentioned before, various research suggests performing sentiment analysis in order to augment business process management models. Mustansir et al. [28] conducted a study concerning sentiment analysis of feedback of process participants on business processes. The authors developed a NLP task for extracting sentiment across the four process performance dimensions of the devil's quadrangle [38]. The performance dimensions are cost, flexibility, time, and quality. In order to do so, the feedback regarding the business processes was annotated at three classification levels. The first-level classification attempts to classify whether the feedback is relevant to business processes or not. The second-level classification aims to classify the corresponding performance dimension of the feedback. The third-level classification performs sentiment analysis in order to classify whether the sentiment of the feedback is positive, neutral or negative. An example criterion for annotating a sentence as positive is that the sentence expresses a positive sentiment for all the aspect terms. Finally, the authors illustrated that a deep learning machine learning method outperformed traditional, supervised machine learning methods in the three-level classification of the feedback.

Recently, Lüftenegger and Softic [25, 26, 43] introduced a tool that allows various stakeholders to provide comments on activities within process models. Subsequently, the tool performs sentiment analysis to illustrate the stakeholders' sentiment opposed to the activity. However, using the tool still requires knowledge on process models as the tool allows comments on activities in process models. As mentioned before, the lack of modeling skills is one of the key barriers for process participants to participate in the process analysis and improvement phases of the BPM life cycle [15]. In this research project, however, we opt for an approach that enables process participants to provide feedback on their day-to-day business activities without the need to understand formal terminology or the ability to analyze process models.

2.4 Natural Language Processing Techniques

The final section literature review covers the exploration of various Natural Language Processing (NLP) techniques that could be used in this research to analyze feedback provided by process participants. This investigation aims to identify the most suitable techniques for integrating process participants into the process analysis and redesign phase effectively. Appendix G presents the protocol used for this study. The first technique that holds potential for this research is topic modeling, which enables the extraction of underlying topics from the collected feedback. Secondly, sentiment analysis has been identified as a relevant technique to integrate the feedback of process participants to enhance the process analysis and redesign phase. Lastly, prompt modeling, a relatively new and promising technique in the field of NLP, deserves exploration for its potential application in this research.

2.4.1 Topic Modeling

Topic modeling is a popular technique for analyzing large datasets [47]. The origins of topic modeling can be traced back to the field of information retrieval, where researchers were working on methods for automatically indexing and organizing large collections of text documents. In this research, topic modeling could be used to identify the underlying

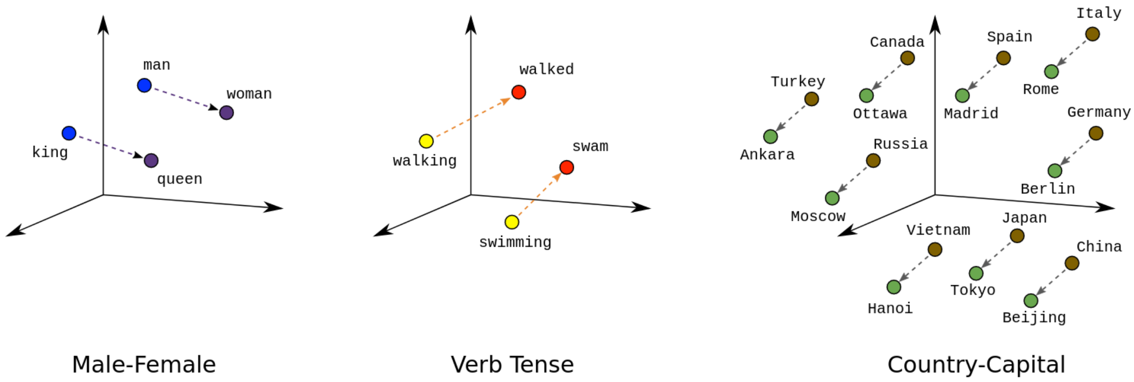


Figure 2.1: Relationships in a word embedding space, source: developers.google.com

topics that are present in the feedback from the process participants. This can help to gain a better understanding of the common issues, concerns, or suggestions that are being expressed by process participants, and to identify patterns or trends in the feedback. Topic modeling utilizes the principles of NLP to identify and extract the underlying topics or themes in a collection of documents or pieces of text [19, 47]. It is an unsupervised machine learning technique that automatically groups similar text documents into clusters based on the common topics or themes they share. This is one of the advantages as it does not require pre-labeled data. Furthermore, topic modeling algorithms are able to handle large datasets and are able to identify hidden patterns and trends [19]. Traditionally, this is done by looking for co-occurence in text documents [34]. Recently developed models use deep learning models and word-embedding algorithms to achieve better results [13, 34]. Word embeddings are a type of NLP technique that represents words as vectors of numbers, which can capture the meaning and context of the words in a high-dimensional space [21]. The basic idea behind word embeddings is to map each word to a high-dimensional vector in such a way that words with similar meanings are close to each other in the vector space. One of the key advantages of word embeddings is their ability to capture semantic and syntactic relationships between words. For example, the vector representation of "king" may be closer to "queen" than to "cat" or "car", reflecting the fact that "king" and "queen" are semantically related words. Word embeddings can also capture syntactic relationships, such as the fact that "running" is related to "run" in the same way that "swimming" is related to "swim". The topic models using word embeddings achieve significantly better

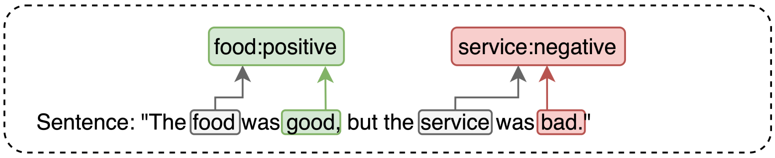


Figure 2.2: An example illustrating Aspect-Based Sentiment Analysis

results on short text messages [3, 34]. This is highly relevant as we expect to receive short feedback in this research, which could very realistically concise of short messages as well. However, topic models perform poorly on small datasets including less than 10.000 documents [3, 22]. Moreover, as topic modeling is an unsupervised approach, it is often difficult to understand the created topics. Therefore, it could be hard to integrate the obtained insights in the organizational knowledge-base.

2.4.2 Sentiment Analysis

Sentiment Analysis (SA) or Opinion Mining (OM) is the computational study of people's opinions, attitudes, and emotions toward an entity [27]. It entails extracting and classifying the sentiment of text as positive, negative, or neutral using NLP techniques and machine learning algorithms. There are three main classification levels in sentiment analysis: document-level, sentence-level, and aspect-level sentiment analysis [27, 37, 52]. The algorithms take a document, sentence, or aspect as input, and returns the sentiment. This can be either as a categorical value, or as a normalized value between -1 and 1, with -1 being negative and 1 being positive. Generally, we talk about document-level or sentence-level sentiment analysis when speaking about sentiment analysis [37]. With document-level sentiment analysis, the sentiment is calculated over a whole document. Sentence-level sentiment analysis, however, involves splitting the document into sentences and calculating the sentiment over a single sentence. More recently, aspect-based sentiment analysis (ABSA), or fine-grained sentiment analysis, is gaining more traction [52]. ABSA focuses on identifying and analyzing sentiment toward specific aspects or features of a product, service or entity. This involves identifying the different aspects or features discussed in a piece of text, such as a review or feedback, and then analyzing the sentiment toward each aspect or feature. For example, figure 2.2 illustrates ABSA on the sentence: "The food was good, but the service was bad". The sentence contains two aspects: "food" and "service". Document-level or sentence-level sentiment analysis would classify this sentence as neutral as it includes both positive and negative aspects. As illustrated in figure 2.2, ABSA implies that the sentiment regarding the food is positive, but the sentiment regarding the service is negative. Sentiment analysis, on all levels, can provide insights into how process participants perceive the daily performed business activities, and can aid to visualize the overall sentiment towards the processes and identify areas of improvement. However, the technique still has some limitations, such as detecting sarcasm, recognizing negations, handling context-specific language, language ambiguity, data quality, and requiring large amounts of data [37]. Moreover, accuracy of multilingual (models that can interpret more than one language) aspect-based sentiment analysis models is significantly lower than the accuracy of native English models [37]. This can be an issue for this research, as the collected feedback will be in Dutch.

2.4.3 Prompt Engineering

Prompt engineering is the process of designing and creating prompts for the use in NLP models, such as those for text classification, question answering, and language generation [39]. A prompt is a piece of text, or input, given to an AI system that utilizes a pre-trained language model (PLM). Based on the prompt, the language model (LM) generates a response, often a sentence or phrase. The technique has received considerable attention in recent years because it is a powerful tool for improving the performance of NLP models, reducing the need for extensive training data, and improving the interpretability of the output of a large LM. Providing a PLM with a few examples is called few-shot learning. This significantly increases the performance of a PLM [16, 18, 24, 39]. GPT-3, a PLM developed by OpenAI, was able to achieve state-of-the-art performance on a wide range of text-classification tasks without the need for fine-tuning, solely using few-shot prompts where solved task examples (shots) are given as input to the trained model [18]. Compared to current model analysis techniques, prompting is non-invasive, as it does not entail tuning specific parameters or necessitate direct examination of a model's representations [18, 41]. As such, prompting offers a baseline for what the model "knows", making it a more practical tool for analysis. Liu et al. [24], exemplify the importance of providing context to PLMs before assigning a task to the model. The authors illustrate that a PLM performs significantly better when it is provided with a few examples. The authors refer to this as prompting the PLM. While prompting the model, prompts are usually designed to provide the NLP model with specific information, guidance or context that can improve its accuracy, efficiency or effectiveness. Prompts usually consist of both giving instructions to the PLM, such as: "Perform sentiment analysis on the input messages", and providing

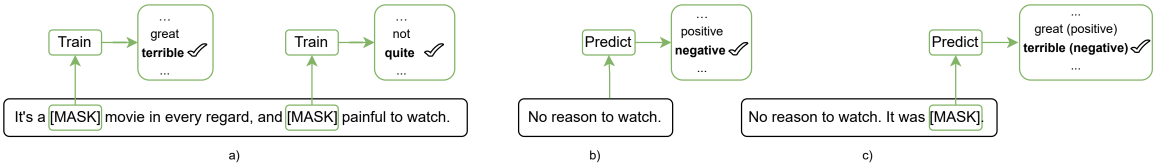


Figure 2.3: a) Masked learning, b) traditional fine-tuning, and c) prompt-based fine-tuning with examples, introduced by Gao et al. [16].

the model with some examples. Finally, the model can be used to make predictions on unseen data. Examples include sentiment analysis on document-, sentence-, and aspect-level, topic detection, text summarization, and translation [24, 41].

However, it can be difficult to create effective prompts to fine-tune a PLM. Therefore, Gao et al. [16] created a template for the creation of prompts to fine-tune models. PLMs are trained using a technique called masked learning [16, 24, 35, 41]. Masking a sentence implies leaving out one or more words. The model predicts which word should be in place. Figure 2.3a illustrates how a LM is trained using masked learning. During the training phase of a LM, the model predicts which word should be used instead of [MASK]. After doing so, the correct word is given to the model. Figure 2.3b exemplifies how models are traditionally fine-tuned. The model receives an example, it predicts the label, and finally it receives the correct answer. Gao et al. [16], however, propose using examples in a way that replicates their training by using a template that simulates masked learning. Figure 2.3c presents an example of the proposed template. The research illustrates that the usage of their proposed template either rivals or surpasses prompts created by experts. Furthermore, the results indicate that integrating a couple of demonstrations is useful for fine-tuning and that it improves few-shot performance. Li et al. [23] compared state-of-the-art methods for aspect-level sentiment analysis with (few-shot) prompt-based learning. The authors found that current techniques for ABSA require huge amounts of annotated data. Since annotated data is often very scarce in real-world implications, the authors investigated whether PLMs can be used to perform ABSA without the need for large annotated corpora. Similar to Gao et al. [16], the authors converted the problem into a masked learning task and created a template to generate prompts. The templates are constructed in the same way. The findings demonstrate that prompt-based methods outperform the current state-of-the-art models in all scenarios whenever the model was fine-tuned with fewer than 1000 examples.

Moreover, the prompt-based method performed significantly better when 16 examples were provided instead of 4. The results of both Li et al. [23] and Gao et al. [16] indicate that prompt-based methods are well suited for cases where limited (labeled) data is available.

2.5 Conclusion

The traditional, top-down manner in which BPM is implemented leads to drawbacks such as lost innovation and a gap between the model and reality [1, 5, 10, 15]. Social BPM aims to overcome these challenges by lowering the barrier to participate in modelling activities by using social software to involve more process participants in the modelling process [9, 10, 15, 30, 40]. One of the key challenging remains to analyze and process all the input received from process participants [15, 36, 40]. NLP offers a way to analyze large amounts of unstructured data. Furthermore, NLP can be used to enrich existing models with the help of feedback from process participants [29]. The existing literature recommends future research to perform sentiment analysis on process participants' feedback, and to include this in process models [2, 29]. Few recent studies experimented by performing sentiment analysis on comments given on activities in process models [25, 26, 43]. However, the concept of continuously processing and analyzing insights from process participants at the instance level, and incorporating them into the organizational knowledge-base, has not yet been explored [7, 48]. Three NLP techniques were examined for the analysis of textual data, namely, topic modeling, sentiment analysis, and prompt engineering. According to the literature, both sentiment analysis and topic modeling require large amounts of annotated data [3, 22]. Prompt engineering, however, achieves remarkable results on small data sets in various text classification tasks [16, 18, 23, 24].

Chapter 3

Research Methodology

The following chapter discusses the research problem, and methods. Moreover, the designed pipeline, data collection procedure, validation methods, and the threats to validity are presented.

3.1 Research Problem

Existing process analysis methods lack an effective tool to incorporate the feedback of process participants into the organizational knowledge-base in a time-efficient manner. This makes it unlikely for process participants to submit their feedback in a systematic manner. The ability to incorporate process participant feedback is critical for organizations as it enables them to leverage the valuable insights of process participants and improve the effectiveness of process analysis and process redesign. Therefore, the main objective of this research is to to develop a tool that collects, processes, and analyzes the feedback of process participants, and effectively incorporates the obtained insights into the organizational knowledge-base in a time-efficient manner. This research addresses the current gap in tools and methodologies, enabling organizations to initiate bottom-up process improvements. Using the design science template, introduced by Wieringa [49], the following technical research problem is constructed:

> *Improve process analysis by developing a tool that analyzes the process participants' feedback such that their insights are incorporated in the organizational knowledge-base in order to initiate process redesign in a time-efficient, bottom-up manner.*

Furthermore, several sub-questions are raised in order to come to an understanding of the research problem.

SQ1: What is the optimal social software for retrieving the process participants' feedback?

SQ2: What is the most suitable NLP technique for analyzing the process participants' feedback?

SQ3: How well does the developed tool aid to incorporate the process participants' insights in the organizational knowledge-base?

3.2 Design Science Methodology

Design science is a research approach focused on the design and investigation of Information Technology (IT) artifacts in a specific context [49]. In the context of this research, design science methodology is particularly suitable as it involves the design and evaluation of an IT artifact, namely a tool for processing and analyzing feedback from process participants regarding their daily executed business activities. The design science methodology follows a design cycle, as presented in Figure 3.1. According to this methodology, the artifact, in this case the tool that processes and analyzes feedback from process participants, interacts with the problem context, the process analysis phase. This interaction is called the treatment. The design cycle encompasses three tasks, which are presented below.

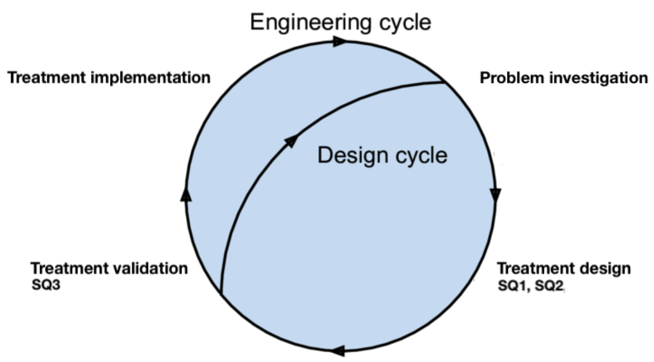


Figure 3.1: The design cycle, introduced by Wieringa [49].

Problem Investigation During the problem investigation, the problem context is examined. This includes allocating the stakeholders, their goals, and the research problem.

Step	Expert Interview	Literature Review	Case Study
Problem Investigation	✓	✓	✗
Treatment Design	✓	✓	✗
Treatment Validation	✗	✗	✓

Table 3.1: Research methods for each step.

Treatment Design The knowledge acquired during the previous step, the problem investigation, is implemented in the design of the artifact. The result is the pipeline presented in Section 3.4.

Treatment Validation The final phase of the design cycle consists of the treatment validation phase. In this research, the treatment is validated by conducting a single case study. The results are evaluated quantitatively and qualitatively. The quantitative evaluation measures the performance of the tool, whilst the qualitative evaluation assesses the ease of use, efficiency, generality, and operationality of the proposed treatment.

3.3 Research Methods

First of all, expert interviews were conducted to investigate which state-of-the-art NLP techniques can be used while designing the tool. Secondly, a literature review was conducted to provide a theoretical basis for the techniques mentioned by the experts. Third, the tool was validated by conducting a single case study. Table 3.1 presents the research methods for each step of the design cycle.

3.3.1 Expert Interviews

Subsequently, experts were consulted for the following topics:

1. The selection of a partner organization.

2. The selection of a activity, process, or department that can be monitored.

3. The selection of the social software technology, to retrieve the data from the process participants.

4. The selection of a suitable NLP technique, that can be used to analyze the data.

Selection of a Partner The research was conducted in collaboration with BPM Consult. BPM Consult is a consultancy firm, operating in the field of BPM. The company has a diverse client base, ranging from educational institutions to large housing corporations. To validate our treatment, it was essential to secure the participation of one of BPM Consult's clients. The consultants of BPM Consult have their own business relations with their clients. Therefore, multiple consultants were approached in order to ask them whether an case study could be performed at one of their clients. The consultants proposed this to their clients. Two clients expressed their interest in participating in the study. A meeting was arranged with both clients. During the meeting, the researcher presented the problem statement, research methods, and the estimated effort for the client to participate in the research. After meeting with both clients, one client immediately agreed to participate. The second client's decision-making process took several weeks. Due to the time constraints of this research project and the limited availability of the consultants, we decided to proceed with the first partner organization. The client wants to remain anonymously, so from now on the company is called CorpX. CorpX is a large Dutch company that operates in the food industry. The firm is sources its raw materials from around the globe, ensures high quality standards, and runs its operations in a sustainable manner. It is active in dozens of countries, with the Dutch division encompassing hundreds of employees.

Selection of a Process The interviews took place at the headquarters of CorpX, with three experts being physically present. The goal of the interview was to determine which activity, process, or department would be monitored for feedback. In total, three managers participated in the interview. The experts consisted of the Improvement manager, R&D manager, and the Quality manager. In advance, the experts were asked to propose an activity, process or department which they considered appropriate as the chances of a successful project are higher when the task to be performed is perceived as relevant by the management of the organization [14]. The experts preferred monitoring the product development process, which handles the development of new products. CorpX receives approximately 2000 requests for new product developments each year. Fig. 3.2 presents the process model in BPMN 2.0 format. The manager of the Quality department preferred to group the first three activities of their workflow, labeled as "Step 1", as those are very similar and difficult to separate. The process includes the following steps:

1. *Initialization*: The process starts when a customer submits a product development request.

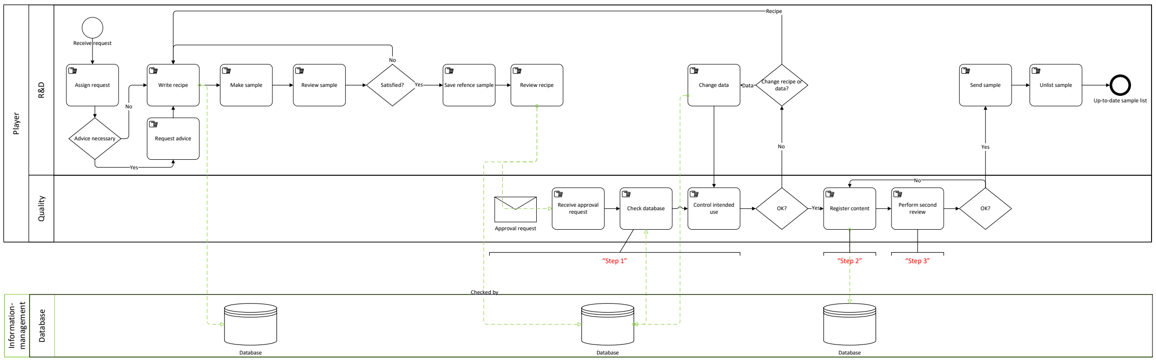


Figure 3.2: Product development process

2. *Assignment*: The request is assigned to a Research and Development (R&D) staff member, who assesses whether advice from a fellow staff member is needed.

3. *Development*: The assigned collaborator proceeds to develop a sample. This includes writing the recipe, making the sample, and reviewing the sample.

4. *Transmission*: If the collaborator is satisfied with the sample, a reference sample is stored and the recipe is sent to the Quality department.

5. *Quality Assessment*: The Quality Department receives the recipe and assesses whether it complies with the company's quality requirements. They also assess the intended use of the product. If discrepancies are found in the data or recipe and requirements, they are send back to the R&D department.

6. *Registration*: If the recipe meets all quality requirements, it is registered in CorpX's database. Additionally, a second review of the recipe is carried out.

7. *Approval*: If the recipe successfully passes the second review, the Quality department sends an approval to the R&D department. The R&D department then sends a sample to the customer.

8. *Completion*: After the sample is sent to the customer, it is removed from the sample list and the process is completed.

Retrieving the Feedback The interview took place after choosing which process would be monitored. The same experts participated in the interview. The goal of this interview was to choose the best social software for retrieving the feedback. During the interview, we opted to select a social software technology that is easy to use. This is one of the critical success factors within a social BPM project [14]. The process managers of the R&D and Quality departments stated that process participants in their departments spend a significant portion of their workday without a phone. Besides, the process participants enter data into a computer all day long. They also indicated that it would be ideal if process participants could enter data on a keyboard, instead of their phones. Hence, for retrieving the feedback, both managers recommended employing software that is easily accessible on a computer. The last expert, an improvement manager who works across multiple departments, confirmed the aforementioned conditions. Thus, we decided to employ a survey that can be completed utilizing natural text responses. Google Forms was chosen as the survey software, as it meets the previously mentioned conditions. Moreover, some process participants have morning shifts, while others have shifts during the afternoon. Sending an email daily would exclude either the process participants from the morning or the afternoon shift. Therefore, we agreed to send an email twice a day to solicit feedback, one in the morning and one in the afternoon. In addition, the online survey could be completed whenever process participants experience difficulties or feel the need to provide feedback on their performed activities.

Selection of a NLP Technique Two experts in the field of NLP were consulted to allocate candidate techniques for this research via email. The rationale behind consulting experts is that experts generally possess a vast amount of knowledge on the current state-of-the-art techniques that are currently used in practice. We briefly elaborated on the content of the research, and asked which techniques are potentially suited. Appendix H contains the initial email which was sent to the two experts: Dr. D.P. Nguyen and Prof. Dr. H. Leopold. Both experts proposed using prompt engineering, sentiment analysis, and creating our own language model. Additionally, one of the experts suggested using Topic Modeling and Aspect-Based Sentiment Analysis. After exchanging more information about the expected quantity of the data, the experts advised to use either topic modeling, prompt engineering, or (aspect-based) sentiment analysis, and to disregard creating our own language model as this requires more data than we expected to receive. The three candidate techniques are presented in Table 3.2. To select one of the three techniques, a literature study was conducted, as mentioned in Section 3.3.2.

Technique	Expert 1	Expert 2
Prompt engineering	✓	✓
(Aspect-based) sentiment analysis	✓	✓
Topic modeling	✓	✗

Table 3.2: Recommended NLP techniques by experts.

3.3.2 Literature Review

Two literature reviews were carried out, each serving its own purpose. The results can be found in Chapter 2. The literature reviews concerned a literature review concerning the background of the problem and a literature review investigating different NLP techniques at hand.

Background Literature Review First of all, a background literature review was conducted in order to gain a broader understanding of the problem at hand, the existing literature on this topic, and the current gap in the literature. The obtained information helped to identify what was already known about the topic, and it highlights the gap in current knowledge that is yet unexplored. Moreover, the literature review helped to formulate the research questions and objectives for this study. The literature review embeds the study in a broader context and underlines the importance of the research problem. The study was done in a systematic manner by following a pre-defined protocol. The literature protocol can be found in Appendix E. This resulted in Sections 2.1, 2.2, and 2.3.

Selection of a NLP Technique To ensure that the most appropriate NLP technique for the purpose of processing and analyzing feedback was selected, a literature review was conducted to further explore the NLP techniques obtained during the expert interviews in Section 3.3.1. Although NLP can be applied to many use cases, the techniques we specifically searched for were those that can process and incorporate feedback into the organizational knowledge base and that were mentioned during the expert interviews. The review aimed to gain a deeper understanding of the use cases, benefits and limitations of these techniques. Although there are numerous NLP techniques available, we focused on those that were most relevant to our needs. Several comparative studies were examined to identify the most appropriate techniques in the field of NLP. However, it is important to note that the main objective was not to list and compare all possible NLP techniques that exist. Instead, findings from previous studies were used as a theoretical basis for

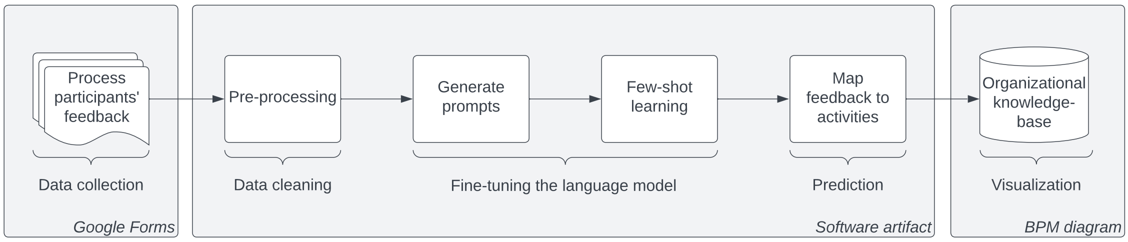


Figure 3.3: Pipeline for the software artifact

selecting the most appropriate technique. The literature protocol utilized for this study is presented in Appendix G. Section 2.4 presents the results. The results of the expert interviews, combined with the findings of the literature review, contributed to an informed decision on the most appropriate NLP technique to include in the tool's pipeline.

3.3.3 Single Case Study

After designing and building the tool, a single case study was conducted at CorpX [49] to empirically validate the effectiveness of the tool. Section 3.4 presents the pipeline of the tool. In this case study, the tool was applied to collect, process, and analyze feedback from process participants from two departments regarding their executed business activities in the product development process. The case study was designed to collect empirical evidence about how the tool performs within a real-world setting. Finally, the performance of the tool was evaluated both quantitatively and qualitatively, using the evaluation metrics presented in Section 3.6.

3.4 Pipeline

In the following section, the steps and the rationale behind the selection of each step of the pipeline are presented. Fig. 3.3 visualizes the pipeline. The code is available on https://github.com/BenjaminKleppe/TheOrganizationalThermometer.

3.4.1 Data Collection

As mentioned earlier, experts at CorpX were consulted for this study. They indicated a preference for using Google Forms as the social software to collect the feedback. Therefore, this was chosen as the medium to solicit feedback. Appendix I presents the Google Form used for this research. Additionally, Appendix J presents one of the reminders to solicit feedback. First, process participants are asked to which department they belong. Moreover, they are asked to provide a positive experience, negative experience, or both. Additionally, the process participants received an email twice a day by email as a reminder to submit feedback. This resulted in a total of 90 feedback messages. After the data collection period, the data was send to the process managers who manually labeled the feedback with the corresponding BPMN activities. This resulted in a valuable contribution: a dataset to test the performance of the proposed tool on its ability to map the feedback to the corresponding BPMN activities.

3.4.2 Software Artifact

The NLP technique was selected based on the obtained information by consulting experts and by expanding on these findings by conducting a literature review. The experts suggested three techniques for analyzing the collected data: topic modeling, sentiment analysis, and prompt engineering. The literature indicated that topic modeling requires at least 10.000 documents (feedback messages) to produce reliable and sound results [3, 22]. Therefore, topic modeling is not a valid technique as the data collection resulted in 90 documents. Document-level and sentence-level sentiment analysis are irrelevant choices as the feedback is already divided into positive and negative feedback by the process participants. Aspect-Based Sentiment Analysis (ABSA) is a relevant choice because it is a more fine-grained analysis as it provides the sentiment scores on all aspects within a document. However, as seen in the literature, ABSA models lack performance on non-English datasets while the language of the collected feedback in this research is Dutch. Prompt engineering can be utilized for a range of text classification problems [16, 18, 23, 24, 41] and has been shown to outperform traditional Aspect-Based Sentiment Analysis (ABSA) models on predicting the sentiment scores of datasets containing less than 1.000 documents [16, 18, 23, 24]. Therefore, the tool employs prompt engineering to leverage a pre-trained language model (PLM) in order to map the feedback to the corresponding BPMN activities. ChatGPT-3.5 Turbo, a PLM developed by OpenAI (March 2023), is used in this research as one of the experts recommended using this model.

Preprocessing First of all, the data was cleaned by removing all invalid values. Secondly, data was randomly split into two sets: the training dataset to fine-tune the model and the test dataset to test its performance. A recent study illustrated that the performance of prompt engineering is raised significantly when the model is provided with 16 examples [23]. Therefore, 16 messages were used for training the model. The 48 remaining messages were used to test the model. During the preprocessing phase, the software artifact generates prompts to fine-tuning the model with the training data and for validating the model with the test data. As seen in the literature, PLMs are pre-trained by predicting missing parts of a sentence, which is called masked learning [16, 24, 35, 41]. The ability of a PLM to classify text increases significantly when the model is fitted with a few examples (few-shot learning). To improve the performance of few-shot learning, the examples were reformatted in a format that mimics masked learning [16, 18, 24, 39]. Essentially, this entails allowing the language model to predict omitted parts of sentences. Therefore, the examples were reformatted to this format. The format imitates the templates introduced by Li et al. [23] and Gao et al. [16] by masking the parts we want to predict. For example, the feedback message: "It was difficult to order new ingredients" correlates to the BPMN activity "ordering ingredients". In this case, the tool would generate the following prompt: "It was difficult to order new ingredients. This message belongs to: [MASK]". The model predicts the masked phrase, which is the BPMN activity name. The correct output is: "ordering ingredients".

Fine-tuning During this step, the model was fine-tuned using the training data. Initially, a list containing all activity names from the BPMN model was provided as potential aspects, supplemented by a brief textual description of the process. Subsequently, the prompts, which were generated during the preprocessing step, were sequentially fed to the model. For each prompt, the model was tasked to predict the most likely BPMN activity name(s). After each prediction, the correct answer was given to the model in order to fine-tune the model for this specific task [16, 18, 23, 39]. This was done iteratively until the model has seen all the training data. Given that some feedback messages were mapped to multiple labels by one of the process managers, this task represents a multi-label classification problem [45].

Prediction The final step of the software artifact was to predict the most likely BPMN activity names for the unseen, test data. This resulted in two tables. The first table contains the total sentiment score for each activity of the BPMN model. Negative comments are counted as -1, while positive comments are counted as +1. The sentiment score for each BPMN activity is calculated by adding up all the corresponding sentiment scores. For example, if an activity received three negative and one positive comment, the sentiment score for this activity would be -2. The second table includes the each feedback message, the correct label(s), and the predicted label(s).

3.4.3 Visualization

Finally, the results were visualized as a heat map on top of the process model. The goal of the visualization was to provide a comprehensive view on the insights of process participants towards their daily executed business activities. To achieve this, the activities were be colored according to the number of positive and negative feedback messages regarding that activity. The heat map uses a color code to indicate the degree of positive (green) and negative (red) feedback on each activity. The intensity of the color varies based on the magnitude of the sentiment score, with highly polarized scores resulting in more intense colors (e.g., an activity with a sentiment score of -5 will appear prominently red, while an activity with a sentiment score of -1 shows a subtle orange tint). This highlights the activities that received multiple positive or negative feedback messages. The mapping of the feedback by the process managers is used for the visualization as they possess domain expertise and a deep understanding of the business process. The visualization is extended by an additional table, presenting the obtained feedback for the activities. This enables process managers to drill-down to the feedback of the issues that are visualized by the tool. Additionally, a word cloud is included for each activity, to quickly analyze the main findings. *By visualizing the sentiment of process participants in a clear and concise way, the tool should enable process managers to make data-driven decisions based on the insights of process participants to improve the process in a time-efficient manner.*

3.5 Data Collection

The following section presents information regarding the participants, data characteristics, number of responses, and ethical considerations. The data was collected over a period of 15 days using the data collection method presented in Section 3.4.1.

3.5.1 Participants

Feedback was collected from a total of 13 process participants at CorpX. This group of process participants, consisting of 10 women and three men, makes up the entire population of the key departments involved in processing process requests for product development at CorpX. All participants were asked to provide feedback on new product development

Department	Positive	Negative	Total
Quality	9	12	21
R&D	13	30	43
Total	22	42	64

Table 3.3: Distribution of feedback by department

process applications. Of the participants, eight were part of the Research and Development (R&D) department. This department writes the recipes and physically tests the requested products by making samples. The Quality Department checks the samples for compliance with quality requirements. This department consists of five process participants, four of whom work full-time and one part-time on Mondays, Thursdays and Fridays.

3.5.2 Data Characteristics

The collected data in this study is unstructured, as it was collected without predetermined response options, resulting in a diverse range of insights. In addition, participants were given the freedom to provide multiple answers. The dataset consists of textual descriptions of emotions faced by process participants in their daily business activities. It is important to recognize that this type of data involves subjectivity, as it captures individual perceptions and experiences, and these may differ among process participants.

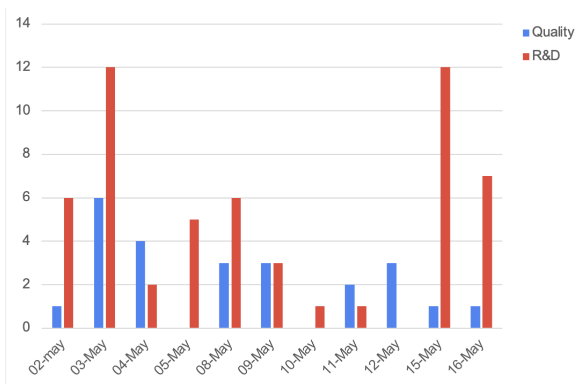


Figure 3.4: Distribution of feedback by day

3.5.3 Responses

In total, we collected 87 responses. However, out of the 87 responses collected, only 79 were considered usable for analysis due to the presence of values such as 'n/a'. The distribution of both departments is presented in Table 3.3. R&D contributed about 70% of the feedback, while the Quality department contributed about 30% of the feedback. The difference in the amount of feedback between the two departments may be due to several factors. First, R&D has more participants than the Quality department. In addition, R&D has a greater workload within the process. Fig 3.4 presents the amount of responses for each day. It should be noted that process managers were contacted on May 3, May 8 and May 15 to motivate process participants from their departments. The response rate increased after contacting the process managers. A total of 15 responses were excluded as they lacked sufficient information, which led to the process managers not associating them with any activities. Therefore, 64 responses were used in this study.

3.5.4 Gold Standard Dataset

After obtaining the responses, the process managers labeled each response with the corresponding BPMN activity names. The process manager of the R&D department mapped all responses to no more than one activity. The process manager of the Quality department, however, mapped some responses to multiple activities. This implies that the classification problem at hand is an multi-label classification problem as it is possible to assign multiple labels to a single responses.

3.5.5 Ethical Considerations

All process participants received and signed an informed consent form two weeks before participating in the study, which can be found in Appendix D. Moreover, all data was anonymized and no personal data was stored.

3.6 Validation

To ensure the validity and usability of the developed tool, it was validated with the help of two distinct approaches. The first approach comprises of a quantitative evaluation, measuring the accuracy, precision, recall, and F1-score. The second approach focuses on the tool's ease of use and efficiency from the perspective of both process participants and

Criteria	Process participants	Process managers
Ease of use	✓	✓
Efficiency	✓	✓
Generality	✗	✓
Operationality	✗	✓

Table 3.4: Evaluation matrix

process managers. Moreover, the generality and operationality will be reviewed from the perspective of three process managers.

3.6.1 Quantitative Evaluation

One of the most common methods to quantitatively evaluate a software tool is to calculate the accuracy, precision, recall, and F1-score of the tool [32]. However, accuracy is not a valid method for measuring the tools performance as the problem at hand is a multi-label classification problem [45]. Sorower [45] performed a literature survey regarding the evaluation metrics of multi-label algorithms. The authors found that the best way to evaluate an algorithm that solves multi-label classification problems is to measure the weighted precision, recall, and F1-score. Moreover, weighted average of these metrics seem to perform well on imbalanced datasets. This is an important feature as our dataset is heavily imbalanced. Therefore, the aforementioned weighted metrics are used to evaluate the tool. The tool is tested against the gold standard dataset, which was composed by the process managers as described in Section 3.5.4. Additionally, an error analysis is carried out in order to delve deeper into the results of the tool. The error analysis highlights common mistakes made by the tool, and which areas still need improvement. This approach allows for an objective evaluation of the tool's predictive capabilities and reliability.

3.6.2 Qualitative Evaluation

Validating a treatment implies verifying whether it contributes to stakeholder goals when implemented in the problem context. There are various models to validate information system design methods. Sonnenberg and Vom Brocke [44] proposed three principles to validate design methods. Moreover, the authors state that the evaluation criteria of an artifact is not independent from the type of an artifact. Therefore, a distinction is made between construct, model, method, and instantiation artifacts. In this research, we propose a method for involving the process participants' insights in the organizational knowledge-base. With respect to Sonnenberg and Vom Brocke [44], the evaluation criteria of a designed method are the ease of use, efficiency, generality, and operationality. Table 3.4 presents the evaluation criteria measured for process participants and process managers.

Ease of Use Ease of use refers to the amount of effort required to use the tool. This is measured by both process participants and process managers, as it is interesting to measure how much effort it takes to use the tool to submit feedback and to use the tool to analyze the feedback.

Efficiency Efficiency measures whether the tool improves the time taken to submit, process and analyze feedback from process participants on their daily performed business activities. Efficiency regarding the time required to submit feedback is measured among process participants, while efficiency regarding the time required to process and analyze feedback is measured among process managers.

Generality Generality measures whether the tool can be used to analyze different activities, processes or departments. This is measured by validating whether the instrument can be applied in another department using a survey amongst the process managers.

Operationality Finally, operationality measures whether the instrument can actually achieve its goal of incorporating the insights of process participants into the organizational knowledge base. This is validated by measuring whether the tool is able to integrate insights into the organizational knowledge-base and whether the insights gained by the tool are new to the process managers.

The evaluation statements we used for the tool in our study were derived from previous research conducted by Gonzalez-Lopez et al. [17]. Their evaluation statements were followed, modifying it as necessary to align with our specific needs and goals. The resulting survey is presented in Appendix K.

Chapter 4

Results

Section 3.4 presents the design and function of the tool which leverages Natural Language Processing (NLP) techniques to collect and analyze feedback from process participants. Data was collected using Google Forms to solicit feedback from process participants on their day-to-day experiences. The tool then mapped the feedback to the corresponding BPMN activities with the ultimate goal of integrating this feedback into the organizational knowledge-base. The training set comprised 16 feedback messages, while the test set consisted of 48 feedback messages. The output from this prediction process is a table providing a sentiment score for each activity of the process model. The following chapter presents the results obtained from the application of this tool within CorpX.

This chapter is further divided into sections that provide a detailed analysis of the tool's results. Section 4.1 presents a visualization of the tool's output, followed by Section 4.2 which delves into the tool's performance, discussing its evaluation metrics and analysing its errors. Finally, Section 4.3 provides a qualitative assessment of the results.

4.1 Visualization

Table 4.1 provides an overview of the feedback sentiments categorized by activity. The feedback mapping of the process managers is used for the visualization as they possess domain expertise and a deep understanding of the business processes. For each activity, the number of positive and negative feedback messages is presented, along with the total sentiment score. The sentiment scores are calculated by subtracting the number of negative feedback messages from the number of positive feedback messages. Negative scores indicate activities where negative feedback outweighed the positive, and vice versa for the

Activity	Positive	Negative	Sentiment
request advice	0	9	-9
review sample	2	0	+2
make sample	2	6	-4
send sample	0	3	-3
write recipe	5	1	+4
save reference sample	0	1	-1
step 1	2	8	-6
step 2	2	3	-1
step 3	4	4	0
assign request	1	2	-1

Table 4.1: Class distributions: number of positive feedback messages, negative feedback messages, and the corresponding sentiment score for each activity.

positive scores. Key observations include that the 'request advice' activity had the highest number of negative feedback messages, and therefore the lowest sentiment score (-9). In contrast, the 'write recipe' activity received both negative and positive feedback but had the highest positive sentiment score (+4). Moreover, 'step 3' received 4 positive and negative feedback messages, resulting in a neutral sentiment score (0).

The sentiment scores are visualized in Fig. 4.1. Here, activities are color-coded: green signifies activities with more positive feedback, while red indicates those receiving more negative feedback. This visualization allows for a quick, intuitive understanding of the

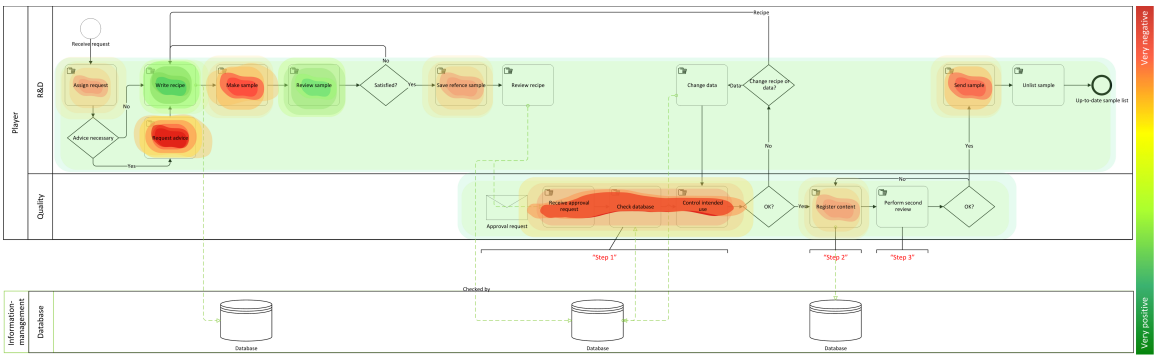


Figure 4.1: The organizational thermometer: a heatmap that visualizes the insights of process participants onto a BPMN diagram (presented in Fig. 3.2).

Choose elements below

Activity

- assign request
- make sample
- request advice
- review recipe
- review sample
- save reference sample
- send sample
- step 1
- step 1, step 2
- step 1, step 2, step 3
- step 2

Feedback for each activity and department

Activity	Department	Sentiment	Feedback message
step 1	QS	Positive	Swift response from R&D
assign reques	R&D	Positive	It was a quiet day for emails, so I had time and space for development.
request advic	R&D	Negative	It's very cumbersome when we're asked to provide a stick length format for something we haven't made sticks of (often the weight per stick is still unknown). If we're lucky, there's a sample jar, but you're certainly fiddling around for half an hour if they ask this question. It always irritates me.
request advic	R&D	Negative	The process was hindered by a bitter response from support
make sample	R&D	Positive	Weighing.
step 1	QS	Negative	In the future, it would be more practical to create a new a/b/c/d version for version 3 of a code to avoid confusion.
step 1	QS	Negative	I had to send a product back to R&D because it didn't meet the requested criteria
send sample	R&D	Negative	Support and Export storing up requests and all sending them on Friday.
review sampl	R&D	Positive	I tried out new flavor directions together and they were also very delicious.
send sample	R&D	Negative	A lot of requests where the shipping address is incomplete or incorrect, causing shipping delays.
make sample	R&D	Negative	Making a sample for the customer [NAME], where we always have to make 50 sample sticks.
step 3	QS	Positive	It was very nice that when checking a declaration by Quality department, I could easily find out that it was the second version. Everything else looked neat and I found the request email from R&D very funny.
step 3	QS	Positive	Second check approvals processed- Frames were correct and declarations were accurate.
write recipe	R&D	Positive	The use of the database being online again so we could write our recipes.
step 1	QS	Negative	I had to put back 2 items to R&D
save referenc	R&D	Negative	I was checking a new dry mix. For that, I ensured reference material in the basement. But after rifling through all the containers, I couldn't find anything.
step 2	QS	Positive	Multiple declarations made without any system issues.
step 3	QS	Negative	I was in the 2nd approval step and there was talk in the request email of just an administrative error. I was curious what the administrative error was because the impact could be quite large but would feel like a nuisance if I started it again. After all, the development had already been delayed.
make sample	R&D	Negative	Making samples; for the umpteenth time shortage of raw materials because people don't order this (on time) when it's almost gone.
step 1, step 2	QS	Negative	Checking specifications/creating declaration. Annoyance about formulation why a request has urgency; customer has been waiting for specs since early April while we are on variant 3 of the recipe and the delay is not on our end.
write recipe	R&D	Positive	The freedom we have to do our own developments.
request advic	R&D	Negative	Request with too low a volume received had to be sent back because of the low volume. Then back to us because it was approved by the head of Support.
step 3	QS	Negative	Second check approvals processed- Mixed up AB/AC version.
step 3	QS	Positive	No errors discovered when checking the specification (step 3).
make sample	R&D	Positive	The tasks I managed to complete.
assign reques	R&D	Negative	The lead-up to the urgent case; colleague indicated she didn't have time and couldn't make the development.
make sample	R&D	Negative	Raw material was out so we had to walk to the factory.
make sample	R&D	Negative	Raw material was out so I had to go to the factory to finish my mix.
request advic	R&D	Negative	It was not clearly stated in the request that it had to be [NAME] and therefore I have to adjust the recipe and resend the samples. Extra work and it takes me time.
step 1	QS	Negative	Approval of PR-code returned because [NAME] was included in the name but the recipe was not [NAME].
step 1	QS	Negative	Sent recipes back to R&D because the name mentioned ingredients that were not used as an ingredient.
make sample	R&D	Negative	Raw materials were out and had to go to the factory.
step 1, step 2,	QS	Positive	Did a number of approvals and encountered no issues.
request advic	R&D	Negative	Received incorrect info from PM regarding a development.
step 1, step 2,	QS	Negative	Many requests that fall below the minimum volume for an NPD but are still pushed through as standard.
write recipe	R&D	Positive	New project where I can put my creativity to use.
write recipe	R&D	Positive	The distribution of workload; I had enough time and possibility to get my work done for today.
request advic	R&D	Negative	Checking 300 recipes manually for raw materials is not efficient.
write recipe	R&D	Positive	I'm glad I was able to help a colleague so he has more breathing room.
write recipe	R&D	Negative	Colleagues' requests not completed on time, so there is now an urgent need to work.
send sample	R&D	Negative	As the daytaker, I received an email with a question from [COUNTRY] about where the samples are and whether they have already been shipped. The email did not mention which PR-code it was about and I can't really find out whether the samples have already been sent. Now it turns out there's a whole other story playing out w
step 1, step 2,	QS	Negative	Checking approval; recipe and PO-form are not in line and it's vague what's correct. So, I had to send an email with follow-up questions.
request advic	R&D	Negative	Colleague overlooked emails resulting in miscommunication; this landed on my plate to solve.
review sampl	R&D	Positive	The sample I made turned out better than expected.
make sample	R&D	Negative	Raw material that was out and not ordered.
request advic	R&D	Negative	Request not fully filled out.
request advic	R&D	Negative	A sample request of 10 KG mix is requested. That's almost impossible to do.
assign reques	R&D	Negative	Many requests with long ingredient lists; distribution of requests not equal for example: I received 7 in recent days and a colleague 1.
review recipe	R&D	Negatief	Upon reviewing a development, it turned out the correct attachment (application form) had not been added. When I tried to correct this, the correct application was not saved on the disk. The application could be found in the email, but the wrong code was mentioned in it. The purchase department makes the mistake of not sa
make sample	R&D	Negatief	Helped a colleague in distress because others were negligent.
request advic	R&D	Negatief	The unclearness of emails. Received an email on Friday with missing information which required me to consult the original source first, causing irritation because the email was vague.
request advic	R&D	Negatief	Some daily task moments were forgotten by a colleague, which burdened another.
step 1, step 2,	QS	Positief	No special incidents encountered along the way.
step 2	QS	Positief	Found no issues while reviewing declarations and the system.
step 2	QS	Positief	Could not find any issues while reviewing a declaration.
step 1, step 2,	QS	Negatief	Approval, in this case, there was version confusion with the PR-code the final product is based on. Resulting in confusion between AS and QS, the EP itself went back to the purchase department.
send sample	R&D	Negatief	It is quite difficult to have an urgent order delivered. We apparently do not really have a procedure for this. I have already spent an hour figuring out how it should be arranged.
write recipe	R&D	Positief	The moment we collaborated as a team when we had a number of urgent matters.
make sample	R&D	Positief	Help from colleagues.
write recipe	R&D	Positief	I was able to put all my recipes in the database without being called or disturbed by others. Now I've been able to process all my recipes.
review sampl	R&D	Negatief	I was bombarded with emails about the Friday inspection. Whether products can be squeezed in that were not even requested as a development.
request advic	R&D	Negatief	Dropped coffee over my desk, also communication with a [COUNTRY] customer ([NAME]) which is quite difficult.
assign reques	R&D	Negatief	Many requests that are submitted too late or with a short deadline. As a result, the process does not always go smoothly or things are overlooked.
step 1	QS	Negatief	Why are # raw materials used in new developments without notifying us that it is okay?

Figure 4.2: Drill-down of the feedback.

feedback sentiments across different activities. The specific details of the feedback are available in an accompanying table, which includes all feedback messages for each activity. Figure 4.2 presents a table, serving as a drill-down function, providing a detailed exploration of the sentiment scores.

This approach, however, may become impractical if the frequency of feedback messages dramatically increases, for instance, up to 1000 feedback messages for each activity. Manual analysis of such a vast amount of feedback using the drill-down function could become exceedingly tedious. Therefore, Figure 4.3 presents an additional solution, namely, an automatically generated word cloud that illustrates the most frequent words within the feedback that was associated with the BPMN activity: "Request advice". It is important to note that the efficacy of the solution is somewhat hindered by the limited number of messages in our dataset, thus potentially resulting in a less optimal solution. To illustrate, the word "stick" appears notably large in the word cloud, despite it was only mentioned in a single feedback message (three times). Hence, this underlines the importance of a larger dataset. It is plausible that the word cloud would only be included in larger datasets, eliminating the aforementioned problem. Furthermore, when the dataset would become

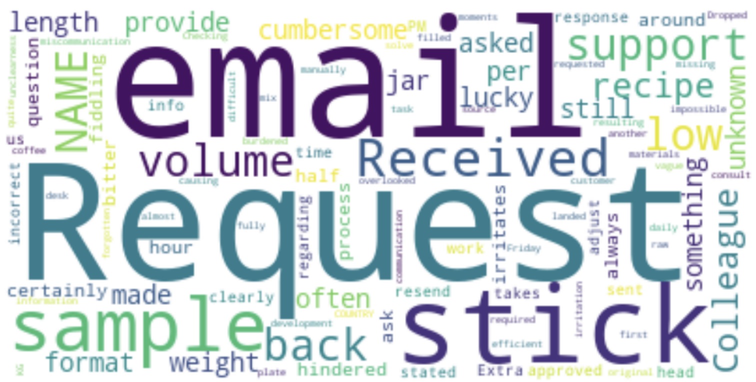


Figure 4.3: Word cloud of the feedback regarding the BPMN activity: "Request advice".

larger, advanced analytical approaches such as topic modeling could be considered for more detailed investigation of the feedback. Topic modeling allows for the discovery of the underlying topics and themes within large amounts of messages [19, 47] when there are over 10.000 messages [3, 22]. Therefore, as the volume of feedback messages increases, topic modeling could offer a more insightful analysis compared to a word cloud, plotting patterns and trends of the process participants' feedback in an efficient manner.

4.2 Performance

The performance of our tool is measured in three ways: precision, recall, and the F1-score. Precision is the proportion of correct predictions in comparison with all the predictions made by the tool. In this case, the average precision is 0.82, meaning that 82% of the activities predicted by our tool were correct. On the other hand, recall measures the proportion of actual activities that were correctly identified. The tool had an average recall of 0.64, implying that it identified 64% of the total correct activities. Finally, the F1-score is the harmonic mean of precision and recall, providing a balanced measure of the tool's performance. The average F1-score of the tool was 0.66. The precision is higher than the recall. This means that a high proportion of the predictions were correct, but that it has missed some activities. This implies that the model is conservative; it tends to make fewer mistakes at the cost of not always being able to identify the correct outcome. It should be noted that the feedback often includes little information, which can pose a challenge for the model's classification performance and can explain the lower recall. Moreover, with a total of 14 distinct classes to predict with some activities belonging to

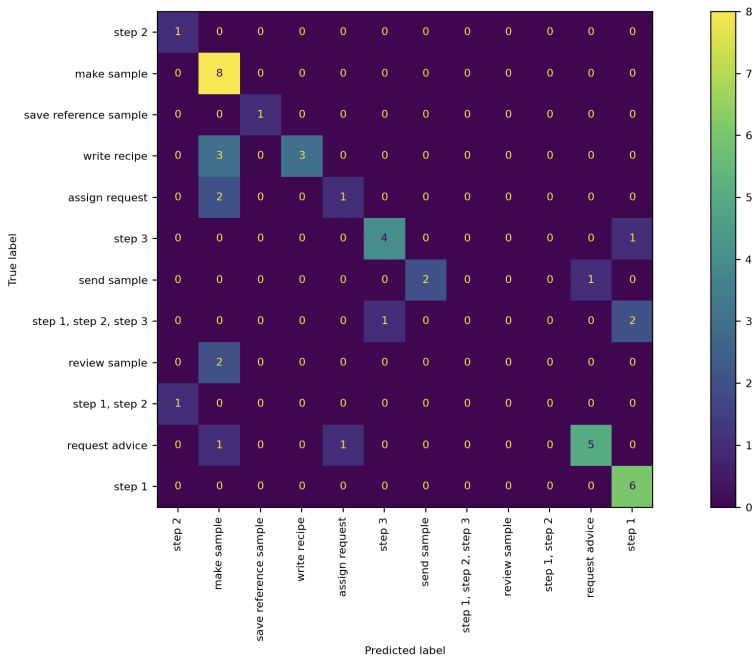


Figure 4.4: Confusion matrix of the true and predicted classes

multiple labels, the complexity of the task at hand is significantly increased. Despite these challenges, the model exhibits a respectable F1-score [45].

4.2.1 Error Analysis

The following subsection delves deeper into the model's performance by performing an error analysis. This analysis aims to examine the instances where the model failed to correctly classify the feedback to the corresponding BPMN element. The output of the model is presented in Appendix L. The incorrect classifications, including the feedback message, the corresponding activity, and the predicted activity, is presented in Appendix M. Fig. 4.4 presents the confusion matrix of the predicted and the true labels. In the confusion matrix, each row represents instances of an actual class, while each column represents instances of a predicted class. The values in the diagonal line of the confusion matrix correspond to correct predictions, the actual class matches the predicted class. The off-diagonal entries in the confusion matrix represent the incorrect classifications made by the model. The confusion matrix illustrates that the model tends to predict the class 'make sample' over any other class (a total of 16 times). However, only 50% of those predictions are correct. Five out of the eight false positives correspond to 'write recipe' and 'review sample'. These three activities - writing, creating and reviewing samples - are the core activities of the development process, and it is likely that feedback related to these steps is ambiguous. To illustrate, consider the feedback message: "The sample I made turned out better than expected". While the process manager attributed this to the activity: 'review sample', the tool categorized it under the activity: 'make sample'. Additionally, the model falls short in identifying feedback associated with multiple steps from the Quality department, despite four messages being labeled with two or three steps. The training dataset consisted of two examples with more than one activity.

The remaining errors seem to be due to the ambiguity or lack of information within a message. For example, consider the message, "Checking 300 recipes manually for raw materials is not efficient." This is classified as "request advise" by the process manager, whilst the tool classifies it as "review recipe". Moreover, the message: 'A sample request of 10 KG mix is requested. That is almost impossible to do.' was classified by the tool as 'make sample' while the process manager classified this as 'request advice'. It can be argued whether the correct activity can be derived from the limited context of the messages. For future research, it would be interesting to investigate whether increasing the size of the training dataset would increase the performance of the tool.

4.3 Qualitative Evaluation

Feedback was collected from process managers concerning the four evaluation metrics of the tool: ease of use, efficiency, generality, and operationality [44]. Moreover, feedback was collected from process participants regarding the ease of use and efficiency. Responses were collected using a 5-point Likert scale (1: totally disagree, 2: disagree, 3: neutral, 4: agree, 5: totally agree). Appendix K presents the statements and the results. In this section, the results are presented.

4.3.1 Process Managers

Table 4.2 presents the average ease of use, efficiency, generality, and operationality for each process manager. In the following subsection, the individual responses are evaluated as each manager has used the tool in a different context. Then, the results are summarized to present an overall evaluation.

Manager	**EoU**	**E**	**G**	**O**
R&D	5	5	5	5
Quality	3.5	3.5	4	3.5
Improvement	4.5	5	4	4
Average	**4.33**	**4.5**	**4.33**	**4.17**

Table 4.2: Average perceived ease of use, efficiency, generality, and operationality for each manager.

R&D Department The process manager of the R&D department manages one of the two participating departments of this study. The R&D department has the biggest workload of the monitored process and contributed 70% of all the feedback. The manager fully agreed with every statement (mean: 5/5). This indicates a high satisfaction with the tool and the process regarding all dimensions.

Quality Department The process manager of the Quality department provided mixed responses. The Quality department contributed to 30% of all feedback. The manager found the tool itself easy to use (4/5), but was neutral about the ease of use of the study as a whole (3/5). In addition, the manager indicated an improvement in the time and effort required to analyze the results (4/5), but that the time and effort required to analyze the feedback were not necessarily improved (3/5). Furthermore, the manager agreed that both the tool and the study could be modified to better understand feedback from other departments (mean: 4/5). Finally, the manager strongly agreed that the tool's insights were novel (5/5), but disagreed that the tool's insights lead to actionable improvement ideas (2/5).

Improvement Manager The improvement manager, who works in several departments, gave generally positive feedback. The manager perceived that the tool was easy to use (mean: 4.5/5). Moreover, the manager considered the tool efficient (mean: 5/5). In addition, the manager thought that both the survey and the tool could be used to understand feedback from other departments (4/5). The manager was neutral regarding the novelty of the tool's insights (3/5). Yet, the manager did fully endorsed that the tool's insights lead to actionable improvement ideas (5/5).

Department	EoU (Form)	EoU (Study)	Ef (Form)	Ef (Form)
R&D	5	5	4	4
R&D	5	4	5	3
R&D	5	4	4	5
R&D	5	3	3	4
R&D	3	3	4	4
Average R&D	*4.6*	*3.8*	*4*	*4*
Quality	5	4	5	3
Quality	4	4	2	2
Quality	2	3	3	2
Average Quality	*3.67*	*3.67*	*3.33*	*2.33*
Average overall	**4.25**	**3.75**	**3.75**	**3.38**

Table 4.3: Individual responses of the ease of use and efficiency of the form and the tool for each department.

Overall Evaluation To conclude, the managers found the tool easy to use (mean: 4.33/5), and perceived the entire study, encompassing feedback form, discussions, briefing and reminders, easy to follow (mean: 4.33/5). The time and mental effort required to analyze the results were reasonable (mean: 4.67/5), and the tool deemed to enhance the efficiency of feedback collection and analysis (mean: 4.33/5). Overall, managers agreed that both the tool and the study could be adapted to different departments (mean: 4.33/5). Two managers stated that the insights derived from using the tool were novel (mean: 4.33/5), and two managers stated that the tool's insights lead to actionable improvement ideas (mean: 4/5).

4.3.2 Process Participants

Table 4.3 presents the individual responses to the feedback form for each department. In the following subsection, the responses regarding the ease of use and the efficiency of both the tool and the study are evaluated for each department. Finally, the results are summarized.

R&D Department In total, five out of the eight process participants from the R&D department filled in the evaluation form. The participants generally showed a positive response towards both the tool and the study. The feedback form was easy to use (mean:

4.6/5), and the entire study was easy to follow (mean: 3.8/5). Additionally, feedback on the efficiency was also favorable. The time and effort needed to fill out the form was reasonable (mean: 4/5), and the participants experienced a reduction in the time and effort required to provide feedback (mean: 4/5).

Quality Department Three out of the five participants filled in the evaluation form. The results of the evaluation from the Quality department were mixed. The participants generally indicated that the feedback form and the study were easy to use and follow (mean: 3.67/5). However, one participant strongly agreed, one participant neither agreed nor disagreed, and one participant disagreed that the time and effort required to fill in the form was reasonable (mean: 3.33/5). Moreover, the participants generally disagreed that the tool decreases the required time and effort to give feedback by using the tool (mean: 2.33). One participant questioned the applicability of the feedback form for the Quality department, as they did not experience complications.

Overall Evaluation Overall, process participants found the feedback form relatively easy to use (mean: 4.25/5) and the overall study was generally perceived as easy to follow (mean: 3.75). The perceived time and effort to fill out the form (mean: 3.75/5) and the efficiency of soliciting feedback by using the tool (mean: 3.38/5) varied between the two departments. The R&D department experienced the time and effort needed to use the feedback form as reasonable and the tool made soliciting feedback more efficient. The Quality department, however, generally did not agree that the time required to fill in the feedback form was reasonable, and they did not experience any increases in efficiency.

Chapter 5

Discussion

The tool presented in this paper was designed to improve process analysis by developing a tool that analyzes the process participants' feedback such that their insights are incorporated in the organizational knowledge-base in order to initiate process redesign in a time-efficient, bottom-up manner.

5.1 Interpretation of the Results

> *"I found it especially difficult to write down positive things. Writing down irritations is much easier." - process participant R&D*

Ease of Use The process participants strongly agreed that the feedback form was easy to use. Most of the participants agreed that the study was easy to follow. Furthermore, the managers of the R&D and Quality departments strongly agreed that the visualizations presented by the tool were easy to understand. In particular, an improvement manager, despite not being directly involved in the single case study, found the visualizations easy to understand. This indicates that the design of the visualization is easy to use, even by managers who did not cooperate during the single case study. *Thus, the tool successfully meets the research aim to incorporate the insights of process participants into the organizational knowledge-base.*

> *"I found it difficult to see the usefulness of the feedback form for the Quality Department, considering little went wrong" - process participant Quality*

Efficiency The process participants of the R&D department strongly agreed that the time needed to complete the feedback form was reasonable. Additionally, the outcomes indicate that the tool improved the efficiency of gathering feedback within the R&D department. In contrary, most of the participants from the Quality department did not experience an increase in efficiency. Both the manager as multiple participants of the Quality department mentioned not encountering a lot of issues during the data collection period and that it was difficult to provide daily feedback. This can explain the mixed results as allocating time to provide feedback can be perceived as inefficient when there barely any issues arise. *The findings indicate that the efficiency benefits of the tool is dependent on the level of problems encountered by process participants and managers.* It would be interesting for future research to investigate the specific conditions or scenarios in which the feedback tool has positive effects on the efficiency of soliciting feedback, such as the process participants' perception of the necessity for improving the process or the frequency of feedback collection.

Both the improvement manager and the manager of the R&D department reported experiencing an increased efficiency in collecting and analyzing feedback by using the tool. *This confirms the tool's time-saving capability for the process managers and indicates that the research aim of incorporating the insights of process participants into the organizational knowledge-base in a time-efficient manner is achieved.*

Generality The process managers generally agreed on the the tool's capability to adapt to different departments. *This implies that the tool is generalizable, which highlights its potential applicability in other use-cases.* However, further research should investigate how the tool performs in a different context.

> *"It is very interesting to see that there are many concerns, that are quite easy to fix, but you normally never hear about!" - manager R&D*

Operationality The improvement manager was neutral regarding the novelty of the obtained insights by applying the tool. This may be due to the fact that this manager is not involved in the day-to-day activities of the department, and perhaps lacks the context to make the judgement whether the insights are novel or not. The managers of the R&D and Quality department strongly agreed that the tool's insights are novel. Moreover, the manager of the R&D department was very surprised about the feedback, and mentioned that a lot of concerns were easy to solve. *The novelty of the obtained insights underlines the ability of the tool to improve process analysis.*

The manager of the Quality department disagreed that the tool led to actionable insights. This can be explained by the fact that some process participants from the Quality Department communicated that they did not encounter significant difficulties in their work, resulting in fewer feedback and, consequently, fewer actionable improvement ideas. However, both the improvement manager and the manager of the R&D department strongly agreed that the insights obtained by applying the tool result in actionable improvement ideas. *This validates the tool's ability to improves process analysis and its potential to initiate process redesign in a bottom-up manner, if there are improvement opportunities within a process.*

5.2 Implications

This research presents a novel tool that aims to integrate the feedback of process participants into the organizational knowledge-base by visualizing their sentiment towards their daily performed business activities onto a BPMN model in the form of a heat map. The established research method effectively draws upon state-of-the-art Natural Language Processing (NLP) techniques, using the ChatGPT-3.5 Turbo model, to process, analyze, and classify the insights of process participants. The results highlight the potential of utilizing NLP techniques to enhance process analysis and initiate bottom-up process redesign, thus providing both a design and the proof of concept. Organizations could further optimize and implement such tools to innovate and improve their business processes by enabling more stakeholders to participate in the BPM life cycle [9, 10, 15, 30, 40].

Secondly, there has been insufficient research on improving business processes using unstructured, tacit knowledge [7, 48]. The single case study resulted in a unique, anonymized dataset consisting of feedback pertaining to an actual business process. The absence of specific guidelines for feedback led to a variety of responses, illustrating how process participants provide feedback on their experiences and concerns under these conditions. Therefore, both this study and the resulting dataset could serve as a foundation for developing methods and tools to capture, analyze, and utilize such unstructured feedback effectively. This is particularly relevant in today's business environment, where agile and adaptive processes are increasingly dependent on feedback from participants [50].

Thirdly, the tool, including the pipeline along with its corresponding Python code, is included in this research. This contribution not only ensures reproducibility of the study but also serves as a basis for further refinement in future research. The results of the tool act as a baseline for future research as it is the first study to demonstrate the capability of large language models to map feedback to the corresponding BPMN activities.

Lastly, the tool gives a voice to process participants by enabling them to express their insights, feelings, experiences, and concerns regarding their day-to-day activities. By addressing their issues, the tool could potentially empower process participants while also improving their daily workflow in a data-driven way. Empowering process participants generally results in more innovation, creativity, motivation and instills shared values to promote and atmosphere for learning and accomplishment [11].

5.3 Limitations

First of all, the study was performed on a specific process within a single organization. Only 13 process participants and three process managers were involved. In total, this resulted in 64 usable feedback messages. This limits the generalizability of the results. To enhance the generalizability, a process manager, who is not operationally involved within the two departments, was involved in the evaluation.

Secondly, the final visualization was created manually due to the absence of available tools that are capable of generating an appealing visualization automatically.

Thirdly, the feedback was manually mapped to corresponding BPMN activities by the process managers. Despite their domain knowledge, there might be a degree of subjectivity involved in this mapping process. This, in turn, could have influenced the results.

Fourthly, while the primary objective of this research was the design and development of the tool, it should be noted that the tool has not undergone extensive optimization.

Finally, the researcher had to contact the process managers several times to motivate the process participants to deliver more feedback messages as the volume was decreasing over time. This need for manual intervention to sustain adequate feedback decreases the time-saving capabilities of the tool.

5.4 Threats to Validity

The performed research includes potential threats to validity. This may impact the interpretation and generalization of the findings. In this section, the potential threats to validity are discussed to enhance the credibility of the research by looking into four categories: construct validity, internal validity, external validity, and reliability [51].

5.4.1 Construct Validity

Construct validity concerns the degree to which the tests or measures accurately reflect the concepts they are supposed to measure [51]. In the context of this research, there are several threats to construct validity. First of all, a pre-trained language model, ChatGPT-3.5 Turbo, is used for mapping the feedback messages to the corresponding BPMN activities. The performance of such a model is heavily dependent on the quality of the prompts. To increase the consistency and mitigate the risks associated with prompt variations, this research leveraged a systematic approach to prompt engineering. A prompt template, proposed by Gao et al. [16] and Li et al. [23], that is specifically designed to fine-tune pre-trained language models for text classification tasks was utilized.

Secondly, the process managers manually mapped the feedback to the corresponding BPMN activities. This might have caused subjectivity and bias in the correct labels, making the approach prune for errors and misinterpretations, posing a threat to the construct validity. Nonetheless, this approach was chosen as the process managers posses more domain knowledge on the process than the researcher does.

Thirdly, a survey was employed to evaluate the tool qualitatively. Inherent for such a method, however, is the varying individual interpretation of the statements. This poses a threat to the construct validity. To reduce the subjectivity as much as possible, clear definitions of the statements were ensured. Additionally, participants were given the opportunity to provide a textual comment which was carefully examined to enhance the comprehensiveness of the evaluation.

5.4.2 Internal Validity

Internal validity concerns whether the results of your study are truly due to the factors you're investigating [51]. The qualitative evaluation was measured by conducting a survey consisting of Likert-scale statements. However, the insights derived from these questions are susceptible to multiple interpretations, introducing a potential threat to the internal validity. To address this concern, the process participants were asked for additional comments on their statements in order to reduce the ambiguity of the results.

5.4.3 External Validity

External validity concerns the generalizability of the findings of the study [51]. As the data was collected for a process within one organization, the results might not be applicable to other organizations or departments. This poses a threat to the external validity.

Moreover, initially, the tool was going to be validated by the two process managers that are responsible for the process that was monitored. To mitigate the aforementioned risks, an improvement manager, who was not involved throughout the single case study, was involved in the validation of the tool.

5.4.4 Reliability

Reliability refers to the reproducibility and the consistency of the results [51]. A key component of this study is the use of prompt engineering. The performance of this technique relies heavily on how the prompts are constructed. With the goal of enhancing both the results and the reproducibility of this study, the prompts were generated using a template proposed by Li et al. [23] and Gao et al. [16]. Furthermore, all the relevant materials, including the anonymized feedback, anonymized process model, and the code are included to enhance the reliability and transperancy of the research.

5.5 Future Research

This research illustrates the potential of utilizing pre-trained language models within the realm of Business Process Management (BPM) by demonstrating a practical use-case of state-of-the-art AI models to improve a real-world business process. The developed model serves as a baseline for future studies in this area.

Future research can focus on enhancing the generalizability of this approach by examining its applicability in different contexts. Additionally, it would be valuable to compare the performance of various NLP techniques in mapping feedback to corresponding process model activities.

Furthermore, the scalability of the tool warrants investigation. Currently, the drill-down function is employed for manual feedback analysis after allocating bottlenecks, which limits the scalability of the tool. Future research should explore alternative analysis techniques, such as topic modeling, to efficiently analyze large volumes of feedback.

In addition to the technical aspects, it would be intriguing to explore the psychological benefits of providing process participants with a means to express their feelings regarding their daily business activities.

Moreover, future research could focus on the design and development of an automated process for generating heat map overlays.

Lastly, motivating process participants to consistently provide feedback is an important area that merits further investigation.

Chapter 6

Conclusion

This research presents the design and development of a tool that collects, processes, and analyzes the feedback of process participants. The research problem is as following:

> *Improve process analysis by developing a tool that analyzes the process participants' feedback such that their insights are incorporated in the organizational knowledge-base in order to initiate process redesign in a time-efficient, bottom-up manner.*

The sentiment of the process participants towards their daily executed business activities are visualized in the form of a heat map overlay on the corresponding process model. This clear and concise representation of the process participants' insights enhances the process analysis and redesign phase.

A key contribution is the development of a novel dataset that includes a process model and anonymized feedback messages that were manually annotated by the process managers. Moreover, this research contributes the tool's design and code, both available for future investigations.

Finally, the tool enables process participants to express their feelings and share insights regarding their daily performed activities, fostering a bottom-up process redesign approach. The results demonstrate that the tool successfully incorporates the insights of process participants in the organizational knowledge-base. Moreover, process managers indicate that the tool provides novel, actionable insights.

Additionally, the evaluation highlights that the tool's ease of use, time-saving potential in the process analysis and redesign phase, and its applicability to other departments. However, in situations where the process participants encounter few to no issues, the tool's time-saving benefits are perceived to be minimal.

Future research should examine in which contexts the tool proves the most effective, compare the performance of other NLP models on this task, and explore which analysis techniques could further increase the scalability of the tool.

References

[1] Ruth Sara Aguilar-Savén. Business process modelling: Review and framework. *International Journal of Production Economics*, 90, 2004. ISSN 09255273. doi: 10.1016/S0925-5273(03)00102-6.

[2] Sanam Ahmed and Khurram Shahzad. Augmenting business process model elements with end-user feedback. *IEEE Access*, 10:115635–115651, 2022. doi: 10.1109/access.2022.3216418.

[3] Rania Albalawi, Tet Hin Yeap, and Morad Benyoucef. Using topic modeling methods for short-text data: A comparative analysis. *Frontiers in Artificial Intelligence*, 3:42, 2020.

[4] Pedro Antunes, David Simões, Luis Carriço, and José A. Pino. An end-user approach to business process modeling. *Journal of Network and Computer Applications*, 36: 1466–1479, 11 2013. ISSN 10848045. doi: 10.1016/j.jnca.2013.03.014.

[5] Pedro Antunes, Jose A. Pino, Mary Tate, and Alistair Barros. Eliciting process knowledge through process stories. *Information Systems Frontiers*, 22, 2020. ISSN 15729419. doi: 10.1007/s10796-019-09922-0.

[6] Mohammed Bahja. Natural language processing applications in business. *E-Business - Higher Education and Intelligence Applications*, 2020. doi: 10.5772/intechopen.92203. URL `https://www.intechopen.com/chapters/71990`.

[7] Patricia Bazan and Elsa Estevez. Social business process management: Assessing the state of the art and outlining a research agenda. *Business Process Management Journal*, 26, 2020. ISSN 14637154. doi: 10.1108/BPMJ-09-2017-0257.

[8] Ana Cláudia De Almeida Bordignon, Vinicius Stein Dani, Lucinéia Heloisa Thom, Marcelo Fantinato, Thanner Soares Silva, and Renato Cesar Borges Ferreira. Natu-

ral language processing in business process identification and modeling: A systematic literature review. 2018. doi: 10.1145/3229345.3229373.

[9] Marco Brambilla, Piero Fraternali, Carmen Vaca, and Stefano Butti. Combining social web and bpm for improving enterprise performances: The bpm4people approach to social bpm. 2012. doi: 10.1145/2187980.2188014.

[10] Giorgio Bruno, Frank Dengler, Ben Jennings, Rania Khalaf, Selmin Nurcan, Michael Prilla, Marcello Sarini, Rainer Schmidt, and Rito Silva. Key challenges for enabling agile bpm with social software. volume 23, 2011. doi: 10.1002/smr.523.

[11] JD Tony Carter et al. Managers empowering employees. *American Journal of Economics and Business Administration*, 1(2):41–46, 2009.

[12] KR1442 Chowdhary and KR Chowdhary. Natural language processing. *Fundamentals of artificial intelligence*, pages 603–649, 2020.

[13] Rob Churchill and Lisa Singh. The evolution of topic modeling. *ACM Computing Surveys*, 54(10s):1–35, 2022.

[14] Marlon Dumas, Marcello La Rosa, Jan Mendling, and Hajo A. Reijers. *Fundamentals of Business Process Management.* 2013. doi: 10.1007/978-3-642-33143-5.

[15] Selim Erol, Michael Granitzer, Simone Happ, Sami Jantunen, Ben Jennings, Paul Johannesson, Agnes Koschmider, Selmin Nurcan, Davide Rossi, and Rainer Schmidt. Combining bpm and social software: Contradiction or chance? *Journal of Software Maintenance and Evolution*, 22, 2010. ISSN 15320618. doi: 10.1002/smr.460.

[16] Tianyu Gao, Adam Fisch, and Danqi Chen. Making Pre-trained Language Models Better Few-shot Learners. 8 2021. doi: 10.18653/v1/2021.acl-long.295. URL `https://aclanthology.org/2021.acl-long.295.pdf`.

[17] Fernanda Gonzalez-Lopez, Niels Martin, Rene de la Fuente, Victor Galvez-Yanjari, Javiera Guzmán, Eduardo Kattan, Marcos Sepúlveda, and Jorge Munoz-Gama. Prodem: A process-oriented delphi method for systematic asynchronous and consensual surgical process modelling. *Artificial Intelligence in Medicine*, 135:102426, 2023.

[18] Shengding Hu, Ning Ding, Huadong Wang, Zhiyuan Liu, Jingang Wang, Juanzi Li, Wei Wu, and Maosong Sun. Knowledgeable prompt-tuning: Incorporating knowledge into prompt verbalizer for text classification. *arXiv preprint arXiv:2108.02035*, 2021.

[19] Pooja Kherwa and Poonam Bansal. Topic modeling: a comprehensive review. *EAI Endorsed transactions on scalable information systems*, 7(24), 2019.

[20] Barbara Kitchenham. Guidelines for performing systematic literature reviews in software engineering. *Technical report, Ver. 2.3 EBSE Technical Report. EBSE*, 2007.

[21] Omer Levy and Yoav Goldberg. Dependency-based word embeddings. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 302–308, 2014.

[22] Loet Leydesdorff and Adina Nerghes. Co-word maps and topic modeling: A comparison using small and medium-sized corpora (n$<$ 1,000). *Journal of the Association for Information Science and Technology*, 68(4):1024–1035, 2017.

[23] Guowei Li, Fuqiang Lin, Wangqun Chen, Diwen Dong, and Bo Liu. Prompt-based learning for aspect-level sentiment classification. In *Neural Information Processing: 29th International Conference, ICONIP 2022, Virtual Event, November 22–26, 2022, Proceedings, Part III*, pages 509–520. Springer, 2023.

[24] Pengfei Liu, Weizhe Yuan, Jinlan Fu, Zhengbao Jiang, Hiroaki Hayashi, and Graham Neubig. Pre-train, prompt, and predict: A systematic survey of prompting methods in natural language processing. *ACM Computing Surveys*, 55(9):1–35, 2023.

[25] Egon Lüftenegger and Selver Softic. Crowdsourced sentiment-driven process re-design with sentipromoweb: Towards enterprise social information systems. In *IFIP International Conference on Advances in Production Management Systems*, pages 171–178. Springer, 2022.

[26] Egon Lüftenegger and Selver Softic. Sentipromo: A sentiment analysis-enabled social business process modeling tool. volume 397, 2020. doi: 10.1007/978-3-030-66498-5_7.

[27] Walaa Medhat, Ahmed Hassan, and Hoda Korashy. Sentiment analysis algorithms and applications: A survey. *Ain Shams engineering journal*, 5(4):1093–1113, 2014.

[28] Amina Mustansir, Khurram Shahzad, and Muhammad Kamran Malik. Sentiment analysis of user feedback on business processes. 2021. doi: 10.1109/ASEW52652.2021.00048.

[29] Amina Mustansir, Khurram Shahzad, and Muhammad Kamran Malik. Towards automatic business process redesign: an nlp based approach to extract redesign suggestions. *Automated Software Engineering*, 29, 2022. ISSN 15737535. doi: 10.1007/s10515-021-00316-8.

[30] Nicolas Pflanzl and Gottfried Vossen. Human-oriented challenges of social bpm: An overview. volume P222, 2013.

[31] Nicolas Pflanzl and Gottfried Vossen. Challenges of social business process management. 2014. doi: 10.1109/HICSS.2014.480.

[32] David MW Powers. Evaluation: from precision, recall and f-measure to roc, informedness, markedness and correlation. *arXiv preprint arXiv:2010.16061*, 2020.

[33] Michael Prilla and Alexander Nolte. Integrating ordinary users into process management: Towards implementing bottom-up, people-centric bpm. volume 113 LNBIP, 2012. doi: 10.1007/978-3-642-31072-0_13.

[34] Jipeng Qiang, Zhenyu Qian, Yun Li, Yunhao Yuan, and Xindong Wu. Short text topic modeling techniques, applications, and performance: a survey. *IEEE Transactions on Knowledge and Data Engineering*, 34(3):1427–1445, 2020.

[35] Guanghui Qin and Jason Eisner. Learning how to ask: Querying lms with mixtures of soft prompts. *arXiv preprint arXiv:2104.06599*, 2021.

[36] Fajar Ramadhani and E. R. Mahendrawathi. A conceptual model for the use of social software in business process management and knowledge management. volume 161, 2019. doi: 10.1016/j.procs.2019.11.225.

[37] Kumar Ravi and Vadlamani Ravi. A survey on opinion mining and sentiment analysis: tasks, approaches and applications. *Knowledge-based systems*, 89:14–46, 2015.

[38] Hajo A. Reijers and S. Liman Mansar. Best practices in business process redesign: An overview and qualitative evaluation of successful redesign heuristics. *Omega*, 33, 2005. ISSN 03050483. doi: 10.1016/j.omega.2004.04.012.

[39] Laria Reynolds and Kyle McDonell. Prompt programming for large language models: Beyond the few-shot paradigm. In *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems*, pages 1–7, 2021.

[40] Rainer Schmidt and Selmin Nurcan. Bpm and social software. volume 17 LNBIP, 2009. doi: 10.1007/978-3-642-00328-8_65.

[41] Taylor Shin, Yasaman Razeghi, Robert L Logan IV, Eric Wallace, and Sameer Singh. Autoprompt: Eliciting knowledge from language models with automatically generated prompts. *arXiv preprint arXiv:2010.15980*, 2020.

[42] António Rito Silva and Michael Rosemann. Processpedia: An ecological environment for bpm stakeholders' collaboration. *Business Process Management Journal*, 18, 2012. ISSN 14637154. doi: 10.1108/14637151211214993.

[43] Selver Softic and Egon Lüftenegger. Towards empowering business process redesign with sentiment analysis. In *Proceedings of Sixth International Congress on Information and Communication Technology*, pages 119–126. Springer, 2022.

[44] Christian Sonnenberg and Jan Vom Brocke. Evaluations in the science of the artificial–reconsidering the build-evaluate pattern in design science research. In *Design Science Research in Information Systems. Advances in Theory and Practice: 7th International Conference, DESRIST 2012, Las Vegas, NV, USA, May 14-15, 2012. Proceedings 7*, pages 381–397. Springer, 2012.

[45] Mohammad S Sorower. A literature survey on algorithms for multi-label learning. *Oregon State University, Corvallis*, 18(1):25, 2010.

[46] Han van der Aa, Josep Carmona, Henrik Leopold, Jan Mendling, and Lluís Padró. Challenges and opportunities of applying natural language processing in business process management. 2018.

[47] Ike Vayansky and Sathish AP Kumar. A review of topic modeling methods. *Information Systems*, 94:101582, 2020.

[48] Dalia Suša Vugec, Katarina Tomičić-Pupek, and Vesna Bosilj Vukšić. Social business process management in practice: Overcoming the limitations of the traditional business process management. *International Journal of Engineering Business Management*, 10, 2018. ISSN 18479790. doi: 10.1177/1847979017750927.

[49] Roel J. Wieringa. *Design science methodology: For information systems and software engineering*. 2014. doi: 10.1007/978-3-662-43839-8.

[50] Laurie Williams and Alistair Cockburn. Agile software development: It's about feedback and change. *Computer*, 36(6):39–43, 2003.

[51] Chong-ho Yu and Barbara Ohlund. Threats to validity of research design, 2010.

[52] Lei Zhang, Shuai Wang, and Bing Liu. Deep learning for sentiment analysis: A survey. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 8(4):e1253, 2018.

Appendix A

Summary Shortlisted Article

Shortlisted articles were summarized for the literature review. This was mainly done using bullet points. When reading the article, the main focus was to determine: a) what is the purpose of the research, b) what is the research methodology, c) what are the unique findings, and d) what is the intersection with the pillar to which this article belongs? This appendix presents a summary to illustrate how this was done in practice.

The Summary of Prilla and Nolte [33]:

- Currently, BPM relies too much on experts –> result: process specifications often differ from real processes as perceived by process stakeholders.
- Expert driven means that stakeholders involvement is currently limited. People-centric (bottom-up) approach actively involves stakeholders and their contributions, and therefore, close the gap between stakeholders and BPM experts (and speed up documentations and changes).
- The paper states that this can be tackled by involving/integrating the stakeholders of processes into process management –> enable more stakeholders to contribute
- Expert driven nature hinders the ability of stakeholders to proactively contribute to process management
- One of the findings is that BPM would benefit from a more bottom-up, user-centric approach where more stakeholders can submit their input
- The main goal of integrating more stakeholders is to 'enhance the quality of, and commitment to, process models'
- study 1 (interaction of non-modelers with models): Stakeholders are immediately cut from process development after their information is captured. Moreover, they would like to participate in the process models but they don't know how.

- study 2 (how non-modelers specify processes): people can contribute to processes if they do not have to express themselves in modeling language (!).
- Five Proposals for the implementation of stakeholder involvement into BPM:
  - Make models available
  - Redefine roles in BPM
  - **Provide suitable interactions for non-modelers**
  - **Make models tools of everyday use**
  - Intertwine top-down and bottom-up strategies

Appendix B

Progress Discussion Example Notes

To discuss progress, a progress meeting was held every other week. Since these meetings lasted for only 30 minutes, the meeting was well prepared. Throughout the two weeks prior to the discussion, a list of unclear points regarding the research was maintained. Two days before the meeting, these points were sent to the professor by mail. This gave the professor time to prepare for the discussion. During the meeting, the questions were addressed, and notes were taken. This appendix contains one of the notes regarding a progress meeting where the professor, Hajo Reijers, gives feedback on the questions of the author. The notes are written in Dutch as the meetings were in Dutch as well.

The Progress Meeting Notes: Voor het literatuuronderzoek: Per pilaar de stappen doorlopen, nu op bijna 50% –> Hajo Reijers: goed idee om het per pilaar te doen. Maakt het mogelijk om van de voorgaande cycli te leren.

Vragen onderzoeksplan:

1. Hoe vaak feedback? Eens per dag genoeg? —> Hajo Reijers: eens per dag is een goed plan
2. Zal ik selectie criteria voor proces en partner opstellen?

    - Hajo Reijers: Goed plan. Suggestie: issues/nog niet echt gefinetuned, maar geen chaos (dus niet recentelijk ontwikkeld). Niet helemaal mis, maar wel bekend dat er wat verbeterpunten zijn.

3. Is het mogelijk om feedback automatisch te mappen op procesmodel? Evt. medewerker het proces laten kiezen (zie bijlage mail).

   - Hajo Reijers: Gebaseerd op specialisme medewerker (maar wat als iemand in meerdere processen zit?) —> niet laten specificeren, hierdoor wordt de eindgebruiker beperkt. Geen onderdeel van dit type onderzoek.

4. Bij BPM Consult willen ze graag dat ik de feedback op een proces model in de hoogste vorm van abstractie map, is dit oké?

   - Hajo Reijers: Dit is een goed idee, even het goede jargon opzoeken in het boek

Overige Feedback van de Professor:

1. Wat we het liefst willen is de eindgebruikers betrekken
2. Zo makkelijk mogelijk, zo dicht mogelijk bij de mensen
3. Het ene spectrum is alles automatiseren, het andere spectrum is het zo makkelijk mogelijk maken voor de eindgebruiker.
4. Kies eerst voor het zo makkelijk mogelijk maken, met de hand mappen en testen in hoeverre je het kan automatiseren. Dit zou een mooie vervolgstap voor het onderzoek zijn.

Houd het volgende in je gedachte:

- Het blijft een onderzoek
- Lean startup methode
- MVP, prototype, kijken of het werkt
- Voor ons (Hajo & Benjamin): belangrijk om te onderzoeken of het werkt
- Voor jullie (BPM Consult): belangrijk om te weten of het werkt
- Voor de klant (van BPM consult): krijgt een mooie rapportage

Appendix C

Preliminary Research Announcement

Geachte,

Op verzoek van Renco Bakker stuur ik u een vooraankondiging van een onderzoek dat in het voorjaar van 2023 gaat plaatsvinden. Ons idee is om dit onderzoek in samenwerking met Verstegen Spices & Sauces B.V. te laten plaatsvinden, omdat wij denken dat dit onderzoek ook voor uw organisatie interessant kan zijn.

Waar gaat het om? Op dit moment is het erg lastig om medewerkers consistent te betrekken bij het lokaliseren van bedrijfsprocessen die geoptimaliseerd moeten worden. Dit is niet alleen frustrerend voor de medewerkers, maar ook een gemiste kans voor degenen die de processen optimaliseren. Daarom willen we onderzoeken of het mogelijk is om medewerkers te betrekken bij het verbeteren van werkprocessen.

Voor dit onderzoek wordt een algoritme ontwikkeld dat aanwijzingen en emoties uit natuurlijke tekst (bijvoorbeeld Whatsapp berichten of e-mails) kan interpreteren, om deze vervolgens te converteren in een zogenoemde *organisatiethermometer*. Deze thermometer zal visueel weergeven welke processen de medewerkers positief of negatief beoordelen. Doordat u meer inzicht krijgt in welke processen stroef verlopen voelen medewerkers zich beter gehoord, en worden de knelpunten in essentiële processen voor u zichtbaar. Overigens zal er ook onderzocht worden wat voor feedback er gegeven wordt, of dit relevant is voor het verbeteren van de processen.

Om dit voorstel te realiseren zal er een feedbackmoment plaatsvinden met een aantal geïnteresseerden om hun mening over het idee te peilen en de eisen voor de dataverzameling samen te stellen. Vervolgens zal er circa twee weken proefgedraaid worden om de data te verzamelen. Tijdens de dataverzamelingsperiode zal een deel van uw medewerkers een laagdrempelig bericht krijgen om feedback te geven op de processen waar ze dagelijks mee bezig zijn. Als laatste zal er een terugkoppeling van de resultaten zijn.

Aangezien het projectplan nog in de maak is, zijn de volgende tijden indicatief. Begin 2023 zal er een specifiekere tijdlijn aangeleverd worden. Het eerste feedbackmoment zal in februari 2023 plaatsvinden. De data zal in april/mei 2023 worden verzameld. Tot slot zal de terugkoppeling medio juni 2023 plaatsvinden.

Naast dat dit onderzoek voordelen oplevert voor uw medewerkers en u meer inzicht geeft in knelpunten in bedrijfsprocessen, levert u met uw deelname een belangrijke bijdrage aan de wetenschap!

Vriendelijke groet,

Benjamin Kleppe

Figure C.1: Preliminary research announcement

Appendix D

Informed Consent Form

Universiteit
Utrecht

Toestemmingsformulier voor deelname in onderzoeksproject

Organisatorische thermometer

Lees de onderstaande stellingen door, en vink het laatste boxje aan om te bevestigen dat je de stellingen begrepen en gelezen hebt, en dat je akkoord gaat met de deelname in het project.

Ik bevestig dat ik 18 jaar of ouder ben.

Ik bevestig dat het onderzoeksproject "Organisatorische thermometer" aan mij uitgelegd is. Ik heb de mogelijkheid gehad om vragen over het project te stellen, en deze zijn voldoende beantwoord. Daarnaast heb ik genoeg tijd gehad om hierover na te denken.

Ik geef toestemming dat het materiaal dat ik bijdraag wordt gebruikt om inzichten voor het onderzoeksproject "[Organisatorische thermometer]" te genereren.

Ik begrijp dat er persoonlijke gegevens van mij worden verzameld en dat deze informatie vertrouwelijk wordt behandeld, zodat alleen BPM Consult toegang heeft tot deze gegevens en de informatie tot mij persoonlijk kan herleiden. De informatie wordt versleuteld en gedurende maximaal 14 dagen op een beveiligde locatie bewaard. In overeenstemming met de General Data Protection Regulation (GDPR) heb ik toegang tot mijn informatie en kan ik op elk moment tijdens deze periode verzoeken mijn gegevens te verwijderen.

Ik begrijp dat mijn deelname aan dit onderzoek vrijwillig is en dat ik mij te allen tijde zonder opgaaf van reden uit het onderzoek kan terugtrekken, en dat bij terugtrekking alle reeds van mij verzamelde persoonsgegevens zullen worden gewist.

Ik sta toe dat de <u>volledig geanonimiseerde</u> gegevens worden gebruikt in toekomstige publicaties en andere wetenschappelijke middelen om de bevindingen van het onderzoeksproject te verspreiden.

Ik begrijp dat de verkregen gegevens veilig zullen worden opgeslagen door onderzoekers, maar dat naar behoren geanonimiseerde gegevens in de toekomst beschikbaar kunnen worden gesteld aan anderen voor onderzoeksdoeleinden. Ik begrijp dat de universiteit naar behoren geanonimiseerde gegevens kan publiceren in geschikte gegevensbewaarplaatsen voor verificatiedoeleinden en om ze toegankelijk te maken voor onderzoekers en andere onderzoek gebruikers.

Ik begrijp dat ik kan verzoeken dat alle persoonlijke gegevens die van mij zijn verzameld, worden gewist.

☐ Ik bevestig dat ik de bovenstaande verklaringen heb gelezen en begrepen, en ga akkoord met deelname aan het onderzoek. (Vink het vakje aan).

Naam participant:

Datum:

Figure D.1: Informed consent form for the participants

Appendix E

Literature Research Protocol

A literature review was conducted in order to gain further understanding regarding the area of knowledge, current state, and validity of the research. The three main pillars on which the theoretical foundation is build are the traditional bottom-up BPM approach, social BPM, and the current state of research concerning the combination of BPM and NLP.

The literature review followed a protocol outlined in this chapter. The literature review was divided into three parts. Each part was performed sequentially for each of the pillars. First, a long-list was created. In this study, this consists of all relevant articles selected by title. The search terms used during this step are presented in E.1. The search engines used to retrieve the articles are provided in Table E.2. Additionally, several inclusion and exclusion criteria were created [20].

Inclusion criteria:

IC-1 The paper is an academic paper

IC-2 The publishing date is before December 2022

Search terms
"social bpm"
"bpm" & "nlp"
"bpm" & "bottom-up"
"bpm" & "sentiment"
"bpm" & "social software"

Table E.1: Search terms used for the literature review

Search engine	Description
Scopus	The Scopus search engine was primarely used for finding articles. The search terms, which are presented in Table E.1, were entered in this search engine.
Google Scholar	The Google Scholar search engine was consulted for the Snowballing method. The method was used for snowballing forward and backward to find relevant related papers.

Table E.2: Search engines used for the literature research

IC-3 The article focuses on process improvement
IC-4 The article focuses on process redesign (or optimization)
IC-5 The content of the article relates explicitly to one of the three pillars

Exclusion criteria:
EC-1 The language of the paper is not English
EC-2 The article focuses on automatic process model generation
EC-3 The content of the article is not related to one of the three pillars

The result was a short-list with approximately 20-25 articles. The long-list, consisting of the short-list with the included articles and the excluded articles, can be found in Appendix F. The steps and number of resulting articles are presented in Table E.3. During the second step, all articles from the short-list were summarized. The final step was to write a critical synthesis, taking into account questions such as: "What insights did the articles read provide?", "What are the similarities?" and "What are the differences?". The literature review was based on the resulting synthesis.

Step	Candidate articles
Initial search	887
Title screening	30
Snowballing method	71
Abstract & conclusion	30
Full-text review	23

Table E.3: Total number of candidate articles for each step during the literature search

Appendix F

Long-list

Included

[1] Sanam Ahmed and Khurram Shahzad. Augmenting business process model elements with end-user feedback. *IEEE Access*, 10:115635–115651, 2022.

[5] Pedro Antunes, Jose A. Pino, Mary Tate, and Alistair Barros. Eliciting process knowledge through process stories. *Information Systems Frontiers*, 22, 2020.

[4] Pedro Antunes, David Simões, Luis Carriço, and José A. Pino. An end-user approach to business process modeling. *Journal of Network and Computer Applications*, 36:1466–1479, 11 2013.

[4] Alia Ayoub and Amal Elgammal. Utilizing twitter data for identifying and resolving runtime business process disruptions. volume 11229 LNCS, 2018.

[7] Patricia Bazan and Elsa Estevez. Social business process management: Assessing the state of the art and outlining a research agenda. *Business Process Management Journal*, 26, 2020.

[9] Marco Brambilla, Piero Fraternali, Carmen Vaca, and Stefano Butti. Combining social web and bpm for improving enterprise performances: The bpm4people approach to social bpm. 2012.

[10] Giorgio Bruno, Frank Dengler, Ben Jennings, Rania Khalaf, Selmin Nurcan, Michael Prilla, Marcello Sarini, Rainer Schmidt, and Rito Silva. Key challenges for enabling agile bpm with social software. volume 23, 2011.

INCLUDED

[15] Selim Erol, Michael Granitzer, Simone Happ, Sami Jantunen, Ben Jennings, Paul Johannesson, Agnes Koschmider, Selmin Nurcan, Davide Rossi, and Rainer Schmidt. Combining bpm and social software: Contradiction or chance? *Journal of Software Maintenance and Evolution*, 22, 2010.

[9] Henrik Leopold, Jan Mendling, and Artem Polyvyanyy. Generating natural language texts from business process models. volume 7328 LNCS, 2012.

[10] Henrik Leopold, Jan Mendling, and Artem Polyvyanyy. Supporting process model validation through natural language generation. *IEEE Transactions on Software Engineering*, 40, 2014.

[26] Egon Lüftenegger and Selver Softic. Sentipromo: A sentiment analysis-enabled social business process modeling tool. volume 397, 2020.

[28] Amina Mustansir, Khurram Shahzad, and Muhammad Kamran Malik. Sentiment analysis of user feedback on business processes. 2021.

[29] Amina Mustansir, Khurram Shahzad, and Muhammad Kamran Malik. Towards automatic business process redesign: an nlp based approach to extract redesign suggestions. *Automated Software Engineering*, 29, 2022.

[30] Nicolas Pflanzl and Gottfried Vossen. Human-oriented challenges of social bpm: An overview. volume P222, 2013.

[31] Nicolas Pflanzl and Gottfried Vossen. Challenges of social business process management. 2014.

[33] Michael Prilla and Alexander Nolte. Integrating ordinary users into process management: Towards implementing bottom-up, people-centric bpm. volume 113 LNBIP, 2012.

[36] Fajar Ramadhani and E. R. Mahendrawathi. A conceptual model for the use of social software in business process management and knowledge management. volume 161, 2019.

[40] Rainer Schmidt and Selmin Nurcan. Bpm and social software. volume 17 LNBIP, 2009.

[42] António Rito Silva and Michael Rosemann. Processpedia: An ecological environment for bpm stakeholders' collaboration. *Business Process Management Journal*, 18, 2012.

[46] Han van der Aa, Josep Carmona, Henrik Leopold, Jan Mendling, and Lluís Padró. Challenges and opportunities of applying natural language processing in business process management. 2018.

[21] Han van der Aa, Henrik Leopold, and Hajo A. Reijers. Comparing textual descriptions to process models – the automatic detection of inconsistencies. *Information Systems*, 64, 2017.

[48] Dalia Suša Vugec, Katarina Tomičić-Pupek, and Vesna Bosilj Vukšić. Social business process management in practice: Overcoming the limitations of the traditional business process management. *International Journal of Engineering Business Management*, 10, 2018.

Excluded

[1] Sippy Aggarwal. Guidelines for the use of ai in bpm systems: a guide to follow to use ai in bpm systems. 2022.

[2] Pedro Antunes and Mary Tate. Business process conceptualizations and the flexibility-support tradeoff, 4 2022.

[3] Hanane Ariouat, Chihab Hanachi, Eric Andonoff, and Frederick Benaben. A conceptual framework for social business process management. volume 112, 2017.

[4] M. A. Barcelona, L. García-Borgoñón, M. J. Escalona, and I. Ramos. Cbg-framework: A bottom-up model-based approach for collaborative business process management. *Computers in Industry*, 102, 2018.

[5] Patrizio Bellan, Mauro Dragoni, and Chiara Ghidini. Process extraction from text: state of the art and challenges for the future. *arXiv preprint arXiv:2110.03754*, 2021.

[8] Ana Cláudia De Almeida Bordignon, Vinicius Stein Dani, Lucinéia Heloisa Thom, Marcelo Fantinato, Thanner Soares Silva, and Renato Cesar Borges Ferreira. Natural language processing in business process identification and modeling: A systematic literature review. 2018.

[7] Stephan Bögel, Stefan Stieglitz, and Christian Meske. A role model-based approach for modelling collaborative processes. *Business Process Management Journal*, 20, 2014.

[8] Andrea Delgado, Laura González, and Daniel Calegari. Towards setting up a collaborative environment to support collaborative business processes and services with social interactions. volume 10797 LNCS, 2018.

[9] Luis Delicado, Josep Sànchez-Ferreres, Josep Carmona, and Lluís Padró. Nlp4bpm - natural language processing tools for business process management. volume 1920, 2017.

[10] Fery Ardiansyah Effendi and Yuliant Sibaroni. Sentiment classification for film reviews by reducing additional introduced sentiment bias. *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, 5, 2021.

[11] Ali Mert Ertugrul and Onur Demirors. An exploratory study on role-based collaborative business process modeling approaches. volume 23-24-April-2015, 2015.

[12] Thomas Freytag, Benjamin Kanzler, Nils Leger, and Daniel Semling. Nlp as a service: An api to convert between process models and natural language text. volume 2973, 2021.

[13] Fabian Friedrich, Jan Mendling, and Frank Puhlmann. Process model generation from natural language text. volume 6741 LNCS, pages 482–496, 2011.

[14] Vanisha Gokaldas and Mohammad Ehson Rangiha. A framework for improving user engagement in social bpm. volume 308, 2018.

[15] Thomas Grisold, Jan Vom Brocke, Steven Gross, Jan Mendling, Maximilian Röglinger, and Katharina Stelzl. Digital innovation and business process management: Opportunities and challenges as perceived by practitioners. *Communications of the Association for Information Systems*, 49, 2021.

[16] Xue Han, Yabin Dang, Lijun Mei, Yanfei Wang, Shaochun Li, and Xin Zhou. A novel part of speech tagging framework for nlp based business process management. 2019.

[17] Andreas Hermann, Hendrik Scholta, Sebastian Bräuer, and Jörg Becker. Collaborative business process management – a literature-based analysis of methods for supporting model understandability. *Wirtschaftsinformatik 2017*, 2017.

[18] Richard Hull and Hamid R.Motahari Nezhad. Rethinking bpm in a cognitive world: Transforming how we learn and perform business processes. volume 9850 LNCS, 2016.

[19] Jutta Karjalainen. Developing business process automation strategy: case: Finnish financial industry organization. 2022.

[20] Hugo A. Lopez, Morten Marquard, Lukas Muttenthaler, and Rasmus Stromsted. Assisted declarative process creation from natural language descriptions. volume 2019-October, 2019.

[21] Hugo A. López, Søren Debois, Thomas T. Hildebrandt, and Morten Marquard. The process highlighter: From texts to declarative processes and back. volume 2196, 2018.

[22] Carl Mikael Lönn and Elin Uppström. Process management challenges in swedish public sector: A bottom up initiative. volume 8074 LNCS, 2013.

[23] Egon Lüftenegger and Selver Softic. Supporting manufacturing processes design using stakeholder opinions and sentiment analysis. volume 632 IFIP, 2021.

[24] Pablo Mennuto, Julio César Meca Belahonia, and Patricia Bazán. Bpm and socialization tools integrated to improve acquisition and management of information during design and execution of business processes. bpm-social tool: A proposal. *Journal of Computer Science and Technology(Argentina)*, 21, 2021.

[25] Rachid Meziani. Empowering business process users through a pragmatic approach: A case study. 2014.

[26] E. J. Munive-Hernandez, F. W. Dewhurst, M. C. Pritchard, and K. D. Barber. Modelling the strategy management process: An initial bpm approach. *Business Process Management Journal*, 10, 2004.

[27] Syed Irtaza Muzaffar, Khurram Shahzad, Faisal Aslam, Madiha Khalid, and Kamran Malik. Process matching: Performance trade-off between summary and full-length descriptions. *Computing and Informatics*, 38, 2019.

[28] Stephan Poelmans and Hajo A. Reijers. The deployment of business process management systems: A quantitative analysis of end-users' evaluations. 2010.

[29] Luis Quishpi, Josep Carmona, and Lluís Padró. Extracting annotations from textual descriptions of processes. volume 12168 LNCS, 2020.

[30] Mohammad Ehson Rangiha, Marco Comuzzi, and Bill Karakostas. A framework to capture and reuse process knowledge in business process design and execution using social tagging. *Business Process Management Journal*, 22, 2016.

[31] Mohammad Ehson Rangiha and Bill Karakostas. Goal-driven social business process management. 2013.

[32] Aleksandra Revina. Business process management: Integrated data perspective. a framework and research agenda. 2021.

[33] Nina Rizun, Aleksandra Revina, and Vera G. Meister. Analyzing content of tasks in business process management. blending task execution and organization perspectives. *Computers in Industry*, 130, 2021.

[34] Nina Rizun, Aleksandra Revina, and Vera G. Meister. Assessing business process complexity based on textual data: Evidence from itil it ticket processing. *Business Process Management Journal*, 27, 2021.

[35] Leonardo Silva Rosa, Thanner Soares Silva, Marcelo Fantinato, and Lucineia Heloisa Thom. A visual approach for identification and annotation of business process elements in process descriptions. *Computer Standards and Interfaces*, 81, 2022.

[36] Sherif Sakr, Zakaria Maamar, Ahmed Awad, Boualem Benatallah, and Wil M.P. Van Der Aalst. Business process analytics and big data systems: A roadmap to bridge the gap. *IEEE Access*, 6, 2018.

[37] Arik Senderovich, Joop J. Schippers, and Hajo A. Reijers. Socially-aware business process redesign. volume 12168 LNCS, 2020.

[38] Sholiq Sholiq, Riyanarto Sarno, and Endang Siti Astuti. Generating bpmn diagram from textual requirements. *Journal of King Saud University-Computer and Information Sciences*, 2022.

[39] David Simões, Pedro Antunes, and Jocelyn Cranefield. Enriching knowledge in business process modelling: A storytelling approach. *Intelligent Systems Reference Library*, 95, 2016.

[40] Konstantinos Sintoris and Kostas Vergidis. Extracting business process models using natural language processing (nlp) techniques. volume 1, 2017.

[41] Christian Stary. Open organizational learning: Stakeholder knowledge for process development. *Knowledge Management and E-Learning*, 8, 2016.

[42] Tammo Straatmann, Jan-Philip Schumacher, Cosima Koßmann, Ludger Poehler, Frank Teuteberg, Karsten Mueller, and Kai-Christoph Hamborg. Advantages of virtual reality for the participative design of work processes: An integrative perspective. *Work*, (Preprint):1–23, 2022.

[43] Josep Sànchez-Ferreres, Andrea Burattin, Josep Carmona, Marco Montali, and Lluís Padró. Formal reasoning on natural language descriptions of processes. volume 11675 LNCS, 2019.

[44] Josep Sànchez-Ferreres, Andrea Burattin, Josep Carmona, Marco Montali, Lluís Padró, and Luís Quishpi. Unleashing textual descriptions of business processes. *Software and Systems Modeling*, 20, 2021.

[45] Josep Sànchez-Ferreres, Han van der Aa, Josep Carmona, and Lluís Padró. Aligning textual and model-based process descriptions. *Data and Knowledge Engineering*, 118, 2018.

[46] Roman Vaculín, Richard Hull, Maja Vuković, Terry Heath, Nathaniel Mills, and Yutian Sun. Supporting collaborative decision processes. 2013.

[47] Han van der Aa, Adrian Rebmann, and Henrik Leopold. Natural language-based detection of semantic execution anomalies in event logs. *Information Systems*, 102, 2021.

[48] Fadwa Yahya, Khouloud Boukadi, Zakaria Maamar, and Hanêne Ben Abdallah. Enhancing business processes with web 2.0 features. 2015.

Appendix G

Literature Research Protocol NLP Techniques

A literature review was conducted to gain a deeper understanding of four candidate NLP techniques, namely topic modeling, sentiment analysis, aspect-based sentiment analysis (ABSA), and prompt engineering. The purpose of this literature review was to gain a deeper understanding about three aspects: the applications, benefits, and limitations of the techniques, with the overarching goal of gaining knowledge for selecting one of the techniques for this study. Therefore, we searched the literature until we obtained sufficient

Technique	Query	Screened	Used
Topic modeling	TITLE(review AND topic AND modeling)	9	2
Topic modeling	TITLE(topic AND modeling AND ((small AND corpora) OR (short AND text)))	13	4
Topic modeling	TITLE(word AND embeddings AND dependency)	1	1
Sentiment analysis	TITLE(sentiment AND analysis AND(review OR survey))	4	3
Sentiment analysis	TITLE (sentiment AND analysis AND (review OR survey))	4	3
Prompt engineering	TITLE-ABS-KEY(prompt AND (nlp OR natural AND language)) AND (LIMIT-TO(PUBYEAR, (2019, 2020, 2021, 2022, 2023))	48	5

Table G.1: Queries and results for literature review 2 knowledge on these three aspects. To make this study as reproducible and valid as possible, we followed the protocol described in this chapter.

The search terms used during this step are presented in G. Scopus was used as search engine. Additionally, several inclusion and exclusion criteria were created [20].

Inclusion criteria:

- IC-1 The paper is an academic paper
- IC-2 The publishing date is before May 2023
- IC-3 The article focuses on a NLP technique
- IC-4 The article focuses on topic modeling, sentiment analysis, or prompt engineering
- IC-5 The proposed technique should be applicable to a small dataset
- IC-6 In the case of prompt engineering, it should involve few-shot learning
- IC-7 In the case of prompt engineering, the article should be published after 2018
- IC-8 Preferably, the article concerns a survey or comparison of existing literature and techniques

Exclusion criteria:

- EC-1 The language of the paper is not English
- EC-2 The article focuses on a specific domain
- EC-3 The article focuses on Natural Language Generation (NLG)
- EC-4 The article focuses on generating a (pre-trained) language model

Appendix H

Email NLP Experts

Dear,

My name is Benjamin Kleppe, and I am a MSc. Business Informatics student at the University of Utrecht. Together with Hajo Reijers, I am conducting research on the involvement of process participants in allocating process redesign opportunities using NLP. We are currently developing a tool that utilizes NLP techniques to convert feedback from process participants about their daily business activities into meaningful insights that can be visualized on BPM models. This should help with allocating bottlenecks and opportunities for improvement.

The feedback will be collected by asking questions during the working day, such as "Did you encounter any difficulties today? If so, please explain them". Furthermore, participants are encouraged to send feedback at any time during the day.

There are numerous NLP techniques that could be used. Several studies recommend using sentiment analysis to examine participant feedback. Hajo and I are confident that, as an NLP expert, you have a thorough understanding of which NLP technique(s) are appropriate for this work. As a result, I'd like to ask you two questions:

1. Which word-vector algorithm(s) would you suggest for processing the feedback?
2. Which NLP technique(s) would you suggest using to examine the feedback?

If you would prefer, I am willing to explain the study in more detail in a personal call, or I can come by your office. Please do not hesitate to contact me if you have any questions or suggestions about the research.

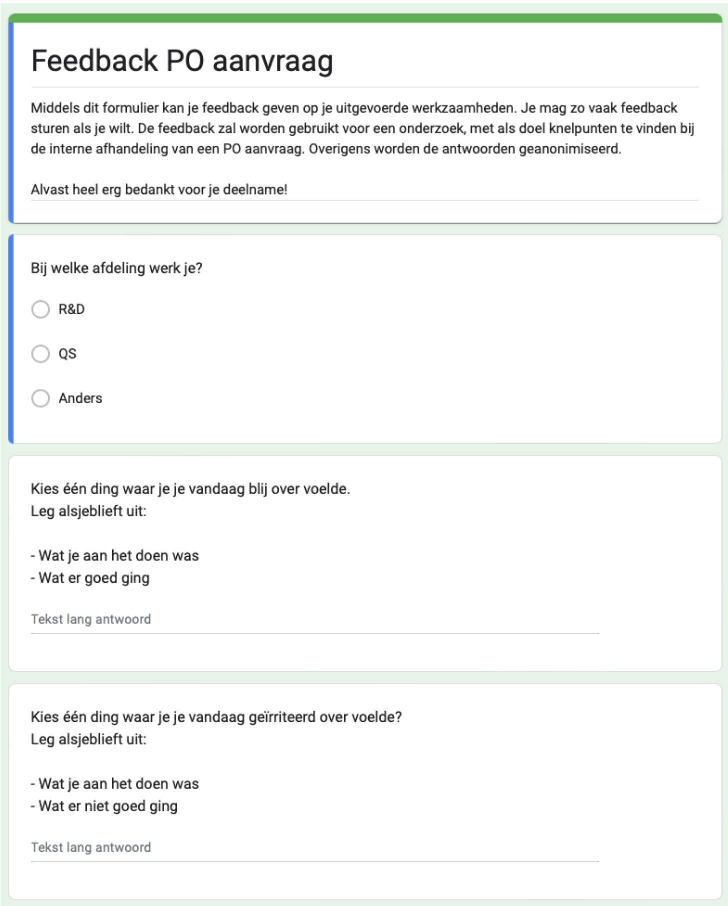
Thank you in advance for your assistance!

Best regards,

Benjamin Kleppe

Appendix I

Google Form



Feedback PO aanvraag

Middels dit formulier kan je feedback geven op je uitgevoerde werkzaamheden. Je mag zo vaak feedback sturen als je wilt. De feedback zal worden gebruikt voor een onderzoek, met als doel knelpunten te vinden bij de interne afhandeling van een PO aanvraag. Overigens worden de antwoorden geanonimiseerd.

Alvast heel erg bedankt voor je deelname!

Bij welke afdeling werk je?

- R&D
- QS
- Anders

Kies één ding waar je je vandaag blij over voelde.
Leg alsjeblieft uit:

- Wat je aan het doen was
- Wat er goed ging

Tekst lang antwoord

Kies één ding waar je je vandaag geïrriteerd over voelde?
Leg alsjeblieft uit:

- Wat je aan het doen was
- Wat er niet goed ging

Tekst lang antwoord

Figure I.1: Google Form used for retrieving the feedback

Appendix J

Reminder

Hello everyone,

This is a friendly reminder that we are collecting feedback on the PO process.

You can fill in the form again at: http://bit.ly/feedback-po

Thanks in advance!

Best regards,

Benjamin

Appendix K

Evaluation Statements

Dim	Statement	Act	5	4	3	2	1
EoU	The feedback form was easy to answer.	PP	62.5	12.5	12.5	12.5	0
EoU	The study as a whole (feedback form, briefing, reminders, among others) was easy to follow.	PP	12.5	37.5	50	0	0
EoU	The visualizations presented by the tool are easy to understand.	PM	67	33	0	0	0
EoU	The study as a whole (feedback form, briefing, reminders, among others) was easy to follow.	PM	67	0	33	0	0
Ef	The time and mental effort needed for answering the feedback form was reasonable.	PP	25	37.5	25	12.5	0
Ef	The time and effort needed to submit feedback regarding my daily executed activities improved by using the tool.	PP	12.5	37.5	25	25	0
Ef	The time and mental effort needed to analyze the results was reasonable.	PM	67	33	0	0	0
Ef	The time and effort needed to analyze the feedback of the process participants is improved by using the tool.	PM	67	0	33	0	0
G	The software tool can be adapted to gain insights about the feedback of another department.	PM	33	67	0	0	0
G	The study as a whole can be adapted to gain insights about feedback of another department.	PM	33	67	0	0	0
O	The insights obtained through the tool are novel.	PM	67	0	33	0	0
O	The tool insights obtained by the tool lead to actionable improvement ideas.	PM	67	0	0	33	0

[1] EoU: ease of use, Ef: efficiency, G: generality, O: operationality.

[2] PP: process participant (N=8), PM: process manager (N=3).

[3] 5: strongly agree, 4: agree, 3: neither agree nor disagree, 2: disagree, 1: strongly disagree.

Table K.1: Evaluation results in %.

Appendix L

Output of the Model

Activity	**Positive**	**Negative**	**Sentiment**
assign request	0	1	-1
make sample	7	9	-2
request advice	0	6	-6
review recipe	0	2	-2
save reference sample	0	1	-1
send sample	0	2	-2
step 1	2	7	-5
step 2	1	1	0
step 3	3	2	+1
write recipe	3	0	+3
TRAINING DATA	6	11	-

Table L.1: Class distributions: number of positive feedback messages, negative feedback messages, and the corresponding sentiment score for each activity.

Appendix M

Incorrect Predictions

Feedback message	Predicted	Correct
It was a quiet day for emails, so I had time and space for development.	make sample	assign request
Support and Export storing up requests and all sending them on Friday.	request advice	send sample
It was very nice that when checking a declaration by Quality department, I could easily find out that it was the second version. Everything else looked neat and I found the request email from R&D very funny.	step 1	step 3
Request with too low a volume received had to be sent back because of the low volume. Then back to us because it was approved by the head of Support.	assign request	request advice
The lead up to the urgent case; colleague indicated she didn't have time and couldn't make the development.	make sample	assign request
Checking 300 recipes manually for raw materials is not efficient.	review recipe	request advice

Feedback message	Correct	Predicted
It was not clearly stated in the request that it had to be [NAME] and therefore I have to adjust the recipe and resend the samples. Extra work and it takes me time.	review recipe	request advice
A sample request of 10 KG mix is requested. That's almost impossible to do.	make sample	request advice
I tried out new flavor directions together and they were also very delicious.	review sample	make sample
The distribution of workload; I had enough time and possibility to get my work done for today.	write recipe	make sample
I'm glad I was able to help a colleague so he has more breathing room.	write recipe	make sample
Colleagues' requests not completed on time, so there is now an urgent need to work.	write recipe	make sample
The sample I made turned out better than expected.	review sample	make sample