

Asian Summer School in Information Access (ASSIA 2017)

<http://goassia.github.io/assia2017/>

Makoto P. Kato
Kyoto University, Japan
mpkato@acm.org

Takehiro Yamamoto
Kyoto University, Japan
tyamamot@dl.kuis.kyoto-u.ac.jp

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

Masatoshi Yoshikawa
Kyoto University, Japan
yoshikawa@i.kyoto-u.ac.jp

Abstract

The third Asian Summer School in Information Access (ASSIA 2017) was held between 2nd and 5th August, 2017 in Kyoto, Japan. The summer school offered 9 lectures on information retrieval and related topics, along with a poster session and a panel discussion. This report introduces the successful international summer school in Asia attracting 63 participants from a wide range of countries.

1 Introduction

The first Asian Summer School in Information Access (ASSIA) was held in 2013 in Tsukuba, Japan, and second ASSIA was held in 2015 in Taipei City, Taiwan. The success in the past ASSIA motivated us to organize the third ASSIA between 2nd and 5th August, 2017 in Kyoto, Japan, right before the 40th international ACM SIGIR conference in Tokyo, Japan¹.

Like our fellow summer schools in Europe, Russia, and India, ASSIA aims to facilitate the research and education in Information Access and related fields in Asian regions by offering the lectures given by leading researchers in the domain. It also provides participants with an opportunity to meet other students or researchers from various countries with similar interests, and to have discussion with the lecturers who are authoritative in the fields. These opportunities are particularly valuable in Asian regions since the students in Asian countries tend to have a limited chance to meet leading researchers from Europe and North America. Furthermore, submissions to top conferences from Asian regions still have a room to catch up to Europe and North America. Therefore, increasing awareness of the fields and providing foundations to young students and researchers are important objectives of ASSIA. Although we have mainly targeted Asian students in the past ASSIA, ASSIA 2017 attracted more

¹<http://sigir.org/sigir2017/>



Figure 1: Group photo of ASSIA 2017 (Day 3).

participants from various countries even outside of Asia since it was held a few days before SIGIR 2017.

A total of 63 people registered to the summer school (52 participants, 7 lecturers, and 4 organizers). Among the participants, 39 are students, 4 are researchers from industry, and 9 are university faculties. The majority of the participants came from Japan (18) and China (16), but there were also participants from a wide variety of countries such as Australia, India, Indonesia, Korea, Thailand, and USA. Figure 1 is a group photo of ASSIA 2017 taken in the front of the summer school venue, Kyoto University Clock Tower.

2 Summer School Program

The ASSIA 2017 program consisted of 9 lectures, a poster session, and a panel on paper writing. In the following subsections, we describe each of the sessions.

2.1 Lectures

A total of 9 lectures were offered in ASSIA 2017. The organizers first selected four major topics in information access and allocated three hours for each, namely, **Offline Evaluation in Information Retrieval**, **User Modeling for Information Retrieval**, **Social**

Media Analysis for Information Retrieval, and **Machine Learning for Information Retrieval**, based on our interests and estimated interests of summer school participants. We also listed up recent trends of information access in a similar way to the major topics, and allocated one or two hours for each, namely, **Social and Collaborative Search**, **Multimedia Information Retrieval**, **Crowdsourcing for Information Retrieval**, and **Online Evaluation in Information Retrieval**. Below, we describe a brief summary of each lecture in the order of the program.

Introduction to Information Retrieval

— **Mark Sanderson (Royal Melbourne Institute of Technology)**

The first lecture of ASSIA 2017 introduced the overview and basic techniques of information retrieval (IR). The lecture was started with a focus on the problem of matching underspecified requests with documents, and then explained basic techniques such as term frequency, inverse document frequency, stemming, and PageRank, as tools for addressing the underspecification problem in IR. This lecture also presented the usage of query logs in IR for identifying relevant documents or generating query suggestions.

Offline Evaluation in Information Retrieval

— **Charles L. A. Clarke (Facebook)**

This lecture introduced the way of evaluating information access systems in an offline manner. Cranfield paradigm was first explained with TREC as an example of its implementation. The lecture then introduced basic evaluation measures such as average precision, normalized discounted cumulative gain (nDCG), and expected reciprocal rank (ERR), by interpreting them as probabilistic models of user behaviors on search engine result pages. Novelty and diversity were also covered by this lecture, in which the intent aware measures and α -nDCG were introduced as diversity-aware evaluation measures. The last half of the lecture focused more on the development of evaluation measures by means of search user modeling. Time-biased gain was used to explain the procedure of designing evaluation measures based on user behavior data. The last topic of this lecture was introduction to an advanced topic, session-based evaluation, in the context of TREC Tasks tracks.

User Modeling for Information Retrieval

— **Yiqun Liu (Tsinghua University)**

This lecture introduced user modeling for IR on three type of search results: homogeneous search results, heterogeneous search results, and image search results. The lecture first showed several kinds of statistics on search queries such as length and topic distributions, and then explained recent studies on the search result examination behavior by highlighting some complex patterns captured by eye-tacking devices. Furthermore, recent advances in click models were also described. The lecture also covered user behaviors on heterogeneous search results including images, videos, and news articles. The lecturer introduced his recent studies on effects of heterogeneous search results on the examination behavior and relevance judgments. Image search was another environment presented in this lecture, in which some image-specific intents were proposed and user behaviors on image search result layout with a preview function were examined.

Social and Collaborative Search

— **Chirag Shah (Rutgers University)**

This lecture began with a quiz on DARPA Red Balloon Challenge, which helped participants understand the importance of collaboration and social connections for search. The lecturer first gave us the idea of social and collaborative search by using different kinds of dimensions such as the number of collaborators, strength of social ties, and level of collaborations. In the topic of collaborative search, search systems designed for collaboration were introduced, and algorithmic mediation methods were explained in details. In the topic of social search, some examples of questions posted in social network services, and categorization of online Q&A were presented. This lecture also discussed the evaluation methodology for collective outcomes in collaborative and social search.

Multimedia Information Retrieval

— **Gareth Jones (Dublin City University)**

This lecture mainly focused on keyword search over multimedia contents such as images, audio clips, and videos, and explained some challenges unique in multimedia IR settings. The first half of this lecture addresses the problem of spoken content retrieval by emphasizing its difficulty due to acoustic ambiguity, and used TREC spoken document retrieval and CLEF cross-language speech retrieval as examples. The last half introduced image and video content retrieval, where the lecturer focused on the difficulty of high-level understanding of images and ambiguity of units in video clips.

Social Media Analysis for Information Retrieval

— **Ee-Peng Lim (Singapore Management University)**

The lecture first introduced interesting characteristics of social networks as well as a real application of social media data to a traffic information service in Singapore, and then explained the details of user profiling and biases in social media. Modeling social networks by graphs, the lecturers gave us some characteristics or phenomena of social networks such as preferential attachment process, small-world-ness, and homophily. As one of the important components in social media analysis, profiling of social media users was extensively described in the lecture. User profiles targeted in the lecture included demographic attributes, personality values, and home locations. This lecture also paid attention to negative aspects of social media analysis, *e.g.* population, content production, and linking biases. To deal with these biases, the lecturer talked about techniques to filter out bots in social networks, and in-depth analysis on selective self disclosure of social media users.

Crowdsourcing for Information Retrieval

— **Gareth Jones (Dublin City University)**

This lecture provided basics of crowdsourcing and the reason why it is useful in information access research. The lecturer first gave us the concept of crowdsourcing and advantages, and talked about disadvantages such as a population bias and difficulty of controlling the quality of task outcomes. The lecturer then introduced the reputation mechanism of crowdsourcing services and some techniques to detect spams (*e.g.* honey pots). A detailed use case in MediaEval Rich Speech Retrieval was introduced as an example of applications of crowdsourcing to information access evaluation.

Online Evaluation in Information Retrieval

— **Charles L. A. Clarke (Facebook)**

Topics of this lecture included logging, click interpretation, and A/B testing for information

access evaluation. The lecturer first put great emphasis on logging of user actions, as it is the foundation of online evaluation, and then explained how to interpret clicks by focusing on the effect of the position bias. The lecture also contained practical suggestions on the design of A/B testing including partition of users in different scenarios, difficulty of interpreting A/B testing results, and choice of evaluation metrics in A/B testing, by referring to a SIGIR 2016 industry paper from Bing.

Machine Learning for Information Retrieval — Hang Li (Noahs Ark Lab, Huawei Technologies)

The last lecture of ASSIA 2017 was on applications of machine learning approaches to IR, and consisted of three parts, learning to rank, learning to match, and deep learning for IR. First, the progress of learning to rank models was explained (*i.e.* from point-wise to list-wise), and details of learning to rank models such as Ranking SVM and AdaRank were briefly introduced. Second, the problem of matching between heterogeneous data was introduced, and addressed by transformation of two different feature spaces into the space where two different kinds of data were comparable. Third, applications of neural networks to IR were explained with basic components of neural networks such as word embedding, recurrent neural networks, and convolutional neural networks. Finally, the lecturer introduced various question answering methods based on neural networks utilizing relational databases and knowledge bases.

2.2 Posters

On the first day of ASSIA 2017, 23 participants presented their posters, some of which were *previews* of their SIGIR 2017 presentations. There was extended discussion among participants, which lasted for one and half hours. The presented topics were diverse and included mobile search, conversational search, social media analysis, IR evaluation, search behavior analysis, *etc.*

2.3 Panel on Paper Writing

Panelists of the panel on paper writing were Mark Sanderson, Charles L. A. Clarke, Yiqun Liu, Chirag Shah, Gareth Jones, and Ee-Peng Lim². The ASSIA 2017 organizers prepared some questions and also asked participants to post questions via an online form or on site. Questions discussed in the panel were

- How do you usually find motivation or ideas of research?
- What is the basic strategy to come up with solutions to a given problem?
- I got comments that my writing is not clear and not easy to read. What should I do?
- How do you sell your idea well?
- Is it true that some reviewers decide to accept/reject based on partial reading?
- How can we plan/predict what will happen during research (for engineering oriented work)?
- Do you have any concerns about arXiv.org?

²Unfortunately, Hang Li could not participate due to his schedule constraint.

The participants were very active to ask questions: most of the questions, five out of seven, came from participants.

3 Organization

ASSIA 2017 was sponsored by Yahoo Japan Corporation (gold sponsor), ACM Special Interest Group on Information Retrieval, and Kyoto City and Kyoto Convention & Visitors Bureau. Graduate School of Informatics at Kyoto University was a supporter of ASSIA 2017. The members of the organizing committee were Makoto P. Kato (Kyoto University), Takehiro Yamamoto (Kyoto University), Hideo Joho (University of Tsukuba), and Masatoshi Yoshikawa (Kyoto University).

4 Summary

ASSIA 2017 was held between 2nd and 5th August, 2017 in Kyoto, Japan, attracting 63 participants from a wide range of countries. We think that the collaboration with the 40th ACM SIGIR conference was successful to attract participants, and most of the participants were satisfied with the summer school according to our survey results: 57.1% and 42.9% of the participants (the sample size is 21) answered “strongly agree” and “agree” with “I am happy with ASSIA 2017”. The 4th ASSIA will be held in China in 2019. If you’re interested in hosting a future ASSIA in 2021 onwards, you can contact the steering committee via assia-sc@googlegroups.com.

Acknowledgements

We sincerely appreciate ASSIA 2017 lecturers, Mark Sanderson, Charles L. A. Clarke, Yiqun Liu, Chirag Shah, Gareth Jones, Ee-Peng Lim, and Hang Li, for giving us great lectures. We express our gratitude to our sponsors and supporters, Yahoo Japan Corporation (gold sponsor), ACM Special Interest Group on Information Retrieval, Kyoto City and Kyoto Convention & Visitors Bureau, and Graduate School of Informatics at Kyoto University, for their tremendous support. We thank the organizing committee of the 40th ACM SIGIR conference for collaboration with ASSIA 2017. Finally, thank all the ASSIA 2017 participants for coming to our summer school.