# Managing Multi-Lingual User Feedback: the SUPERSEDE project experience

Fitsum Meshesha Kifetew
kifetew@fbk.eu

Anna Perini
perini@fbk.eu

Angelo Susi
susi@fbk.eu

Fondazione Bruno Kessler, Trento, Italy

## Abstract

[Context & Motivation] In the SUPERSEDE project, methods and tools have been developed to collect and analyze user feedback, to identify relevant information for deciding which are the most important requirements to be considered for the next release of a product. [Question/problem] Even if the project proposal was to analyze feedback in the English language only, later it emerged that there was a need to analyze multi-lingual (German, English) feedback. [Principal ideas] We considered two different solutions: 1) translating user feedback from German to English, and processing it with the techniques developed for the English language; 2) exploiting Natural Language Processing (NLP) techniques for German to analyze directly the feedback in German. [Contribution] In this short report we describe this project experience, summarizing main commonalities and differences between the aforementioned solutions.

The SUPERSEDE[1] (SUpporting evolution and adaptation of PERsonalized Software by Exploiting contextual Data and End-user feedback) project is an H2020 research and innovation project that proposes a feedback-driven approach to the life cycle management of software services and applications, with the ultimate purpose of improving users' quality of experience. The SUPERSEDE tool-suite, provides feedback-gathering and monitoring tools that allow to collect data concerning user experience. These data are then analyzed with the purpose of obtaining relevant information for deriving software requirements, to be taken into account by developers when making decisions on software evolution, such as which (new) requirements to consider and with which priority [MMK+17, BKM+17]. and to plan for the next release.

The tools developed in the project are validated on three industrial use cases, representing software applications for different application domains, whose users can provide textual feedback in different languages. When performing the industrial validation of the analysis tools for textual user feedback developed in the project, a multi-lingual (i.e. German, English) issue arose and needed to be taken into account. This multi-lingual aspect was not initially part of the objective of the project, i.e., all user feedback was assumed to be in English; however as the project progressed, it became clear that we needed to address it anyways. In this short report, we focus on SEnerCon, one of the use cases, which runs a web application in the domain of household energy saving, called interactive Energy Saving Account - iESA, currently deployed on the German market. In particular, we summarize key aspects of the two different solutions implemented in the project to address the issue of multi-lingual feedback analysis, namely (i) building analysis tool for textual feedback in English, and then validating them on

[1]Project started in May 2015 and will end in April 2018. Website: www.supersede.eu

textual feedback from the industrial case study translated from German to English; (ii) building analysis tools directly for German textual feedback.

The implementations of both solutions rest on a similar process that includes the following steps: (1) *dataset preparation*, where manual annotation of feedback messages by *type* and *sentiment* is performed by a domain expert. Type includes the following labels: *Bug Report*, *Feature Request*, *Enhancement Request*, and *Other*, while sentiment is labeled as *negative*, *neutral*, and *positive*; (2) *pre-processing*, where uninformative tokens are removed; (3) *feature extraction*, where different linguistic properties and sentiment are extracted; (4) *feedback classification*, where machine-learning techniques are employed to train a classifier on a (portion of the) dataset.

Among the main differences in the implementations of the two solutions are: (a) an additional activity for the *dataset preparation* step was requested for the first solution, that is the feedback was translated from German to English by a domain expert in SEnerCon; (b) in the *feature extraction*, different type of features were extracted in the two solutions, in particular for the first solution combinations of the speech-acts used in the messages were extracted, by applying a novel technique that was developed for English text [MKP17]. Moreover, since feedback data were scarcely available at the beginning of the project, we have used openly available datasets that closely mimic the characteristics of the feedback data we analyze. In particular, we used user feedback from the issue tracking system of the OpenOffice Writer application, which were available in English. Since the second solution was implemented later in the project, we used directly the dataset of feedback in German from the SEnerCon use case which were collected during the second year of the project.

Applying the two approaches, we were able to obtain reasonable results, considering the fact that the datasets available were very much limited in size. In particular, for the first solution (translating to English), the dataset from SEnerCon was composed of 575 messages translated to English from German and annotated by domain experts. On this dataset, we obtained classification accuracy of 83%. Similar results were also achieved for sentiment. On the other hand, for the second solution (directly analyzing feedback in German), the dataset was composed of 600 messages in German annotated by domain experts. The accuracy of the analysis was 59.20% for classification and 65.81% for sentiment. It is important to note here that the underlying machine learning techniques applied in the two approaches are also different. However, in both cases the size of the dataset is quite small. Hence, when in the future when more user feedback data becomes available, the accuracy of the trained models is expected to improve.

In conclusion, the decision regarding the two approaches depends, among other things, on availability of resources and the intended application of the tool. If the required expertise, domain and language knowledge are available in house at the time of the development of the analysis tools and potentially in future use of the tool, then implementing the analysis tools to work directly on the feedback messages in the original language (e.g., German) is the optimal choice. Otherwise it is useful to consider the application scenario of the analysis tool as well. If models are built once from the dataset and then used afterwards without the need for continuous update, then adopting the option of translating to English may be considered.

# References

[BKM⁺17] Paolo Busetta, Fitsum Meshesha Kifetew, Denisse Muñante, Anna Perini, Alberto Siena, and Angelo Susi. Tool-supported collaborative requirements prioritisation. In *COMPSAC (1)*, pages 180–189. IEEE Computer Society, 2017.

[MKP17] Itzel Morales-Ramirez, Fitsum Meshesha Kifetew, and Anna Perini. Analysis of online discussions in support of requirements discovery. In *Advanced Information Systems Engineering - 29th International Conference, CAiSE 2017, Essen, Germany, June 12-16, 2017, Proceedings*, pages 159–174, 2017.

[MMK⁺17] Itzel Morales-Ramirez, Denisse Muñante, Fitsum Meshesha Kifetew, Anna Perini, Angelo Susi, and Alberto Siena. Exploiting user feedback in tool-supported multi-criteria requirements prioritization. In *25th IEEE International Requirements Engineering Conference, RE 2017, Lisbon, Portugal, September 4-8, 2017*, pages 424–429, 2017.