ICON 2020

17th International Conference on Natural Language Processing

Proceedings of the TermTraction 2020 Shared Task

December 18 - 21, 2020
Indian Institute of Technology Patna, India

©2020 NLP Association of India (NLPAI)

Introduction

These shared task proceedings concluded the shared task on Unsupervised Technical Domain Terms Extraction, also named as TermTraction - 2020, launched on 7th October 2020. The shared task was collocated with the 17th International Conference on Natural Language Processing (ICON 2020), held at IIT-Patna, India. The goal of the shared task was to extract domain-specific terms given an English document.

Four technical domains were chosen for this task - BioChemistry, Communication, Computer Science and Law. The documents that were part of the training corpus did not contain any marked or annotated domain terms. The participants were tasked to develop an unsupervised algorithm for the extraction of the terms.

We received three system submissions and system description papers. Each system description paper was reviewed by two members of the reviewing committee – all papers were accepted. Macro F1-score was used as the evaluation metrics.

Two major categories of algorithms were used by the participants - Graph-based and Filter based. Textrank was the most used graph-based technique. Rapid Automatic Keyword Extraction (RAKE) and PYthon Automated Term Extraction (PYATE) were the filter based techniques employed by the teams. We would like to thank the ICON-2020 organizers, the shared task participants, the authors, and the reviewers for making this shared task successful.

Shared task page: `http://ssmt.iiit.ac.in/TermTraction`
Main conference page: `https://www.iitp.ac.in/~ai-nlp-ml/icon2020/index.html`

Organizing Committee:

Dipti Misra Sharma (IIIT-Hyderabad)
Asif Ekbal (IIT-Patna)
Karunesh Arora (C-DAC, Noida)
Sudip Kumar Naskar (Jadavpur University)
Dipankar Ganguly (C-DAC, Noida)
Sobha L (AUKBC-Chennai)
Radhika Mamidi (IIIT-Hyderabad)
Sunita Arora (C-DAC, Noida)
Pruthwik Mishra (IIIT-Hyderabad)
Vandan Mujadia (IIIT-Hyderabad)

Table of Contents

Graph Based Automatic Domain Term Extraction
Hema Ala and Dipti Sharma 1

Unsupervised Technical Domain Terms Extraction using Term Extractor
Suman Dowlagar and Radhika Mamidi 5

N-Grams TextRank A Novel Domain Keyword Extraction Technique
Saransh Rajput, Akshat Gahoi, Manvith Reddy and Dipti Mishra Sharma 9

Shared Task Program

Monday, December 21, 2020

+ 14:00 - 14:30 **Talk by Sobha L, AUKBC-Chennai**

+ 14:30 - 14:45 Shared Task Overview

Presentations

16:00 - 16:10 *Graph Based Automatic Domain Term Extraction*
Hema Ala and Dipti Sharma

16:13 - 16:23 *N-Grams TextRank A Novel Domain Keyword Extraction Technique*
Saransh Rajput, Akshat Gahoi, Manvith Reddy and Dipti Mishra Sharma

16:26 - 16:36 *Unsupervised Technical Domain Terms Extraction using Term Extractor*
Suman Dowlagar and Radhika Mamidi