Natural Language Processing for Requirements Engineering: The Best is Yet to Come

Fabiano Dalpiaz[1], Alessio Ferrari[2], Xavier Franch[3,¥], Cristina Palomares[3,¥]

1 Utrecht University, The Netherlands
2 CNR-ISTI, Pisa, Italy
3 Universitat Politècnica de Catalunya, Spain

1. Motivation

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 (e.g., *shall* requirements, user stories, use cases), but also throughout the whole development process (e.g., code comments, documentation, 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 in order to obtain an overview, find inconsistencies, duplicates, and missing requirements.

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

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

The RE research community has exploited this opportunity and is trying to make impactful research on the use of NLP tools and techniques in RE practice. As part of the growing interest in this field, RE researchers have met in the first edition of a dedicated workshop called *NLP4RE*, which has also seen the participation of computational linguists and industry practitioners. We offer a summary of such workshop, and present an overview of the discussion held on the future of the field.

2. Workshop summary

---

[1] Kassab, M., Neill, C., Laplante, P.: State of Practice in Requirements Engineering: Contemporary Data. Innovations in Systems and Software Engineering 10 (4), 235–241 (2014)

[2] Berry, D.M., Kamsties, E., Krieger, M.M.: From Contract Drafting to Software Specification: Linguistic Sources of Ambiguity. Tech. rep., School of Computer Science, University of Waterloo, ON, Canada (2001)

[3] Ryan K. The role of natural language in requirements engineering. In Proc. of the IEEE International Symposium on Requirements Engineering, pp. 240-242 (1993)

[4] Ferrari, A., Dell'Orletta, F., Esuli, A., Gervasi, V., Gnesi, S.: Natural Language Requirements Processing: A 4D Vision. IEEE Software, 34(6), 28-35 (2017)

The First Workshop on Natural Language Processing for Requirements Engineering (NLP4RE'18, http://fmt.isti.cnr.it/nlp4re/) was co-located with the 24th International Working Conference on Requirements Engineering: Foundation for Software Quality (REFSQ 2018), held in Utrecht, The Netherlands, in March 2018. It was celebrated in one single day and attracted around 20-30 attendees at different moments of the day.

The main goal of NLP4RE'18 was to set up a dedicated venue for researchers and practitioners interested in discussing advances, challenges and barriers when applying NLP to RE problems. Despite the aforementioned relevance of NL in writing requirements, such a venue did not exist until now. Given this community-building objective, the call for papers included not only scientific results and empirical studies, but also vision papers about the future, and short reports for research groups to present their past, current, and future work. The workshop attracted 19 papers, a good number for a first-time event held in a conference like REFSQ which has around 80-100 attendees.

The workshop included a keynote by Dan Berry from University of Waterloo (Canada) on the evaluation of the effectiveness of NLP tools in RE, in particular when used by requirements engineers to automate cumbersome and repetitive tasks. The talk discussed the fundamentals for measuring the information retrieval effectiveness (in terms of precision, recall, and F-score) and described which data should be gathered and used during tools evaluation. Tejaswini Deoskar from Utrecht University (the Netherlands) gave an invited talk concerning the challenges for adapting NLP tools (like part-of-speech taggers and parsers) to a given domain. Deoskar proposed semi-supervised learning to adapt supervised NLP models to perform better over new domains of text such as those used in RE documents.

The 19 submitted papers were independently reviewed by three program committee members. Eleven papers were accepted by the co-organizers for presentation at the workshop. Figure 1 shows a word cloud generated from the title, abstract, and body text of these 11 papers. It is worth highlighting how the words *tool*, *similar*, *quality*, *review*, *tagging* and *group* appear in this cloud, representing recurring topics in the accepted submissions.

Figure 1 - Word cloud extracted from the NLP4RE accepted papers

The papers were diverse according to several criteria. For instance, if we consider field of application, several papers were proposed with one specific field of application in mind. Recurrent business domains were railway, telecommunications, and automotive. The proposed approaches support different types of systems such as apps, information systems, and dynamic service applications. On the other hand, papers covered different topics such as quality improvement and information extraction. Last, the presented works applied different techniques, as text and semantic processing, rule-based techniques and classification and clustering machine learning approaches. In addition, some of the papers incorporate extra knowledge, such as synonyms dictionaries or hyperonims ontologies, to improve the results of other techniques.

3. The future of NLP at RE

A brainstorming session was organised at the end of the workshop to discuss to identify future directions of research in the field.

3.1 Resources Availability

Creation of Reliable Data Corpora. NLP applications, especially those using machine learning techniques, need large amounts of data to achieve optimal performance. For NLP4RE techniques, the data needed by the community are generally requirements from companies. Furthermore, domain experts are necessary for a realistic annotation of the requirements, by manually identifying defects or trace links, depending on the task to be solved. The annotations are needed for training machine learning algorithms, and to validate the proposed techniques. Many users from the industry can assist in these tagging tasks; however, it is crucial to identify what kind of NLP tasks in RE can be successfully out- or crowd-sourced, by considering the necessary time and task complexity. The long-term aim is to obtain reliable and 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 (ranging from rigorous NL specifications to diagrammatic models to bug reports), which may require the use of different NLP approaches, and policies for integration of heterogeneous information.

Validation Metrics and Workflows. To properly assess NLP4RE techniques, correct performance metrics, and correct validation workflows, have to be established. As Berry et al. point out[5], RE has traditionally borrowed validation approaches from information retrieval, but these techniques alone do not assess the actual impact of (in)accuracy on the tasks of RE practitioners.

3.2 Context Adaptation

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 of crucial importance. A relevant research direction concerns the creation of techniques for the semi-automated construction of ontologies. This requires also to find strategies to address the problem of tacit knowledge, which is the information that is concealed in the mind of experts, and is not written down. Eliciting this knowledge is necessary to have the appropriate contextual information that NLP4RE tools can leverage to perform their tasks.

Big NLP4RE. Requirements are part of the software process, and NLP4RE tools need to take into account also the other artifacts, such as architecture, design diagrams and software, produced during the process, and their evolution over time. Although public requirements data are scarce, companies have very large amounts of artifacts, and NLP4RE tools are particularly needed to help making sense, cross-referencing, and reuse these artifacts.

Human-in-the-Loop. NLP technologies are not expected to replace experts in their RE tasks, but to empower them. Clearly defining the human scope and the machine scope for the different tasks is another challenge that researchers need to address.

---

[5] Berry, Daniel M., et al. Panel: Context-Dependent Evaluation of Tools for NL RE Tasks: Recall vs. Precision, and Beyond. In: Requirements Engineering Conference (RE), 2017 IEEE 25th International. IEEE, 2017. p. 570-573.

Language Issues. Existing NLP tools perform fairly well for English, but non-English datasets and associated NLP tools are not as good. Researchers need to find ways to deal with this issue, to come to a multi-lingual, multi-cultural and multi-context NLP4RE.

3.3 Players' 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 have to be established between the different parties.

RE Awareness. An overview of the available NLP technologies is needed. Several technologies exist, and RE researchers need to know which ones to use, and for which specific RE tasks. This requires a tight collaboration with computational linguists, who can provide informed answers on the suitability of the various NLP techniques for a given task.

Industry Awareness. Stronger interaction with industries is needed, and it should be made clear to industrial practitioners what NLP can do for them, what it *cannot* do -- finding conceptual defects, for example -- and what is required on their side, such as requirements, data annotation, and domain knowledge. NLP tools for RE need to grow in maturity: adoption in industry requires these tools to exhibit a sufficient quality for use in production.

Tool Vendors' Awareness. Industries normally use IBM Rational® DOORS®, Jira, GitLab or similar tools for managing requirements, and the fast adoption of NLP technologies is highly dependent on the embedding of the developed tools into the existing platforms. Asking industry practitioners to switch to a different environment will not work.

NLP Awareness. RE has a lot of potential for NLP researchers, and making them aware of RE challenges is a key element, to let them develop tools that are tailored to the RE context. For example, NLP is focusing on technologies that require large amounts of data, while RE tasks deal with a low amount of resources, and "doing more with less" can be an interesting challenge for NLP research. This goal can be pursued by including RE-related competitions in NLP conferences, and participating to NLP venues to clarify RE problems.

4. Conclusions and future actions

NLP is becoming a cornerstone technology for different areas and domains (think of chatbots, spam filtering, etc.) and RE is no exception to the rule. Given the current availability of NLP tools and the increasing amounts of available data, more and more researchers are attempting to solve real-world problems that the requirement engineers face. With this picture in mind, we have organized the first edition of the NLP4RE workshop as a community-building event, aiming to make NLP4RE become 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 organize NLP4RE in the following years, also seeking tighter integration with other communities, especially with researchers in NLP and computational linguistics. In 2019, we are discussing the option to hold the workshop as an associate event of a conference as the Association of Computational Linguistics (ACL), the International Conference on Computational Linguistics (COLING) or the Empirical Methods in Natural Language Processing (EMNLP). We expect that this change could raise awareness in the classical NLP community about RE as a field of application.

More generally, the success of NLP4RE workshop and field requires researchers to accelerate the progress in the field: we need data sets and tools to be publicly available, we desperately seek for 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 do believe that the societal pressure for the success of the field will attract resources and accelerate progress. The best is yet to come!