Process Mining and the ProM Framework: An Exploratory Survey

Jan Claes and Geert Poels

Department of Management Information Science and Operations Management,
Faculty of Economics and Business Administration
Ghent University, Belgium
{jan.claes, geert.poels}@ugent.be

Abstract. In the last decade the field of process mining gained attention from research and practice. There is, however, not much known about the use and the appreciation of the involved techniques and tools, many of which are integrated into the well-known ProM framework. Therefore a questionnaire was sent out to ask people's opinions about process mining and the ProM framework. This paper reports on the answers and tries to link them to existing knowledge from academic literature and popular articles. It must be seen as a first, exploratory attempt to reveal the adoption of process mining and the actual use of the ProM framework.

Keywords: Process Mining, ProM Framework, Survey Research

1 Introduction

In the recently published Process Mining Manifesto [1] 11 challenges and 6 guidelines for future development of the process mining field are listed. The paper was authored by 77 researchers and practitioners in the context of the IEEE Task Force on Process Mining[1] and is therefore assumed to reflect the opinion of the process mining community. This provided the inspiration to compose a questionnaire[2] to be able to ask the community for their opinion on related topics. The survey comprised 5 questions about process mining and 5 questions about the most popular process mining framework ProM[3]. Another 5 questions covered the demographical background of the respondents.

This is how the paper is structured: Section 2 explains the methodology. Section 3 provides an overview of the main results of the questionnaire. Section 4 discusses the impact on research and practice.

[1] For more information we refer to http://www.win.tue.nl/ieeetfpm
[2] The questions can be consulted at http://processmining.ugent.be/survey.php
[3] For information and download we refer to http://www.promtools.org

2 Methodology

The intention of the research was to perform an *exploratory* study to reveal *perceptions* of process mining in general (i.e., the concept, its techniques and tools) and the ProM framework in particular. Therefore, we decided to not derive hypotheses from theory, but to formulate open, optional questions that we deemed relevant. In our opinion, this approach would result in getting more respondents. In total 90 people completed all 15 survey questions (43 more than a recent survey about process mining use cases [2]). Another advantage of open questions is that participants are less influenced to give certain predefined answers than in a multiple choice questionnaire. Getting more respondents and less influenced answers provides a certain degree of face validity to what can be learned from the survey.

The questionnaire was put online at 2012-3-18 and was closed at 2012-5-1. We approached possible respondents by mail and by social media (i.e., LinkedIn and Twitter). 90 respondents completed all questions. The survey had a maximum of 119 answers, a minimum of 28 answers, and an average of 97 answers per question. At 2012-4-7 we added 3 additional questions about the most popular plug-ins according to the results so far (see Section 3.6 and 3.8), for which we counted 48 respondents.

We refer to http://www.janclaes.info/papers/PMSurvey for an extended report. The dataset with the provided answers can be downloaded via the same link.

3 Results

3.1 Demographics

The collected demographical data shows that the respondents form a heterogeneous group (see Fig. 1).

- Almost half of the respondents *study* process mining (researchers and students), and half of them *use* the techniques for practical, commercial purposes (other categories).
- The age of respondents varies between 21 and 60 year, but has a high concentration between 25 to 35 year. It is possible that this correctly reflects the process mining community if mainly novice (younger) researchers and practitioners joined the community, because the field only exists for some 15 years.
- There were almost as much respondents that indicated having good and excellent knowledge as the number that indicated having intermediate and bad knowledge about process mining.
- Respondents use process mining techniques mainly for analyzing process quality and performance and for performing process audits.
- The survey attracted respondents from 26 countries with a high concentration in the Netherlands and Belgium. This high concentration can be explained by the fact that the authors are located in Belgium, but can also be a consequence of the community having a high concentration of members in these areas.

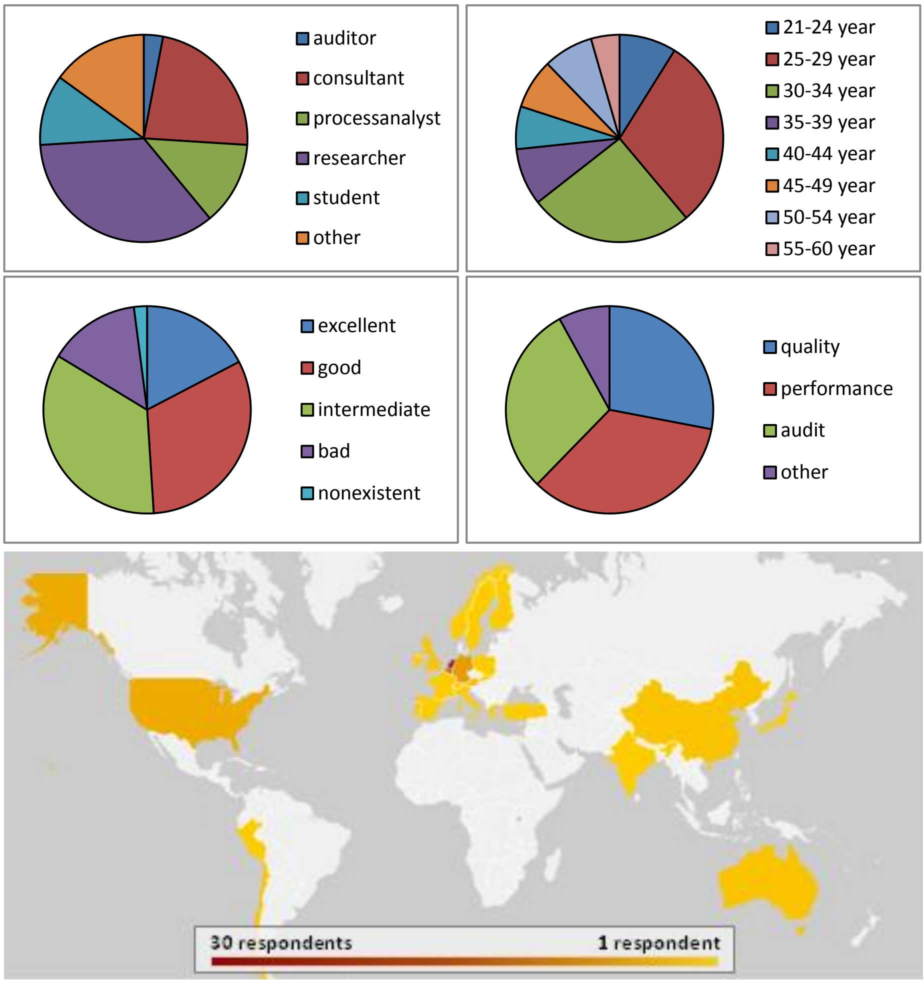

Fig. 1. Demographics of the respondents
(question 11 (98), 12 (87), 13 (98), 14 (90), 15 (90 respondents) [2], map by ammap.com)

3.2 Benefits and drawbacks of process mining techniques

Results

The chart in Fig. 2 shows the indicated *benefits* of process mining techniques. We grouped similar answers in clusters and picked a term from within the group of answers to define the cluster. The clear main perceived benefit is *objectivity* (the use of real process data assures a certain degree of objectivity of analyses). Some respondents focused on the application of the techniques and highlighted *conformance checking* and the possibility to *find causes* of certain characteristics of the process models as the most appreciated applications.

Discussion

A recent survey [2] asked the same question ("What do you think is the biggest benefit of process mining"). Apart from functional qualities, that study concluded that the main benefits are related to objectivity, accuracy, speed, and transparency. Our survey seems to confirm these answers.

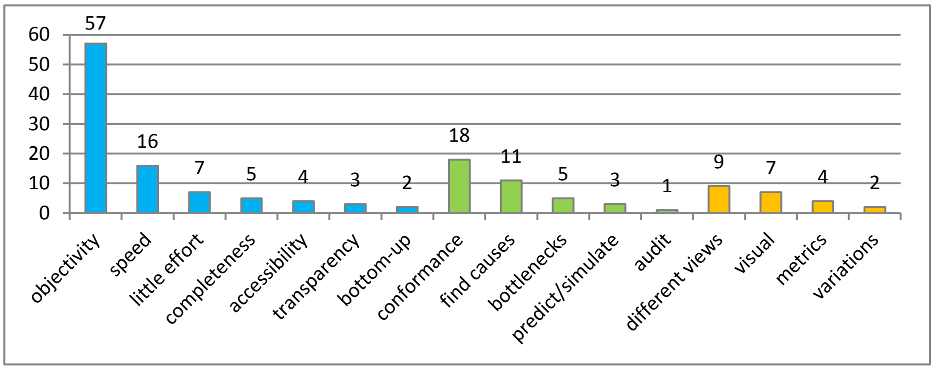

Fig. 2. Benefits of process mining techniques (question 4 [2], 94 respondents) (blue: characteristic, green: application, orange: representation)

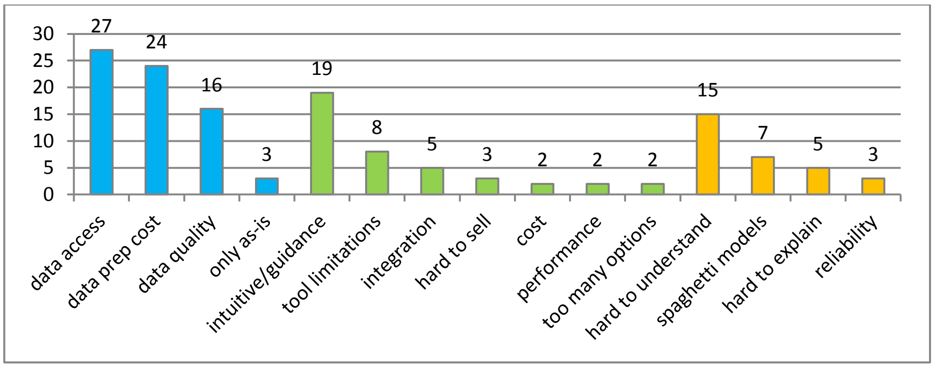

Fig. 3. Drawbacks of process mining techniques (question 5 [2], 90 respondents) (blue: input, green: techniques, orange: output)

Results

Perhaps more interesting for further research are the perceived *drawbacks* of process mining techniques (see Fig. 3). Most indicated drawbacks relate to data properties. It seems to be problematic to *find* and *prepare* the *right* data for process mining. The existing tools for process mining form another important indicated problem: They provide too little guidance and suffer from severe limitations. The current process mining tools also need to be (more) integrated with other tools and techniques. The output of process mining techniques (mostly discovered process models) is hard to understand (e.g., spaghetti models) and hard to explain.

Discussion

These main drawbacks are also addressed by the process mining manifesto [1]. Challenge 1 reports on the difficulties in finding, merging, and cleaning event data. Challenges 8 and 9 indicate that techniques and tools need to be more integrated with other analysis approaches. Challenges 10 and 11 concern the difficulty for non-experts to use and understand the techniques.

3.3 Tools for process mining

Results

The most popular process mining tool for research and practice is the ProM framework (see Fig. 4). Notice that the next three tools in the ranking are tools that help prepare event logs for the ProM framework. Another remarkable conclusion to draw is that the Disco tool - that was not officially released and only available to beta testers at the time of the survey - completes the top 5.

Discussion

We found three documented case studies in academic papers [3–5], they mentioned only ProM as a tool used in the study.

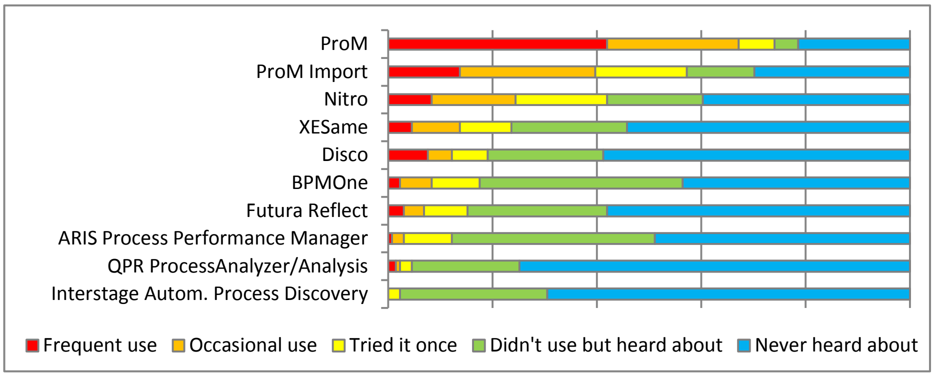

Fig. 4. Tools for process mining (question 2 [2], 119 respondents)

3.4 Benefits and drawbacks of ProM

Results

The main indicated benefit of the ProM framework is that it comes with many plug-ins (see Fig. 5). Also the fact that it is open source is perceived as a main benefit (whether this relates to the possibility to change or extend the software or to the fact it is a free tool is not clear). Another suggested benefit is the clear interface of the framework. The limited ease of use of the software is the main indicated drawback.

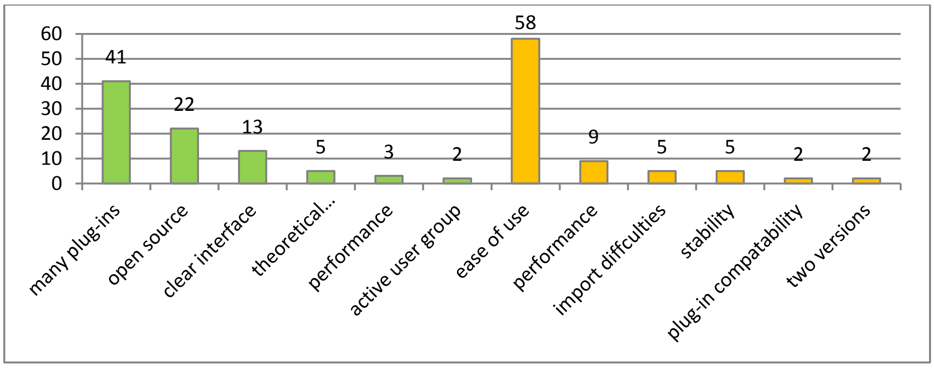

Fig. 5. Benefits and drawbacks of ProM (question 9 [2], 78 respondents) (green: benefits, orange: drawbacks)

3.5 Used versions of ProM

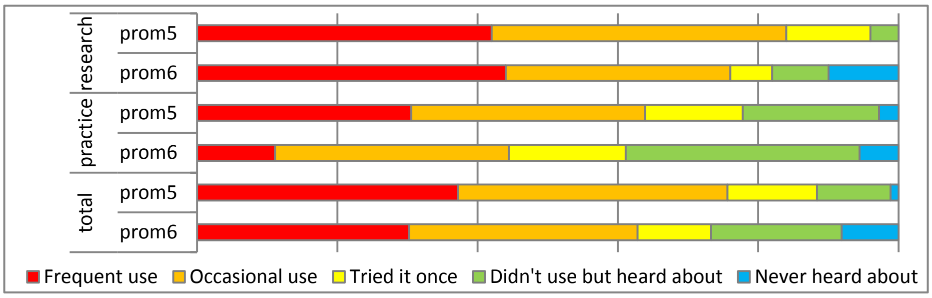

Fig. 6. Used versions of ProM (question 6 [2], 114 respondents)

Results

Fig. 6 shows information about the usage of the latest major ProM versions ProM 5 and ProM 6. Observe that for research (researchers and students) both tools are almost equally used, but 10% (5 out of 50 researchers) indicated to have never heard of ProM 6. For practice (consultants, process analysts and others) ProM 5 is still more used than ProM 6. A blog post from Fluxicon might reveal the reasons for practitioners to not switch to the newest version of ProM [6, 7]: “bugs are still found and fixed” and “a lot of plugins from ProM 5.2, (…), are missing at this point” in ProM 6 (see also Section 3.10).

3.6 Used process mining techniques in ProM 5

Results

A list of the most used plug-ins of ProM 5.2 is provided in Fig. 7. The most popular plug-ins are Fuzzy Miner [8], Heuristics Miner [9], Social Network Miner [10], Dotted Chart Analysis [11] and Alpha algorithm plugin [12].

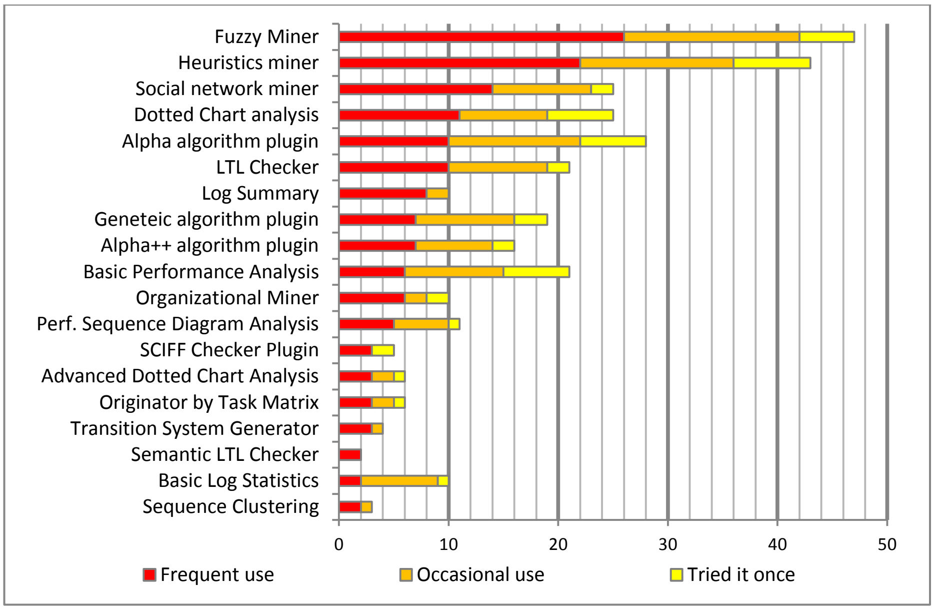

Fig. 7. Used process mining techniques in ProM 5.2 (question 7 [2], 115 respondents)

3.7 Evaluation of most used process mining techniques in ProM 5

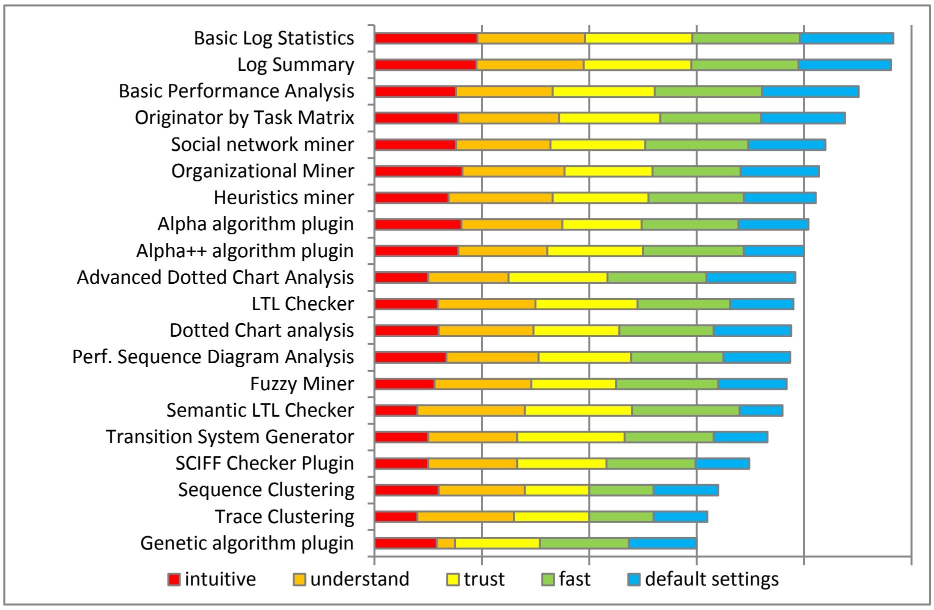

Fig. 8. Evaluation of most used process mining techniques in ProM 5.2 (question 16 & 17 [2], 48 respondents)

Results

Fig. 8 shows the perception of the respondents concerning 5 (possible) characteristics: intuitiveness of the technique, understandability of the results, trust in the result, speed of the technique and whether users adapt parameter settings. Each colored bar in Fig. 8 represents how many percent of the respondents classified the technique as belonging to the category that corresponds to the color of the bar. Notice that there seems to be no clear relation between the usage of a technique (Fig. 7) and its perceived characteristics (that, perhaps except for using default parameter settings, might be seen as quality indicators).

Discussion

In [13] strong indications are found that Heuristics Miner is "*the most appropriate and robust technique in a real-life context in terms of accuracy, comprehensibility, and scalability*" from a set of 8 investigated miners (see [13]). It scores better than Alpha miner and Genetic miner, which is also the case for our survey, although there are 6 miners that score even better than Heuristics Miner in our survey (see Fig. 8).

3.8 Used process mining techniques in ProM 6

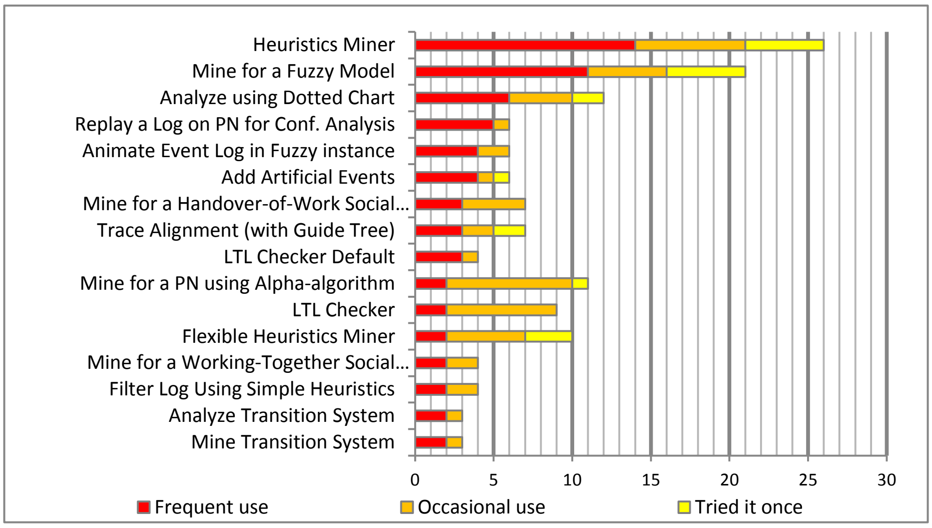

Fig. 9. Used process mining techniques in ProM 6.1 (question 8 [2], 115 respondents)

Results

Also in ProM 6.1 Heuristics Miner, Fuzzy Miner and Dotted Chart analysis are the most popular techniques. Furthermore, we observe a lot of popular plug-ins of ProM 5.2 (Fig. 7) are not in the list of ProM 6.1 (e.g., Genetic algorithm plug-in [14], Basic Performance Analysis). For some of them the reason is simply because they do not exist in ProM 6.1 (e.g., Basic Performance Analysis, see Section 3.10). For the others (e.g., Genetic algorithm plug-in) it is not clear why they are not popular in ProM 6.1.

3.9 Evaluation of most used process mining techniques in ProM 6

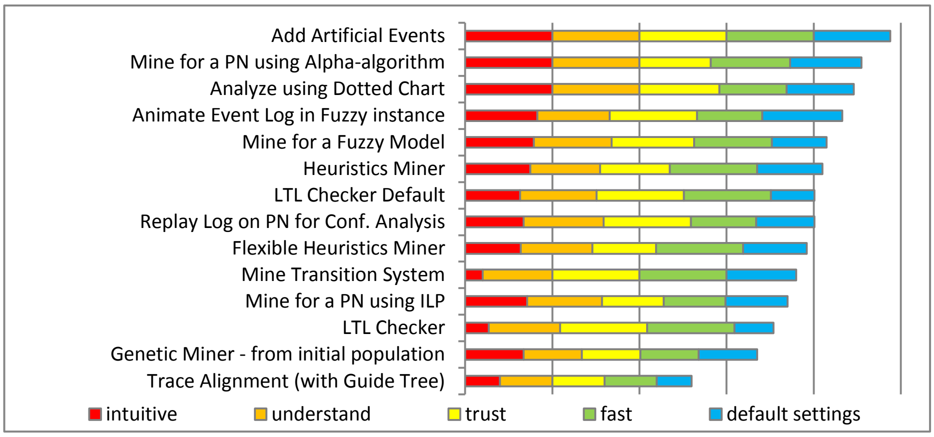

Fig. 10. Evaluation of most used process mining techniques in ProM 6.1 (question 18 [2], 48 respondents)

Results

Equally as for ProM 5.2 respondents were asked to indicate for ProM 6.1 if they feel the techniques and results are (i) intuitive, (ii) easy to understand, (iii) trustworthy, (iv) fast, and (v) if default options can be used. Next to the (subjective) perceived *trust* in the correctness of results, it is also important to investigate the (objective) theoretical correctness of the results. The need for a process mining evaluation framework for research and practice is discussed in [16].

Discussion

In Fig. 10 each colored bar shows how many percent of the respondents classified a plug-in in the category as indicated by its color. Add Artificial Events scored best, but this is only a very simple plug-in that adds an artificial start and/or end event in each trace. Trace Alignment (with Guide Tree) [15] scores poorly on all 5 categories. We recommend that involved developers take a closer look at the data in the extended report[4] to find out how their plug-in is evaluated and what can be improved.

[4] See http://www.janclaes.info/papers/PMSurvey/

3.10 Missing process mining techniques in ProM 6

Results

The question about which plug-ins are missing in ProM 6.1 suggested that this could be existing plug-ins from ProM 5.2 that are not included in ProM 6.1 or new plug-ins that never existed in the ProM framework. Table 1 summarizes the answers. Notice that only 28 respondents answered this question. Some specific plug-ins of ProM 5.2 (see left column of Table 1) are missed in ProM 6.1 (e.g., advanced filters and performance plug-ins). Respondents also would like to have better versions of some techniques (e.g., performance and discovery plug-ins) and a log or model editor is requested.

Table 1. Plug-ins that are missed in ProM 6.1 (question 10 [2], 28 respondents)

Existing plug-ins from ProM 5.2	*New or enhanced plug-ins*
Advanced filters (5x)	Robust performance analysis (2x)
Conformance Checker (3x)	Log/Model editor (2x)
Basic Performance Analysis (3x)	Security analysis (2x)
Performance Sequence Diagram Analysis (2x)	*Better* process discovery techniques
Alpha Algorithm(s)	*Better* performance analysis plugin
Trace Clustering	Medical analysis plug-in
Region Miner	Self organising maps
Pattern Sequence Analyser	Export to image option in *all* plugins

4 Discussion

This paper presents the results of a survey that was conducted as an exploratory study of the perceptions of process mining in general and the ProM framework in particular that are held by the process mining community of researchers and practitioners. The focus is clearly on relevance rather than rigor. "*Overemphasis on rigor in behavioral IS research has often resulted in a corresponding lowering of relevance*" [17]. We think the relevance of this work is proven (i) by the number of visits to the web page (336 different browser sessions were registered between 18 March and 30 April), (ii) by the number of respondents (there was no incentive to participate, yet 90 respondents completed the whole questionnaire) and (iii) by the number of people that indicated they like to receive a report on the results (87 respondents). Especially the data on which plug-ins are used seems to be most interesting (115 respondents).

However, there is still a need to focus on a number of methodological issues that should be taken into account when this exploratory study is replicated in a more rigorous way. First, the questions were formulated without specific hypotheses in mind. Most questions are formulated as open question with the risk of misinterpretation of individual answers. We argue that in a multiple choice setting, the interpretation is done by the respondent and therefore the risk of misinterpretation exists on the side of the respondent.

Second, all questions were optional. The result is that some questions were answered by many respondents (a maximum of 119) and others had few answers (a minimum of 28). This means that not all questions offer the same confidence in their results. This also causes difficulties when linking the answers of different questions to each other. For example, in section 3.5 the answers were divided in two groups according to the indicated profession of the respondent (research and practice). For this question only 86 of the 93 answers of respondents could be included, because the other 7 respondents did not indicate their profession.

Third, some anomalies were determined. We discovered that at least two respondents filled in some questions more than once. Because no personal information was collected, only the respondents that used the same browser session to complete the survey could be detected, so in reality there might be more than two duplicate sets of answers.

The results of the survey have several important implications. For *research*, it provides preliminary insights in the perception of the process mining domain and the ProM framework. Researchers can derive hypotheses from the results of this survey that have to be examined by other, more rigorous research. At least two very relevant research questions emerge from this study. Section 3.4 shows the need for a better, faster and cheaper support of the preparation phase of a process mining effort (i.e., finding and structuring the necessary data). The comparison of the results described in Section 3.6 and 3.8 and the results described in Section 3.7 and 0 indicates that *popular* process mining techniques are not considered to be *better* and vice versa. Is this correct? Why is this so?

For *practice*, this study can help users of the ProM framework to find their way in the long list of available plug-ins. Fig. 7 and Fig. 9 summarize which plug-ins are most used and Fig. 8 and Fig. 10 provide insight in which plug-ins are most appreciated.

In particular for *developers*, the survey points to a number of very specific shortcomings of the currently available ProM plug-ins. First, Table 1 contains a list of programming updates that are desired by the respondents. Second, from the chart of indicated benefits and drawbacks of the ProM framework in Fig. 5 it is clear the ease of use of the plug-ins should be improved. Also Section 3.2 points to the lack of intuitivism and guidance. We suggest more effort can be made to create a user friendly user interface (e.g., tooltip help for each parameter setting) and to provide (better) documentation.

To conclude, we would like to stress that this survey must be seen in the right context. Although we admit that we cannot guarantee full reliability of the data, we are convinced that this paper offers an interesting novel view on the community's perception of process mining and the ProM framework. This exploratory study forms the base to formulate hypotheses to be rigorously corroborated by future research.

Acknowledgements. We would like to thank Joel Ribeiro for his help with the selection of questions for this survey.

5 References

1. Van der Aalst, W.M.P., Adriansyah, A., Karla, A., de Medeiros, A.K.A., Arcieri, F., et al.: Process Mining Manifesto. Proc. BPM ’11 Workshops, LNBIP 99. pp. 169-194 Springer (2011)
2. Ailenei, I., Rozinat, A., Eckert, A.: Definition and Validation of Process Mining Use Cases. Proc. BPM ’11 Workshops, LNBIP 99. pp. 75-96 Springer (2011)
3. Mans, R., Schonenberg, M., Song, M., Van der Aalst, W.M.P., Bakker, P.J.M.: Process mining in healthcare: a case study. Proc. BIOSTEC ’08. pp. 425-438 Springer (2008)
4. Van der Aalst, W.M.P., Reijers, H.A., Weijters, A.J.M.M., Van Dongen, B.F., de Medeiros, A.K.A., Song, M., Verbeek, H.M.W.: Business process mining: An industrial application. Information Systems. 32 (5), pp. 713-732 (2007)
5. Rozinat, A., De Jong, I., Günther, C.W., Van der Aalst, W.M.P.: Process mining applied to the test process of wafer scanners in ASML. Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on. 39 (4), pp. 474-479 (2009)
6. Rozinat, A., Günther, C.W.: Why We Hate ProM 6, http://fluxicon.com/blog/2010/11/why-we-hate-prom-6
7. Rozinat, A., Günther, C.W.: Why We Love ProM 6, http://fluxicon.com/blog/2010/11/why-we-love-prom-6
8. Günther, C.W., Van der Aalst, W.M.P.: Fuzzy mining–adaptive process simplification based on multi-perspective metrics. Proc. BPM ’07, LNCS 4714. pp. 328-343 Springer (2007)
9. Weijters, A.J.M.M., Van der Aalst, W.M.P.: Process mining with the heuristics miner-algorithm. Technische Universiteit Eindhoven, Tech. Rep. WP 166. pp. 1-34 (2006)
10. Van der Aalst, W.M.P., Song, M.: Mining Social Networks: Uncovering interaction patterns in business processes. Proc. BPM ’04, LNCS 3080. pp. 244-260 Springer (2004)
11. Song, M., Van der Aalst, W.M.P.: Supporting process mining by showing events at a glance. Proc. WITS ’07. pp. 139-145 (2007)
12. Li, J., Liu, D., Yang, B.: Process Mining: Extending α-Algorithm to Mine Duplicate Tasks in Process Logs. Proc. APWeb-WAIM ’07, LNCS 4537. pp. 396-407 Springer (2007)
13. De Weerdt, J., De Backer, M., Vanthienen, J.: A multi-dimensional quality assessment of state-of-the-art process discovery algorithms using real-life event logs. Information Systems. 37 (7), pp. 654-676 (2012)
14. de Medeiros, A.K.A., Weijters, A.J.M.M.: Genetic process mining: an experimental evaluation. Data Mining and Knowledge Discovery. 14 (2), pp. 245-304 (2007)
15. Bose, J.C., Van der Aalst, W.M.P.: Process diagnostics using trace alignment: opportunities, issues, and challenges. Information Systems. 37 (2), pp. 117-141 (2011)
16. Rozinat, A., de Medeiros, A.K.A., Günther, C.W.: The need for a process mining evaluation framework in research and practice: position paper. Proc. BPM ’07 Workshops, LNCS 4928. pp. 84-89 Springer (2008)
17. Hevner, A.R., March, S.T., Park, J., Ram, S.: Design science in information systems research. Mis Quarterly. 28 (1), pp. 75-105 (2004)