Automated Software Engineering (2022) 29:12
https://doi.org/10.1007/s10515-021-00316-8

Check for updates

Towards automatic business process redesign: an NLP based approach to extract redesign suggestions

Amina Mustansir[1] · Khurram Shahzad[2] · Muhammad Kamran Malik[3]

Received: 30 December 2020 / Accepted: 8 December 2021 / Published online: 3 January 2022

Abstract

Business process redesign (BPR) is widely recognized as a key phase of the business process management lifecycle. However, the existing studies have focused on proposing theoretical models, methodologies, and redesign patterns, whereas, the BPR activity remains dependent upon domain experts with little or no consideration to end-user feedback. To facilitate these experts, in this study, we have proposed a natural language processing (NLP) based approach to identify redesign suggestions from end-user feedback in natural language text. The proposed approach includes a novel set of annotation guidelines that can be used to generate computational resources for business process redesign. Secondly, to demonstrate the effectiveness of the proposed approach, we have generated computational resources which are composed of three real-world business processes and end-user feedback containing 8421 sentences. Finally, we have performed 270 experiments using six traditional and three deep learning techniques to evaluate their effectiveness for the identification of redesign suggestions from raw text. The classified suggestions can be used by domain experts to prioritize the redesign possibilities, without going through the details of the customer feedback.

Keywords Software engineering · Business process redesign · Application of NLP in BPR · Supervised learning techniques · Deep learning techniques

✉ Amina Mustansir
amina.mustansir@pucit.edu.pk

Khurram Shahzad
khurram@pucit.edu.pk

Muhammad Kamran Malik
kamran.malik@pucit.edu.pk

[1] Department of Computer Science, Department of Software Engineering, University of the Punjab, Lahore, Pakistan

[2] Department of Data Science, University of the Punjab, Lahore, Pakistan

[3] Department of Information Technology, University of the Punjab, Lahore, Pakistan

1 Introduction

Business processes represent the workflow of an organization. More specifically, a business process is defined as an organized set of activities to achieve an enterprise goal (Dumas et al. 2018). Typically, the specifications of a business process are documented in graphical form, formally called business process models. These models play a pivotal role in the BPM life cycle which is composed of five phases: process discovery, process analysis, process redesign, process implementation, and monitoring & controlling of business processes (Dumas et al. 2018). Business Process Redesign (BPR) is a key phase of the life-cycle that focuses on 'how to articulate a process in, e.g. terms of its interdependent tasks and resources' (Mansar and Reijers 2007). The influential researchers of the domain have conducted a comprehensive survey of the progress in research made by the BPM community during the last two decades (Recker and Mendling 2016). The study revealed that BPR is yet to attract the necessary attention of the research community. Furthermore, the study has put-forth five recommendations for future progress in the field which includes the development and open data sharing of the benchmarks (Recker and Mendling 2016).

Prominent existing studies that have reviewed the state-of-the-art revealed that a large majority of the BPR focuses on developing methodologies, framework, or patterns for redesigning activity (Zellner 2013; Cho et al. 2017; Vanwersch et al. 2016), with no attention to the development of benchmarks and use of disruptive technology. However, despite the presence of these artifacts, BPR is still a manual activity which is heavily dependent upon domain experts. Typically, these experts primarily rely on their expertise for the redesign activity, with little or no consideration to the end-user feedback. That is, these experts use their experience and skills to evaluate or prioritize the redesign alternates with limited or no input from the end-users. Typically, these experts have knowledge in the field, however, they may not be equally fluent in disruptive technologies. Consequently, many attractive redesign possibilities of a process, which are desired by end-users, may be ignored. Typically, these missed redesign possibilities may be of higher significance for service-oriented organizations which maintain customer satisfaction as their primary goal.

Digital Innovation (DI) is a disruptive technology which has yielded a new arena of businesses by describing new products and services, as well as revolutionizing the existing ones. For instance, Airbnb and Uber has transformed the hoteling and traveling industry by changing the way we move and lodge. The leading researchers who stipulate the future discourse of the BPM domain have pronounced to converge DI and BPM in order to leverage the benefits of the contemporary development (Mendling et al. 2020; Grisold et al. 2021). In particular, the disruptive technologies, including Machine Learning (ML) and Natural Language Processing (NLP), can pave the way to go beyond exploitative approach for improving business processes by suggesting innovative pathways for redesigning processes (Grisold et al. 2021).

The feedback on business processes may also include pointers to new value propositions that are desired by the end-users (Rosemann 2020). Therefore, it is desired to dynamically management customer expectations by continuously improving business processes. To that end, this study proposes an NLP based approach to automatically extract suggestions for redesigning business processes by using AI for classifying end-user feedback available in natural language. This study builds on the numerous existing studies that proposed to combine the development in ML, NLP and BPM domains. The initial work that combined the two domains performed business process model validation through generating textual description from BPMN process models (Leopold et al. 2012, 2014). The generated textual descriptions have been used to detecting inconsistencies between process models and their textual descriptions (van der Aa et al. 2017), as well as to align the two types of process specifications (Sànchez-Ferreres et al. 2018). Also, the textual process descriptions have been used for searching relevant process models from a process repository (Muzaffar et al. 2019) and to detect anomalies during process execution (van der Aa et al. 2021).

This study has several contributions: firstly, this study proposes an approach to AI based approach to automatically extract BPR suggestions by classifying end-user feedback available in natural language. A second novel contribution is the set of classification guidelines which have manifold benefits. For instance, the guidelines can be used to enhance our developed resources and they can also be used to develop new resources for other processes. Also, in this study, we have developed computational resources for automatic classification of text to identify redesign suggestions. The benefits to the developed resources are manifolds. For instance, it will commence R&D in the area of automated BP redesign. Furthermore, it can be used to evaluate the effectiveness of forthcoming techniques for identifying exploitative and explorative suggestions to change suggestions from huge volume of customer feedback. The developed benchmark is composed of three elements: a) a real-world business process model designed in BPMN, which is the de jure modeling notation, b) real-world feedback about the designed processes, and c) three-level classification of the collected feedback i.e. Suggestion & Non-suggestion, Redesign & Non-redesign suggestion, and classification with respect to four process elements.

The rest of the paper is organized as follows: Sect. 2 presents the background of the study, including research problem, illustration of the research problem, and extensions to a workshop paper. Section 3 presents an overview of the proposed approach along with the novel guidelines that we have developed for the three level classification. Section 4 presents the details of the resources that we have developed, whereas, the effectiveness of supervised learning techniques, both traditional and deep-learning techniques, is evaluated in Sect. 5. The results of the experiments are presented in Sect. 6. The research context of the work, key findings and directions of future work as presented in Sect. 7. Finally, the study concludes in Sect. 8

2 Background

As a background to the study, firstly, the problem statement is presented. Subsequently, the problem statement is illustrated with the help of an example process. Finally, the extensions of this study with respect to the workshop paper are presented.

2.1 Problem statement

Customer feedback about business processes is of higher significance for service providers. In particular, the customer feedback in natural language text is of additional significance as it may include diverse content, such as problems faced by users, user sentiments about activities, and suggestions to improve end-user experience about a process. In the presence of a large volume of textual feedback that may be spread across multiple sources, it is not possible for domain experts to manually review this feedback and evaluate its relevancy with respect to the process redesign activity. Consequently, several attractive alternatives to reduce cost and increase customer satisfaction may be ignored. Therefore, it is desired to develop an approach for automatic extraction of process redesign suggestions.

2.2 Problem illustration

Consider the example of a course registration process at a university having students as the primary users of the process. Figure 1 depicts a simplified excerpt of a real-world process having a start node, an end node, four activities, and two gateways. As shown in the figure, the example process starts with the collection of registration form and it includes activities about completing the application and fee submission. Finally, the submission of the registration form activity concludes the process.

For the example process, it is desired to automatically identify the feedback sentences that contain redesign suggestions. Table 1 shows the example comments on the course registration process. Note, these examples are selected from the real-world feedback discussed in the later part of the paper. For instance, C1 in the table expresses the satisfaction of a user about the process. However, the feedback is clearly irrelevant to the process model shown in the Fig. 1. On the contrary, the user comments C2 and C3 are related to the example process and both comments contain

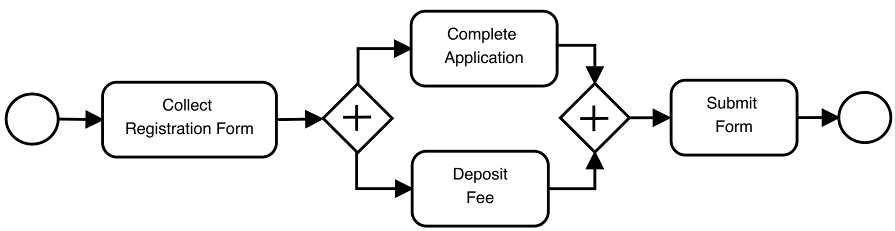

Fig. 1 University course registration process, an excerpt version

Table 1 Sample comments for problem illustration

Comments	Category
Changing password is an easy task (C1)	Non-suggestion
	NA
	NA
Solve the course registration issue (C2)	Suggestion
	Non-redesign suggestion
	NA
I suggest you to make others life easier (C3)	Suggestion
	Non-redesign suggestion
	NA
The system should send email to students (C4)	Suggestion
	Redesign suggestion
	Technology
Registration should be done by the teacher (C5)	Suggestion
	Redesign suggestion
	Actor

a suggestion, however, these comments are non-redesign suggestions. The comment C4 proposes to redesign the process by automating the form collection activity using technology, whereas C5 proposes to shift the control between actors i.e. from student to the teacher.

2.3 Advancements to the existing study

This paper builds on a previous study (Mustansar et al. 2020) having the following notable extensions.

- *Diverse processes* This extended version includes three diverse business processes, including admission process of a public sector and a semi- government university, whereas, the workshop paper includes a single course registration process of a private university. The process model in the workshop paper is composed of 174 elements, whereas, the two additional models included in this paper contains 153 and 100 elements, respectively.
- *Enhanced benchmark* The extended version contains 8421 (3727 + 3058 + 1636) feedback sentences about three processes, which are at least two folds more than the ones used in the workshop paper. Note, this is a significant extension because the additional 4694 sentences were annotated at three levels which involved 14082 additional manual annotations.
- *Use of deep learning techniques* In this extended version of the paper, we have used three state-of-the-art deep learning techniques and four types of word embeddings, which is a state-of-the-art representation of text. The deep learning techniques used in this study are Convolutional Neural Network (CNN), Long

Short-Term Memory (LSTM) and Bidirectional LSTM (BiLSTM). And the four types of word embeddings are fastText, GloVe, Word2vec, and BERT.
- *Additional experiments* The workshop paper included results of merely 54 experiments, whereas, this extended version includes results of 270 experiments. That is, it includes results of 162 experiments using six classical supervise learning techniques and 108 experiments using three deep learning techniques.
- *Thorough analysis of results* This extended version includes a more comprehensive analysis of the results. In particular, we have added an error analysis subsection in which we have identified the key causes of misclassification.

3 Overview of the approach

As discussed in an earlier section, BPR is widely acknowledged as a manual activity and it is heavily dependent upon domain experts. Furthermore, these experts rely on their domain knowledge and skills to redesign the process, with little or no consideration to the end-user feedback. One possible reason for such withstanding is that end-user feedback is in natural language and processing large amount of text to identify redesign suggestions requires substantial effort. To that end, Fig. 2 presents an overview of our proposed approach for the identification of redesign suggestions from end-user feedback. In essence, the proposed approach classifies end-user feedback into three levels to identify process redesign suggestions, as well as the elements about which the suggestions are proposed.

It can be observed from the figure that the first step of the proposed approach is to collect end-user feedback about processes. The feedback may be spread across multiple sources, such as social media, micro blogs, user comments, etc. In such case, the feedback is collected and integrated to form a unified corpus. In the second step, the collected feedback should be pre-processed to clean the text. For instance, if the user feedback is collected from a twitter stream, the hashtags need to be omitted and the text should be converted into the standard English. This step also involves breaking down the feedback into sentences which are used in the subsequent steps. In the third step, the sentences are classified into three levels to identify the redesign suggestions. These guidelines can be used to develop training resources for machine

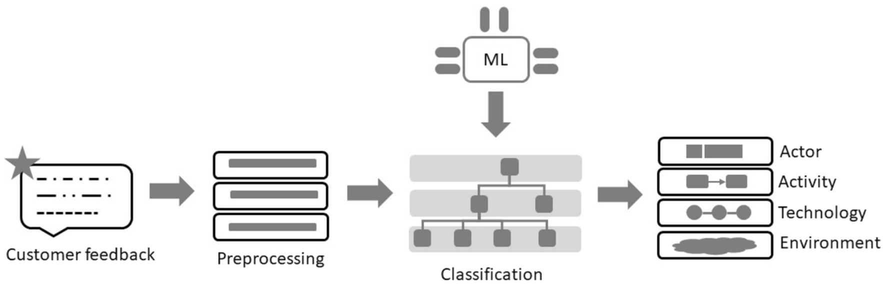

Fig. 2 The proposed approach

learning techniques for the automatic classification. The classification details are as follows:

The first level classification aims to identify suggestions by distinguishing between suggestion and non-suggestion sentences. As an output to the first level classification, the sentences that include an advise are identified. In contrast, the aim of the second level classification is to identify the sentences that include a redesign suggestion. The outcome to the second level classification is the set of sentences that includes a process redesign suggestion. Finally, the third level classification identifies the target of classification i.e. the process elements about which the redesign suggestions are proposed.

A novel contribution of this study is that we have also developed a comprehensive set of annotation guidelines that can be used for the three level classification. The principle reasons for preparing guidelines are as follows: a) the guidelines are helpful in developing a clear understanding of the annotation, b) the use of guidelines leads to consistent annotations i.e. independent researchers can generate similar annotations, c) the guidelines can be used to enhance the developed benchmark in the future. Corresponding to the three-level classification, we have developed guidelines for all the three levels.

Recognizing the manifold benefits of the guidelines discussed above, we have employed a systematic and rigorous approach for developing novel guidelines for each level classification. That is, for each level, we randomly selected a few hundred sentences from three case studies (the case studies are discussed in Sect. 4) and asked two independent researchers to perform the classification along with the development of an initial set of guidelines. Subsequently, the differences were discussed and the guidelines were merged to generate a unified set of guidelines for the data annotation.

The process was repeated to further refine the guidelines in order to ensure that a higher inter-annotator agreement is achieved. Accordingly, we generated separate guidelines for the three levels of annotation. Finally, a randomly selected 10% sentences were performed to validate the quality of the developed guidelines. The finalized guidelines for the three levels are presented in Tables 2, 3, 4. The extensive list of guidelines presented in the three tables clearly represent that we were thorough in the process of developing guidelines. Furthermore, the comprehensiveness of the guidelines represents the effort that we have put into developing the guidelines.

4 Computational resources for BPR classification

This section presents the details of the systematic protocol that we have used for the development of computational resources for the automatic extraction of BPR redesign suggestions from user feedback. Figure 3 presents an overview of the three-step process that we have used for developing the resources. The principle reason for using a systematic protocol is to ensure the quality of resources by minimizing the bias that may be induced due to the human involvement. It can be observed from the figure that the protocol is composed of three steps, business processing modeling,

Table 2 Guidelines for the first-level classification

Guidelines for suggestions
If a sentence proposes any idea about the system or processes under consideration using explicitly the word suggestion, recommendation, advice, opinion, tip or its synonym having similar semantics
Any advice or suggestion mentioning the word "should", "could", "might", "may"
Suggestion for himself or for any stakeholders, within or outside the scope
Suggestion in the form of a directive, or instruction, including imperatives
A conditional recommendation
Suggestion in the form of questions, however, such a question should have certain hints of changes
A recommendation for a specific community
A warning or an explicit awareness about the process or any feature of the system
Intention to change the schedule of activity, control flow of activity, convenience of activity related to quantity or quality, ambiance of the environment, quantity of actor, any role within the scope of the system, behavior of actor, any role within the scope of the system
Intention to change the controls and rights of actor, any role within the scope of the system, technology including tangible or intangible aspects is a recommendation
An explicit expression that recommends a change in the policy of the institute
Guidelines for not suggestions
A sentence that does not explicitly advise in favor of any process or entity i.e. a sentence that highlights the positive or negative features of the system
A sentence that depicts the problems of the system
That highlights the needs of the system
A sentence highlighting the current status or the situations without giving recommendations
A sentence highlighting the comparison with person, thing, system without giving recommendations
Any comment that appreciates the good points without giving recommendations
Any personal experience without any recommendations

Table 3 Guidelines for the second-level classification

Guidelines for redesign suggestions
A comment that provides a tip with reference to changing the process model under consideration
Guidelines for non-redesign suggestion
Advice about any process outside the scope defined
A comment is about the overall process without specifying which part of the process should be improved
A sentence that suggests to improve the process, without specifying what part or which actor should be improved

feedback collection, and data annotation, where each step corresponding to separate resources that we have generated.

As the name represents, the aim of the first step is to design business process models, whereas, the aim of the feedback collection is to collect end-user feedback about business processes. In contrast, the aim of the third step is to employ an iterative approach to develop novel guidelines that can achieve consistent annotation.

Table 4 Guidelines for the third-level classification

Guidelines for Actor
The internal customers of the business process
The participants in the business process considering the organization structure (elements: roles, users, groups, departments, etc)
The participants in the business process considering the organization population (individuals: agents which can have tasks assigned for execution and relationships between them)
Guidelines for Activity
The operation in which the focus is that what is a workflow and how is the operation implemented?, including the number of tasks in a job, relative size of tasks, nature of tasks, degree of customization
The focus is that when is a workflow executed?, including the sequencing of tasks, task consolidation, scheduling of jobs, etc
Collection of activities
Guidelines for Technology
They are tangible products and information resources. Tangible products generally include the hardware
Information resources is concerned with data, maybe the sorting, control, or even display of it
Guidelines for Environment
The environment external to the scope of the case study under consideration
It concerns the ambiance, associated processes with particular scope of the processes under study, and also external human resources
Any actor, process or activity not mentioned in the process model that we are studying are all external to our environment

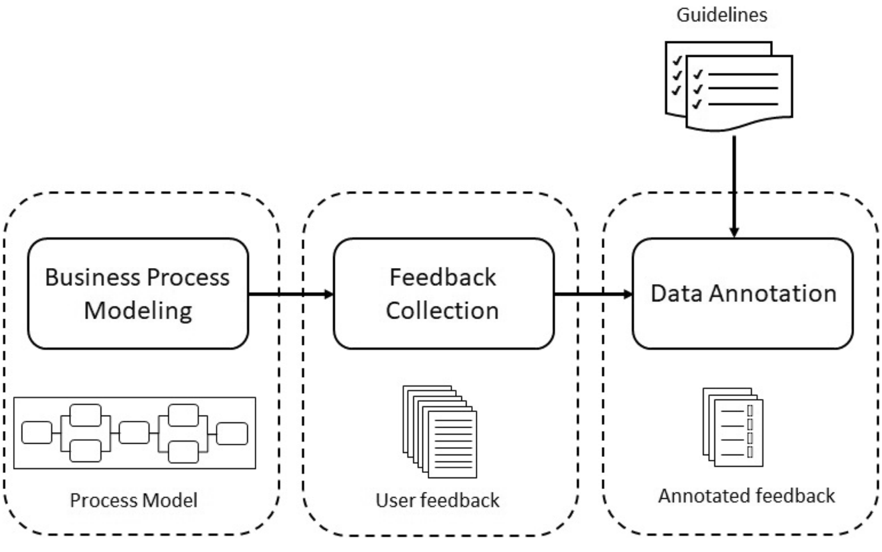

Fig. 3 Benchmark generation process

The aim of the final step is to use the developed guidelines for manually classifying the feedback at the three levels. The classified feedback can then be used by domain experts for prioritizing redesign alternatives. The details of each step are discussed in the following subsections:

4.1 Business process modeling

In the first step, we have designed diverse process models of three different universities to mirror the real-world settings. The key reasons for the choice of university processes in the benchmark are as follows: a) the university processes are public without proprietary issues, b) it is conveniently possible to collect end-user feedback about university processes, and c) university processes have been used for other BPM tasks, such as Process Model Matching Contest 2015 Antunes et al. (2015), and Process Model Matching at OAEI (2016).

The three diverse processes that we have designed in this study includes a student course registration process and two admission processes from three different universities. Where, the course registration process P1 belongs to a private university, one admission process (P2) belongs to a public sector university, and the second admission process (P3) belongs to a semi-government sector university. Note, the two admission processes are substantially different from each other as the admission to one university is managed by the academic department, whereas, the admission to the other university is managed by an independent office of the university.

The designing of the business processes was a non-trivial task as it involved multiple stakeholders from three academic institutes. Furthermore, the benefits of modeling these processes are manifolds. For instance, the designed models were useful to identify the objectives of the processes, document the correct scope of the processes, identify problems, and highlight priorities. For designing the three processes, we employed a two-pruned strategy, observatory and traditional technique (Phellas et al. 2011). The observational technique involves experiencing the process, whereas, the traditional approach includes taking interviews, and gathering requirements using questionnaires. In particular, firstly, we conducted preliminary interviews to develop a perception of the process. Secondly, the observatory approach was employed to develop an initial model. Finally, the second round of interviews

Table 5 Overview of the specifications of the three processes

Items	P1	P2	P3
No. of elements	174	153	100
Diameter	30	46	27
No. of edges	89	79	52
No. of actors	5	6	1
No. of split gateways	12	6	6
No. of activities	56	52	36

was conducted to refine the models. Accordingly, we designed comprehensive process models using BPMN [20]- the de jure modeling language (vom Brocke and Rosemann 2014).

The complete process models are very large and cannot be presented in this article. Alternately an overview of the specifications of the developed models is presented in Table 5. Based on the specifications presented in the table we observe the following about the designed process models.

- The three process models are composed of 174, 153, and 100 elements, respectively. These higher number of elements represent the substantial amount of effort involved in designing these models. Also, these numbers indicate that that we were thorough in our approach.
- The number of gateways (23, 13 and 9) and the values of diameters (30, 46 and 27) represent that the designed models are not merely sequential but there are ample gateways between them which represent the breadth of the model.
- Finally, the presence of a substantial amount of split gateways represent the presence of adequate alternate pathways.

4.2 Feedback collection

The aim of the second step is to collect the end-user feedback about the collected processes. For the P1 process, the end-users are the university students who intend to attend a course, whereas, for P2 and P3 processes the end-users are the applicants interested in getting admission to a university.

For the process P1, we used two sources for collecting end-user feedback, a social media platform and student reviews collected through a survey. For the social media platform, a facebook page was created and shared with students of the private sector university. The facebook page was used to guide students for providing feedback about their course registration experience over a period of multiple months. Note, students were merely encouraged to post their honest feedback, without any restriction of positive or negative feedback. Similarly, a google form was created and shared with the students to provide their feedback in textual form. Subsequently, we scrapped the user comments from the facebook page and downloaded the comments from the google forms.

Following that, we tokenized each comment to generate a large corpus of sentences. Although the manual evaluation of such a large corpus is a cumbersome job, yet the accuracy and authenticity of such an approach is of higher significance. Therefore, we used a manual approach to prepare quality data. The process was also repeated for P2 and P3 processes. Thereafter, the collected feedback was analyzed to identify several grammatical errors and spelling mistakes. Furthermore, the formatting issues and ambiguities were resolved to ensure quality text. Finally, the punctuation marks, special signs, and html tags were also removed. Accordingly, we

Table 6 Specifications of the feedback corpora

University	Sentences	Avg. length	Max. length
P1	3727	13.38	82
P2	3058	10.05	41
P3	1636	12.22	45

generated three corpora of end-user feedback which includes one corpus for each of the three process models.

Table 6 provides an overview of the specifications of the three corpora that we have developed. It can be observed from the table that the three corpora contains 8421 sentences, including 3727 sentences feedback for P1, 3058 sentences feedback for P2, and 1636 sentences feedback for P3. Furthermore, the average lengths P1, P2, and P3 corpora are 13.38, 10.05 and 12.22, whereas, the maximum lengths of the three corpora are 82, 41, and 45, respectively. These higher numbers indicate that each corpus includes a mixture of short and long sentences.

4.3 Data annotation

This is a vital step in the development of resources as it includes three-level classification of the end-user feedback. The key motivations for using the three-level classification are the following: a) it segments the redesign suggestion extraction problem, hence, simplifying the extraction task, b) it offers an opportunity to get a more holistic view of the end-user feedback, and c) it offers a step-by-step approach to extract the redesign suggestions.

4.3.1 Annotation levels

The aim of the first-level classification is to annotate that a sentence is a Suggestion or Non-suggestion. As discussed in the guidelines discussed in Sect. 3, a sentence is annotated as a suggestion, if it proposes some change to the process, whereas, a sentence that does not give any advice or opinion is a non-suggestion. For instance, the sentence 'guide the students to change the default password later' gives an advice about the process. On the contrary, the sentence 'the speed of net and computers are very good' does not include any advice. Table 2 presents the detailed guidelines that we have used for the first-level annotation.

The second-level annotation aims at annotating appropriate class, Redesign & Non-redesign suggestion, to a sentence. Where, a sentence that proposes change to the process under consideration is annotated as a redesign suggestion, whereas, a sentence that does not suggest a change that can be mapped to a process element is annotated as non-redesign suggestion. For instance, consider the two example sentences, 'they should keep it opened' and 'I think there is need to give some relaxation about this'. These are the examples of the sentences that are suggestions but cannot be mapped to any process element. Therefore, they are not annotated as non-redesign suggestions. Table 3 presents the detailed guidelines that we have used for the second-level annotation.

The third-level annotation aims at identifying the elements about which the redesign suggestion is proposed. For that, we have used key elements of the generic business process redesign framework which includes Actor, Activity, Technology, and External Environment (Mansar and Reijers 2005). Where, Actor is the internal customer of the business process. It can also be treated as the participant of the business process considering the organization structure (elements: roles, users, groups, departments, etc.) or the organization population (agents which can have tasks assigned for execution and relationships between them) (Dumas et al. 2013). Activity is another element of a business process which has two important views, where, the first view is the operation view which defines the workflow and how is the operation implemented? It includes, the number of tasks in a job, relative size of tasks, nature of tasks, and degree of customization.

The behavior view is the second view of activity that focuses on when a workflow executes? It includes the sequencing of tasks, task consolidation, scheduling of jobs, etc. External Environment is another key element which concerns the ambiance, associated processes with particular scope of the processes under study, and also external human resources. Any actor, process or activity not modeled in the business process falls in this category. In the sentence ’they should improve their bank staff’ the bank staff is external to the system as it is not modeled in our course registration system. Technology is another important element of business process that has two main parts, tangible products and information resources. Where, tangible products include the hardware or technology infrastructure, and the information resource concern the data associated with activities. An example of redesign suggestion about tangible products is ’there should be good computers in the lab’, whereas, an example redesign suggestion about information resource is ’no, just change the font and size for display’. Table 4 presents the detailed guidelines that we have used for the identification of the third level annotation.

4.3.2 Annotation procedure

Several existing studies, such as Vanwersch et al. (2011) and Khan et al. (2021), argue that a combination of clearly defined annotation guidelines and the seemly application of annotation procedure ensure the quality and consistency of annotations. Recall, from Sect. 3, we employed a systematic and rigorous approach to develop a comprehensive set of annotation guidelines for the three level classification. According to the approach, two independent researchers iteratively developed and refined annotation guidelines using a randomly selected sample until Kappa score ≥ 0.75 was achieved. Accordingly, the generated guidelines for the three level classification are presented in Tables 2, 3, and 4.

For seemly application of the annotation guidelines, an expert reviewed each sentence to determine if the given sentence is a suggestion or not suggestion, using the first-level guidelines. The cross-validation of the first-level annotations was performed by the second expert on a random sample. The sentences that were classified as suggestions were further classified into redesign suggestions or non-redesign suggestions. Subsequently, the cross-validation of the second-level annotations was also performed. Finally, the third level classification of the redesign suggestion sentences was performed by one expert and cross-validated by the second expert.

4.4 Specification of the benchmark

This section summarizes the specifications of the resources that we have developed. Specifically, the resources include, formal specification of a process model, end-user feedback about the process, and three-level manual annotations. We have developed the three benchmarks for each business process, P1, P2, and P3. The summary specifications of the process models are presented in Table 5. The summary specifications of the collected feedback are presented in Table 6, whereas, the summary of the manual annotations are presented in Table 7.

Based on the specifications presented in the three tables, we make the following two key observations: firstly, P1 process model is substantially larger than P2 and P3 process models, as it is composed of 174 elements, compared to 153 and 100 elements in the other two models. Furthermore, the P1 feedback corpus having

Table 7 Specifications of the three-level feedback corpora

Case	Classification	Sent.	Avg. length	Max. length
P1	Suggestion	772	15.9	82
	Non-suggestion	2954	12.7	63
	Redesign	688	16.9	82
	Non-redesign	84	7.6	42
	Actor	80	18.1	56
	Activity	321	17.3	51
	Technology	101	15.8	82
	Environment	186	15.7	77
P2	Suggestion	191	12.15	36
	Non-suggestion	2867	9.91	41
	Redesign	168	12.62	36
	Non-redesign	23	8.74	25
	Actor	22	11.91	20
	Activity	88	12.74	36
	Technology	23	12.04	25
	Environment	42	12.5	33
P3	Suggestion	58	14.15	29
	Non-suggestion	1578	10.47	45
	Redesign	50	14.48	27
	Non-redesign	8	14.2	29
	Actor	2	15	17
	Activity	34	14.2	24
	Technology	8	16.4	27
	Environment	6	13.5	19

3727 comments is larger than P2 and P3 corpora, which contain 3058, and 1636 comments, respectively. These different sized models and corpora represent that we have generated diverse resources, which is a valuable contribution for the BPM community.

Secondly, from the specifications of the first-level classification, it can be observed that the number of suggestions are significantly less than non suggestions. That is, the count of suggestions and non suggestions in the three corpora are (772, 2954), (191, 2867), and (58, 1578), respectively. In contrast, from the specifications of the second-level classification, it can be observed that the number of redesign suggestions are significantly higher than the non-redesign suggestions i.e. (688, 84), (168, 23), and (50, 8), respectively. Furthermore, from the specifications of the third-level classification, it can be observed that for the three processes, the feedback comments are not evenly distributed across the four elements of the process i.e. (80, 321, 101, 186) for P1, (22, 88, 23, 42) for P2, and (2, 34, 8, 6) for P3.

Based on the above discussion, as well as the following two reasons, we conclude that we have developed a valuable and a challenging resource for advancement towards automatic redesign of business process models.

- The unbalanced distribution of feedback for all three processes across all the three levels makes it a harder problem to solve.
- The fewer number of comments for some classes makes it an even harder problem, particularly for the supervised learning techniques.

5 Effectiveness of ML techniques

To demonstrate the effectiveness of the developed resource, in this section, we discuss the details of the experiments performed using six traditional machine learning techniques and three state-of-the-art deep learning techniques. Below, we present the dataset used for the experimentation, followed by an overview of the techniques used for experimentation. Finally, the experimental settings are presented.

5.1 Dataset

For the experimentation, we have used the resources that we have discussed in the preceding section. That is, we have used the three feedback corpora and three-level manual annotations for each corpus. Note, for the experiments we have not used the three process models, P1, P2, and P3. However, the non-use of the process models for the experiments does not represent the insignificance of the process models that consumed a substantial effort. We rather content that these models are mandatory in the following ways: a) the models are desired to determine the relevance of the redesign suggestions with the processes, b) the models are mandatory to interpret the classification results, and c) the models are also mandatory to perform the redesign task.

5.2 Techniques

For the experiments, we have used six traditional and three state-of-the-art deep learning techniques. The traditional supervised learning techniques are, Support Vector Machines (SVM), Decision Trees (DT), K-Neighbours (KN), Naïve Bayes (NB), Random Forest (RF), and Linear Regression (LR). Whereas, the deep learning techniques includes, Convolutional Neural Network (NN), Long Short-Term Memory (LSTM) and Bidirectional Long Short-Term Memory(BiLSTM).

For the experiments, the dataset was divided into three parts, training, validation and testing. Where, the validation data was used to manually tune the parameters in order to determine the parameters settings that allowed the model to achieve highest performance. For CNN, LSTM and BiLSTM 300 dimensions were used along with four additional layers. Furthermore, a dropout layer of 50% was added for all the embedding types except for BERT embedding which a dropout layer of 20% was used. Also, Relu activation function was used in the hidden layer, and Sigmoid function was used for binary classification, whereas, Softmax activation was used in

Table 8 Specifications of the techniques

Classification	Technique	Layer	Dimension-Size
Binary	CNN	Embedding layer	300
		Convolutional Layer	32, 8
		Global max pooling layer	32
		Dense layer	Number of unit - 1, 10
	LSTM	Embedding layer	300
		Dropout Layer	0.5
		LSTM Layer	300
		Dense Layer	Number of unit - 1, 10
	BILSTM	Embedding Layer	300
		Dropout Layer	0.5
		Bidirectional Layer	600
		Dense Layer	Number of unit - 1
Multiclass	CNN	Embedding Layer	300
		Convolutional Layer	32, 8
		Global max pooling Layer	32
		Dense Layer	Number of unit - 5, 50
	LSTM	Embedding Layer	300
		Dropout Layer	0.5
		LSTM Layer	300
		Dense Layer	5, 50
	BILSTM	Embedding Layer	300
		Dropout Layer	0.5
		Bidirectional Layer	600
		Dense Layer	Number of unit - 5, 50

the output layer for multiclass classification. The summary of the specifications are presented in Table 8.

5.3 Features and embeddings

For the classical techniques, we have used three types of features, unigram, bigram, and a combination of both unigram and bigram. This key reason for choosing bigram features is that it contains two continuous words from a given text, hence, capturing the words and their order. In contrast, for the deep learning techniques, we have used four types of word embeddings, including state-of-the-art embeddings. The types of embeddings we have used are Glove, Word2Vec, fastText, and BERT. Where, GloVe embeddings are generated by the leading NLP research group from Stanford University. Word2Vec embeddings are generated by Google Inc., fastText embeddings are released by Facebook Inc., which is a prominent social media platform. Whereas, BERT embeddings are state-of-the-art embeddings that have been recently developed and widely used for various NLP tasks. A key feature of these embeddings is that they capture complex relationships in text with the help of hundreds of dimension, that are difficult to capture otherwise.

5.4 Experimentation

In this study, we have performed 270 experiments, which includes 162 experiments using traditional supervised learning techniques and 108 experiments using deep learning techniques. That is, experiments are performed using the six traditional techniques for all of the three levels using unigram, bigram, and tri-gram features. Furthermore, the experiments are repeated for all of the three processes P1, P2, and P3. Also, three deep learning techniques are used for experimentation using four types of embeddings. The experiments are also repeated for all of the three processes. For each experiment, training to testing ratio of 75–25 is used. That is, 75% of the data are used for training and the remaining 25% data for the testing. The effectiveness of the techniques is computed in terms of widely used evaluation measures, Precision, Recall, and F1 scores. However, these large number of results cannot be presented in this paper due to the space limitations. Therefore, we have only reported F1 scores in the results section.

6 Results

Table 9 presents the F1 scores of all the 162 experiments that we have performed using the traditional techniques. It can be observed from the results that RF achieved a nearly perfect F1 score of 0.99 for the first-level classification. This nearly perfect F1 score represents the ability of RF to distinguish between Suggestion & Non-suggestion sentences. Similarly, RF achieved a perfect F1 score of 1.00 for the second-level classification. This perfect F1 score represents the effectiveness of the techniques for the second level classification i.e. to distinguish between Redesign

Table 9 F1 scores of the classical techniques

		First-level			Second-level			Third-level		
Tech.	Case	U	B	UB	U	B	UB	U	B	UB
SVM	P1	**0.90**	**0.90**	**0.90**	0.95	0.95	0.95	0.70	0.63	**0.75**
	P2	0.98	0.97	0.98	0.76	**0.86**	0.81	0.57	0.39	**0.64**
	P3	0.98	**0.99**	0.96	0.97	0.71	0.74	0.32	0.35	0.59
DT	P1	0.89	0.89	0.89	0.94	0.94	0.94	0.67	0.57	0.70
	P2	0.98	0.97	0.98	0.81	**0.86**	0.82	0.49	0.33	0.55
	P3	0.98	0.98	0.98	0.97	0.71	0.71	0.34	0.35	0.56
KN	P1	0.78	0.78	0.78	0.91	0.91	0.91	0.59	0.55	0.56
	P2	0.95	0.9	0.93	0.72	0.18	0.79	0.39	0.32	0.28
	P3	0.95	0.96	0.93	0.89	0.71	0.36	0.4	0.14	0.47
NB	P1	0.84	0.81	0.84	0.95	0.95	0.95	0.52	0.63	0.64
	P2	0.97	0.92	0.97	0.72	0.53	0.79	0.38	0.34	0.43
	P3	0.97	0.93	0.94	0.85	0.34	0.34	0.34	0.25	0.29
RF	P1	0.88	0.87	0.89	**0.96**	0.95	0.95	0.68	0.60	0.64
	P2	**0.99**	0.96	0.98	0.77	0.82	0.83	0.54	0.35	0.53
	P3	0.98	0.97	0.96	**1.00**	0.71	0.71	0.38	0.14	0.47
LR	P1	0.89	0.89	0.89	**0.96**	**0.96**	**0.96**	0.72	0.68	0.74
	P2	**0.99**	0.96	0.98	0.77	0.82	0.80	0.39	0.37	0.57
	P3	0.98	0.97	0.96	0.40	0.14	0.47	0.40	0.14	0.47

& Non-redesign suggestions. In contrast, for the third-level classification, SVM achieved the highest F1 score which is merely 0.75. This indicates that the third-level classification is a complex sub-task compared to the first two sub-task. The reasons to that are: a) the training dataset of the third sub-task is much smaller than the first two sub-tasks (see Table 6), which has thwarted the learning ability of the supervised learning techniques, b) the first two sub-tasks involve a binary decision, whereas, the third sub-task involves a non-binary decision, which is a more challenging sub-task.

Based on the comparison of results of the three types of features, unigram, bigram, and uni + bigram, for all the techniques, it can be observed that a large majority of the techniques achieved comparable F1 scores. This indicates that all the features are equally effective for the classification task, particularly for the first two level. This represents that there is a substantial difference between the vocabulary used in the sentences of the two classes. It further represents that the formulation of sentences in the classes is substantially different, which facilities the classification task.

Table 10 presents the results of the 108 experiments that we have performed using the deep learning techniques. It can be observed from the results of the first-level classification that all the three deep learning techniques achieved nearly perfect F1 scores. These nearly perfect F1 scores represent the effectiveness of deep learning techniques of the first level classification task. From the comparison of results of classical and deep learning techniques it can be observed that deep learning

Table 10 F1 scores of deep learning techniques

Tech.	Measure	First-level				Second-level				Third-level			
		FT	GV	WV	BR	FT	GV	WV	BR	FT	GV	WV	BR
CNN	P1	0.96	0.96	0.95	0.95	0.89	0.90	0.92	0.88	0.71	0.63	0.71	0.59
	P2	**0.99**	0.97	**0.99**	0.97	0.77	0.76	0.76	0.80	**0.64**	0.58	0.52	0.39
	P3	0.95	0.96	0.96	**0.99**	0.64	0.64	0.64	0.62	**0.57**	**0.57**	**0.57**	0.47
LSTM	P1	**0.97**	0.94	0.96	0.95	0.91	0.90	0.92	**0.96**	0.67	0.63	0.62	0.64
	P2	**0.99**	0.98	**0.99**	0.98	0.81	0.81	0.81	0.87	0.51	0.35	0.43	0.62
	P3	0.97	0.96	0.97	0.98	0.64	0.64	0.64	**0.67**	**0.57**	**0.57**	**0.57**	0.49
BiLSTM	P1	0.96	0.95	0.96	0.95	0.90	0.92	0.90	**0.96**	**0.72**	0.69	**0.72**	0.57
	P2	**0.99**	0.98	0.98	**0.99**	0.77	0.81	0.83	**0.88**	0.58	0.57	0.54	0.73
	P3	0.98	0.97	0.97	**0.99**	0.64	0.64	0.64	**0.67**	0.53	0.53	0.53	0.47

techniques either outperform the classical techniques or their F1 scores are comparable with the classical techniques. Similarly, for the second and third-level classifications, the F1 scores achieved by the deep learning techniques are higher than the classical techniques. Based on these results we conclude that the deep learning techniques are more effective than the classical techniques.

From the comparison of the F1 scores of all the four types of embeddings, it can be observed that there is no significant difference between their scores. These comparable scores represent that all the types of embeddings are equally effective for the classification tasks. Another notable observation is that in general, the F1 scores achieved by all the techniques are the highest for the first-level classification and lowest for the third-level classification. A key reason for these lower scores stem from the small number of training data i.e. the training data is so small that it does not provide ample opportunities for learning and prediction. Recall, we observed a similar trend for the classical supervised learning techniques. Based on the two observations we conclude that the third-level classification is more challenging than the second-level classification, and the second-level classification is more challenging than the first-level classification.

6.1 Error analysis

In order to understand the underlying reasons for the misclassification, we have performed an error analysis of the misclassified results. For the error analysis, we identified all the sentences that were misclassified by a majority of the deep learning techniques and manually grouped those sentences to ascertain the reasons of misclassification.

In order to perform an error analysis of the deep learning techniques for the first-level classification, we combined the predictions of all the deep learning techniques for all types of embeddings. Subsequently, we identified the sentences that were misclassified in a majority, in half of the cases. Finally, we analyzed the misclassified sentences to identifying the reasons for misclassification. Accordingly, the identified key reasons are as follows:

- Several sentences were classified as suggestions in the human benchmark, whereas, a majority of the techniques classified them as non-suggestions. Our examination of these sentences revealed that a majority of these sentences are longer in length and are composed of two or more parts. Furthermore, we observed that several of these sentences had a suggestion in the later half.
- There was another case in which suggestion sentences were misclassified as non-suggestions. That is, the sentences that were smaller in length and had indirect speech or a brief directive. The examples of such sentences are: let it be mediocre, and make life easier.
- Several sentences classified as non-suggestions in the benchmark, were classified as suggestions by a majority of the techniques. A careful examination reason revealed that these sentences included one or more words that are typically used in the context of advising or suggestion. However, they were not classified as suggestion according to our developed guidelines.

Similarly, we performed error analysis of the deep learning techniques for the second-level classification. We observed that a large majority of misclassified sentences made a suggestion about changing the process without proposing any specific change to the process. However, according to our annotation guidelines, the sentences that do not propose a specific change to the process were annotated as non-redesign suggestions. The examples of such sentences are, 'need to do something' and 'should be less time consuming'.

7 Discussion and outlook

This section discusses how this work contributes towards the traditional approaches, as well as the latest trends, in the field of BPM. Finally, the notable directions for the future research are also presented to the benefit of the community.

7.1 Research context

BPR is an established field of research and a plethora of studies have been conducted for developing diverse artifacts for redesigning business processes (Danilova 2019). These studies vary from high-level theoretical frameworks (Womack et al. 2007; Mansar and Reijers 2005) to the fine grained redesign patterns Zellner (2013); Lohrmann and Reichert (2016), and from process redesign heuristics (Reijers and Mansar 2005) to process performance dimensions as Devil's quadrangle (Reijers and Mansar 2005). However, the recent discourse by leading researchers have distinguished between two types of BPR approaches, exploitative and explorative (Grisold et al. 2021; Rosemann 2020). The *exploitative* approaches are reactive approaches which aim to improve processes by making them cost-effective, faster, flexible and compliant. These are incremental approaches that concentrate on the identification of issues in the processes to generate a streamlined version that can achieve operational effectiveness and efficiency (Rosemann 2020). In contrast, the *explorative* approaches are proactive which aims to exploit the emerging disruptive technologies, such as digital transformation, for finding new value propositions (Grisold et al. 2021). Typically, they are revolutionary approaches that provide guidance on exploring new revenue opportunities, rather than tweaking the existing process.

This study proposes the use of an automated approach for extracting process redesign suggestions from end-user feedback about the given process. We contend that the proposed approach can be beneficial for both exploitative, as well as explorative, approaches.

- *Explorative approach.* The suggestion sentences extracted from the classification of the feedback are useful for the explorative approaches. It is due to the reason that these sentences include advises, opinions, and tips for improving the processes. A careful examination of the sentences has revealed that several sentences propose new value propositions for changing the process. For instance, 'C1: payment should be made possible through bitcoin', and C2: 'for saving the traveling cost, payment system should be integrated with the banking system'.
- *Exploitative approach.* The sentences that are classified as non-suggestions include the problems reported by the end-users. These sentences represent the issues experienced by the end-users of the process that need to be addressed during business process redesign. The examples of end-user feedback that reports problem are: 'C3: advisor takes more time to process application' and 'C4: credit hour restrictions causes delays in degree time'. The idea of this study can be extended to perform a second-level classification of the non-suggestion sentences into two classes. Where, once class represents the issues, problems and difficulties reported by end users and the second class represents the non-problem reporting sentences.

7.2 BPM and digital innovation

DI has not only revolutionized the existing products and services but it has also introduced new products and services. For instance, Airbnb and Uber have transformed the way we lodge and travel, whereas and Bitcoin has introduced a new digital currency without any central administrator. Similarly, other services, such as Blockchain, Azure, and Dropbox, have disrupted the way we characterize computing. A recent study (Mendling et al. 2020) has synthesized the impact of these innovations have also altered the underlying business process. For instance, Uber has changed the mechanism of booking and cancelling a trip, communicating with the rider, rating the trip and paying for it.

The new arena of the emerging digital technologies have a significant impact on the entire BPM life cycle. However, this is impact is more evident on the business process redesign phase. It is because these technologies have paved the way to go beyond exploitative approach for redesigning processes. That is, they have unearthed new opportunities for changing processes, as well as to create new value propositions for businesses. This study relies the advancements in machine learning and natural language processing to analyze customer feedback and identify the redesign suggestions for process redesign. The study can be extended to investigate the interest of customers in disruptive technologies and predict their readiness for adapting to new processes, products and services. Hence, creating an opportunity for dynamically manage customer expectations.

8 Conclusion

Business Process Redesign (BPR) is a key phase of the BPM life cycle which aims at articulating the to-be process in terms of tasks and resources. Despite the fact that several artifacts have been developed for BPR, however, it is still a manual activity which is heavily dependent on domain experts, with little or no attention to the end-users feedback. Consequently, many attractive redesign suggestions or preferences of end-users may be ignored. To that end, in this study, we have proposed a novel approach for extracting redesign suggestions from the end-user feedback available in natural language text, hence, facilitating the domain experts for exploitative, as well as explorative improvement of business process.

The proposed approach pre-processes the raw text and subsequently classifies it into three levels. The first level classification identifies a suggestion from raw text by classifying it into suggestions and non-suggestion. The second level classification further classifies suggestion into redesign suggestions and non-redesign suggestions. Whereas, the aim of the third level classification identifies the target of the redesign suggestion, which could be Actor, Activity, Technology, or Environment.

We have also applied the proposed approach on three university admission processes. More specifically, we designed real-world process models for each process, collected feedback of over 8421 sentences and manually classified these sentences into three levels. The classified sentences and their three level classification is made publicly available for the advancement of research and development in the area of automatic BPR. Also, we encourage the BPM community to develop novel techniques and evaluate their effectiveness for the identification of redesign suggestions from natural language text. Furthermore, the community may define new tasks over our developed dataset. For instance, extraction of specific process elements about which the redesign suggestion is proposed. Another novel contribution is that we have proposed generic guidelines for the three level classifications. The guidelines can be used to extend our developed resources and may also be used to develop new resources.

Lastly, we have used the developed resource to perform 270 experiments. It includes experiments using six traditional supervised learning techniques and three state-of-the-art deep learning techniques. For the traditional supervised learning techniques, we used three types of features, and the experiments are performed separately for the three levels. The experiments are repeated for all the three university admission cases. For each deep learning technique, we have performed experiments using four types of word embeddings and for the three level classification. We have analyzed the results of experiments to identify notable observations. Also, we have performed a comprehensive error analysis of the results to identify the key reasons for the misclassification. Our analysis of the results revealed that all the deep learning techniques outperformed the traditional supervised learning techniques for all the three level classification tasks. Another notable observation is that the third-level classification task is more challenging that the second-level classification task, and the second-level classification task is more challenging than the first-level classification task.

This novel idea of using end-user feedback for automatic business process redesign has several future research directions.

- A key direction for future research is to develop NLP based approaches to automatically extract the change proposals for BPR. Furthermore, once the these suggestions are extracted, they can be used to identify the relevant process redesign patterns for the BPR activity.
- The second notable direction for future research is to perform the sentiment analysis on the extracted suggestions and prioritize the suggestions. Furthermore, the suggestions can be grouped by employing a clustering approach to examine the similarity between proposals, as well as their frequency.
- Finally, the study can be extended to develop the next-generation of intelligent approaches that can identify the problem areas in business processes, which are not identified in this study, and propose suggestions to overcome these problems.

References

Antunes, G., Bakhshandeh, M., Borbinha, J., Cardoso, J., Dadashnia, S., Di Francescomarino, C., Dragoni, M., Fettke, P., Gal, A., Ghidini: The process model matching contest 2015. GI-Edit. Lect. Notes Inf. **248**, 127–1553 (2015)

Cho, M., Song, M., Comuzzi, M., Yoo, S.: Evaluating the effect of best practices for business process redesign: an evidence-based approach based on process mining techniques. Decis. Support Syst. **104**, 92–103 (2017)

Danilova, K.B.: Process owners in business process management: a systematic literature review. Bus. Process Manag. J. Bus. Process Manag. J. **25**, 1377–1412 (2019)

Dumas, M., la Rosa, M., Mendling, J., Reijers, H.A.: Introduction to business process management. In: Fundamentals of business process management. Springer, The Netherlands (2013)

Dumas, M., la Rosa, M., Mendling, J., Reijer, H.A.: Fundamentals of Business Process Management. Springer, Berlin (2018)

Grisold, T., vom Brocke, J., Gross, S., Mendling, J., Röglinger, M., Stelzl, K.: Digital innovation and business process management: opportunities and challenges as perceived by practitioners. Communications of the Association for Information Systems (2021)

Grisold, T., Grob, S., Stelzl, K., Brocke, J., vom, Mendling, J., Roglinger, M., Rosemann, M.: The five diamond method for explorative business process management. Bus. Inf. Syst. Eng. **63**, 1–18 (2021)

Khan, M.M., Shahzad, K., Malik, M.K.: Hate speech detection in roman urdu. ACM Trans. Asian Low-Resour. Lang. Inf. Process. (TALLIP) **20**, 1–19 (2021)

Leopold, H., Mendling J., Polyvyanyy, A.: Generating natural language texts from business process models. In: Proceedings of the International Conference on Advanced Information Systems Engineering, 64–79 (2012)

Leopold, H., Mendling, J., Polyvyanyy, A.: Supporting process model validation through natural language generation. IEEE Trans. Softw. Eng. **40**, 818–840 (2014)
Lohrmann, M., Reichert, M.: Effective application of process improvement patterns to business processes. Softw. Syst. Model. **15**, 353–375 (2016)
Mansar, S.L., Reijers, H.A.: Best practices in business process redesign: validation of a redesign framework. Comput. Ind. **56**, 457–471 (2005)
Mansar, S.L., Reijers, H.A.: Best practices in business process redesign: use and impact. Bus. Process Manag. J. **13**, 193–213 (2007)
Mendling, J., Pentland, B.T., Recker, J.: Building a complementary agenda for business process management and digital innovation. 208–219 (2020)
Mustansar, A., Shahzad, K., Malik, M.K.: AutoEPRS-20: extracting business process redesign suggestions from natural language text. In: Proceedings of the 35th IEEE/ACM International Conference on Automated Software Engineering Workshops, 118–124 (2020)
Muzaffar, S.I., Shahzad, K., Aslam, F., Khalid, M., Malik, M.K.: Process matching: performance tradeoff between summary and full-length descriptions. Comput. Inf. **38**, 851–882 (2019)
Phellas, C.N., Bloch, A., Seale, C.: Structured methods: interviews, questionnaires and observation. Res Soc Cult **3**, 181–205 (2011)
Process Model Matching at Ontology Alignment Evaluation Initiative (OAEI) (2016). http://web.informatik.uni-mannheim.de/oaei/pm16/. Accessed 10 July 2020
Recker, J., Mendling, J.: The state of the art of business process management research as published in the BPM conference. Bus. Inf. Syst. Eng. **58**, 55–72 (2016)
Reijers, H., Mansar, S.L.: Best practices in business process redesign: an overview and qualitative evaluation of successful redesign heuristics. Omega **33**, 283–306 (2005)
Rosemann, M.: Explorative process design patterns. In: Proceedings of the International Conference on Business Process Management. 349–367 (2020)
Sànchez-Ferreres, J., van der Aa, H., Carmona, J., Padró, L.: Aligning textual and model-based process descriptions. Data Knowl. Eng. **118**, 25–40 (2018)
van der Aa, H., Leopold, H., Reijers, H.A.: Comparing textual descriptions to process models-the automatic detection of inconsistencies. Inf. Syst. **64**, 447–460 (2017)
van der Aa, H., Rebmann, A., Leopold, H.: Natural language-based detection of semantic execution anomalies in event logs. Inf. Syst. **102**, 1–13 (2021)
Vanwersch, R., Shahzad, S., Vanderfeesten, I., Vanhaecht, K., Grefen, P., Pintelon, L., Mendling, J., van Merode, G.G., Reijers, H.A.: Methodological support for business process redesign in health care: a literature review protocol. Int. J. Care Pathw. **15**, 119–126 (2011)
Vanwersch, R., Shahzad, S., Vanderfeesten, I., Vanhaecht, K., Grefen, P., Pintelon, L., Mendling, J., van Merode, G..G., Reijers, H..A.: A critical evaluation and framework of business process improvement methods. Bus. Inf. Syst. Eng. **58**, 43–53 (2016)
vom Brocke, J., Rosemann, M.: Handbook on Business Process Management 1: Introduction, Methods, and Information Systems. Springer, Berlin (2014)
Womack, J.P., Jones, D.T., Roos, D.: The Machine that Changed the World: the Story of Lean Production. Simon and Schuster, New York (2007)
Zellner, G.: Towards a framework for identifying business process redesign patterns. Bus. Process Manag. J. **19**, 600–623 (2013)