

ICON 2020

**17th International Conference on Natural Language
Processing**

Proceedings of the TechDOfication 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 Technical Domain Identification, named as TechDOfication 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 automatically identify the technical domain of a given text in a specified language. The languages included in this task were English, Bangla, Gujarati, Hindi, Malayalam, Marathi, Tamil, and Telugu.

Two subtasks were part of this shared task. The first subtask was to identify a Coarse-grained technical domain like Computer Science, Communication Technology, Management, Math, Physics, Life Science, Law etc. The second subtask involved the task of identification of fine-grained subdomains for the Computer Science domain like Operating System, Computer Network, Database etc. The first subtask was conducted in all the mentioned languages while the second subtask only involved English.

We received nine 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 scores were used to evaluate the systems.

Both Machine Learning and Neural Network methods were used by different teams in this shared task. Support Vector Machine and Voting Classifiers were the predominantly used Machine Learning models. Different Neural architectures like Multi-Layer Perceptron, BiLSTM, BERT based models, Graph Convolutional Neural Networks, XLM-RoBERTa, Multichannel LSTM-CNN were also used. 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/techdofication>

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

<i>MUCS@TechDOfication using FineTuned Vectors and n-grams</i> Fazlourrahman Balouchzahi, M D Anusha and H L Shashirekha	1
<i>A Graph Convolution Network-based System for Technical Domain Identification</i> Alapin Kuila, Ayan Das and Sudeshna Sarkar	6
<i>Multichannel LSTM-CNN for Telugu Text Classification</i> Sunil Gundapu and Radhika Mamidi	11
<i>Multilingual Pre-Trained Transformers and Convolutional NN Classification Models for Technical Domain Identification</i> Suman Dowlagar and Radhika Mamidi	16
<i>Technical Domain Identification using word2vec and BiLSTM</i> Koyel Ghosh, Dr. Apurbalal Senapati and Dr. Ranjan Maity	21
<i>Automatic Technical Domain Identification</i> Hema Ala and Dipti Sharma	27
<i>Fine-grained domain classification using Transformers</i> Akshat Gahoi, Akshat Chhajer and Dipti Mishra Sharma	31
<i>TechTexC: Classification of Technical Texts using Convolution and Bidirectional Long Short Term Memory Network</i> Omar Sharif, Eftekhari Hossain and Mohammed Moshikul Hoque	35
<i>An Attention Ensemble Approach for Efficient Text Classification of Indian Languages</i> Atharva Kulkarni, Amey Hengle and Rutuja Udyawar	40

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

- 16:45 - 16:55 *Automatic Technical Domain Identification*
Hema Ala and Dipti Sharma
- 16:58 - 17:08 *Technical Domain Identification using word2vec and BiLSTM*
Koyel Ghosh, Dr. Apurbalal Senapati and Dr. Ranjan Maity
- 17:11 - 17:21 *A Graph Convolution Network-based System for Technical Domain Identification*
Alapan Kuila, Ayan Das and Sudeshna Sarkar
- 17:24 - 17:34 *Fine-grained domain classification using Transformers*
Akshat Gahoi, Akshat Chhajer and Dipti Mishra Sharma
- 17:37 - 17:47 *MUCS@TechDOfication using FineTuned Vectors and n-grams*
Fazlourrahman Balouchzahi, M D Anusha and H L Shashirekha
- 17:50 - 18:00 *TechTexC: Classification of Technical Texts using Convolution and Bidirectional Long Short Term Memory Network*
Omar Sharif, Eftekhar Hossain and Mohammed Moshiul Hoque
- 18:03 - 18:13 *An Attention Ensemble Approach for Efficient Text Classification of Indian Languages*
Atharva Kulkarni, Amey Hengle and Rutuja Udyawar
- 18:16 - 18:26 *Multilingual Pre-Trained Transformers and Convolutional NN Classification Models for Technical Domain Identification*
Suman Dowlagar and Radhika Mamidi
- 18:29 - 18:39 *Multichannel LSTM-CNN for Telugu Text Classification*
Sunil Gundapu and Radhika Mamidi

