Asian Summer School in Information Access (ASSIA 2013)

http://www.kc.tsukuba.ac.jp/assia2013/

Hideo Joho

University of Tsukuba

hideo@slis.tsukuba.ac.jp

Noriko Kando

National Institute of Informatics

kando@nii.ac.jp

Tetsuya Sakai

Waseda University

tetsuyasakai@acm.org

Yohei Seki

University of Tsukuba

yohei@slis.tsukuba.ac.jp

Shigeo Sugimoto

University of Tsukuba

sugimoto@slis.tsukuba.ac.jp

Abstract

The first Asian Summer School in Information Access (ASSIA 2013) was held between 22nd and 24th June, 2013 in Tsukuba, Japan. The summer school offered 9 lectures in Information Retrieval, Web Search, and related topics, along with two panel discussions and a poster session. This reports a successful international summer school in Asia attracting a total of 63 participants from the range of countries in Asia, Europe, and North America.

1   Introduction

The Asian Summer School in Information Access (ASSIA 2013) was held between 22nd and 24th June, 2013 in Tsukuba, Japan. Although there has been a series of Information Retrieval (IR) summer school in India (Yahoo! Summer School), this was the first international summer school for students, faculty members, and industrial researchers in Asian regions to learn basic and advanced topics in Information Access. IR summer schools in Europe are traditionally organized in a relatively distant location where lectures and participants can have relaxed and casual discussion opportunities during the program. Tsukuba is one such location in Japan, located in 80km north of Tokyo, with easy access from major international airports such as Narita and Haneda. The main hall and open seminar room in the Union of Library and Information Science Building, University of Tsukuba, were used as the main venue of the summer school.

Figure 1: Group photo of ASSIA 2013 (Day 1).

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 taught by leading researchers in the domain. It also provides participants with an opportunity to meet other participants from various countries with similar interests, and to have discussions 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.

As indicated by the name of the summer school (i.e., Information Access), ASSIA has a wider scope than traditional IR summer schools. This reflects our long-term intention of covering not only IR topics but also related topics such as NLP, Digital Libraries, Visualizations in future summer schools. Having said that, as the first ASSIA, ASSIA 2013 mainly focused on Information Retrieval and Web Search, along with applicational topics such as E-Discovery and Crowdsourcing.

A total of 63 people registered to the summer school (54 participants, 6 lectures, and 3 organisers). We had more applications than we had planned to accommodate (which was 50), and thus, had to close the application before the original deadline. This indicated that there was a strong interest and need to have a summer school in Asian regions. The background of participants also showed a great diversity.

Thanks to generous support from our sponsors, we were able to offer a travel support to seven students participating from China (incl. Hong Kong), India, Indonesia, and Thailand. The travel support was awarded during the banquet on Day 2 of the summer school. We were grateful for the sponsorship of Microsoft Research (thanks to Noboru Kuno) and Catalyst Repository Systems (thanks to Jeremy Pickens).

2　Summer School Program

ASSIA 2013 had a 3 days program which included 9 lectures, 2 panels, and a poster presentation. The summer school opened its program by having a welcome message from Shigeo Sugimoto, the Director of Research Center for Knowledge Communities (RCKC), University of Tsukuba, followed by the introduction of the summer school from one of the organisers. This section presents a summary of each program.

2.1　Lectures

Each day had three lectures. The title with a star is short lectures (60 min) while others are standard lectures (90 min). A total of 9 lectures offered over 160 carefully selected references, which can represent a great part of IR literature, for participants to revisit even after the summer school. A brief summary of individual lectures is as follows.

Introduction to Information Retrieval* (Mark Sanderson, RMIT University)　The main program started with the introductory lecture on IR by Sanderson. Sanderson first briefly defined IR as "the study of finding relevant information in unstructured collections of material based on a poorly specified query". The lecture then described the early history of IR, the range of approaches used by various IR systems to retrieve relevant materials. The lecture also provided the basic knowledge about text processing, link analysis, retrieval models, and extraction of data from interaction logs.

Modern Information Retrieval Models (Iadh Ounis, University of Glasgow) The second lecture given by Ounis provided a comprehensive description of IR models, which has been "at the forefront of IR research in the past three decades". Ounis first introduced a retrieval model as something that specifies how documents and queries should be represented and how relevance should be decided. Then, the lecture explained the representative IR models based on a series of major paradigms emerged during the history of the field. Also, the participants learnt recent models such as search results diversification.

Evaluation of IR Systems (Mark Sanderson, RMIT University)　After the general introduction and comprehensive discussion of IR models, Sanderson was back to teach evaluation as the final lecture of Day 1. The lecture first introduced the concept of test collections, its history, the components, and typical protocol. Then, issues related to scaling of test collections were discussed along with performance measure metrics. The last part of the lecture was about examination of test collections themselves, and statistical tests in IR experiments.

User-Oriented Factors in Information Retrieval (Kalervo Järvelin, University of Tampere)　Day 2 of the summer school started with the lecture on user-oriented factors given by Järvelin. First, the lecture introduced the concept of information interaction consisting of a dozen of actions taken to communicate with information. Then, the lecture described how users are modeled in a typical test collection based evaluation. Järvelin then discussed user-centered evaluation and its design issues. The final part of the lecture further looked into the study of user-oriented factors in IR and representative variables that one might want to consider in such studies.

Multimedia Information Retrieval (Gareth Jones, Dublin City University)
Up till now, most lectures had assumed text contents, while Jones provided the lecture on multimedia contents. The lecture first looked at retrieval of multimedia contents using texts, which can be familiar to many participants. Then, it described the retrieval of spoken contents, followed by content-based approaches to multimedia retrieval. The last parts of the lecture covered multimedia search on personal collections, exploitation of affect in multimedia search.

Cross-Lingual Information Retrieval (Douglas W. Oard, University of Mayland) The last lecture of Day 2 was cross-lingual IR (CLIR), which was of great interest to many participants due to Asian context, given by Oard. Oard brought a copy of Rosette Stone he found in the venue building to highlight the importance of accessing different languages to increase our knowledge. The lecture described the range of issues in CLIR such as query translation, document translation, available lexical resources, bilingual corpora, bilingual query expansion.

Web Search: Information Retrieval in Practice (Fernando Diaz, Microsoft Research) The last day of the summer school started with the lecture on Web search by Diaz. Diaz defined his lecture's objective as to understand the importance of relevance estimation in ranking, indexing, and evaluation of web search. The lecture then described the range of signals used in raking, how to prioritize pages to index, and finally evaluation methods used in production such as click analysis and A/B testing. The lecture ended with a reminder to be very careful about interrelation of their results.

E-Discovery (Douglas W. Oard, University of Mayland) As the one of the timely lectures, Oard was back to give a lecture on E-Discovery. The lecture first highlighted the characteristics of E-Discovery search as many participants, long duration, large budget, recall-oriented goals, and authoritative relevance. Based on the TREC Legal Track, the lecture provided several key issues in the evaluation of E-Discovery technologies such as relevance estimation and authoritative judgements. The lecture concluded with future research questions of this new domain.

Crowdsourcing for Information Retrieval* (Gareth Jones, Dublin City University) As the second timely lecture, Jones was back to provide a lecture on crowd sourcing for IR. Although crowdsourcing has been well known to experienced researchers, this was still a new topic to many participants. Jones first described the concept and benefits of crowdsourcing carefully. Then, the lecture discussed several issues raised when applying crowdsourcing to IR research such as recruitment, management, and payment for workers. The last part of the lecture gave a step-by-step tutorial of MTurk in the context of MediaEval 2011.

2.2 Panels

Panels were designed to offer interactive sessions for participants to gain deep insights into common issues. The first panel was focused on paper writing, while the second post-school panel was focused on teaching and learning of Information Access. Both panels lasted 90

minutes, where each of the lecturers first gave a short advice in turn as a panelist, followed by free discussions with participants.

How to write a solid paper   The first panel was held on Day 2 and asked the lecturers for advices to write a solid paper, which was a prominent issue to most participants, and the panel was very lively. A great emphasis made by the panel was the importance of reading literature to familiarize the field, find a good research topic, formulate interesting hypotheses, which all contribute to the writing of solid papers. Many answers to participants' questions were also about how to find literature, how to read papers, or how to manage reading groups. Advices specific to paper writing were also given such as establishing your own paper structure, having informal peer reviewing before submission, how to avoid self-plagiarism, how to report negative results, how to sell a local topic in international conferences/journals, and so on. Test collections such as TREC, NTCIR, and CLEF were suggested as a good resource to identify future research in IR and related domains.

How to teach and learn Information Access   The second panel was held right after the closing of the official program on Day 3, and asked the lectures to demonstrate some of their teaching courses and materials. The idea was to share and exchange teaching and learning methods in this field among participants. Panelists shared their course website, lecture video recording and broadcasting system, position of their course within a broader curriculum, or comparison of IR courses in CS department and LIS department, and exchanged their insight into incentive system to encourage students to learn abstract ideas (e.g., models, algorithms, etc.), how to create exercise driven by local business needs, how to design classes with CS and LIS background students, and more. Several references were mentioned to specifically discuss these issues during the panel session.

2.3   Poster Presentation

Of 54 participants, 30 gave a poster presentation at the summer school. The poster presentation was a 2 hours session on Day 1 of the program, and presented topics were fairly diverse: Semantic Web, Multimodal interaction, E-Discovery, Multimedia Retrieval, Machine Translation, Collaborative search, Math Retrieval, NLP, Interface design, Information needs mining, Decision making. This diversity demonstrated a great potential research in Information Access can be developed and presented at international venues from Asian regions even more. Many commented that they had great feedback from the lecturers and other participants of the summer school. Please visit the summer school website to view the full list of poster presentations.

3   Participants' Feedback

We asked participants to fill in our questionnaires to capture their feedback on the summer school. We had 37 answers to our questionnaires. Overall feedback was very positive, and the majority (over 80%) of participants rated all the lectures either "Very Useful" or "Useful". The first panel on writing also had a very good rating of over 80% said "Very Useful". These suggested that the design of ASSIA 2013 program was satisfactory to most participants. Specific comments highlighted that participants appreciated the opportunity to meet and

have a discussion with the world-class researchers, to learn theoretical aspects of IR topics, and to exchange ideas with other participants.

4   Organisation

ASSIA 2013 was co-hosted by RCKC and ASSIA 2013 Organising Committee. The members of the organising committee were Hideo Joho (University of Tsukuba), Noriko Kando (National Institute of Informatics), Tetsuya Sakai (Waseda University), and Yohei Seki (University of Tsukuba). The organisation of ASSIA 2013 was further strengthened by a number of organisations and people. Please refer to the acknowledgements for the full list of our partnership.

5   Conclusion

The first Asian Summer School in Information Access was successfully held in June 2013 at Tsukuba, Japan. The initial development of the second ASSIA is under way. We were also considering to form a steering committee for long-term organisation of the Information Access summer schools in Asia. As a final note, it should be emphasised that we were also able to build a positive relationship with the organisers of other summer schools such as European Summer School in Information Retrieval (thanks to Juan Manuel Fernández Luna) and Russian Summer School in Information Retrieval (thanks to Pavel Braslavski). We all listed the URLs of other summer schools in our website to help international students be aware of other opportunities and select appropriate venues to study Information Retrieval and Information Access. We look forward to developing more collaborations with relevant summer schools in the future.