NLP for RE

Introduction of RE and NLP

Dr. Gouri Deshpande, Dr. Mariana Bento*, and Mitra Mirshafiee (Graduate Student)

Electrical and Software Engineering; *Biomedical Engineering

Tutorial outline

A. Introduction of RE and NLP (not interactive) (25 mins)

B. Data gathering (interactive)(1hr)

C. Pre-processing using NLP operations for a dataset from Bugzilla (interactive) (1hr)

D. Modeling and evaluation (interactive) (30 mins)

Intros

Mariana Bento

Assistant Professor,
University of Calgary
mariana.pinheirobent@ucalgary.ca

Website:
https://profiles.ucalgary.ca/mariana-bento

LinkedIn:
https://www.linkedin.com/in/mariana-bento-b7a75010b/

Gouri Deshpande

Assistant Professor,

University of Calgary
Gouri.deshpande@ucalgary.ca

Website:
https://pages.cpsc.ucalgary.ca/~gouri.deshpande/

LinkedIn:
https://www.linkedin.com/in/deshpandegouri/

Mitra Mirshafiee

Master of CS Student,
University of Calgary
Mitrasadat.mirshafie@ucalgary.ca

Website: https://mitramirshafiee.ir/

LinkedIn:
https://www.linkedin.com/in/mitra-mirshafiee-data-scientist/

Learning Goals

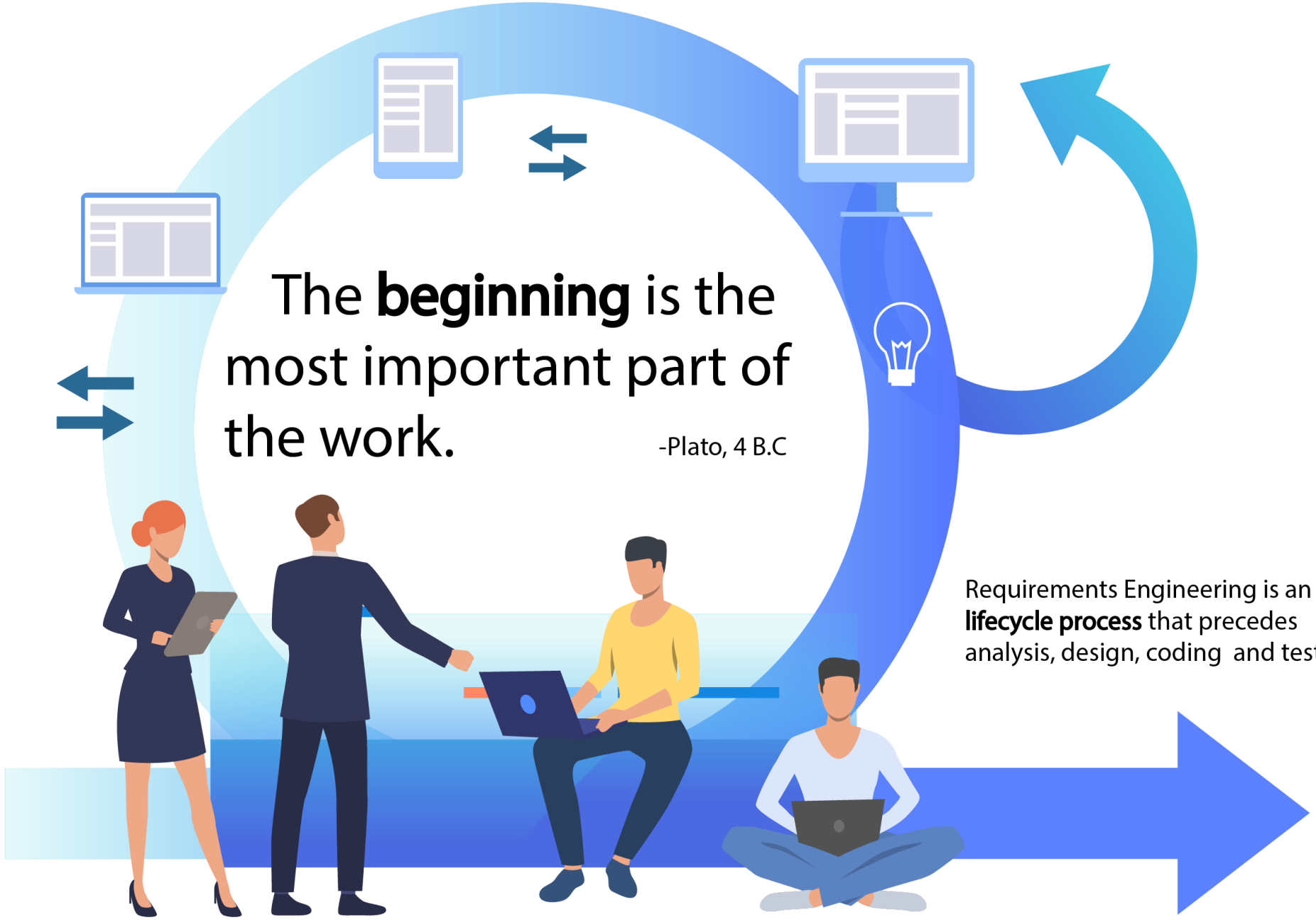
UNIVERSITY OF CALGARY

Introduction to RE

Introduction to NLP

UNIVERSITY OF CALGARY

Introduction of RE



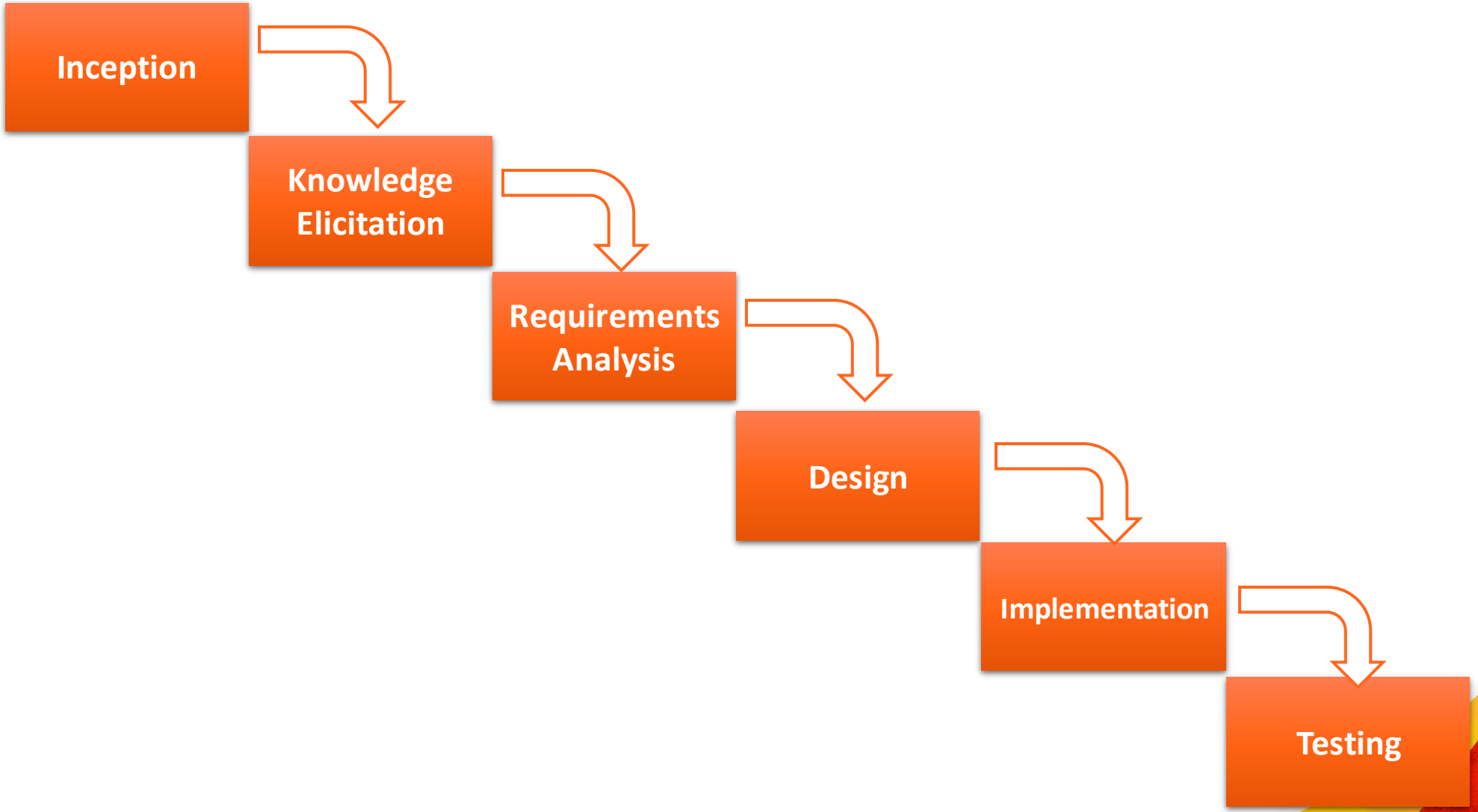
Requirements Engineering is an **early lifecycle process** that precedes analysis, design, coding and testing.[1]

[1] Bencomo, Nelly, Jin LC Guo, Rachel Harrison, Hans-Martin Heyn, and Tim Menzies. "The Secret to Better AI and Better Software (Is Requirements Engineering)." *IEEE Software* 39, no. 1 (2021): 105-110.

Brief Introduction to Software Development Process

UNIVERSITY OF CALGARY

- A simplified Linear model of software development process:

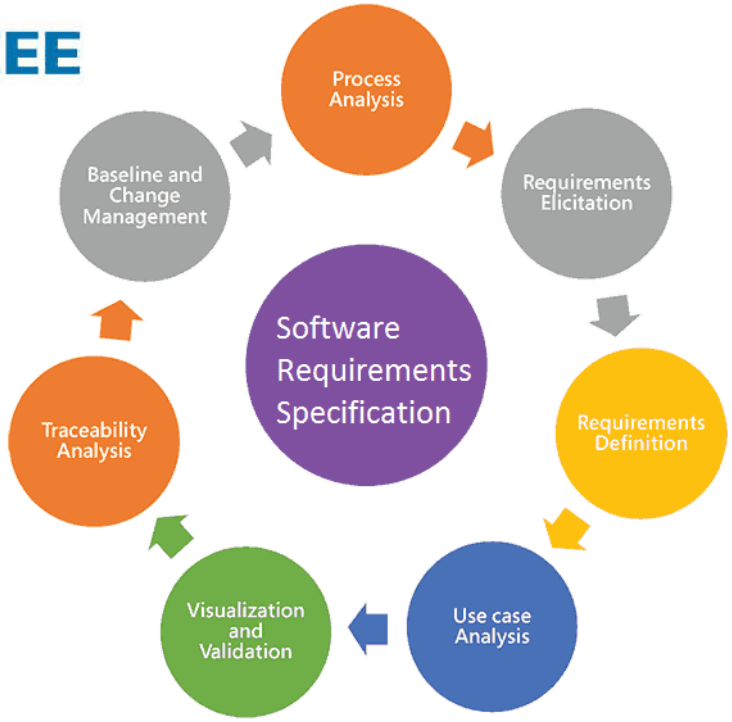


Software requirements Specification

Introduction: For Software product and defined by the customer

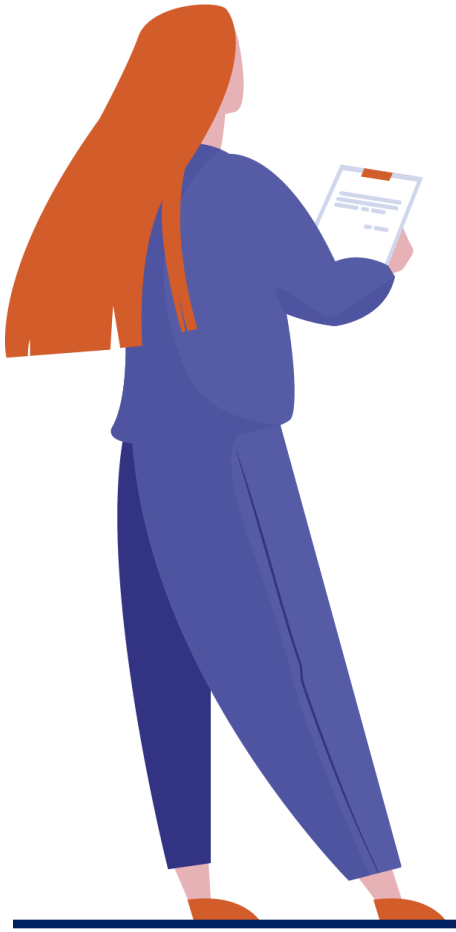
Overall Description: Functions, Users, Constraints, etc.

Overall Description: Functional, non-functional, interfaces, etc.



Requirements Elicitation Techniques

- Brain Storming
- Document Analysis
- Focus Group
- Interface Analysis
- Interviews
- Observation
- Process Modeling
- Prototype
- Workshops
- Surveys/Questionnaire



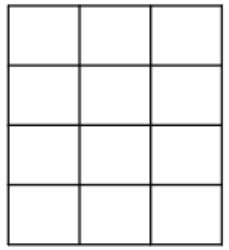
https://systemsplusgroup.blogspot.com/2017/02/requirement-elicitation-techniques.html

UNIVERSITY OF CALGARY

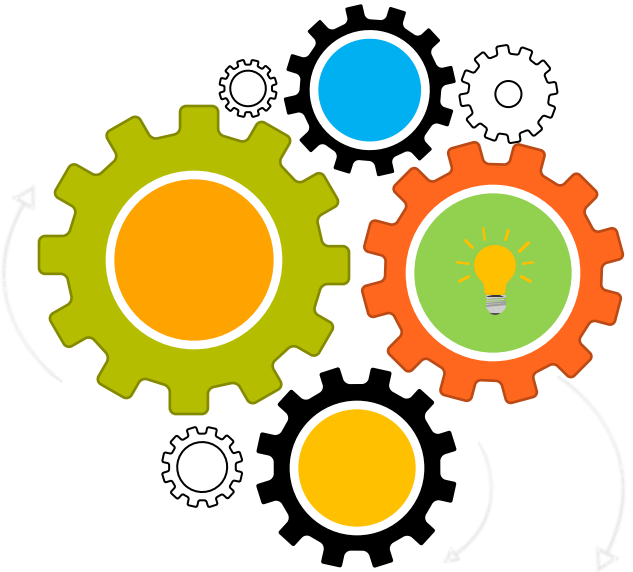
Introduction to NLP

NLP and Machine learning

Count
TFIDF
Hashing Vectorizer
On all documents



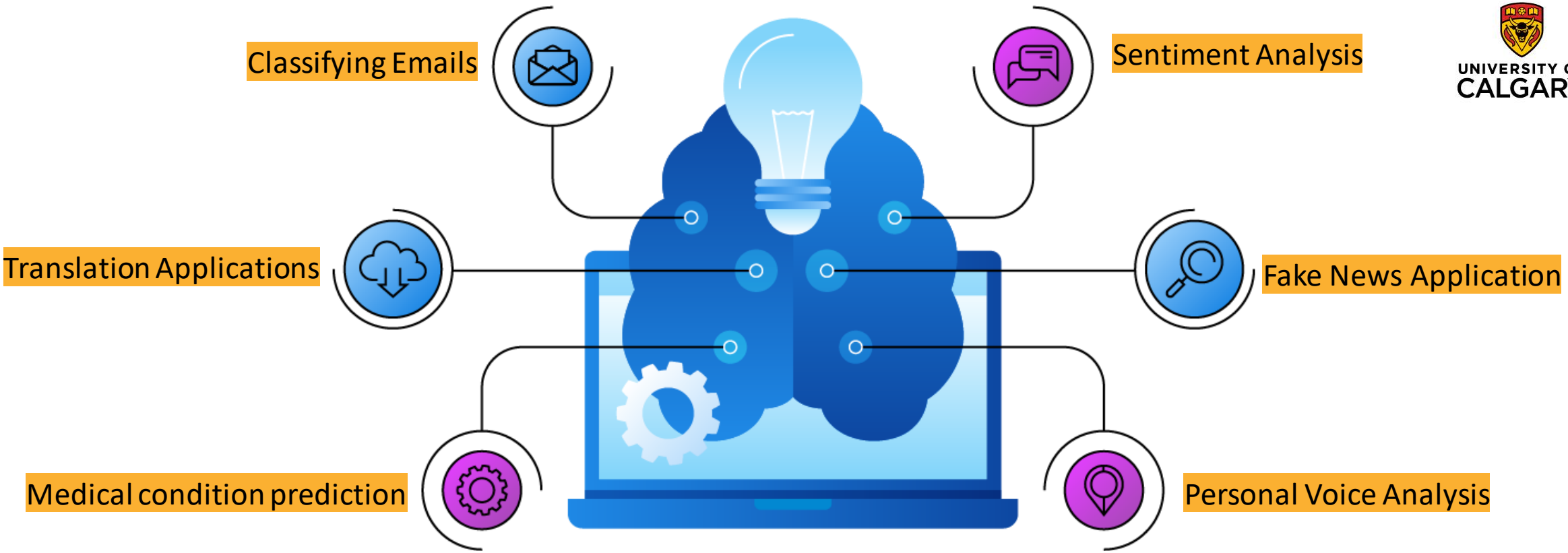
Huge Matrix with Numbers



ML/DL Model

Classification
Language Generation
Part of Speech Tagging
Question Answering
Etc.

Source: https://www.datajango.com/deep-learning-and-nlp/



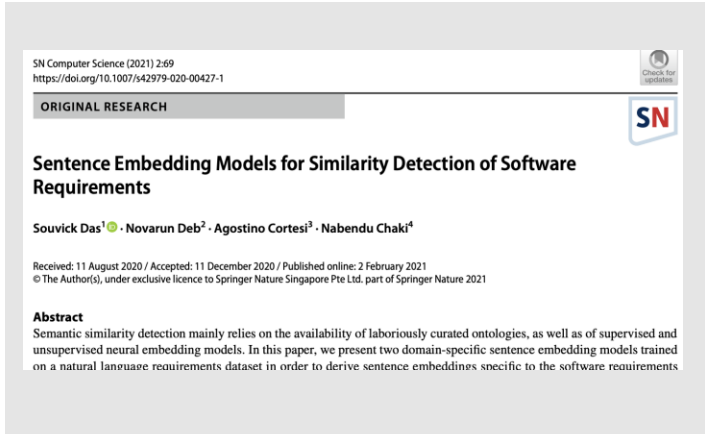
Natural Language Processing (NLP)

NLP is a subfield of **artificial intelligence** that deals with the link between computers and human language.
NLP involves programming computers to process and analyze a massive amount of **natural language data**.

What are Some Applications of NLP in the Industry?

- Search engines
- Advanced text editors: Such as Grammarly.com
- Computational advertising
- Fraud detection
- Sentiment analysis
- Opinion mining
- Text summarization
- Context analysis

Example Papers



SN Computer Science (2021) 2:69
https://doi.org/10.1007/s42979-020-00427-1

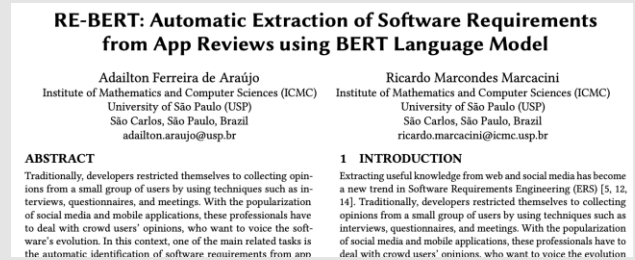
ORIGINAL RESEARCH

Sentence Embedding Models for Similarity Detection of Software Requirements

Souvick Das[1] · Novarun Deb[2] · Agostino Cortesi[3] · Nabendu Chaki[4]

Received: 11 August 2020 / Accepted: 11 December 2020 / Published online: 2 February 2021
© The Author(s), under exclusive licence to Springer Nature Singapore Pte Ltd. part of Springer Nature 2021

Abstract
Semantic similarity detection mainly relies on the availability of laboriously curated ontologies, as well as of supervised and unsupervised neural embedding models. In this paper, we present two domain-specific sentence embedding models trained on a natural language requirements dataset in order to derive sentence embeddings specific to the software requirements



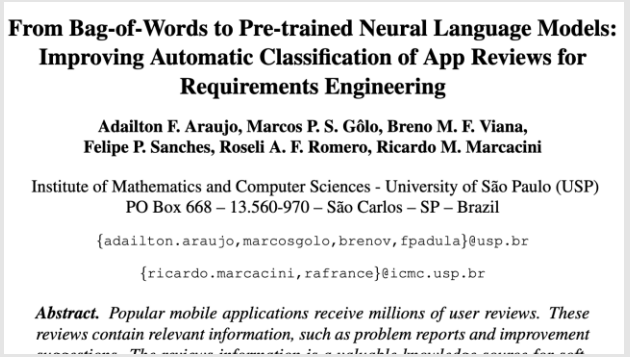
RE-BERT: Automatic Extraction of Software Requirements from App Reviews using BERT Language Model

Adailton Ferreira de Araújo
Institute of Mathematics and Computer Sciences (ICMC)
University of São Paulo (USP)
São Carlos, São Paulo, Brazil
adailton.araujo@usp.br

Ricardo Marcondes Marcacini
Institute of Mathematics and Computer Sciences (ICMC)
University of São Paulo (USP)
São Carlos, São Paulo, Brazil
ricardo.marcacini@icmc.usp.br

ABSTRACT
Traditionally, developers restricted themselves to collecting opinions from a small group of users by using techniques such as interviews, questionnaires, and meetings. With the popularization of social media and mobile applications, these professionals have to deal with crowd users' opinions, who want to voice the software's evolution. In this context, one of the main related tasks is the automatic identification of software requirements from app

1 INTRODUCTION
Extracting useful knowledge from web and social media has become a new trend in Software Requirements Engineering (ERS) [5, 12, 14]. Traditionally, developers restricted themselves to collecting opinions from a small group of users by using techniques such as interviews, questionnaires, and meetings. With the popularization of social media and mobile applications, these professionals have to deal with crowd users' opinions, who want to voice the evolution



From Bag-of-Words to Pre-trained Neural Language Models: Improving Automatic Classification of App Reviews for Requirements Engineering

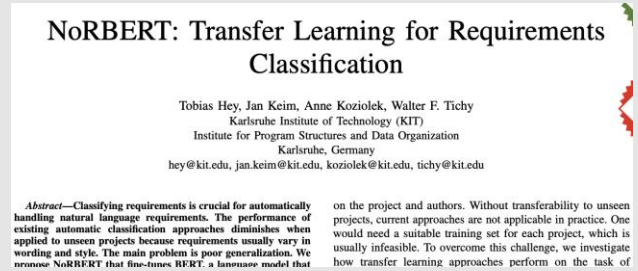
Adailton F. Araujo, Marcos P. S. Gôlo, Breno M. F. Viana, Felipe P. Sanches, Roseli A. F. Romero, Ricardo M. Marcacini

Institute of Mathematics and Computer Sciences - University of São Paulo (USP)
PO Box 668 – 13.560-970 – São Carlos – SP – Brazil

{adailton.araujo,marcosgolo,brenov,fpadula}@usp.br

{ricardo.marcacini,rafrance}@icmc.usp.br

Abstract. *Popular mobile applications receive millions of user reviews. These reviews contain relevant information, such as problem reports and improvement*



NoRBERT: Transfer Learning for Requirements Classification

Tobias Hey, Jan Keim, Anne Koziolek, Walter F. Tichy
Karlsruhe Institute of Technology (KIT)
Institute for Program Structures and Data Organization
Karlsruhe, Germany
hey@kit.edu, jan.keim@kit.edu, koziolek@kit.edu, tichy@kit.edu

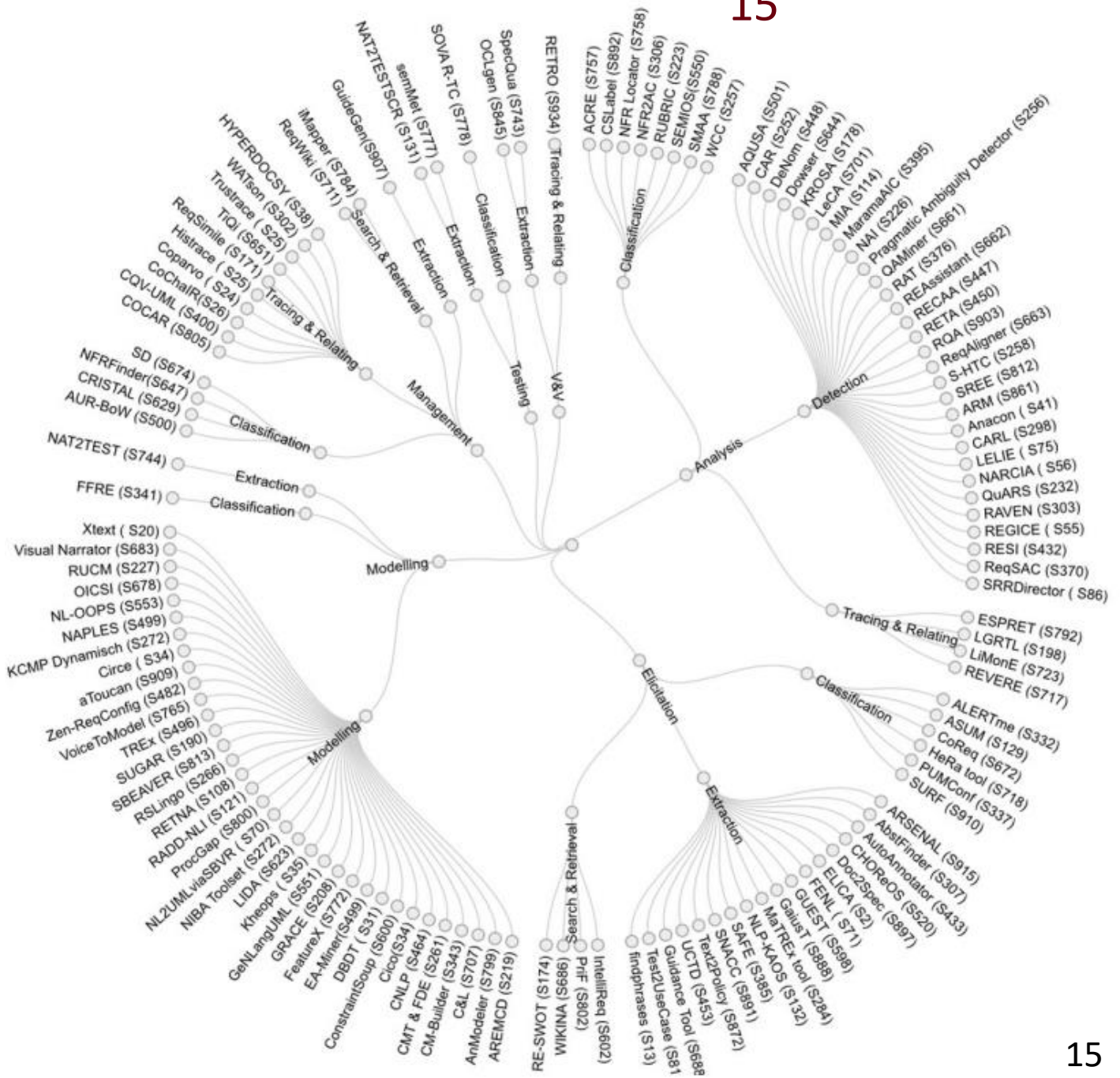
***Abstract*—Classifying requirements is crucial for automatically handling natural language requirements. The performance of existing automatic classification approaches diminishes when applied to unseen projects because requirements usually vary in wording and style. The main problem is poor generalization. We propose NoRBERT that fine-tunes BERT, a language model that**

on the project and authors. Without transferability to unseen projects, current approaches are not applicable in practice. One would need a suitable training set for each project, which is usually infeasible. To overcome this challenge, we investigate how transfer learning approaches perform on the task of

Applications of NLP in RE

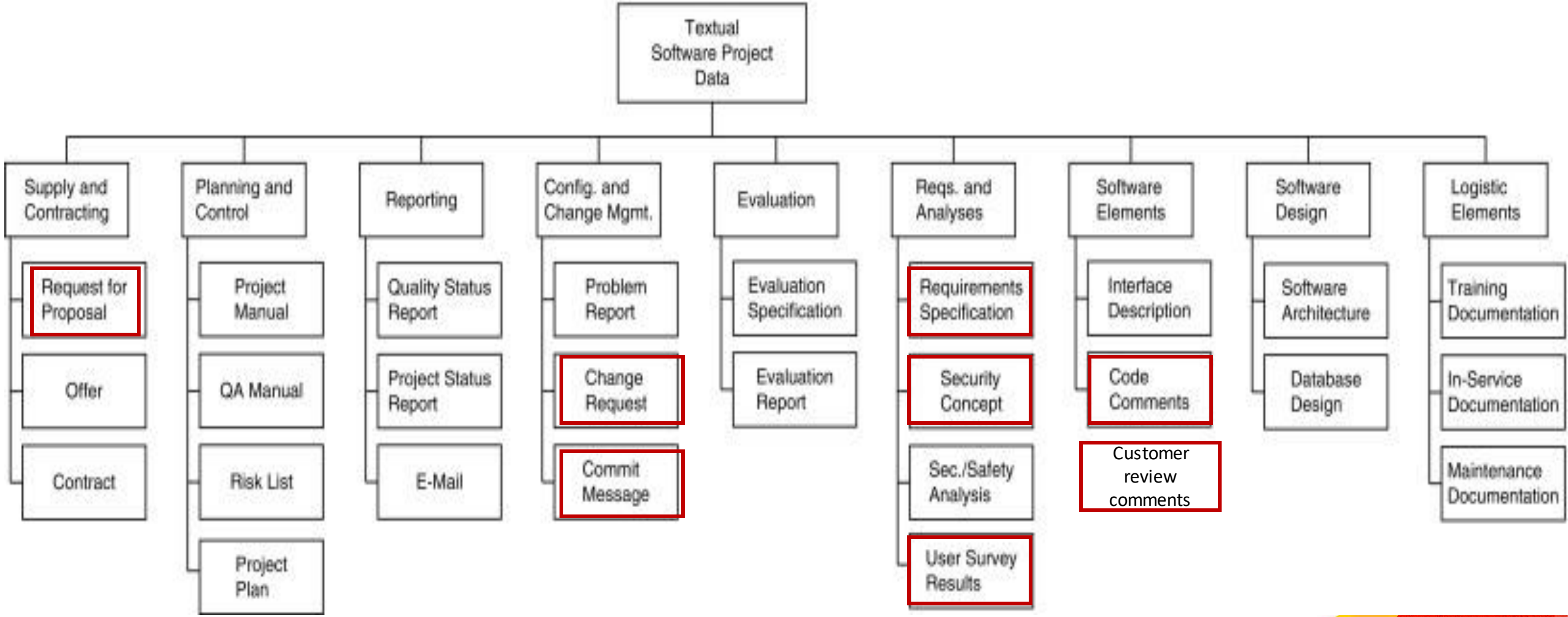
NLP for RE Tools

Zhao, Liping, et al. **"Natural language processing (NLP) for requirements engineering: A systematic mapping study**." *arXiv preprint arXiv:2004.01099* (2020).



Software Projects are Textual in Nature

Textual Data in Software Projects



Bird, Christian, Tim Menzies, and Thomas Zimmermann, eds. The art and science of analyzing software data. Elsevier, 2015.

Textual Software Project Data and Retrieval

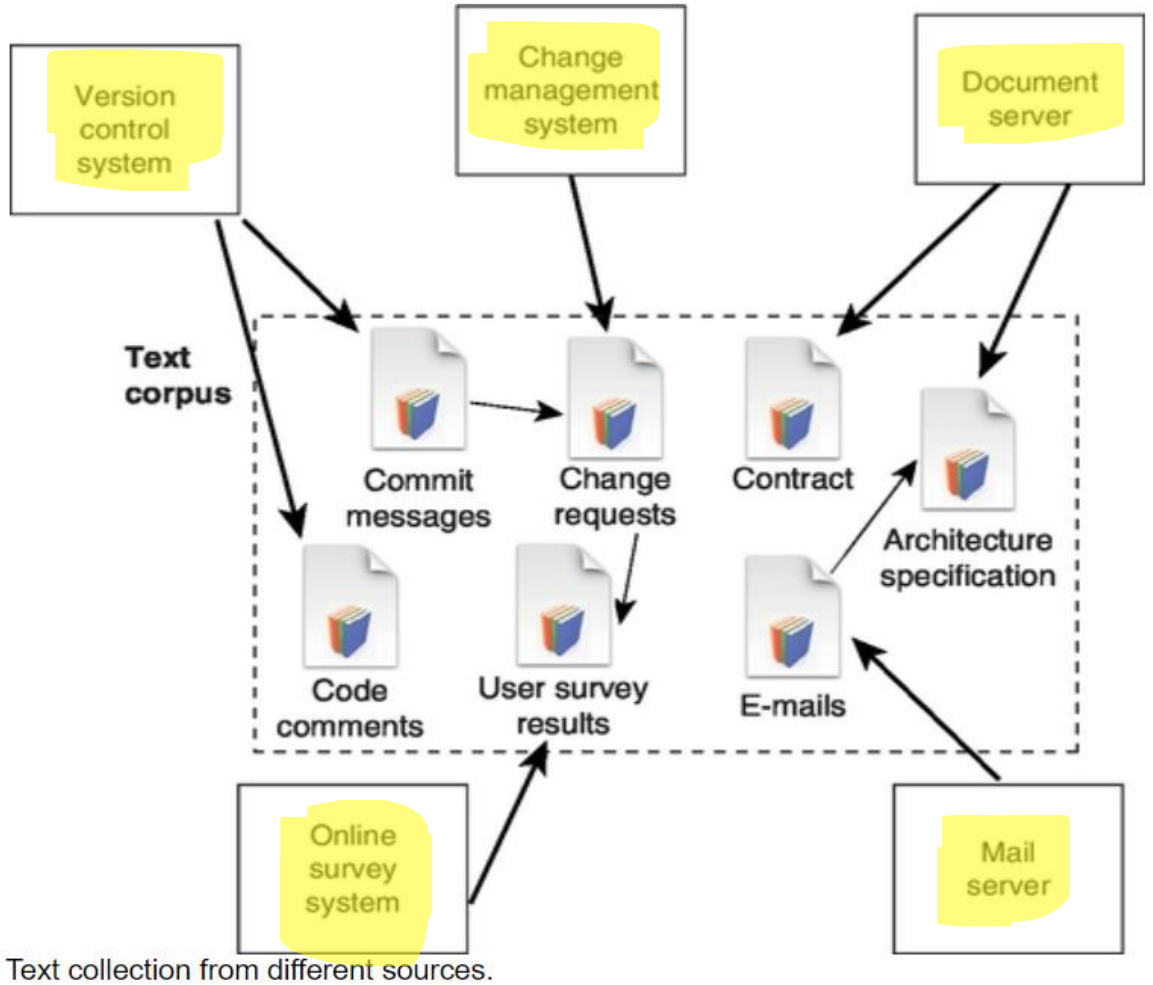
Custom extraction tool

API's

Formatted files

Anonymize

Web crawling



3.2 Text collection from different sources.

Arnaoudova, Venera, et al. "The use of text retrieval and natural language processing in software engineering." *Proceedings of the 37th International Conference on Software Engineering-Volume 2.* 2015.

Importance of NLP in Software Engineering

- Text Retrieval (TR) and NLP in software is one of the fastest growing areas of research in SE. [2]
- Exposing the SE community to these techniques and their applications in SE would help to fill a gap in their current background and allow them to immediately use TR and NLP to advance their research.
- In particular, for TR, approaches such as Vector Space Model, Latent Semantic Analysis, Latent Dirichlet Association, Language Models will be covered. NLP techniques covered will include part-of-speech tagging, stemming, stopwords elimination, semantics analysis, sentiment analysis, etc.

References

- Bird, Christian, Tim Menzies, and Thomas Zimmermann, eds. The art and science of analyzing software data. Elsevier, 2015.
- Arnaoudova, Venera, et al. "The use of text retrieval and natural language processing in software engineering." Proceedings of the 37th International Conference on Software Engineering-Volume 2. 2015.
- Juergens, Elmar, et al. "Do code clones matter?." 2009 IEEE 31st International Conference on Software Engineering. IEEE, 2009.
- Juergens, Elmar, et al. "Can clone detection support quality assessments of requirements specifications?." Proceedings of the 32nd ACM/IEEE International Conference on Software Engineering-Volume 2. ACM, 2010.
- https://medium.com/square-corner-blog/topic-modeling-optimizing-for-human-interpretability-48a81f6ce0ed
- Reference to a few pictures designed by : vectorjuice / Freepik, pch.vector / Freepik, Omnibus-Type / Freepik, katemangostar / Freepik

SURVEY

https://forms.gle/TuVVuAkf4JWHFDp17