Preface

Ian Soboroff¹ · Audrey Tong¹

Published online: 28 December 2017

© Springer Science+Business Media B.V., part of Springer Nature (outside the USA) 2017

The area of Natural Language Processing (NLP) has made significant progress over the last 30 years as a result of the development of machine learning techniques that employ statistical inferences from large training data. These training data were often carefully hand-crafted at huge expense, collected over many years, and targeted a small set of commonly used languages. For various application scenarios, these techniques work quite well. However, they tend to quickly fall apart when these data resources are not readily available. Being able to rapidly retarget existing technologies developed for one language to a new language is tremendously useful in many contexts such as timely access of information from non-English sources, assessing and responding to emergent situations and threats from around the world. The U.S. Defense Advanced Research Projects Agency (DARPA) wants to address this area of need through the Low Resource Languages for Emergent Incidents (LORELEI) program. The goals of the program are to promote and develop language technologies that will work for many languages, with less dependency on large, careful manually annotated data. This special issue, *NLP in Low Resource Languages*, compiles selected papers from the LORELEI program and wider community that tackle various challenges in machine translation (MT), named entity recognition (NER), and situation frame (SF) when applied to languages with limited resources. The SF task consists of choosing between different types of frames and filling in the appropriate information for those frames.

The issue begins with a brief background of the LORELEI program, describing its motivation, research thrust objectives, and concept of operations from DARPA's per-

✉ Audrey Tong
audrey.tong@nist.gov

Ian Soboroff
ian.soboroff@nist.gov

¹ National Institute of Standards and Technology, 100 Bureau Drive, Stop 8940, Gaithersburg, MD 20899, USA

spective. A summary of the Low Resource Human Language Technology (LoReHLT) component evaluation that was conducted in support of the LORELEI program follows and details the evaluation process and results. The bulk of the issue is devoted to descriptions of systems developed for the LoReHLT evaluation as well as novel or interesting approaches that were developed to address challenges of under-resourced languages.

We would like to thank all of the authors and reviewers who contributed and helped prepare this special issue. We hope that this issue serves to support and promote research in NLP technologies in low-resource settings.