

Report on the First Summer School on NLP and IR in Beijing

**Ken Church¹, Jian-Yun Nie², Le Sun³,
Maosong Sun⁴, Haifeng Wang⁵,
Endong Xun⁶**

¹ Johns Hopkins University,
Kenneth.Church@jhu.edu

² University of Montreal, nie@iro.umontreal.ca

³ Institute of Software, Chinese Academy of
Sciences, lesunle@gmail.com

⁴ Tsinghua University,
sunmaosong@gmail.com

⁵ Baidu, wanghaifeng@baidu.com

⁶ Beijing Language and Culture University,
edxun@126.com

Abstract

The first Summer School on NLP and IR has been organized in China in July 2011 just before SIGIR conference. This report describes the event.

1 The Summer School

Between July 21 and 23, 2011, just before the SIGIR conference, the first Summer School on NLP and IR was held in the Beijing Language and Culture University, in cooperation with ACL and SIGIR.

This summer school was organized following a strong and long-standing demand of the Chinese community in NLP and IR, and especially the graduate students working in these areas. Researches in NLP and IR in China have been rapidly expanding during the last decade. We see more and more Chinese scholars and students participating in conferences such as ACL and SIGIR. Yet many young Chinese researchers and students have fewer opportunities to listen to and to interact with the leading researchers in these areas. A summer school would offer them such an opportunity and the holding of the SIGIR conference in Beijing created a perfect momentum to organize the summer school.

The summer school was organized by the Chinese Information Processing Society of China, Tsinghua University and Beijing Language and Culture University, in cooperation with ACL and SIGIR. It was sponsored by Baidu and ACL. The general chairs were Ken Church (John Hopkins University) and Maosong Sun (Tsinghua University). The program chairs were Jian-Yun Nie (University of Montreal) and Haifeng Wang (Baidu), and the organization committee chairs were Le Sun (Chinese Information Processing Society) and Endong Xun (Beijing Language and Culture University).

2 The Presentations

The summer school offered a series of half-day classes on different topics in NLP and IR.

Jian-Yun Nie presented an overview on IR models. He described different IR models, discussed on their problems and possible improvements, and presented a perspective on the future development on IR modeling.

Doug Oard talked about cross-language information retrieval. He described the underlying problems and the solutions developed by researchers in the past two decades. He concluded with some remarks on the current state of CLIR and its future.

Haifeng Wang and Shiqi Zhao described the uses of NLP techniques in Web applications. From an industrial perspective, they presented the practical NLP problems underlying various Web applications such as QA, information extraction, summarization, machine translation, and so on. In particular, their talk described how industry has to deal with the huge amount of noisy data.

Ken Church offered an NLP lab course. He showed how different NLP tasks can be accomplished with the R and NLTK toolkits. Examples he demonstrated include concordance analysis, SVD and LSI, clustering, logistic regression, etc.

Jianfeng Gao presented a family of approaches to Web search using statistical machine translation methods. The talk covered not only subjects on statistical machine translation, but also how Web search can benefit from these methods.

The slides of the presentations can be downloaded from the website of the summer school (<http://nlp.blcu.edu.cn/others/nlpir-ss/>).

After the presentations, Baidu offered an onsite visit to all the participants, and presented demos and talks by Baidu researchers.

3 Feedback and the Future

The summer school has been a great success. The total number of participants was 205, which was the maximum number allowed by the classroom. All the participants are from China. The enthusiasm was beyond the expectation of the organizers - the maximum number of participants was reached within only a week after the opening of registration site. Many regretted not being able to register. Among the participants, 158 are from universities and research institutes, and 47 are from industry. 19.5% of the participants are females.

The feedback from the audience during and after the summer school has been unanimously positive. Many participants shared their experience at the summer school on their Chinese microblogs.

The great success clearly demonstrates how attractive the areas and the summer school are to young researchers and students in China. It also strongly suggests that this summer school should continue in the future. Would this be the beginning of a new tradition to hold a series of regular summer schools in these areas in China and Asia, as it is the case for the European Summer School on IR? The answer seems to be enthusiastic.