
SUMMER SCHOOL REPORT

2nd Asian Summer School in Information Access (ASSIA 2015)

<http://assia2015.lib.ntnu.edu.tw>

Hao-Ren Ke
National Taiwan Normal University
clavenke@ntnu.edu.tw

Hideo Joho
University of Tsukuba
hideo@slis.tsukuba.ac.jp

Abstract

The second Asian Summer School in Information Access (ASSIA 2015) was held between 24th and 27th August, 2015 in Taipei City, Taiwan, R.O.C. The summer school offered 10 lectures in Information Retrieval, Social Media Search, and related topics, along with one panel discussion, a poster session, and a session from commercial sectors and young scholars. This reports a successful international summer school in Asia attracting a total of 45 participants from the range of countries in Asia and Europe.

1 Introduction

The summer schools in Information Retrieval have been very successful in Europe where research students, young faculty members, or industrial researchers can gather together to learn the core and advanced topics from the world-class experts in the field. However, no such opportunity was available in most Asian countries. Following the success of the first Asian Summer School in Information Access (ASSIA 2013), which was held in 2013 at Tsukuba, Japan, the 2nd Asian Summer School in Information Access (ASSIA 2015) was held between 24th and 27th August, 2015 in the capital of Taiwan, R.O.C., the Taipei City. The International Conference Hall of the Library of National Taiwan Normal University was used as the main venue of the summer school.

A total of 45 people registered to the summer school (35 participants, 9 lecturers, and 1 organizer). The majority of participants came from Taiwan, but there were also 12 participants from China, Japan, Malaysia, and Thailand. See Figure 1 for the group photo.

2 Summer School Program

ASSIA 2015 had a 4 days program which included 10 lectures, 1 panel session, 1 session from commercial sectors and young scholars, and a poster presentation. The summer school opened its program by having a welcome message from Hao-Ren Ke, the University Librarian of National Taiwan Normal University and also the Organizing Committee Chair of ASSIA 2015.



Figure 1 Group Photo of ASSIA 2015 (Day 1)

2.1 Lectures

A total of 10 lectures were offered. **Introduction to Information Retrieval** was a 60-minute lecture for providing the foundation. **Advanced Information Retrieval Models**, **Evaluation I: system oriented**, and **Search Interface** were 90-minute lectures. All others are standard lectures (120 min). The lectures had a good balance between IR theories and applications. A brief summary of individual lectures is as follows.

Introduction to Information Retrieval (Iadh Ounis, University of Glasgow) The main program started with the introductory lecture on IR by Ounis. The lecture provided an overview of classical search systems, their applications, and the challenges. In particular, the lecture described the core text processing and retrieval techniques for search engines to find relevant documents for a user query. Finally, the lecture concluded with current topics in information retrieval, as well as a summary of current resources useful for learning about and developing IR systems.

Advanced Information Retrieval Models (Iadh Ounis, University of Glasgow) The second lecture was given by Ounis as well. Firstly, the lecture introduced the purpose of a retrieval model. Then a variety of IR models including theoretically founded models, sophisticated probabilistic models, and more complex machine-learned ranking models were introduced. This lecture covered a number of advanced and effective search models, such as field-based models, proximity models and state-of-the-art learning to rank approaches, focusing on their intuitions, assumptions, their salient characteristics and their uses.

Evaluation I: system oriented (Tetsuya Sakai, Waseda University) The final lecture of Day 1

given by Sakai described some evaluation measures for modern information access tasks, how to choose from available measures, and how to report experimental results in technical papers. The measures in the lecture covered average precision, normalized discounted cumulative gain, expected reciprocal rank, normalized cumulative utility measures, diversified search measures, time-biased gain, U-measure, and statistical significance tests.

Search Interface (Hideo Joho, University of Tsukuba) Day 2 of the summer school started with the lecture on search interface delivered by Joho. This lecture looked at human factors in Information Retrieval with focus on interactive features of search user interfaces. First, it provided background knowledge such as major challenges faced by search engine users, and their behavioral patterns in search. Second, the lecture looked at major components such query box and search result presentation, and expanded them to query expansion techniques and result diversification methods. Third, the lecture introduced some of the advanced topics such as multimodal interfaces and collaborative search interfaces. Finally, the lecture discussed future directions of search interfaces.

Search User Behavior Modeling (Yiqun Liu, Tsinghua University) Liu gave the lecture on modeling search user behavior. The first half of this lecture introduced existing efforts for understanding the cognitive behavior of search users and focused on the characteristics of search users' querying, clicking-through and examination behaviors. The second half of the lecture described some recent trends in search behavior modeling and focused on how to model user behavior for supporting effective IR methods in heterogeneous environments such as information of various media types, structure and semantics, short- and long-term search tasks, and users with different background and preferences.

Social Media Search and Analytics (Ee-Peng Lim, Singapore Management University) Day 2 ended with the lecture on social media search and analytics delivered by Lim. Following the social media and its impact, the lecture presented a framework for social media search and analytics. The lecture classified social analytics research roughly into text mining, network mining, network dynamic mining, user profiling, and business analytics. Several social media search and analytics studies that used Twitter data were introduced. Finally it described the challenges and future directions.

Building digital archive systems for historians (Jieh Hsiang, National Taiwan University) The morning lecture of Day 3 focused on using IR techniques to build digital archive systems for historians. This lecture given by Hsiang pointed out that historians usually consider documents as related, not independent as assumed by document retrieval systems. This lecture presented the design of a digital archives system for scholarly use. Instead of ranking the documents (as is done in the precision/recall model), the proposed methodology returns a list of documents when issued a query, together with a choice of textual contexts of the returned sub-collection as well as visual mechanisms to observe, explain, and refine the contexts. Some typical contexts are chronological distribution, geographical distribution, term frequencies and co-occurrences of people and places, relevance factors, appositional term analysis, holistic comparisons of query returns, land transition graphs, and co-citation diagrams.

Linguistic Information Retrieval and Interactive Writing Environment (Jason S. Chang, National Tsing Hua University) The last day of the summer school started with a lecture introducing two systems for linguistic information retrieval, Linggle (<http://linggle.com/>) and WriteAhead (<http://writeahead.nlpweb.org/>), designed by Chang to demonstrate the workable

linguistic IR applications. Linggle is a Web-scale linguistics search engine that retrieves short phrases in response to a given query based on Google Web 1T corpus. WriteAhead is an Interactive Writing Environment (IWE) that provides writing suggestions for writers.

Evaluation II: User Oriented (Muh-Chyun Tang, National Taiwan University) The second lecture on evaluation were given by Tang and focused on user-oriented evaluation. First, the lecture reviewed system evaluation from aspects of purposes and context. Second, it described operational evaluation criteria. The main part of this lecture elaborated on basic elements in system evaluation, including construct, criteria, measures, measuring instruments, and methodology.

Multimedia IR (Winston Hsu, National Taiwan University) Up till now, most lectures had assumed text contents while the final lecture of the summer school given by Hsu brought the participants a broad and comprehensive coverage on the foundations and recent developments of content-based and semantic-based image and video retrieval on large-scale image / video collections. The lecture gave a balanced review of the area of content-based and semantic-based visual retrieval for large-scale image and video collections by presenting topics of both practical and theoretical interest. Then it incorporated additional topics on the latest development of local features, coding methods, efficient feature indexing, hash learning, scalable semantic detection, mobile visual search, etc. Furthermore, the lecture briefed recent huge improvements in visual analytics by deep convolutional network (DCN).

2.2 Panel

The panel regarding how to write solid papers was held on Day 3. Lecturers shared their experiences on how to find a good topic, how to know this is a good topic. Reading papers from subject-related journals, monographs, and conferences is essential for finding a good topic. Report improvement on performance (such as efficiency and accuracy) may help the acceptance of a paper. Lecturers also gave their suggestions on common pitfalls and writing pros. Participants raised issues on the review process (such as how to deal with tough reviews), and how to choose a good publication venue to submit papers. Tasks in TREC, NTCIR, and CLEF were suggested as good resources to identify future research in IR and related domains.

2.3 Talk from Commercial Sectors and Young Scholars

In the afternoon of Day 3, the summer school arranged a session that invited two speakers from commercial sectors and one young scholar. Willie Yang from eLand Technologies/Tornado Tech presented his experience in social listening and analysis by using the big data from enterprises and social networks. Yang introduced Tornado Enterprise Search / NLP Platform, comprising data processing & integration, language processing, content analysis, mining & suggestion, and visualization layers. This platform facilitates information retrieval, text mining, and knowledge discovery. eLand's OpView Listening Platform is the largest social listening platform in Taiwan. Annie Yang shared cacaFly's experience of being an authorized advertisement reseller of Facebook, Microsoft, and Google, and how they exploited information access techniques to achieve customized and personalized advertisement. cacaFly's new platform for real-time bidding (RTB) were presented as well. Jilung Hsieh, an assistant professor of Graduate Institute of Library and Information Studies at National Taiwan Normal University, described the use of mobile phones for daily life research. Research methods such as time-diary method, experience sampling method, daily reconstruction

model, and prompted recall method were introduced.

2.4 Poster Session

Eight participants presented their posters in the afternoon of Day 1. The poster session lasted for one and half hours and there was vivid interaction between presenters, participants, and lecturers. The poster topics were quite diverse, covering information seeking behavior, ontology development, automatic conversion from MARC (MACHine Readable Catalog) into FRBR (Functional Requirements for Bibliographic Records), Twitter search, search engine evaluation, and Web site usability.

3 Organization

Chinese Association of Library & Information Science Education (CALISE), Graduate Institute of Library and Information Studies at National Taiwan Normal University, Interlibrary Cooperation Association (ILCA) were the three main hosts of ASSIA 2015. Association for Computational Linguistics and Chinese Language Processing (ACLCLP), Research Center for Knowledge Communities at University of Tsukuba, and the Taipei Chapter of Association for Information Science and Technology (ASIS&T) were the three co-hosts. The members of the organizing committee were Hao-Ren Ke (National Taiwan Normal University), Hideo Joho (University of Tsukuba), Pu-Jen Cheng (National Taiwan University), and Nei-Ching Yeh (Shih Hsin University).

4 Conclusion

The second Asian Summer School in Information Access was held in August 2015 at Taipei City, Taiwan, R.O.C. Comparing with ASSIA 2013, the participants of ASSIA 2015 were fewer. Despite this, all lecturers and participants deemed ASSIA 2015 successful. The organizing committee considers that a right time for the summer school (June, July, or August?), the involvement of local scholars, travel and accommodation support for foreign students, and sponsorship from the industry are essential for a successful summer school. We are currently forming a steering committee for long-term organization of the Information Access Summer Schools in Asia.