The Process Mining Question Forge

Lisa Zimmermann[1]

[1]*University of St. Gallen, St Gallen, Switzerland*

Abstract

This paper introduces the Process Mining Question Forge (PMQF), a tool that supports the design and refinement of questions for process analysis projects. Motivated by the observation that formulating well-defined questions is essential in process analysis, PMQF addresses challenges such as difficulty in designing appropriate questions and issues arising from poorly defined ones, aiming to improve the overall effectiveness of process analysis projects. In particular, it guides users in viewing, selecting, and refining example questions for their own use cases.

Keywords

Process Mining, Question Design, Question Refinement, End-User Support

1. Introduction

When analyzing processes, organizations increasingly resort to process mining (PM) techniques and tools [1]. Several methodologies and case studies are available to guide and inspire the planning and implementation of such projects. [2]. A similarity of them is that they include an initial phase in which questions should be formulated to guide subsequent project phases, like data extraction, preparation, or analysis [3]. For example, a question such as "What are the most common paths taken by cases in the process?" indicates an interest in the process flow. It requires the analyst to compute variants based on the available process data, identify the most frequent ones, and describe their execution. Instead the question "Are there any interesting patterns in the process?" is much broader. Without a keyword like "variant" or "path", it is not restricted to process flow patterns and analysts might also investigate other perspectives, such as time or resources. Additionally, "interesting" might not be equivalent to "most common", as it could refer to edge cases that are less relevant in terms of frequency. While someone familiar with PM techniques can likely envision the analysis for the first question, the second question requires more exploration and may depend on the available data and the analyst's experience. A less experienced project stakeholder interested in control-flow patterns might be disappointed by the outcome if they formulate a question similar to the second example.

Arguably, designing questions is an essential step in setting up PM initiatives, as they elicit requirements and direct the analysis. Research confirms that the execution of this step is a relevant success factor for projects [4]. However, formulating questions is not trivial and analysts lack support from tools and methods [5]. An interview study conducted with 40

analysts confirmed that question formulation is a significant challenge in PM projects, often arising when analysts work with questions that are unclear, overly specific, or too broad (as in the second example above) [6].

In this work, we address this problem and introduce the Process Mining Question Forge (PMQF), a tool that implements guidance for the crucial task of question design in PM projects. We developed PMQF based on findings from [5], that highlight that (experienced) analysts might rely on domain knowledge or analysis templates when confronted with a project that starts without clear questions. In practice, especially less experienced analysts lack this knowledge and templates are not always available. Therefore PMQF leverages categorized example questions and a classification schema to help users in designing their own questions. PMQF can be set up with any custom set of questions and a respective categorization schema. On top of these resources, it guides users to design their questions in a structured way.

2. Tool Description and Features

PMQF has been developed as a web application in Python using *Flask*[1]. When it is set up, it expects a categorized set of analysis questions and the corresponding classification schema as input. The source code of PMQF can be found at https://github.com/promise-ics-hsg/demoApplication-PMQF, and a deployed version can be accessed via http://130.82.168.60:5000/. In the deployed version, PMQF runs on an exemplary set of 405 categorized analysis questions that we gathered from diverse sources, such as the BPI Challenges[2] or Case Studies, and a classification schema that classifies questions across six dimensions. We provide a video demonstrating how the tool can be used: https://drive.switch.ch/index.php/s/Y9cW0Nk3DEmdVOR. PMQF supports users in (i) retrieving an overview of questions and their classification, (ii) designing new questions, and (iii) clarifying and refining existing questions.

2.1. Keyword Search

PMQF features an advanced keyword search that allows for the efficient location of relevant analysis questions. Users can enter keywords related to their areas of interest and the tool returns a list of questions that match this search criteria. To this end, we integrated the computation of synonyms based on *wordnet*[3] (using the *nltk* library[4]). The keyword search is implemented for project teams, learners, or teachers who are interested in a specific concept of PM and aim to retrieve an overview of what kind of questions they could ask in this regard.

2.2. Question Design

The question design feature supports users in formulating analysis questions through three phases. It is particularly suited for less experienced analysts or project teams without extensive expertise who will benefit from guidance on how to iteratively identify areas of interest.
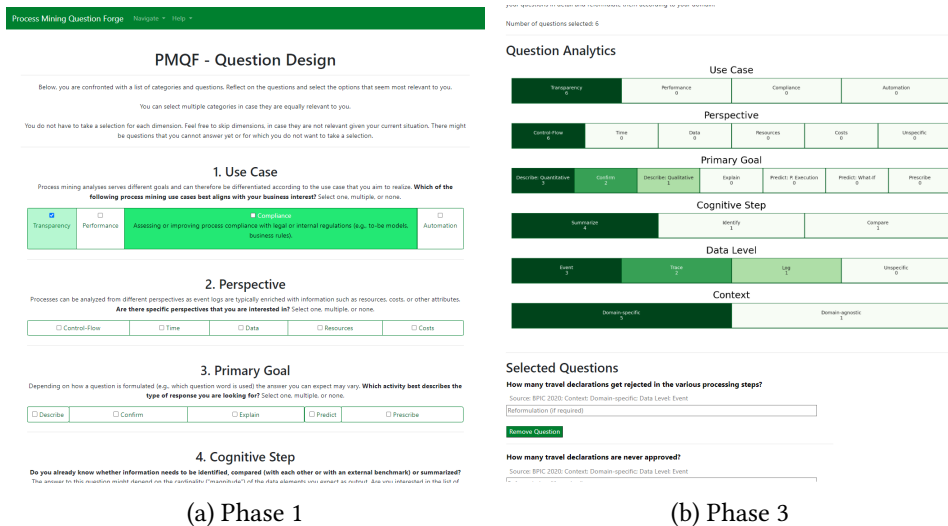
---

[1]https://flask.palletsprojects.com/en/3.0.x/
[2]https://www.tf-pm.org/competitions-awards/bpi-challenge
[3]https://wordnet.princeton.edu/
[4]https://www.nltk.org/

(a) Phase 1   (b) Phase 3

Figure 1: Interface for phases 1 and 3 of the question design feature in PMQF

Category Selection: In the first phase, users are directed to select categories based on the dimensions of the classification schema (Fig. 1a). For each dimension, we suggest a guiding question that helps them to choose categories that are aligned with their project goals. Definitions for all categories are displayed when hovering over the buttons.

Question Filtering: After selecting categories, PMQF filters the available questions accordingly and displays the resulting set. Users are asked to review the questions and select and save those they identify as relevant for further investigation.

Question Customization: In the last phase, users conduct their final review by discarding or reformulating questions to fit their domain-specific terminology (Fig. 1b). We assume that the reformulation maintains the original question categorization. Additionally, PMQF generates a heatmap to visualize the range of selected questions across the classification schema.

After customizing the questions, users can either go back to the category selection (e.g., when they identified the need to cover further categories) or end the question design by exporting the identified and reformulated set of questions.

2.3. Question Refinement

The question refinement feature supports users in formulating concrete and understandable questions based on broad ideas. Users begin by entering their initial questions into an input form. After saving, their input appears on a second screen for reflection, where they are prompted to categorize the questions according to the existing categorization schema. This helps users refine their ideas to fit the categories and formulate them as direct questions.

We find the question refinement especially beneficial for project teams, allowing discussions and consensus on question formulation and categorization. Refined questions can be exported and PMQF stores a copy of the same export, enabling administrators to review and potentially add novel questions to the set of exemplary questions.

3. Evaluation and Maturity of the Tool

The three core features of PMQF run without any known errors. Additionally, we evaluated the tool in two sessions with (1) a leading international commercial vehicle manufacturer from Germany with initial PM experience through the analysis of one of their core manufacturing processes, and (2) a public sector organization with no prior PM experience, exploring the value of integrating it into their new BPM initiative. During the sessions, two representatives from each organization used the tool, aiming to design new analysis questions for their ongoing or planned PM projects. In both cases, the users were able to navigate the tool and use the features as expected. As part of the evaluation, the participants filled out the Technology Acceptance Model (TAM) [7]. Results are provided in Tab. 1. On average, the four participants evaluated the usefulness with 2.46 and the perceived ease of use with 2.33 on a 7-point Likert scale. However, they also pointed out that the usefulness largely depends on the quality and scope of the provided set of questions and the classification schema. For the evaluation, we used the one that is also available in the deployed version of PMQF.

Both organizations were able to successfully derive a set of relevant analysis questions for their projects which they planned to use further. Qualitative feedback addressed smaller aspects such as the use of colors, options to store reformulations at once, and saving the selection of categories for better traceability of the results. The feedback is already implemented in the current version of PMQF. Additionally, participants suggested enhancing the tool by adding more questions for specific domains and including guidelines for analyzing the questions.

TAM Items (average ratings per item are provided in brackets)	Avg. Rating
Usefulness Using PMQF would enable me to accomplish the design of analysis question more quickly. (3.00); Using PMQF would improve my performance in designing analysis questions. (2.75); Using PMQF would increase my productivity in designing analysis questions. (3.00); Using PMQF would enhance my effectiveness in PM project planning and question design.(2.25); Using PMQF would make it easier to design PM analysis questions. (2.00); I would find PMQF useful for designing PM analysis questions. (1.75)	2.46
Ease of Use Learning to operate PMQF would be easy for me. (2.25); I would find it easy to get PMQF do what I want it to do. (2.25); My interaction with PMQF would be clear and understandable. (2.50); I would find PMQF to be flexible to interact with. (2.75); It would be easy for me to become skillful at using PMQF. (2.25); I would find PMQF easy to use. (2.00);	2.33

Table 1
Average responses to the TAM on a 7-point Likert scale ranging from extremely likely (1), to extremely unlikely (7).

4. Conclusion and Outlook

Based on our knowledge, PMQF is the first tool providing practical support for analysis question design and refinement in PM. As such, it contributes to research by suggesting a standardization for these two tasks and thus enables higher consistency and comprehensiveness across projects. We believe that in the future, this can lead to more comparable and reliable project outcomes.

However, in its current version, PMQF is sensitive to the existence of a set of categorized

example questions and the respective classification schema. The deployed version we provide is optimized for one specific schema and runs on top of a collection of 405 analysis questions. Local installations can be adapted to custom input during setup.

In the future, we aim to provide a stable, universal categorization schema applicable for all PM domains. Additionally, PMQF could be further advanced in several directions:

1. **Integration of large language models (LLMs)** to enhance the question design and question refinement features.
2. **Integration of analysis guidance** by linking questions to relevant analysis techniques and provide hints for how to answer them (e.g., supported by GenAI [8] or integrated in PM tools).
3. **Requesting community feedback** in the form of ratings for questions or indications on whether questions were answerable and valuable to project teams in practice. Such information would help identify what constitutes a *good* analysis question and which types of questions are most frequently addressed in projects.

By integrating insights from the community and refining our approach with advanced technological capabilities, PMQF may be able to offer even more sophisticated and tailored support functions in the future.

Acknowledgments

References

[1] M. Dumas, M. La Rosa, J. Mendling, H. Reijers, Fundamentals of Business Process Management, Springer, 2018.

[2] F. Emamjome, R. Andrews, A. ter Hofstede, A case study lens on process mining in practice, in: On the Move to Meaningful Internet Systems: OTM 2019 Conferences, Springer, 2019.

[3] M. van Eck, X. Lu, S. Leemans, W. van der Aalst, Pm2: A process mining project methodology, in: Advanced Information Systems Engineering: CAiSE 2015, Springer, 2015.

[4] A. Mamudu, W. Bandara, M. T. Wynn, S. J. Leemans, A process mining success factors model, in: International Conference on Business Process Management, Springer, 2022.

[5] F. Zerbato, J. Koorn, I. Beerepoot, B. Weber, H. Reijers, On the origin of questions in process mining projects, in: EDOC 2022, Springer, 2022.

[6] L. Zimmermann, F. Zerbato, B. Weber, What makes life for process mining analysts difficult? a reflection of challenges, Software and Systems Modeling (2023) 1–29.

[7] F. D. Davis, Perceived usefulness, perceived ease of use, and user acceptance of information technology, MIS quarterly (1989) 319–340.

[8] A. Berti, D. Schuster, W. M. van der Aalst, Abstractions, scenarios, and prompt definitions for process mining with llms: a case study, in: International Conference on Business Process Management„ Springer, 2023, pp. 427–439.