**CODE-MIXED QUESTION ANSWERING CHALLENGE**

**Webpage:** [http://tts.speech.cs.cmu.edu/codemixedqa/](http://tts.speech.cs.cmu.edu/codemixedqa/)

Code-Mixing (CM) is the phenomenon of “embedding of linguistic units such as phrases, words and morphemes of one language into an utterance of another language”. It is a frequent phenomenon in multilingual communities.

To increase the reach, impact and effectiveness of Question Answering (QA) in multi-lingual societies, it is imperative to support QA in CM languages.

**This is a Question Answering task in Code-Mixed languages from two dominantly spoken Indian languages: Hinglish (Hindi+English) and Tenglish (Telugu+English).**

Consider an English question: “Who wrote the English translation of the preface of Rabindra Nath Tagore’s Gitanjali?” expressed in Hinglish (code-mixing) as follows:

```
Rabindra Nath Tagore ki Gitaanjali ke english translation ka preface kisne likha tha?
```

- Most QA systems expect a single language of interaction.
- This hampers the ability of a multi-lingual user to interact naturally with the QA system.

**OUR GOAL:**

The goal of this Challenge is to:
- Gather dataset for Code-Mixed questions and answers
- Analyze and create Challenge datasets
- Organize open evaluations on Code-Mixed Question Answering

**WE ARE SEEKING ...**

- Informants who are native speakers of Hindi and Telugu to volunteer in the data collection task
- Students/QA researchers to participate in the challenge of Code-Mixed Question Answering

**TASK TIMELINE:**
(tentative)

- Training data release: 30 Jan 2018
- Test Batch 1: 13 Feb 2018
- Test Batch 2: 27 Feb 2018
- Test Batch 3: 13 Mar 2018
- Test Batch 4: 27 Mar 2018
- Test Batch 5: 10 Apr 2018

**Paper Submission for a special session at EMNLP/NIPS (tentative)** - date and venue to be announced soon

Code-Mixing with non-standardized spellings and cross scripted usage of languages is very common in most social-media fora and informal communication. Let’s take the first step towards addressing an important NLP application i.e QA in this eccentric and prevalent form of language usage.

**Organizers:**
- Alan W Black, Carnegie Mellon University
- Khyathi Raghavi Chandu, Carnegie Mellon University
- Manoj Chinnakotla, Microsoft, India
- Eric Nyberg, Carnegie Mellon University

**Contact:**
- awb@cs.cmu.edu
- kchandu@cs.cmu.edu

**Webpage:** [http://tts.speech.cs.cmu.edu/codemixedqa/](http://tts.speech.cs.cmu.edu/codemixedqa/)