

Report on the 2nd ACM SIGIR/SIGKDD Africa School on Machine Learning for Data Mining and Search

Tanya Berger-Wolf
The Ohio State University
berger-wolf.1@osu.edu

Ben Carterette
Spotify
carteret@acm.org

Tamer Elsayed
Qatar University
telsayed@qu.edu.qa

Maria Keet
University of Cape Town
mkeet@cs.uct.ac.za

Fabrizio Sebastiani
ISTI-CNR
fabrizio.sebastiani@isti.cnr.it

Hussein Suleman
University of Cape Town
hussein@cs.uct.ac.za

Abstract

We report on the organization and activities of the 2nd ACM SIGIR/SIGKDD Africa School on Machine Learning for Data Mining and Search, which took place at the University of Cape Town in South Africa January 27–31, 2020.

1 Introduction

In January of 2019, ACM SIGIR and ACM SIGKDD co-sponsored the 1st SIGIR/SIGKDD Africa School on Machine Learning for Data Mining and Search¹, which took place at the University of Cape Town in South Africa. The event was conceived with the aims of increasing opportunities in research from traditionally underserved communities, growing the IR and data mining communities in sub-Saharan Africa, and expanding the horizons of IR and data mining research. The school took place for the second time in January of 2020², again at the University of Cape Town. This report is on the organization and activities of AFIRM 2020. For more background and context for the event, please refer to the report on SIGIR 2019 by [Carterette et al. \[2019\]](#).

AFIRM 2020 grew in many ways from AFIRM 2019. It went from single-track to parallel-track; both the number of students and the number of instructors increased; and topics were more diverse.

¹<http://www.sigir.org/afirm2019>

²<http://www.sigir.org/afirm2020>

2 Event Organization

2.1 Organizing team

The organizing team consisted of the following committees and individuals:

1. Steering committee

- Ben Carterette, Chair (Spotify)
- Diane Kelly, SIGIR Representative (University of Tennessee Knoxville)
- Jian Pei, SIGKDD Representative (Simon Fraser University, British Columbia)
- Abdiganyi Diriye (IBM Kenya)
- Ricardo Baeza-Yates (NTENT / Northeastern University)
- Lighton Phiri (University of Zambia)
- Reuben Dlamini (University of the Witwatersrand)
- Douglas Oard (University of Maryland)

2. General chairs

- Tamer Elsayed, co-Chair (Qatar University)
- Hussein Suleman, co-Chair (University of Cape Town)

3. Local support at the University of Cape Town

- Mary-Jane Antia
- Jecton Anyango
- Catherine Chavula
- Sarah Dsane-Nsor
- Frances Gillis-Webber
- Zola Mahlaza
- Jivashi Nagar
- Toky Hajatiana Raboanary
- Joseph Telemala
- Tezira Wanyana

4. Program committee

- Ricardo Baeza-Yates (NTENT / Northeastern University)
- Maria Maistro (University of Copenhagen, Denmark)
- Douglas Oard (University of Maryland)
- Iadh Ounis (University of Glasgow, UK)

-
- Lighton Phiri (University of Zambia)
 - Barbara Poblete (University of Chile)
 - Matteo Riondato (Amherst College, USA)
 - Mark Sanderson (RMIT University, Australia)
 - Guido Zuccon (The University of Queensland, Australia)

2.2 Event funding

ACM SIGIR and ACM SIGKDD funded the event with USD20,000 each. Additional support for local arrangements came from the University of Cape Town. We also received sponsorship from the South Africa National Research Foundation (Gold), Web4Good (Gold), Microsoft (Silver), Google (Bronze).

2.3 Participation

We issued open calls for course proposals and for general attendance. The CfP for course proposals was published 7 July 2019, and proposals were accepted through 2 August 2019. There were several rounds of calls for attendance, the first for those that wished to apply for travel support, followed by additional calls for unsupported participation.

2.3.1 Courses and instructor participation

We received 31 course proposals, more than twice as many as 2019. These were reviewed by the program committee listed above. Final decisions were made after discussion among the six authors of this report over the course of two conference calls. The decision criteria included having a good mix of data mining and search topics, having some introductory courses as well as more advanced courses, ability to program courses in parallel, and fit with the rest of the program.

The final list of courses accepted through proposal and invitation were:

1. Introduction to Information Retrieval, Charlie Clarke and Zeynep Akkalyoncu Yilmaz
2. Data Mining: From an ad-hoc art form to solid, reproducible processes leading to transparent and accountable outcomes, Andreas Rauber
3. The Role of Evaluation for Search, Nicola Ferro and Maria Maistro
4. Fact Checking, Axel-Cyrille Ngonga Ngomo
5. Crowdsourcing for Language and Multimedia Research, Gareth Jones and Maria Eskevich
6. Spatio-Temporal Data Mining: Methods and Applications to Ocean Data, Stan Matwin and Luis Torgo
7. Learning to Rank for Information Retrieval with Neural Networks, Bhaskar Mitra, Nick Craswell, Daniel Campos, and Emine Yilmaz

-
8. Explainable Machine Learning for Trustworthy Artificial Intelligence, Fosca Giannotto and Dino Pedreschi

There were also three invited keynote lectures:

1. IR4All: Building Search Engines for Everyone, by Doulgas W. Oard
2. IR and DM for Development, by Hussein Suleman
3. NLP for African Languages, by Maria Keet

2.3.2 Student participation

Calls for participation were issued in two phases after the program had been finalized. The first call was for applications for travel support. We received 273 applications.

We allocated USD25,000 for supporting participant travel, including accommodation at the All Africa House just off the University of Cape Town campus. As we wanted to ensure this would be a truly multi-national event, we tried to allocate awards such that those receiving support would receive nearly all of what they needed for their travel—it is not common for African institutions to support international travel, so many of them would not be able to attend at all unless their travel was fully supported. We also wanted a diverse event, in terms of nationality, gender, background, and other considerations.

We offered support to 24 applicants. Nearly all accepted the support by registering for the event. Most were able to attend, but five who accepted support were ultimately unable to attend for various reasons. Nigerian attendees who were offered support the previous year but not able to attend were given priority for support this year, and most of them were able to attend.

A second call, without travel support, was issued closer to the dates of the event. Several dozen additional attendees, mainly from South Africa, registered via this call.

In the end there were 50 participants representing 10 different African countries: Botswana, Cameroon, Ethiopia, Ghana, Malawi, Morocco, Namibia, Nigeria, South Africa, and Zambia. This is an 85% increase in participation over AFIRM 2019. Additionally, gender representation was greatly improved over AFIRM 2019, with 20 women included in the 50 participants.

3 The event

The majority of the event took place in lecture rooms and labs in the Computer Science Building at the University of Cape Town. There were two dedicated lecture halls and two dedicated labs. Each day was split between lectures and labs, with lectures starting at 0900 and going through 1200, at which point lunch was served. Lab sessions picked up at 1300 and went through to 1630. After a break, each day concluded with a reception, keynote, or panel.

The first two days of the event were single-track, featuring lectures and labs to introduce participants to the search and data mining settings for the rest of the week. Each day after that had two courses in parallel, one for a data mining track and one for a search track.

Day 1 The event kicked off with the University of Waterloo team Charlie Clarke and Zeynep Akkalyoncu Yilmaz given an introduction to Information Retrieval via lectures and hands-on labs. The day concluded with a keynote lecture from Doug Oard on building search engines for under-resourced languages.

Day 2 Next on the agenda was an introduction to data mining by Andreas Rauber and Tanya Berger-Wolf. An evening poster reception allowed instructors and students to meet, discuss the students' work, and get to know each other.

Day 3 The third day of the event was the first with parallel tracks. Nicola Ferro and Maria Maistro presented lectures and hands-on labs on topics in search evaluation, while Axel Ngongo gave lectures and labs on fact checking. Wednesday officially closed with a keynote lecture from Hussein Suleman on IR and DM for development. An optional panel on approaches to research took place later at All Africa House, where many participants were staying, moderated by Lighton Phiri and Ben Carterette.

Day 4 The fourth day continued the split, with lectures and labs on crowdsourcing given by Gareth Jones and Maria Eskevich in one room, and lectures and labs on spatio-temporal data mining in another. The final keynote lecture by Maria Keet closed the day's sessions, and this was followed by the gala dinner at the Gold Restaurant in Cape Town.

Day 5 The week ended with lectures and labs on learning to rank given by Bhaskar Mitra, Nick Craswell, Emine Yilmaz and Daniel Campos, and explainable machine learning by Fosca Giannotti and Dino Pedreschi. The school concluded with a panel featuring four of the schools attendees—Ahmed Olanrewaju of the University of Ibadan, Nigeria; Chitungo Chisunka-Mwila of the University of Zambia; Sylvester Hatsu from Accra Technical University, Ghana; and Meryem Janati Idrissi of Mohammed VI Polytechnic University, Morocco—and moderated by Tamer Elsayed, and a closing ceremony.

4 Conclusions

The 2nd ACM SIGIR/SIGKDD Africa School on Machine Learning for Data Mining and Search ran from January 27–31, 2020, in Cape Town, South Africa. We increased diversity of both presents and participants in terms of subject matter, origin, and gender, all of which we had as goals. Response was overwhelmingly positive for the second time, and we are highly motivated to continue the initiative.

4.1 Next steps

We continue to maintain the open-subscription Google Group at afirm-africa@googlegroups.com for communication among those who have attended AFIRM 2019 or 2020 in person as well as any one else who is interested in joining. Unfortunately, it is likely that AFIRM will not happen in January 2021 due to travel restrictions and bans on gatherings as a result of the COVID-19 pandemic.

References

Ben Carterette, Suleman Hussein, and Oard Douglas W. Report on the 1st acm sigir/sigkdd africa school on machine learning for data mining and search. *SIGIR Forum*, 53(1):3–13, 2019.