Handbook on Natural Language Processing for Requirements Engineering

Alessio Ferrari • Gouri Ginde
Editors

Handbook on Natural Language Processing for Requirements Engineering

Springer

Editors
Alessio Ferrari
University College Dublin (UCD)
Dublin, Ireland

Gouri Ginde
Electrical and Software Engineering,
Schulich School of Engineering
University of Calgary
Calgary, AB, Canada

ISBN 978-3-031-73142-6 ISBN 978-3-031-73143-3 (eBook)
https://doi.org/10.1007/978-3-031-73143-3

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Foreword

Natural Language Processing (NLP) has long played a crucial role in Requirements Engineering (RE), since requirements are often (partly) expressed in natural language. This book involves many of the talented experts in the RE community and provides a comprehensive, accessible treatment of the most important aspects of the application of NLP to RE. It can thus serve both as a primer to newcomers and as a reference to experts, both from academia and industry.

Academics will benefit from getting a thorough and structured treatment of the state of the art, as well as clear guidelines and recommendations to improve the way they conduct research. Practitioners will, on the other hand, get access to the most recent solutions and innovations to address their RE problems and clear practical guidelines to apply them.

The various subjects and techniques being presented in this book are core to NLP for RE. Its chapters cover important topics related to the effective engineering of requirements, such as their classification, traceability or quality assurance. They also address specific and important applications of NLP such as privacy, regulatory compliance or issue trackers. Last, they include practical guidelines for selecting RE techniques and tools, as well as performing empirical studies in a rigorous fashion. The former also tackles the rising use of Large Language Models and more generally generative Artificial Intelligence to better support RE. Taken together, these chapters constitute an all-encompassing treatment of modern RE topics.

In summary, this book is truly a tour de force I can strongly recommend to anybody interested in RE as a research subject or its effective application in practice.

August 7, 2024

Lionel Briand

IEEE Fellow
ACM Fellow
Fellow of the Academy of Science of the Royal Society of Canada

University of Ottawa, Canada, and
Lero Centre, University of Limerick, Ireland

Preface

Requirements are both the cornerstone and the pain point of software engineering. They define the functionalities and qualities that the software must exhibit to satisfy stakeholder needs, yet their incorrect analysis and management can lead to costly errors and delays in the development lifecycle. Moreover, requirements can become challenging to handle in large projects due to their tendency to change, proliferation in number, diversification in terms of abstraction levels and stakeholders' perspectives considered. Last but not least, requirements are notoriously expressed in natural language, a medium subject to multiple interpretations. This inherent ambiguity does not align well with the precision and rigour required in software development.

Requirements Engineering (RE) aims to tackle these issues systematically, incorporating tasks such as defect detection, traceability, classification and more. However, performing these tasks manually can be error-prone, as it demands meticulous attention to detail and is sometimes tedious since RE activities can often become repetitive. For these reasons, requirements engineers have long hoped for automated partners to assist them in addressing the challenges associated with natural language requirements, and relieve their daily frustrations.

Natural Language Processing (NLP) is a field of artificial intelligence focusing on developing algorithms and models to process, interpret and understand natural language data. In order to support RE tasks with automated tools and facilitate the work of requirements engineers, researchers have experimented with NLP techniques for over 20 years, tailoring existing approaches to the specific needs of the RE domain. The interest in NLP techniques for RE has recently witnessed significant growth, also driven by the advancements in NLP technologies, which have become increasingly accessible to researchers and even to the general public, with tools that perform complex tasks such as machine translation and question answering with near-human levels of performance. We do not mention ChatGPT because this name may soon become obsolete given the pace of evolution of the NLP field.

This *Handbook on Natural Language Processing for Requirements Engineering* aims to provide a comprehensive guide on how NLP can be leveraged to enhance various aspects of RE, leading the reader from the exploration of fundamental concepts and techniques to the practical implementation of NLP for RE solutions in real-world scenarios.

The motivation for this book is threefold. Firstly, the field of NLP for RE has accumulated a vast body of knowledge over the years, yet this knowledge is dispersed across various publications and venues, making it challenging for interested individuals to efficiently obtain a comprehensive overview. This book aims to consolidate this knowledge into a single, accessible source, providing readers with a clear roadmap through the complexities of integrating NLP with RE. Secondly, NLP is currently experiencing a transformative period, mainly driven

by advancements in Large Language Models (LLMs). These models present unprecedented opportunities to tackle RE challenges with enhanced effectiveness. By clarifying how NLP can support RE and showcasing state-of-the-art techniques and methodologies, this book serves as a gateway for researchers and practitioners to leverage these advancements in practical applications. Lastly, throughout our career, we have witnessed firsthand the struggles practitioners face in handling natural language requirements and the potential of NLP to alleviate these challenges. We thus aim to help them understand how they can apply NLP to enhance their RE process and, ultimately, the quality of their systems.

The book is structured into three parts, each focusing on distinct aspects of applying NLP to RE:

Part I—NLP for Downstream RE Tasks This part delves into the application of NLP techniques to tasks that are typically part of the RE process. It includes chapters on NLP for requirements classification, requirements similarity and retrieval, requirements traceability, defect detection, and automated terminology and relations extraction.

Part II—NLP for Specialized Types of Requirements and Artefacts This part explores how NLP can be tailored to handle specific requirements types and artefacts. The chapters cover legal requirements processing, privacy requirements acquisition and analysis, user feedback intelligence, mining issue trackers, and analysis of user story requirements.

Part III—NLP for RE in Practice The final part addresses practical applications and tools for implementing NLP in RE. It includes a chapter on the different tools that use NLP techniques for RE tasks, followed by chapters on empirical evaluation of tools, practical guidelines for selecting and evaluating NLP techniques, guidelines on using LLMs in RE, and dealing with data challenges in RE.

This book is designed for a diverse audience. Whether you are a researcher, a practitioner, or a student in the field of RE, this volume offers valuable insights and practical guidance. The content is structured to provide a clear progression from fundamental concepts to advanced techniques, ensuring that readers of all levels can reap the benefits.

We would like to express our gratitude to all the individuals and organizations who have supported us throughout this journey. Special thanks to our colleagues, who provided valuable feedback and encouragement; to the Springer team, which has always been present to guide us throughout the publication process; and most of all, to the authors, who meticulously edited their chapters and carefully reviewed each other's manuscripts. Special thanks go to Daniel M. Berry, who helped us substantially improve the first chapter; to Quim Motger, Muhammad Abbas, and Sallam Abualhaija, who provided us with aesthetically pleasing LaTeX code that was beneficial to all the authors; and to Lionel Briand, for kindly writing a foreword to the volume.

This book reflects the outcome of thorough research, practical insights, and collaborative efforts aimed at connecting theoretical advancements in NLP with their practical applications in RE, thereby bridging the gap between theory and

practice, and providing both a foundation for understanding and a catalyst for innovation. We sincerely hope that the readers will appreciate the efforts of all the people involved in making this book comprehensive and accessible.

August 7, 2024

Alessio Ferrari
Pisa, Italy

Gouri Ginde
Calgary, Canada

Table of Contents

II NLP for Specialized Types of Requirements and Artefacts

List of Contributors

Editors

Alessio Ferrari University College Dublin (UCD), Dublin, Ireland

Consiglio Nazionale delle Ricerche (CNR), Istituto di Scienza e Tecnologie dell'Informazione "A. Faedo" (ISTI), Pisa, Italy

Gouri Ginde Electrical and Software Engineering, Schulich School of Engineering, University of Calgary, Calgary, AB, Canada

Authors

Muhammad Abbas RISE Research Institutes of Sweden, Mälardalen University, Västerås, Sweden

Sallam Abualhaija SnT, University of Luxembourg, Luxembourg, Luxembourg

Waad Alhoshan Imam Mohammad Ibn Saud Islamic University (IMSIU), Riyadh, Saudi Arabia

Chetan Arora Monash University, Clayton, VIC, Australia

Sarmad Bashir RISE Research Institutes of Sweden, Mälardalen University, Västerås, Sweden

Daniel M. Berry Cheriton School of Computer Science, University of Waterloo, Waterloo, ON, Canada

Volodymyr Biryuk Department of Informatics, University of Hamburg, Hamburg, Germany

Mitra Bokaei Hosseini University of Texas at San Antonio, San Antonio, TX, USA

Travis Breaux Carnegie Mellon University, Pittsburgh, PA, USA

Sjaak Brinkkemper Department of Information and Computing Sciences, Utrecht University, Utrecht, The Netherlands

Lionel Briand School of EECS, University of Ottawa, Ottawa, ON, Canada; Lero SFI centre for Software Research and University of Limerick, Limerick, Ireland

Marcello Ceci SnT, University of Luxembourg, Luxembourg, Luxembourg

Jane Cleland-Huang University of Notre Dame, Notre Dame, IN, USA

Fabiano Dalpiaz Department of Information and Computing Sciences, Utrecht University, Utrecht, The Netherlands

Eduard Paul Enoiu Mälardalen University, Västerås, Sweden

Jannik Fischbach Netlight Consulting GmbH, Munich, Germany; fortiss GmbH, Munich, Germany

Julian Frattini Blekinge Institute of Technology, Karlskrona, Sweden

Davide Fucci Blekinge Institute of Technology, Karlskrona, Sweden

Smita Ghaisas TCS Research, Pune, India

Jin L.C. Guo McGill University, Montreal, QC, Canada

Erik Kamsties Fachhochschule Dortmund, University of Applied Sciences and Arts, Dortmund, Germany

Clara Lüders Department of Informatics, University of Hamburg, Hamburg, Germany

Walid Maalej Department of Informatics, University of Hamburg, Hamburg, Germany

Daniel Mendez Blekinge Institute of Technology, Karlskrona, Sweden; fortiss GmbH, Munich, Germany

Lloyd Montgomery Department of Informatics, University of Hamburg, Hamburg, Germany

Nan Niu University of Cincinnati, Cincinnati, OH, USA

Fabian Panse Hasso-Plattner-Institut für Digital Engineering gGmbH, Potsdam, Germany

Cristina Ribeiro School of Applied Computing, Sheridan College, Oakville, ON, Canada

Mehrdad Saadatmand RISE Research Institutes of Sweden, Västerås, Sweden

Mehrdad Sabetzadeh University of Ottawa, Ottawa, ON, Canada

Anmol Singhal TCS Research, Delhi, India

Jan-Philipp Steghöfer XITASO GmbH IT & Software Solutions, Augsburg, Germany

Daniel Sundmark Mälardalen University, Västerås, Sweden

Sri Fatimah Tjong School of Computer Science, University of Nottingham Malaysia Campus, Semenyih, Selangor Darul Ehsan, Malaysia

Michael Unterkalmsteiner Blekinge Institute of Technology, Karlskrona, Sweden

Andreas Vogelsang University of Cologne, Köln, Germany

Jialiang Wei EuroMov Digital Health in Motion, Univ Montpellier, IMT Mines Alès, Alès, France

Jianzhang Zhang Hangzhou Normal University, Hangzhou Zhejiang, P.R. China

Liping Zhao University of Manchester, Manchester, UK