Call for Participation

2018 NLP Challenge on Machine Reading Comprehension

2018 NLP Challenge on Machine Reading Comprehension is hosted by **Chinese Information Processing Society of China** (CIPS) and **China Computer Federation** (CCF), and organized jointly by Baidu Inc., Committee on Evaluation of CIPS (CIPS CE), and Technical Committee on Chinese Information Technology of CCF (CCF TCCI). **The official registration to the competition will be open on March 1, 2018. The winning teams will share a total of 100,000RMB bonus** and the competition forum and award ceremony will be held at the Third Language & Intelligence Summit. All researchers and developers are welcomed to this competition.

Background

Language & Intelligence Summit, initiated by Chinese Information Processing Society of China and China Computer Federation, has been held successfully in 2016 and 2017. It aims to gather scholars and experts to discuss and probe into the new technologies and developments in the field of Language and Intelligence. **The Third Language & Intelligence Summit will be held in Beijing on July 28, 2018**, featuring the latest development and innovation of language and intelligence, as well as the 2018 NLP Challenge on Machine Reading Comprehension.

Machine Reading Comprehension (MRC) enables computers to read, process, and understand natural language text, considered to be one of the core abilities of artificial intelligence. It is of great value for next-generation search engines and intelligent agent products. However, MRC is an extremely challenging work since it involves several difficult tasks such as comprehension, inference and summarization. To promote the development of reading comprehension technology, this competition provides large-scale, open-domain Chinese MRC dataset to inspire MRC research in the real world setting. And hopefully, it would also be an academic platform for researchers to exchange ideas and could help in promoting the development of MRC technologies and applications.

Task Description

Given a question *q*, and a set of documents *D = d1, d2, ..., dn*, the participant MRC system is expected to output an answer *a* that best answers *q* according to the evidences in document *D* .

- ■ **Input/Output:**
  Input: Question q and its corresponding evidence document set D.
  Output: An answer a that best answers q according to the document set D.

- ■ **Dataset**

The dataset contains 300k questions, which are sampled from real anonymized user queries of Baidu Search. For each question, 5 evidence documents and human generated answers are provided. The dataset is divided into a training set (280k questions), a developing set (10k questions) and a testing set (10k questions). A subset of 200k questions from the DuReader dataset, now available for free download, can be used as pre-training/validation before the release of the full dataset. Competition participants will get an extra data of 100k questions.

- ■ **Evaluation Metrics**

BLEU and ROUGE-L are used as the basic evaluation metrics to measure the performance of participant system. The result on main task (all test data) will be the final evaluation result.

- ■ **Baseline Systems**

This competition provides two open sourced baseline systems, for details please refer to: source and paper.

Prizes

This competition will award **one First Prize, two Second Prizes and three Third Prizes**. Winners will get the award certificates issued by CIPS & CCF. The prizes and travel grants for attending the competition forum and award ceremony will be sponsored by Baidu.

- ● First prize：   **50,000 RMB** + award certification
- ● Second prize: **20,000 RMB** + award certification
- ● Third Prize: **3,000 RMB** + award certification

Timeline

- ● Mar 1, 2018: Registration opens;
- ● Mar 31, 2018: Competition begins (dataset ready for the public) ;
- ● April 30, 2018: Competition ends (submission deadline);
- ● May 15, 2018: Winners will be notified;
- ● July 28, 2018: Competition forum and award ceremony on Language & Intelligence Summit.

Organization

- **Hosts**
  - Chinese Information Processing Society of China (CIPS)
  - China Computer Federation (CCF)

- **Organizer**
  - Baidu Inc.
  - Committee on Evaluation of CIPS (CIPS CE)
  - Technical Committee on Chinese Information Technology of CCF (CCF TCCI)

- **Steering committee**
  - Le Sun, Institute of Software, Chinese Academy of Sciences
  - Ming Zhou, Microsoft Research Asia
  - Erhong Yang, Beijing Language and Culture University
  - Dongyan Zhao, Peking University
  - Hua Wu, Baidu Inc.

- **Organizing committee**
  - Yajuan Lyu, Baidu Inc.
  - Xianpei Han, Institute of Software, Chinese Academy of Sciences
  - Xiaojun Wang, Peking University
  - Kai Liu, Baidu Inc.

Registration

Pre-registration: Please scan the follow QR-code for pre-registration. All pre-registered teams will be notified by email when official registration opens, and will receive souvenirs as long as official registration is completed.

Official registration: The official registration will be open on March 1, 2018. Please pay attention to the official registration deadline 2018/3/31. All participants who submit valid competition results will receive a competition custom T-shirt.

Contact us

If you have any question or suggestion, please feel free to contact us.
Contact email: MRC2018@126.com