# Requirement Specification Document

Label Refinement by Behavioral Similarity

Document owners:
Bianka Bakullari
Christopher Beine
Nicole Ventsch
Juan

Last edited: April 25, 2019

## Contents

## 5.1 Open Issues

Our team has a rough idea of the steps needed to implement the algorithm and the interface. However, we are unsure of the relative time needed for each of these parts. Many unexpected bugs in our algorithm could hinder the time we have in our disposal to design a user-friendly web-interface.
While we will try to achieve writing an efficient algorithm, there are no concrete performance measures required. Drawbacks will probably be found during implementation and decisions will have to be made on the spot.
Since each of the members has to be fully aware of and actively participate in the coding process, it is still unclear how the functionalities will be divided between the group members. Whether or not some functionalities should be implemented together or individually depends on type and importance of functionality, personal schedules, deadlines and so on.

## 5.2 Off-the-Shelf Solutions

There has already been an implementation of the algorithm in ProM. This might be helpful in the designing steps of our algorithm.

We will use a code written in Java to automatically generate events logs.
@juan: Where did you get the java code you showed us once?

### 5.3 New Problems

The Label Refinement algorithm intends to give the user an alternative event log on which process discovery algorithms can be applied. Whether or not the models resulting from the new refined log have better precision or fitness compared to the original models is beyond our scope. Since the result also depends on the thresholds and set of candidate activities which are chosen by the user, we assume that the user has some background on Process Mining and has clear intentions for trying to work with a new refined log. The algorithm does not filter out events or features in the data.
A common mistake the user could make is upload an event log in the .csv or .xlsx format instead of .xes format.

### 5.4 Tasks

### 5.5 Migration to the New Product

### 5.6 Risks

| Risks | Description | Category | Mitigation |
|---|---|---|---|
| Inaccurate expectations. | Stakeholders develop inaccurate expectations (believe that the project will achieve something not in the requirements, plan, etc). | Stakeholder | Clearly state in the requirement documentation what are the deliverables meant to be done and the scope of the project. |
| Process inputs are low quality. | Inputs from stakeholders that are low quality (e.g. business case, requirements, change requests). | Stakeholder | Kindly ask the stakeholder for a more detailed and clearer version of any input they may provide i.e., requirements, business cases. |
| Misunderstood requirements. | When requirements are misinterpreted by the project team. | Communication | Meet with the stakeholders and discuss the requirements again until the team is sure that they have completely understood them. |
| Learning curves. | Project team needs to acquire new skills for the project. | Team | Motivate the project team, give them the best practices on the IT field and make experts instruct them using their knowledge and own experience. |
| Integration failure. | Product components will fail to integrate with each other. | Integration | Establish standards for product development and make sure that the individual components passed flawlessly the unit test. |
| Requirements are incomplete. | Requirements are not fully captured or are overlooked. | Requirements | Make a peer-review of the requirement documentation and make sure that nothing is being left out. |

## 5.7 Costs

## 5.8 User Documentation

1. Technical documentation:
   - Software code documentation.
   - Technical specifications.
2. User documentation including:
   - How to use the UI.
   - Examples of inputs and outputs.
   - Explanation of error messages.
   - Information to contact the developers (in case of further questions).

## 5.9 Waiting Room

- Additional feature which enables the user to choose a Business Process Model Discovery (BPMD) technique to visualize the resulting process model and to pick the one which is considered to be the best one (according to user's expertise).
- Additional feature that allows for the automatic detection of "imprecise labels" by using properties of the Inductive Miner (IM).

# References

[1] Lu, Xixi, et al. "Handling duplicated tasks in process discovery by refining event labels." International Conference on Business Process Management. Springer, Cham, 2016.