Exploiting Online Discussions in Collaborative Distributed Requirements Engineering

Itzel Morales-Ramirez, Matthieu Vergne, Mirko Morandini, Anna Perini, and Angelo Susi

Fondazione Bruno Kessler
Via Sommarive 18, 38123 Trento, Italy

Abstract. Large, distributed software development projects, like Open Source Software (OSS), adopt different collaborative working tools, including online forums and mailing list discussions that are valuable source of knowledge for requirements engineering tasks in software evolution, such as model revision and evolution. In our research, we aim at providing tool support for retrieving information from these online resources, and for analyzing it. The solution we propose combines natural language processing techniques, machine learning, statistical and search based techniques to address two key problems, namely the so called expert finding problem, and the problem of identifying requests for changing requirements or soliciting new requirements, by exploiting online discussions.
In this paper, we describe the solution approach set up so far with the help of an OSS scenario, discuss some preliminary evaluation, and highlight future work.

1 Introduction

Large and distributed software development projects, such as Open Source Software (OSS), challenge traditional software development paradigms and call for tools to support distributed requirements elicitation, modeling, requirements negotiation, and the management of distributed teams [10]. OSS projects usually adopt online community platforms to support cooperative work. Platforms such as online forums and mailing lists where different types of stakeholders (e.g., users, developers, and analysts) share their knowledge by engaging in discussions, mainly using free or semi-structured Natural Language (NL) text. The level of expertise, on specific topics, of a particular stakeholder may be revealed by an analysis of these discussions. This information is key for an effective and successful software evolution process. While these discussions grow, spreading over a variety of interconnected domain concepts, and involving more and more stakeholders, manual analysis becomes rapidly an effort demanding and error prone task.

Automated support for extracting relevant information from these online discussions, and for identifying and ranking those members who can contribute key knowledge about a given topic, is crucial to improve collaborative Requirements Engineering (RE) and to provide better support to project management [4, 1]. In our research we are combining Natural Language Processing (NLP) techniques, Machine Learning (ML), statistical and search based techniques to provide a tool-supported method for the analysis of large message archives, with the ultimate purpose of supporting a requirements-driven evolution process as sketched in Fig. 1. The input to the process is user feedback expressed in online discussion as depicted in Fig. 2 (left side), which represents an excerpt of a discussion posted in the OpenOffice Bugzilla issue tracking system[1]. These online discussions are considered a source of requirements knowledge (e.g., user's goals and preferences, domain assumptions), which can be extracted using analysis techniques so that experts can assess its impact on an existing requirements model. Key elements of the online discussion are highlighted in Fig. 2 (left side), namely topic, terms, participant and feedback classification. The user *dhpeterson* posted her message as *ENHANCEMENT* and it is accepted by the engineer who was managing the issue tracking system, thus providing input to release planning tasks, including requirements model revision and extension. For instance, the analysis of this user feedback could lead to a revision of the goal model representing the requirements of the OpenOffice Writer component, depicted in Fig. 2 (right side).

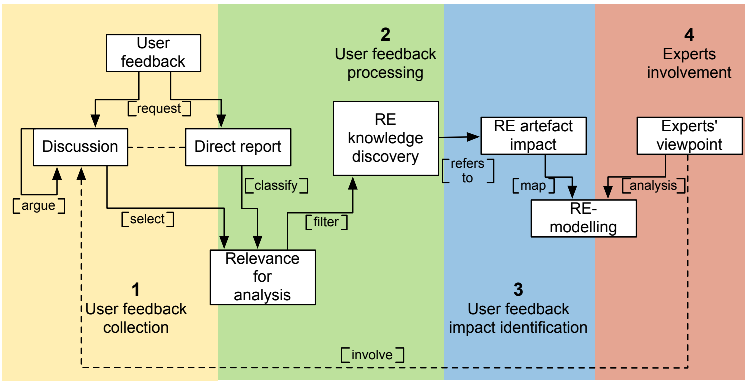

Fig. 1. Overview of our approach to requirements-driven software evolution.

The proposed process for exploiting online feedback and discussions addresses two basic problems, namely how to elicit requirements knowledge from big online discussions (we refer to this problem as *requirements knowledge discovery*), and how to elicit indicators of the level of expertise of the participants to a given discussion (we refer to it as *expert finding* problem). An initial vision of our approach for the case of Tropos requirements models have been presented in a previous iStar Workshop (iStar 2013 [13]). In this paper we give some details on the techniques we exploit (Section 2). We recall the main results achieved so far (Section 2), and ongoing and future work (Section 3).

2 Approach: RE knowledge discovery and Expert Finding

In this section we propose a formulation for the *Requirements knowledge discovery* and the *Expert Finding* problems, introduce the types of analysis and relative implementation techniques that we have defined so far. We will refer to the example taken from online discussions in OSS projects depicted in Fig. 2.

The Requirements Knowledge Discovery Problem. Online discussions can be considered almost synchronous written conversations [12] that can be described in terms of

[1] `https://bz.apache.org/ooo/show_bug.cgi?id=76801`

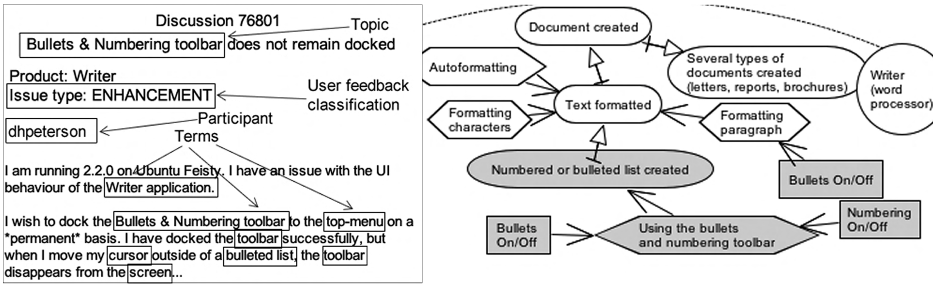

Fig. 2. Left: an excerpt of a discussion posted in the OpenOffice Bugzilla issue tracking system, with its key elements. Right: an excerpt of a requirements model of the OpenOffice-Writer component, in Tropos notation.

speech-acts, according to the Speech Act Theory (SAT) [2]. SAT rests on the following idea –by saying something, we react by doing something–, details of the used theory can be found in [7]. In other words, the speaker's intention can be one of persuading, inspiring or getting a hearer to do something and such an intention is manifested in the *speech-acts* she/he uses. Concretely, speech acts are classified according to specific performative verbs, such as *suggest*, *recommend*, *confirm*, and *advise*, etc., which reveal the speakers' intentions. On these premises, discovering requirements knowledge in online discussions amounts to recognizing those fragments of conversation that contain specific *speech-acts* combinations or patterns, which are found to be commonly used for expressing feature requests, bugs or clarification requests.
The Expert Finding Problem. The problem of expert finding in online discussions can be conceived as a problem of Information Extraction (IE). Formally, given a set P of *Participants* (who are users and developers participating to the mailing list, such as "dhpeterson") engaged in a set of discussions D (e.g., "Discussion 76801") about a set of topics T (e.g., "Bullets and Numbering toolbar"), the problem of expert finding can be stated as the problem of ranking the *Participants* according to their expertise on a topic $t \in T$, based on the emphasis made in a sentence expressing their intentions I (e.g., classifying a message as FEATURE or ENHANCEMENT) and their use of terms related to the topic t (e.g., "bullet list").

We apply two types of analysis to online discussion archives: (a) the analysis of the structure of the archives – i.e., the authors of the messages and the terms found in the messages with their frequency; we refer to this as to the **content** of online discussions; and (b) the analysis of the **intentions** of the participants to an online discussion – i.e., identification of *speech-acts* combinations and patterns.

Our approach to address the Requirements Knowledge Discovery problem rests on the second type of analysis. Specifically, we take online discussions as txt files that we pre-process and clean using NLP techniques. We apply a sentence splitter, a tokenizer, and use the Hepple Tagger (POS tagger). Once each word has been tagged, we use a lemmatizer, gazetteers and lexico-syntactic rules, called JAPE, to finally annotate the intentions [7]. After the annotation is performed we recognize some patterns or combinations of intentions that can lead us to identify a possible feature or bug. For example, feature request indicators correspond to combination of *speech-acts* of type *Requirements* and one among the followings *Positive opinion*$|$*Questions*$|$*Suppositives*$|$*Suggestive* or as a combination of the *speech-act* type *Positive* with an URL link.

To address the Expert Finding problem we defined two approaches. The first exploits the analysis of the content of online discussions, as above defined, the second combines intention and content analysis. Specifically, we first retrieve stakeholders, terms and topics by applying NLP techniques for IE to the considered online discussions. The retrieved elements are represented as nodes of a weighted graph, and by counting their co-occurrences (e.g., how many times a given stakeholder use a given term) we define weights for their connecting arcs. By building a Markov Network (MN), we can compute probabilities for each node of the graph based on the specific topic we are looking experts for. On the basis of these probabilities, stakeholders are ranked along their expertise to a given topic [14]. In the intent- and content-based approach the weights of the arcs in the graph take into account the stakeholder's intentions revealed in her/his messages when talking about a given topic [9].

Results so far. Since the initial vision of our research, which has been presented in [13, 9], we refined and experimentally evaluated the proposed analysis techniques. The online discussion archives we consider are those taken from OSS projects. The overall dataset that we have crawled consists of 713 discussions from the year 2012, containing 2728 messages and 215 participants. For evaluating the analysis techniques we devised an experiment to address the requirements knowledge discovery. We have sampled 20 discussions with 310 messages (1685 sentences) and designed an empirical evaluation aiming at measuring the time effort required for manually annotating sentences w.r.t. their intentions [7]. A first execution of the study has been performed with twenty subjects working in a distributed way using a crowdsourcing-like platform. We observed that the time for annotating has the following distribution: a mean of 35 seconds, a median of 18 and a standard deviation of 56 seconds. On the basis of the participant profiles we derived some insights based on a post-questionnaire filled in by the participants to the study and we applied ANOVA test to identify influencing factors. A technical report discussing the results is available online[2].

Concerning the approach to the expert finding problem, MN techniques revealed scalability problems when considering large OSS discussions, and the attempt to mitigate them using approximate computation is not satisfactory since it gives unstable solution rankings. In order to fix this stability issue, we are now investigating the applicability of search-based techniques. For the implementation, we use NSGAII [5] which allows us to properly deal with the scalability problem, although the functions we tried to optimize so far did not allow us to get meaningful results yet.

Related work. Regarding relevant related work of the requirements knowledge problem, we shall mention the automated identification of intentions presented by Twitchell et al. [12]. This investigation proposes a tool that is based on SAT, dialogue acts and fuzzy logic to analyze transcripts of telephone conversations. The goal of this research is to derive participant profiles based on a map of the intentions expressed in the conversation. The classification of emails using speech acts is investigated by Carvalho et al. [3]. They are interested in classifying emails regarding office-related interactions as negotiation and delegation of tasks. Among the related works that are of particular interest in the expert finding problem, we can mention Serdyukov and Hiemstra [11], in which the content of documents is analyzed to identify the contributions of their different authors

[2] `http://selab.fbk.eu/imramirez/TR_CAiSEDec2014.pdf`

(i.e., used terms) and the probability that a given author relates to a queried topic is computed. Focusing on the social dimensions, the work of Zhang et al. [15] consider the question/answers in an online forum to identify the knowledge seekers (i.e., non-experts) and the knowledge providers (i.e., experts). They compare several algorithms to rank people, starting from the simple counting of answers, then combining it with the number of questions, which should be negatively correlated to the level of expertise, before to propagate the computed values over the community (PageRank-like) to simulate social recognition.

3 Ongoing and Future work

In parallel with the above mentioned improvements of the analysis techniques, we are working at the consolidation of our problem formulation through the elaboration of a conceptual framework that rests on novel ontologies. Specifically, for the purpose of giving an ontological foundation to our intent-based analysis approach, we are extending a communication ontology that relies on the SAT and that accounts for concepts regarding the software development [8].

Analogously, we are working on a more rigorous conceptualization of expert finding systems to help find new indicators to be considered. Indeed, it seems that the indicators used so far (terms and topics, additionally roles) relate to the accessible knowledge and the social recognition of the stakeholders, which seem not the best indicators for inferring expertise [6]. We believe that this conceptualization could be helpful to the research and engineering community, also for supporting a more robust and faster design of expert finding systems.

As for future work, we believe that our intent-based analysis of online discussions could be used in combination with already existing techniques as topic modeling or sentiment analysis to build an enriched classification model for categorizing messages as bug reports or feature request. For example by doing a basic text analysis[3] of the first message posted in the discussion 76801, we can get information on terms and frequencies, as those summarized in Table 1. But the intent-based analysis can provide an emphasis on the importance of the found terms, if the terms are contained in a sentence expressing a certain type of intention. Further work is needed to combine information on frequency and similarity between phrases to rank the relevance of a section of a goal model to be changed.

Table 1. Top 3 phrases (containing nouns) with their frequency and similarity % (Levenshtein distance) w.r.t. the goal *Numbered or bulleted list created* and the task *Using the bullets and numbering toolbar* depicted in Fig. 2.

4 words	F	G-sim%	T-sim%	3 words	F	G-sim%	T-sim%	2 words	F	G-sim%	T-sim%
observe that the toolbar	4	21	35	the bulleted list	6	45	35	the toolbar	10	15	28
of the bulleted list	2	45	38	that the toolbar	4	15	33	bulleted list	9	39	25
bullets & numbering toolbar	2	24	66	a bulleted list	3	42	28	the bulleted	6	30	28

[3] `http://www.online-utility.org/text/analyzer.jsp`

Concerning the evaluation of our intent-based analysis approach, we are also designing a study for evaluating manual annotations versus automatic annotation of our tool, measuring the accuracy in terms of precision and recall.

References

1. T. A. Alspaugh and W. Scacchi. Ongoing software development without classical requirements. In *21st IEEE RE 2013, Rio de Janeiro-RJ, Brazil, July 15-19*, pages 165–174. IEEE Computer Society, 2013.
2. K. Bach and R. M. Harnish. *Linguistic Communication and Speech Acts*. MIT Press, Cambridge, MA, 1979.
3. V. R. Carvalho and W. W. Cohen. On the collective classification of email speech acts. In *Proceedings of the 28th annual international ACM SIGIR conference on Research and development in information retrieval*, pages 345–352. ACM, 2005.
4. C. Castro-Herrera and J. Cleland-Huang. Utilizing recommender systems to support software requirements elicitation. In *RSSE*, RSSE'10, pages 6–10. ACM, 2010.
5. K. Deb, A. Pratap, S. Agarwal, and T. Meyarivan. *A Fast and Elitist Multi-Objective Genetic Algorithm: NSGA-II*. 2000.
6. K. A. Ericsson. The Influence of Experience and Deliberate Practice on the Development of Superior Expert Performance. In K. A. Ericsson, N. Charness, P. J. Feltovich, and R. R. Hoffman, editors, *The Cambridge handbook of expertise and expert performance*, pages 683–703. Cambridge University Press, New York, NY, US, 2006.
7. I. Morales-Ramirez, A. Perini, and M. Ceccato. Towards supporting the analysis of online discussions in OSS communities: A speech-act based approach. In S. Nurcan and E. Pimenidis, editors, *Information Systems Engineering in Complex Environments - CAiSE Forum 2014, Thessaloniki, Greece, June 16-20, 2014, Selected Extended Papers*, volume 204 of *Lecture Notes in Business Information Processing*, pages 215–232. Springer, 2014.
8. I. Morales-Ramirez, A. Perini, and R. S. S. Guizzardi. Providing foundation for user feedback concepts by extending a communication ontology. In *33rd International Conference, ER 2014, Atlanta, GA, USA*, volume 8824 of *LNCS*, pages 305–312. Springer, 2014.
9. I. Morales-Ramirez, M. Vergne, M. Morandini, A. Siena, A. Perini, and A. Susi. Who is the expert? combining intention and knowledge of online discussants in collaborative RE tasks. In *ICSE '14, May 31 - June 07*, pages 452–455. ACM, 2014.
10. B. Nuseibeh and S. Easterbrook. Requirements engineering: a roadmap. In *Proceedings of the Conference on The Future of SE*, ICSE '00, pages 35–46, New York, NY, USA, 2000.
11. P. Serdyukov and D. Hiemstra. Modeling Documents As Mixtures of Persons for Expert Finding. In *Proceedings of the IR Research, 30th European Conference on Advances in Information Retrieval*, ECIR'08, pages 309–320, Berlin, Heidelberg, 2008. Springer-Verlag.
12. D. P. Twitchell and J. Nunamaker, J.F. Speech act profiling: a probabilistic method for analyzing persistent conversations and their participants. In *System Sciences, 2004. Proceedings of the 37th Annual Hawaii International Conference on*, pages 10 pp.–, 2004.
13. M. Vergne, I. Morales-Ramirez, M. Morandini, A. Susi, and A. Perini. Analysing user feedback and finding experts: Can goal-orientation help? In J. Castro, J. Horkoff, N. A. M. Maiden, and E. S. K. Yu, editors, *Proceedings of the 6th International* i* *Workshop 2013, Valencia, Spain, June 17-18, 2013*, volume 978 of *CEUR Workshop Proceedings*, pages 49–54. CEUR-WS.org, 2013.
14. M. Vergne and A. Susi. Expert Finding Using Markov Networks in Open Source Communities. In *CAiSE*, number 8484 in LNCS, pages 196–210. Springer, 2014.
15. J. Zhang, M. S. Ackerman, and L. Adamic. Expertise networks in online communities: structure and algorithms. In *Proceedings of the 16th international conference on World Wide Web*, WWW '07, pages 221–230, New York, NY, USA, 2007. ACM.