

Preface

Since its inception in 2008, the FIRE community has grown in stature and has played a prominent role in conducting Information Retrieval Evaluation in South Asia. FIRE 2016 workshop proceedings is the biggest volume in history of FIRE so far, with a diverse participation of 64 teams.

We hosted the following 7 tracks in 2016.

- **Consumer Health Information Search (CHIS)** : This track aims to investigate complex health information search in scenarios where users search for health information with more than just a single correct answer, and look for multiple perspectives from diverse sources both from medical research and from real world patient narratives.
- **Detecting Paraphrases in Indian Languages (DPIL)** : In this track, given a pair of sentences in the same language, participants are asked to detect the semantic equivalence between the sentences. The shared task was proposed for four Indian languages namely, Tamil, Malayalam, Hindi, and Punjabi.
- **Information Extraction from Microblogs Posted during Disasters** : This track offered IR challenges on tweets collected during a disaster event (here, Nepal earthquake 2015) for broad queries like resource need, resource identification etc.
- **Persian Plagiarism Detection (PersianPlagDet)** : This track was on detecting plagiarism in documents written in Persian (subtask on text alignment) as well as to provide corpora that contain cases of plagiarism in Persian (subtask on corpus construction).
- **Personality Recognition in Source Code (PR-SOCO)** : This track addressed the problem of predicting an author's personality traits from her source code. That is, given a set of source codes written in Java by students who answered a personality test, participants had to predict their personality traits.
- **Shared Task on Mixed Script Information Retrieval (MSIR)** : This track hosted two subtasks - (a) Subtask-1 was on classifying code-mixed cross-script question; b) Subtask-2 was on information retrieval of Hindi-English code-mixed tweets.

- **Shared Task on Code Mix Entity Extraction in Indian Languages (CMEE-IL)** : The objective of the task was to encourage research in entity recognition and extraction for the code mixed social media text. This track aimed at identification of the various entities such as person names, organization names, movie names, location names in a given tweet. The tweets were written in Roman script and has code mix, where an Indian Language is mixed with English.

FIRE 2016 saw the introduction of three new tracks Consumer Health Information Search, Detecting Paraphrases in Indian Language and Information Extraction from Microblogs Posted during Disasters. Other tracks have evolved over some of the previous FIRE iterations offering various facets of a central problem.

We express our heartfelt gratitude to the track organizers for taking pains in organizing these interesting tracks and thus presenting exciting research problems to the community.

FIRE data has been used frequently for empirical studies across the community in information access. Papers in reputed journals like ACM TALIP, IPM, IR etc. and conferences like SIGIR, CIKM etc. have regularly cited FIRE data. The download of FIRE data has increased considerably over the past few years.

We look forward to continuing this endeavour in future. We will consider this effort worthy if readers find the datasets and experiments reported in this volume useful.

Prasenjit Majumder
Mandar Mitra
Parth Mehta
Jainisha Sankhvara
Kripabandhu Ghosh