Editor: **Sarah Gregory**
Intel Corporation
sarah.c.gregory@ieee.org

Natural Language Processing for Requirements Engineering

The Best Is Yet to Come

Fabiano Dalpiaz, Alessio Ferrari, Xavier Franch, and Cristina Palomares

NATURAL LANGUAGE (NL) is traditionally the predominant notation for documenting and specifying software and system requirements.[1] NL is used extensively not only in specifications (for example, "shall" requirements, user stories, and use cases) but also throughout development (for example, code comments, documentation, and bug reports).

Requirements written in NL are easy to write and comprehend, even by stakeholders with limited experience in requirements engineering (RE). On the other hand, NL is inherently ambiguous[2] ("I saw Peter and Paul and Mary saw me"). Besides, large collections of NL requirements are hard to examine manually to obtain an overview and find inconsistencies, duplicates, and missing requirements.

RE researchers have been studying NL's role and the potential offered by natural language processing (NLP) since the early '90s.[3] The community has explored topics such as the identification of quality defects and ambiguity, classification and clustering of large collections of requirements, extraction of key abstractions, generation of models, and traceability between NL requirements and code.[4]

Until recently, many of these applications of NLP have been confined to the academic world owing to NLP tools' inaccessibility and steep learning curve. Luckily, advances in deep learning and the availability of large NL corpora have significantly lowered the entry barriers to using NLP. This creates unprecedented opportunities to apply NLP techniques to RE practice and to help automatically analyze requirements-related documents.

The RE research community has exploited this opportunity and is trying to perform impactful research on the use of NLP tools and techniques in RE practice. As part of the growing interest in this field, RE researchers, computational linguists, and industry practitioners met at the First Workshop on Natural Language Processing for Requirements Engineering (NLP4RE 18; http://fmt.isti.cnr.it/nlp4re). Here, we summarize the workshop and present an overview of the discussion held on the field's future.

Workshop Summary

NLP4RE 18 was collocated with the 24th International Working Conference on Requirements Engineering: Foundation for Software Quality (REFSQ 18), in Utrecht in March 2018. It lasted one day and had 20 to 30 attendees at any given time.

The workshop's main goal was to set up a dedicated venue for researchers and practitioners interested in discussing advances, challenges, and barriers related to applying NLP to RE problems. Despite NL's importance in writing requirements, such a venue hadn't previously existed. Given this community-building objective, the call for papers requested

 0740-7459/18/$33.00 © 2018 IEEE

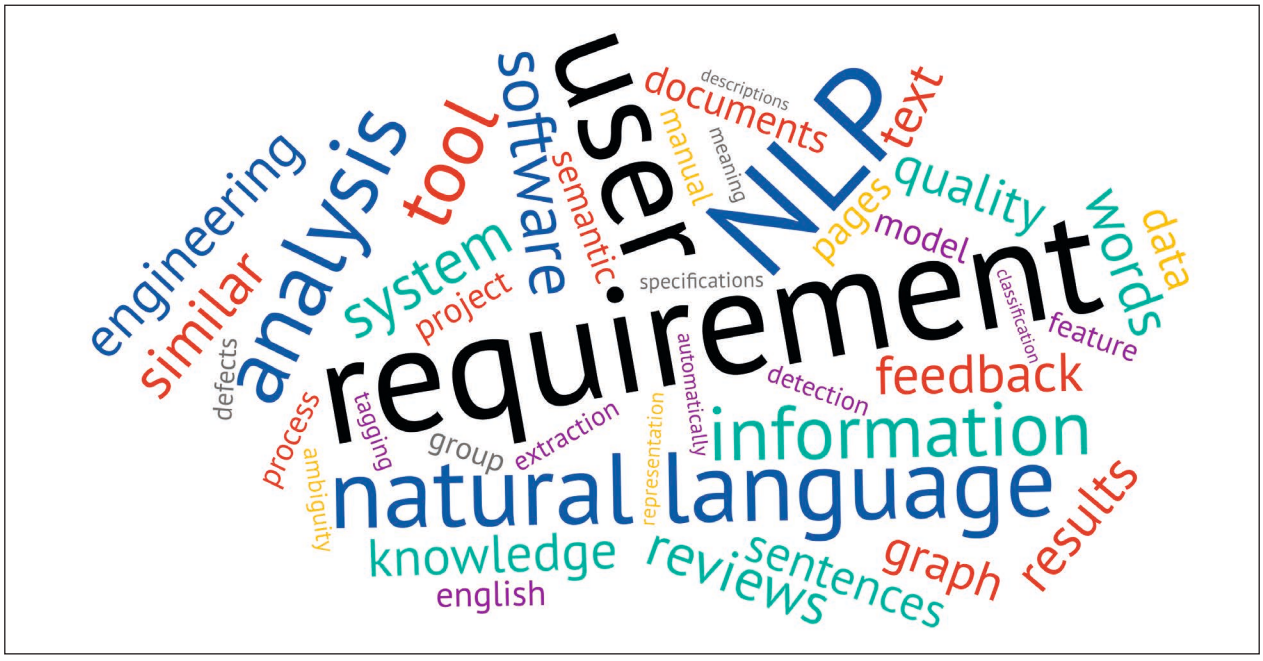

FIGURE 1. A word cloud extracted from the 11 papers accepted for presentation at the First Workshop on Natural Language Processing for Requirements Engineering (NLP4RE 18). Note how the words "tool," "similar," "quality," "reviews," "tagging," and "group" appear in this cloud, representing recurring topics in the accepted submissions.

not only novel techniques and empirical studies but also vision papers about the future and short reports from research groups on their past, current, and future work.

The workshop included a keynote by Dan Berry (University of Waterloo) on the evaluation of NLP tools' effectiveness in RE, particularly when requirements engineers use them to automate cumbersome, repetitive tasks. He discussed the fundamentals for measuring information retrieval's effectiveness (in terms of precision, recall, and the F-score) and described which data to gather and use when you're evaluating tools.

Tejaswini Deoskar (Utrecht University) gave an invited talk concerning the challenges of adapting NLP tools (such as part-of-speech taggers and parsers) to a given domain. Deoskar proposed semisupervised learning to adapt supervised NLP models to perform better over new domains of text such as those used in RE documents.

The workshop attracted 19 papers, a good number for a first-time event held at a conference such as REFSQ, which has around 80 to 100 attendees. Three program committee members independently reviewed the papers. The co-organizers accepted 11 papers for presentation. Figure 1 shows a word cloud generated from the title, abstract, and body text of the 11 papers. Note how the words "tool," "similar," "quality," "reviews," "tagging," and "group" appear in this cloud, representing recurring topics in the accepted submissions.

The presented papers covered diverse topics. Several papers concentrated on a specific application area. Recurring business domains were the railway, telecommunications, and automotive domains. The proposed approaches supported such types of systems as apps, information systems, and dynamic-service applications. Some papers also covered topics such as quality improvement and information extraction. Finally, the papers applied different techniques, such as rule-based techniques for information extraction and machine-learning (ML) approaches for classification and clustering.

In addition, some papers incorporated extra knowledge, such as synonym dictionaries or hyperonym ontologies, to augment techniques.

The Future of NLP for RE

In a brainstorming session at the workshop's end, participants identified the following future research directions.

Resource Availability

NLP techniques' effectiveness depends largely on the processed data's quality and quantity. NLP4RE is no exception to this rule.

Creation of reliable data corpora. NLP applications, especially those using ML, need large amounts of data to perform optimally. Regarding NLP4RE, the necessary data generally is companies' requirements. Furthermore, realistic annotation of the requirements requires domain experts to manually identify defects or trace links, depending on the task to perform. The annotations are needed for training ML algorithms and validating the proposed techniques.

Many users from industry can assist in these tagging tasks. However, it's crucial to identify what NLP tasks in RE can be successfully outsourced or crowdsourced, by considering the necessary time and task complexity. The long-term aim is to obtain reliable, reusable public requirements corpora.

Data quality and heterogeneity. Often, requirements and their NL sources (e.g., app reviews) exhibit poor quality, which impairs the performance of existing NLP tools. Another complexity factor to consider is the variety of formats, which can range from rigorous NL specifications to diagrammatic models to bug reports. Such format variations might require using different NLP approaches and different policies for integrating heterogeneous information.

Validation metrics and workflows. For NLP4RE techniques to be assessed properly, correct performance metrics and validation workflows must be established. As Berry and his colleagues pointed out, RE has traditionally borrowed validation approaches from information retrieval, but these approaches alone don't assess how accuracy or inaccuracy actually affects RE practitioners' tasks.[5]

Context Adaptation

NLP4RE techniques require customizing general NLP techniques to make them applicable for solving the problems requirements engineers face in their daily practice.

Domain specificity. Each domain has its own specific jargon, business rules, and process practices. NLP tools need to handle the domain specificity of NL requirements. The adoption of domain-specific, and even company-specific, ontologies is crucial.

A relevant research direction concerns the creation of techniques for semiautomatic ontology construction. This also requires finding strategies to address the problem of tacit knowledge—information that's concealed in experts' minds and isn't written down. Eliciting this knowledge is necessary to have the appropriate contextual information that NLP4RE tools can leverage to perform their tasks.

Big NLP4RE. Besides requirements, NLP4RE tools need to take into account other artifacts produced during the software process, such as architecture, design diagrams, and software, and their evolution. Although public requirements data are scarce, companies have large numbers of artifacts, and NLP4RE tools are particularly needed to help make sense of, cross-reference, and reuse them.

Human-in-the-loop. NLP technology won't replace experts in their RE tasks; it will empower them. Researchers need to clearly define the human scope and the machine scope for the different tasks.

Language issues. Existing NLP tools perform fairly well for English, but non-English datasets and the associated NLP tools aren't as good. Researchers must find ways to deal with this issue, to come to multilingual, multicultural, and multicontext NLP4RE.

Player Cooperation

The NLP4RE ecosystem involves four main players: RE researchers, NLP experts, vendors of requirements management tools, and industries, the final users of the produced technologies. Mutual cooperation and awareness need to be established between these parties.

RE researchers. An overview of the available NLP technologies is needed. RE researchers need to know which ones to use, and for which RE tasks. This requires close collaboration with computational linguists, who can provide informed answers regarding NLP techniques' suitability for a given task.

NLP experts. RE has much potential for NLP researchers. Making them aware of RE challenges is a key element, to let them develop tools tailored to the RE context. For example, NLP focuses on technologies requiring large amounts of data, whereas RE tasks deal with a low number of resources. Doing more with less can be an interesting challenge for NLP research. This goal can be pursued through RE-related competitions at NLP conferences and participation at NLP venues to clarify RE problems.

Tool vendors. To manage requirements, industries normally use IBM

Rational DOORS, Jira, GitLab, or similar tools. The fast adoption of NLP technologies depends highly on embedding them into these existing platforms. Asking industry practitioners to switch to a different environment won't work.

Industries. Stronger interaction with industries is needed. It should be made clear to industrial practitioners what NLP can do for them, what it can't do (for example, finding conceptual defects), and what's required on their side, such as requirements, data annotation, and domain knowledge. NLP4RE tools need to become more mature: adoption in industry requires them to exhibit sufficient quality for use in production.

NLP is becoming a cornerstone technology for different areas and domains (think of chatbots, spam filtering, and so on), and RE is no exception to this trend. Given the current availability of NLP tools and the increasing amount of available data, more and more researchers are attempting to solve requirements engineers' real-world problems. With this picture in mind, we organized the first NLP4RE workshop as a community-building event, aiming to make the NLP4RE workshop a regular meeting point for the community. In retrospective, the number of submissions and the good participation indicate that the workshop fulfilled this objective.

Sustainability is now the challenge. We plan to seek tighter integration with other communities, especially with NLP and computational-linguistics researchers. For 2019, we're discussing associating the workshop with a conference such as the Association of Computational Linguistics (ACL) conference, the International Conference on Computational Linguistics (COLING), or the Conference on Empirical Methods in Natural Language Processing (EMNLP). We expect that this change could raise awareness in the classic NLP community about RE as an application area.

More generally, the success of the NLP4RE workshop and field requires researchers to accelerate the progress. We need datasets and tools to be publicly available, we desperately seek evidence of the long-term impact on RE practice, and we still rely on fundamental research experimenting with state-of-the-art NLP techniques. Although these are critical challenges, we believe that societal pressure for the field's success will attract resources and accelerate progress. The best is yet to come!

ABOUT THE AUTHORS

FABIANO DALPIAZ is assistant professor in Utrecht University's Department of Information and Computing Sciences. Contact him at f.dalpiaz@uu.nl.

ALESSIO FERRARI is a researcher at CNR-ISTI's Formal Methods and Tools Laboratory. Contact him at alessio.ferrari@isti.cnr.it.

XAVIER FRANCH is a professor of informatics at Universitat Politècnica de Catalunya. Contact him at franch@essi.upc.edu.

CRISTINA PALOMARES is a postdoctoral researcher at Universitat Politècnica de Catalunya, working on the OpenReq project. Contact her at cpalomares@essi.upc.edu.

Acknowledgments

The work of Xavier Franch and Cristina Palomares has been supported by the OpenReq project, which has received funding from the European Union's Horizon 2020 Research and Innovation program under grant agreement 732463.

References

1. M. Kassab, C. Neill, and P. Laplante, "State of Practice in Requirements Engineering: Contemporary Data," *Innovations in Systems and Software Eng.*, vol. 10, no. 4, 2014, pp. 235–241.
2. D.M. Berry, E. Kamsties, and M.M. Krieger, *From Contract Drafting to Software Specification: Linguistic Sources of Ambiguity*, tech. report, School of Computer Science, Univ. of Waterloo, 2001.
3. K. Ryan, "The Role of Natural Language in Requirements Engineering," *Proc. 1993 IEEE Int'l Symp. Requirements Eng.*, 1993, pp. 240–242.
4. A. Ferrari et al., "Natural Language Requirements Processing: A 4D Vision," *IEEE Software*, vol. 34, no. 6, 2017, pp. 28–35.
5. D.M. Berry et al., "Panel: Context-Dependent Evaluation of Tools for NL RE Tasks: Recall vs. Precision, and Beyond," *Proc. IEEE 25th Int'l Requirements Eng. Conf.* (RE 17), 2017, pp. 570–573.