Research on NLP for RE at the FBK-Software Engineering Research Line: A Report

Fitsum Meshesha Kifetew
Fondazione Bruno Kessler (FBK)
kifetew@fbk.eu

Anna Perini
Fondazione Bruno Kessler (FBK)
perini@fbk.eu

Angelo Susi
Fondazione Bruno Kessler (FBK)
susi@fbk.eu

Abstract

In this short report paper, we introduce the Software Engineering research unit at Fondazione Bruno Kessler, and summarise the research carried out in the area of requirements engineering for which natural language processing techniques have been exploited to build tools at support of software engineers. Ongoing and longer term research objectives are briefly outlined.

CEUR-WS.org/Vol-2376/NLP4RE19_paper12.pdf

1 Team Overview

The Software Engineering (SE) research unit located at Fondazione Bruno Kessler (FBK)[1] has been conducting research in the last fifteen years on techniques, tools and methods that aim at supporting practitioners when they perform key software development activities, namely Requirements Engineering (RE), code analysis and testing. The SE unit has been contributing to several projects funded by the European Union, and most of the research activities are performed in collaboration with national and international universities and research centers[2].

Natural language textual artefacts are typically produced during software engineering activities, e.g., from domain and system requirements documents, to comments in the code, thus Natural Language Processing (NLP) techniques have been widely investigated and exploited in the novel techniques developed in the group. Moreover, at FBK a research unit working in the NLP area is active since more than thirty years, thus offering the opportunity to get expert advice on state of the art research results in this area.

2 Past Research on NLP for RE

In this short report we focus on research studies in the area of RE, where we developed methods and tools that exploit NLP techniques. Main works are summarised in Table 1, which reports the type of artefact which is the object of analysis, the RE task that have been considered, the NLP technique used in the research, the application domain and the specific use case.

Copyright © 2019 by the paper's authors. Copying permitted for private and academic purposes.

[1] https://www.fbk.eu
[2] https://se.fbk.eu

Table 1: RE research using NLP techniques at SE-FBK

NL Artefact	RE Task	Technique	Application Domain	Use Case	Ref.
Requirements documents, free textual NL in English	Semi-structured specification of Requirements	Rule-based and Controlled Natural Language	European Railways Signaling System	Validation and verification of requirements specifications	[CRST12, CRST11]
Online discussion, as in user forum. Thread of textual messages in English	Elicitation of Requirements' relevant information	Speech-Act based analysis techniques, ML classification algo.	OSS Software development	Stakeholder feedback analysis for software maintenance and evolution in OSS Requirements management	[MPC14, MRKP18]
User-feedback, short textual messages in English	Elicitation of Requirements relevant information	Sentiment analysis and Speech-Act based analysis techniques, ML classification algo.	Home energy management apps	Requirements management for software evolution	[MRKP18]
User-feedback, short textual messages in German	Elicitation of Requirements relevant information	Sentiment analysis, ML classification algo.	Home energy management apps	Requirements management for software evolution	[KPS18]

2.1 Generating semi-formal requirements specification from requirements descriptions expressed as unstructured text.

The first activity (summarized in the first row in Table 1) refers to the work on the formalization of the European railways signaling system specifications that has been done in the context of the EURAILCHECK ERA project[3]. A key result of this work is a methodology for the analysis and verification of requirements specifications that consists of three main steps. The first step is devoted to the identification of specific patterns in the requirements specification documents, consisting of tens of pages of unstructured textual descriptions. This step allows to identify different categories of specifications such as glossary, describing an entity of the railway domain, or behavior, describing the behavior of the entities in the domain. Dependency relationships among those categories are also identified. The output of this step is a set of categorized requirements fragments. The second step focuses on the formalization of requirements fragments into Linear Temporal Logic (LTL) formulae. The output of this step is a set of formalized requirements fragments. Finally, the third step has the purpose to enable the verification and validation of the formalized requirements fragments (e.g. portions of the signaling system's specifications) via formal techniques, such as model checking [CRST12, CRST11].

The implemented approach includes a set of rules for the manual annotation and categorization of textual chunks, and for the translation of these chunks into LTL formulae. For example, "*The speed of the Train at the end of the Movement Authority must be 0*" is a chunk of specifications that involves the domain entities *Train* and *Movement Authority* and describes a behavioral property of the train, which is translated into the formal counterpart: "**always** (Train.position = MovementAuthority.end **implies** Train.speed = 0)". In case the model-checking identify inconsistencies associated to formalized requirement fragments, they can be traced back to the corresponding categorized requirement fragments, and then up to the original informal requirements, so that the requirements engineer can address the identified flaw.

2.2 Automated analysis of online discussions for RE purposes.

The research work on requirements elicitation from online discussions was started in the context of a PhD work [Mor15], and it has been further extended in the SUPERSEDE project[4], when developing automated anal-

[3] https://es-static.fbk.eu/projects/eurailcheck/

[4] An EU H2020 funded research and innovation project, https://www.supersede.eu

ysis techniques of user-feedback for RE purposes [MRKP18]. The objective of the PHD research was to investigate if online discussions about software, such as open-source software (OSS) user forums, contain requirements relevant information, which could be useful for improving the software; and to develop a novel linguistic technique to automate the analysis of large online discussion datasets in English. Among the main results of this work are an ontology of online user-feedback [MPG15] and a new linguistic technique for the analysis of online discussions which is based on *speech-act theory* [MPC14]. The speech-acts considered for the proposed technique are taken from a taxonomy originally proposed by Bach and Harnish [BH79], for the linguistic community, that we revised and adapted for the software development domain. The speech-act based analysis technique builds on the hypothesis that textual messages sent by software users are meant to suggest a new feature, or to request enhancements of an existing functionality, or simply to complain. The idea is that the speech-acts contained in these messages can be used as indicators for identifying the different types of user requests. For instance, from an experiment on a dataset of 161, 120 messages from online discussions in the Apache OpenOffice issue tracking system, we found that speech-acts of types *Requestive, Requirement,* and *Accept* are strongly present in user-feedback messages concerning the request of new functionalities or the enhancement of existing ones. The implementation of the speech-act based analysis technique for online discussions includes: i) preprocessing steps, which extract online discussion threads from the issue tracking system as XML files which are cleaned, parsed and stored in a MySQL database; ii) a set of 142 lexico-syntactic rules for the annotation of twenty speech-acts, which exploit gazetteers containing verbs related to each speech-act. The rules are implemented in GATE [CMB+11]; iii) properties extracted from the resulting annotated textual messages, which are used for building a classifier that provides an automatic classification of feedback as bug report, enhancement or new feature requests. For the classification, three ML algorithms (J48, SMO, and Random Forest) available in Weka[5] library have been used. These three algorithms were used because performing a preliminary assessment revealed that for the given dataset they gave the best performances.

2.3 Automated analysis of online user-feedback for software maintenance and evolution.

A follow-up of the above described research was developed in the context of the SUPERSEDE project, whose overall objective was the development of a highly configurable toolsuite that enables a user-feedback driven approach to software evolution. To realize such a software evolution approach, one key problem concerns supporting a development team in eliciting bug-fix, new feature or feature enhancement requests from large sets of textual user-feedback. To address this problem, we developed a user-feedback analysis tool that combines the previously mentioned speech-act based analysis technique with sentiment analysis, aimed at automating the analysis of large sets of textual feedback. We obtained encouraging results when validating it on the project's industrial use cases, in particular feedback messages from end users of a household energy management software application. For example, when applied to a dataset containing 575 feedback messages (translated to English from German, as the software has a predominantly German speaking user base) from the users of the energy management software, *enhancement requests* were classified with an accuracy of 86% precision, and 77% recall (F-measure 0.81) [MRKP18].

In order to avoid the effort demanding task of translating messages from German to English, we developed a classifier for textual feedback expressed in German. The classifier has been trained on a dataset containing 600 German messages, which have been annotated with respect to sentiment and feedback category by the software developers. The accuracy of the resulting classifier amounts to 59% [KPS18].

3 Ongoing Research Exploiting NLP for RE

Based on our experience in the SUPERSEDE project we are currently implementing and validating a tool that will support user-feedback driven requirements prioritization. In a nutshell, the idea is to identify which software requirements are affected by user-feedback, by exploiting text similarity techniques enriched with domain ontologies. An estimate of the *value perceived by users* of the software requirements is computed by using properties of the user-feedback that we can compute with the analysis techniques we developed in SUPERSEDE [MRKP18]. Among such properties we consider are: topics, sentiment, intention of the user as determined by the speech-acts, and severity of the issue reported in the feedback.

The longer term research objective is that of integrating this tool in multi-criteria requirements prioritization tools, such as the one we developed in SUPERSEDE [KMP+17]. A research preview is presented in [MMK+17].

[5] https://www.cs.waikato.ac.nz/ml/weka/

A related research is ongoing in the App development domain, where key artifacts are users' App reviews [DLPS19]. Also in this case the idea is to focus on RE decision-making tasks, which can be supported by automated analyses of textual feedback. We will combine NLP techniques to process App reviews with unsupervised learning techniques, design and execute empirical studies with intended users of the proposed tool, i.e. software developers or requirements engineers, to evaluate its effectiveness and usefulness. As a longer term objective, we are considering to revisit our research on conceptual modeling for RE within data-driven engineering perspective. Specifically, we intend to investigate how to exploit the analysis of textual artifacts expressed in natural languages (e.g. documents, and stakeholders' feedback) to elicit model elements as well as to validate models.

References

[BH79] Kent Bach and Robert M. Harnish. *Linguistic Communication and Speech Acts*. MIT Press, Cambridge, MA, 1979.

[CMB+11] Hamish Cunningham, Diana Maynard, Kalina Bontcheva, Valentin Tablan, Niraj Aswani, Ian Roberts, Genevieve Gorrell, Adam Funk, Angus Roberts, Danica Damljanovic, Thomas Heitz, Mark A. Greenwood, Horacio Saggion, Johann Petrak, Yaoyong Li, and Wim Peters. *Text Processing with GATE (Version 6)*. 2011.

[CRST11] Alessandro Cimatti, Marco Roveri, Angelo Susi, and Stefano Tonetta. Formalizing requirements with object models and temporal constraints. *Software and System Modeling*, 10(2):147–160, 2011.

[CRST12] Alessandro Cimatti, Marco Roveri, Angelo Susi, and Stefano Tonetta. Validation of requirements for hybrid systems: A formal approach. *ACM Trans. Softw. Eng. Methodol.*, 21(4):22:1–22:34, 2012.

[DLPS19] Jacek Dabrowski, Emmanuel Letier, Anna Perini, and Angelo Susi. Finding and analyzing app reviews related to specific features: A research preview. In *To appear in 24th International Conference on Requirements Engineering: Foundation for Software Quality (REFSQ 2019)*, 2019.

[KMP+17] Fitsum Meshesha Kifetew, Denisse Muñante, Anna Perini, Angelo Susi, Alberto Siena, and Paolo Busetta. DMGame: A Gamified Collaborative Requirements Prioritisation Tool. In *25th IEEE International Requirements Engineering Conference, RE 2017, Lisbon, Portugal, September 4-8, 2017*, pages 468–469. IEEE Computer Society, 2017.

[KPS18] Fitsum Meshesha Kifetew, Anna Perini, and Angelo Susi. Managing multi-lingual user feedback: The SUPERSEDE project experience. In *Joint Proceedings of REFSQ-2018 Workshops, Doctoral Symposium, Live Studies Track, and Poster Track co-located with the 23rd International Conference on Requirements Engineering: Foundation for Software Quality (REFSQ 2018), Utrecht, The Netherlands, March 19, 2018.*, volume 2075 of *CEUR Workshop Proceedings*. CEUR-WS.org, 2018.

[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.

[Mor15] Itzel Morales-Ramirez. *Exploiting Online User Feedback in Requirements Engineering*. PhD thesis, University of Trento, Italy, 2015.

[MPC14] Itzel Morales-Ramirez, Anna Perini, and Mariano Ceccato. Towards supporting the analysis of online discussions in OSS communities: A speech-act based approach. In *Information Systems Engineering in Complex Environments - CAiSE Forum 2014, Thessaloniki, Greece, June 16-20, 2014, Selected Extended Papers*, pages 215–232, 2014.

[MPG15] Itzel Morales-Ramirez, Anna Perini, and Renata S. S. Guizzardi. An ontology of online user feedback in software engineering. *Journal of Applied Ontology*, 10(3-4):297–330, 2015.

[MRKP18] Itzel Morales-Ramirez, Fitsum Meshesha Kifetew, and Anna Perini. Speech-acts based analysis for requirements discovery from online discussions. *Information Systems*, 2018.