NLP for RE

1st Gouri Deshpande
dept. Electrical and Software Engineering
University of Calgary
Calgary, Canada
gouri.deshpande@ucalgary.ca

2nd Mariana Bento
dept. Electrical and Software Engineering
University of Calgary
Calgary, Canada
mariana.pinheirobent@ucalgary.ca

***Abstract*—NLP for RE tutorial aims to provide the basics of how NLP could be used to process data for automation. It considers various instances that could be used to get started with applied NLP inclusive of data gathering, preprocessing, modeling and evaluation. We teach you how to apply various NLP techniques to a data set you bring with you or example systems we have as a backup. The take-aways from the tutorial are 1) How to gather data from public repositories such as utilizing REST-API's, 2) Pre-processing requirements utilizing various NLP techniques 3) Develop a classification ML model for requirement dependencies.**

I. Motivation

In the recent past automation of RE life cycle stages such as elicitation and dependency extraction have been explored widely using NLP/ML techniques. This tutorial will establish the basics for conducting advanced research in the area of RE for applied NLP and ML.

II. Objectives

Objectives of this tutorial are, 1) providing techniques to gather RE data from public repositories 2) Pre-processing requirements utilizing various NLP techniques 3) Develop a classification ML model for requirement dependencies

Duration: Half day

III. Outline of Topics

Outline of the tutorial is as follows. Goal is to help the newbie RE researchers get started with problem solving beginning with data gathering, followed by showcasing NLP classification schema and hands-on on utilizing few of these techniques through a case study (data from Bugzilla).

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

1) What is NLP and classification taxonomy:
2) Various RE public repositories for data mining:

B. Data gathering (interactive)(1hr)

1) Using REST API's:
2) Using BeautifulSoup for data scraping: a brief introduction only:

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

1) Data selection:
2) Data cleaning:
3) Feature extraction:

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

1) Building ML model for classification:
2) Evaluation metrics:

IV. Target Audience

Attendees are expected to have basic Python coding experience. No prior experience in data gathering or NLP is needed. I hope there will be interest among newbie researchers and students in attending this hands-on tutorial.

V. Tutorial History

Mariana Bento has conducted an undergraduate course (over 200 hundred students): Software Requirements in the past. Introduction (4.1) will be drawn from her knowledge.

Gouri Deshpande has taught part of this tutorial's content to summer students and undergraduates at her lab. She also covered some part of this tutorial during a tutorial for 13 graduate students as part of the graduate-level course that was taught by my supervisor (Dr. Guenther Ruhe).

VI. Presenters' Bios

- **Gouri Deshpande:** Gouri is an Assistant Professor at the dept. of Electrical and Software Engineering, University of Calgary. She graduated in 2022 and her thesis work explored requirements dependency extraction using advanced Machine Learning approaches. Gouri worked as a Software Engineer before switching to academia. She holds PhD in Computer Science from the University of Calgary.
- **Mariana Bento:** Mariana is an Assistant Professor of Biomedical and of Electrical and Software Engineering, a full member of the Hotchkiss Brain Institute, funded by NSERC, who has expertise in medical image processing and machine learning applied to aging and dementia. She develops more robust and reliable open tools in neuroimaging and is engaged in open science activities and EDI activities including conducting workshops.

VII. Publicity

We plan to promote this tutorial on social media such as Twitter, LinkedIn and Slack channels to attract the participants.