A Report on the First Workshop on Natural Language Processing Advancements for Software Engineering (NLPaSE) co-located with APSEC 2020

Saurabh Tiwari[a], Santosh Singh Rathore[b], Lata Nautiyal[c] and Ravindra Singh[d]

[a]DA-IICT Gandhinagar, India
[b]ABV-IIITM Gwalior, India
[c]University of Bristol, UK
[d]Delhi Technological University, Delhi, India

Abstract

NLPaSE 2020[1], 1st International Workshop on Natural Language Processing Advancements for Software Engineering, co-located with 27th Asia-Pacific Software Engineering Conference (APSEC 2020[2]) 1-3 December at Singapore, aims to bring out the existing techniques, their limitations, and discussions which may result in future collaborations and advancements. We also aim to identify the areas where NLP was applied (limitation, challenges), can be applied (possible directions), and bring out the researchers/industry practitioners in a platform to interact, share their experiences and collaborate. The workshop was a half-day workshop – consists of 2 keynote talks, 3 research presentations and an open discussion session – organised on 01 December 2020 virtually using Cisco WebEx platform. The workshop was also hosted live on YouTube[3], and available for the interested participants for watching.

Keywords

Workshop, Annual Event, Software Engineering, Advancements, Research and Practice, natural language processing (NLP)

1. Motivation and Aim

Natural Language Processing (NLP) has played an important role in automating the various software engineering processes (e.g., requirements analysis, requirements elicitation, identifying design artefacts, test automation, maintenance, etc.) [1, 2]. The need for automation in the Software Engineering domain pushed the use of various techniques for reducing the effort and cost-cutting where ever possible. The automation is not only limited to software development it reaches to the automobile, embedded systems and others. Specifically, the requirements are volatile in nature and it requires a huge amount of effort in assessing their quality and so on [3].

[1]https://sites.google.com/view/nlpase2020/
[2]https://formal-analysis.com/apsec/2020/
[3]https://youtu.be/9cnUGUF6Hds

Joint Proceedings of SEED & NLPaSE co-located with APSEC 2020, 01-Dec, 2020, Singapore

saurabh_t@daiict.ac.in (S. Tiwari); santoshs@iiitm.ac.in (S. S. Rathore); lata.nautiyal@bristol.ac.uk (L. Nautiyal); ravindra@dtu.ac.in (R. Singh)
https://sites.google.com/site/saurabhiiitdmj/ (S. Tiwari); https://sites.google.com/site/santoshiiitmdj/(S. S. Rathore)

NLP and their advancements have brought a lot of changes in analysing the software artefacts and processing the information [4].

The main goal of the NLPaSE-2020 workshop is to bring out the NLP-based existing techniques, their challenges, and limitations, which may result in discussions, future collaborations and advancements. The workshop call for research and position papers focused on the following topics of interest include w.r.to the application of NLP techniques and advancements for Software Engineering:

- Requirements analysis, elicitation, and tracing.
- Automation and tool support.
- Generating Software Artefacts by applying NLP techniques (e.g., domain models, entities, test information, test cases, etc.)
- Bug report mining and analysis.
- Test automation.
- Functional and non-functional requirements categorisation.
- Information extraction (abstraction identification, feature extraction).
- Test information and test case generation.
- Text mining.
- Systematic Reviews (SLR, SMS).
- Natural Language is a Programming Language.
- Generation of test oracles using NLP.
- Generating code from natural-language specifications.

NLPaSE-2020 invited submissions under three categories.

1. **Research Papers**, including case studies, reporting on original research results on the use of NLP in the Software Engineering domain.
2. **Experience Reports** describing experiences in the use of NLP, challenges, and lessons learnt in the Software Engineering domain.
3. **Position Papers** sharing the author's insights or proposing an original idea or an opinion on NLP Advancements for Software Engineering.

The workshop received a total of 5 submissions. Each submission was reviewed by three independent reviewers (program committee members) with respect to the overall quality including presentation, the future impact of the research, and the likely benefit to the students, academics, and professionals who will attend the workshop. Finally, 3 papers were accepted for presentation and publication at the workshop.

We aim to set up NLPaSE as a regular meetup event at APSEC future editions starting from this year. The discussions, specifically, focused on identifying the areas where NLP applied, challenges that can be faced, and bring out the researchers/industry practitioners in a platform to interact, share their experiences and collaborate.

2. Workshop Format

The workshop is organised in a half-day session with an aim to keep it interactive, productive and short by having talks, discussions and presentations by the researchers working in the area of NLP for Software Engineering. Therefore, the workshop consists of two keynote talks focused on applying NLP techniques in Software Verification and Requirements Engineering; three research presentations with discussions; and an open discussion session on the workshop goals, takeaways and feedback.

The three hours (half-day) workshop is organised as:

- 2 keynote talks, each of 45 minutes with discussions
- 3 research presentations, each of 15 minutes
- 3 discussion sessions, each of 10 minutes after research presentations
- An open session of 25 minutes

2.1. Keynote Talks

The workshop consisted of 2 keynote talks. The duration of each talk was 40 minutes followed by 5 minutes of question-answer based discussions.

The first keynote talk was given by **Michael Felderer**[1], *University of Innsbruck, Austria.* Michael talks about the current approaches and future directions for NLP in system verification.
Abstract of the talk: Natural language test cases are essential for verification of software systems, products, or services as the intended behavior can neither be fully formalized nor thoroughly be tested automatically. Furthermore, comprehensive natural language system specifications or norms are applied to derive test cases and the number and complexity of test scenarios are ever-increasing. Therefore natural language processes play a central role in keeping system verification effective and efficient. However, the potential of natural language processing for system verification has not been fully exploited so far. In this talk, we first give an overview of the current state of natural language processing in system verification. Then, we present our recent results on the application of natural language processing for detecting dependencies between system test cases, which enables an enormous increase in the efficiency of system testing. Finally, we sketch future directions of research on the application of natural language processing in system verification, especially also with respect to AI-enabled systems in regulated environments.

Second keynote talk was given by **Fabiano Dalpiaz**[2], *Utrecht University, Netherlands.* Fabiano discussed on the effectiveness of NLP for Requirements Engineering.
Abstract of the talk: Requirements Engineering is a natural language heavy phase of Software Engineering. The prevalent notation for expressing requirements is still text; consequently, the research community proposed numerous NLP-powered tools for analyzing requirements-relevant information. Building on the experience gained within the Requirements Engineering

[1] http://mfelderer.at/
[2] https://webspace.science.uu.nl/~dalpi001/index.php

Lab (RE-Lab), the talk discussed the notion of quality when it comes to NLP tools for requirements engineering (NLP4RE tools). How to measure quality? How have we measured quality so far? What does "good enough" mean? Who should determine whether an NLP4RE tool performs well? While answering these questions, the talk focused on the NLP tools for user stories that the members of the RE-Lab have developed, while keeping a keen eye on tools proposed by other research groups. The ultimate goal of the talk is to provide a "where do we stand, where do we go" perspective on NLP4RE research, its results, and its impact.

2.2. Accepted Papers

1. Bahareh Afshinpour, Roland Groz, Massih-Reza Amini, Yves Ledru and Catherine Oriat, *Reducing Regression Test Suites using the Word2Vec Natural Language Processing Tool*
2. Fabian Gilson, Sam Annand and Jack Steel, *Recording Software Design Decisions on the Fly*
3. Kaisei Hanayama, Shinsuke Matsumoto and Shinji Kusumoto, *Humpback: Code Completion System for Dockerfile Based on Language Models*

3. Summary

The first edition of NLPaSE-2020 workshop received active and overwhelming responses from the authors and participants. A total of 25 participants from different parts of the world had attended the workshop. Though it was a bit difficult to have interactive sessions throughout the workshop, the sessions were interactive and discussion-oriented. After each presentation and talk, questions were posed by the participants and organisers. The workshop was also streamed live on YouTube, and available for the interested participants for watching. We intend to continue the next and future editions of NLPaSE workshop in the APSEC conference as a regular meetup event from this year.

Recorded Video of NLPaSE-2020 workshop: https://youtu.be/9cnUGUF6Hds

Acknowledgements

The organisers are thankful for, Prof Lei Ma and Prof Foutse Khomh, the APSEC 2020 committee for giving NLPaSE a platform to host. We are thankful for our keynote speakers, authors and workshop program committee members for making NLPaSE successful. We are also thankful to the participants. Hopefully, you will get some insight on NLP, their challenges, in SE. We are looking for your submissions and participation in the next edition of the workshop.

References

[1] V. Garousi, S. Bauer, M. Felderer, Nlp-assisted software testing: A systematic mapping of the literature, Information and Software Technology (2020) 106321.

[2] F. Dalpiaz, N. Niu, Requirements engineering in the days of artificial intelligence, IEEE Software 37 (2020) 7–10.
[3] C. Arora, M. Sabetzadeh, L. Briand, F. Zimmer, Extracting domain models from natural-language requirements: Approach and industrial evaluation, in: Proceedings of the ACM/IEEE 19th International Conference on Model Driven Engineering Languages and Systems, MODELS '16, ACM, 2016, pp. 250–260. doi:10.1145/2976767.2976769.
[4] F. Friedrich, J. Mendling, F. Puhlmann, Process model generation from natural language text, in: H. Mouratidis, C. Rolland (Eds.), Advanced Information Systems Engineering, Springer, 2011, pp. 482–496. doi:10.1007/978-3-642-21640-4_36.