

2006-07 SIGHAN Report

There is a new slate of officer bearers as indicated in 2 below. Since SIGHAN's focus is on NLP research involving the prime example of a non-Alphabetic language in the world, efforts have been made to improve and augment ties between the ACL community and the large community of researchers and research organizations in places such as China. The following is a summary report.

1. Workshop Organization

One major event for the year was the 5th SIGHAN Workshop, held in conjunction with COLING/ACL-2006 in Sydney, Australia, 22-23 July 2006. The workshop was co-chaired by Hwee Tou Ng and Olivia Kwong. There were 24 submissions received, 8 of which were accepted for oral paper presentations and 6 as posters.

The 3rd International Chinese Language Processing Bakeoff was also organized in conjunction with SIGHAN-5, coordinated by Gina Levow and Olivia Kwong. There were two shared tasks, namely automatic Chinese word segmentation and named entity recognition. Altogether 29 teams representing 29 organizations in 8 territories participated in the Bakeoff, which have set a new record. This year the benchmark corpora used in the evaluation also drew a new contributor: Microsoft [China], in addition to the regulars such as Academia Sinica [Taipei], Peking University, University of Colorado [originating from U Penn], and City University of Hong Kong. The 6th SIGHAN Workshop is intended to be held in conjunction with IJCNLP-08 in Hyderabad, India, 11-12 January 2008. SIGHAN-6 will be co-chaired by Haizhou Li and Olivia Kwong. A workshop proposal has recently been submitted. We also plan to concurrently organise the 4th International Chinese Language Processing bakeoff, and we have reached initial agreement with the Verification Center on Chinese Language and Character Standards of the State Language Commission of PRC, and the Chinese LDC, to co-organise the event. Additionally, advanced discussions are underway for a panel on the status quo of automatic word segmentation of Chinese and related undertakings, to be jointly conducted by the Chinese Academy of Sciences, SIGHAN, and the Chinese Information Processing Society at the biennial conference of the latter in Dalian, China in August 2007.

The issues relevant to tokenization and automatic segmentation of texts in mostly west and southeast Asian languages and transferable experience from competitive events such as the SIGHAN Bakeoffs were shared with participants of the conference on Regional Consultation on Local Language Computing Policy in Developing Asia, organized in Bhutan by the PAN Localization project based in Pakistan.

2. SIG Compliance

The current SIGHAN officers are:

- Benjamin Tsou (Chair)
- Dan Jurafsky (Vice-Chair)
- Maosong Sun (Treasurer)
- Gina Levow (Secretary / Web Master)

The webpage of SIGHAN is located at <http://www.sighan.org>, containing a statement of purpose and the list of officers and information officers from individual regions, and the constitution can be found at <http://www.sighan.org/constitution.htm>. We are proposing a constitutional change on adding an at-large-member and are in the process of holding a vote among members. The issue will be further discussed in the SIGHAN business meeting in the next workshop.