

Report on the 25th Australasian Document Computing Symposium (ADCS 2021)

Joel Mackenzie
The University of Queensland
Australia
joel.mackenzie@uq.edu.au

Damiano Spina
RMIT University
Australia
damiano.spina@rmit.edu.au

Abstract

The 25th Australasian Document Computing Symposium (ADCS 2021) took place as a virtual seminar series from December 2021 through to February 2022, co-hosted by the University of Melbourne and RMIT University, and held in cooperation with ACM SIGIR. Five sessions were dedicated to presenting and discussing nine accepted research papers, book-ended by opening and closing keynotes. The virtual nature of ADCS 2021, coupled with free registration, attracted a broader audience than usual, with 110 registrants from 16 countries including those outside of Australia and New Zealand.

Date: 9 December, 2021 – 24 February, 2022.

Website: <http://adcs-conference.org/2021>.

1 Overview

The Australasian Document Computing Symposium (ADCS) has been a yearly event running since 1996 in locations across Australia and New Zealand. After the unfortunate cancellation of ADCS 2020 due to the ongoing disruption caused by the COVID-19 pandemic, the 25th iteration of ADCS was held in 2021 and was the first to be held virtually. After discussion among the chairs and the steering committee, it was decided to run ADCS 2021 as a seminar series, with short (1–2 hour) sessions spread across a multi-week period between December 2021 and February 2022. The main aim was to minimize the *virtual meeting fatigue* which has become a familiar burden for many of us over the last two years. Running ADCS virtually had the added benefit of attracting participants from outside of Australia and New Zealand — including authors from Ireland, almost certainly an ADCS first. Overall, the first virtual ADCS was deemed to be a success, and was well received by the community.

2 Program

The ADCS 2021 program was made up of two keynotes, nine paper presentations, a business meeting, and a social event. All of these events were held virtually across December 2021 into

February 2022, with breaks for the Christmas/summer holiday period, and a break during the SIGIR submission week. Most sessions ran for one hour, and there was typically only one per week (with one exception, where two seminars were held in the same week) in an effort to avoid fatigue.

2.1 Keynotes

ADCS had two fantastic keynotes:

- The first keynote was held on December 9, 2021, and was presented by Professor Min Zhang from Tsinghua University, China. The talk, titled “*What users tell us: User understanding and modeling for personalized recommendation,*” gave an overview of how user’s behavior and intents can be characterized in a number of information access scenarios, such as news streaming, e-commerce, job seeking, and music recommendation.
- The second keynote was held on February 24, 2022, and was presented by Dr Terrence Szymanski from SEEK Ltd. The talk, titled “*Maximising the potential of documents, NLP, and AI in an online employment marketplace: Technical and non-Technical challenges,*” brought together a number of challenges faced in industry when building AI services to solve information access problems in the context of job/talent search and recommendation.

Both keynotes saw great interest from the attendees, and we are grateful to Min and Terrence for taking the time to share their experience with the community.

2.2 Papers

Previously, ADCS held both *short paper* and *long paper* tracks. However, in 2021, we decided to follow ICTIR 2021 and invite authors to write papers of a length that matched their content. In particular, any length between 2 and 9 pages was allowed, with an unlimited space allowance for references. We hoped that this would encourage more submissions, and provide more flexibility to authors. A total of 14 research papers were submitted, with 9 accepted for publication, across a diverse range of research topics. The length of time allocated for presentations took into account the length of the paper: papers with 6 pages (or less) were allocated 20 minutes, and papers with 7 pages (or more) were allocated 25 minutes. A brief summary of each contribution is as follows:

- *An analysis of the Australian political discourse in sponsored social media content*, by Lei Han, Rudra Sawant, Shaoyang Fan, Glenn Kefford, and Gianluca Demartini, explored how sponsored content on social media is used in political voting campaigns in the context of Australia. They collected a large range of political ads and related media from Facebook, and annotated the ads through a crowdsourcing task. Their resulting dataset is publicly available to promote further research on this area.
- *Cohort-based clinical trial retrieval*, by Bevan Koopman and Guido Zuccon, proposed a cohort-based approach to maximize recruitment opportunities for patients in clinical trial retrieval. While typical approaches use retrieval to match a single patient to a trial, they consider approaches to maximize opportunities across the entire patient cohort.

-
- *Cost-effective updating of distributed reordered indexes*, by Joel Mackenzie and Alistair Moffat, focused on the problem of handling batch update operations on a reordered index, and the resulting efficiency implications for both storage and querying. This work won the SEEK Best Paper award.
 - *Annotation of struck-out text in handwritten documents*, by Hiqmat Nisa, Vic Ciesielski, James Thom, and Ruwan Tennakoon, investigated annotation strategies for training deep learning models on handwriting recognition tasks. In particular, they examined best practices for annotating struck-out text such that the recognition models can learn to ignore the struck-out text.
 - *Passage based answer-set graph approach for query performance prediction*, by Ghulam Sarwar and Colm O’Riordan, revisited the problem of query performance prediction. Given a query, they model the similarity between the top ranking results from a system as a weighted graph, and use graph features such as set-cover to derive novel post-retrieval predictors.
 - *Hierarchical clustering of corals using image clustering*, by Rudra Sawant, Gianluca Demartini, and Tom Bridge, investigated the problem of automatically generating a taxonomy of different species of coral from an image collection. They use different machine learned models to generate feature embeddings from the coral images, and apply clustering to generate a hierarchy.
 - *Crowdsourcing backstories for complex task-based search*, by Manuel Steiner, Damiano Spina, Falk Scholer, and Lawrence Cavedon, explored how descriptive backstories can be generated for complex search tasks. In the context of job search, they describe a crowdsourcing methodology where workers compile their thoughts into a backstory given a job advertisement and the query which retrieved it.
 - *MeSH term suggestion for systematic review literature search*, by Shuai Wang, Hang Li, Harrison Scells, Daniel Locke, and Guido Zuccon, proposed methods for suggesting Medical Subject Heading (MeSH) terms to an initial Boolean query to improve search effectiveness for medical systematic reviews. This work won the SEEK Best Student Paper award.
 - *Document clustering vs topic models: A case study*, by Meng Yuan, Pauline Lin, and Justin Zobel, investigated the relationship between document clustering and topic models. They find that document clustering and topic modeling result in surprisingly cohesive descriptions (with respect to one-another) of a large text collection.

Further details of these interesting works can be found in the ACM digital library [Mackenzie and Spina, 2021],¹ and all talks (including the keynotes) can be found on the ACM’s official YouTube channel.²

3 ADCS Community

We now provide a brief overview of the community who helped make ADCS 2021 a success.

¹ADCS 2021 in the ACM Digital Library: <https://dl.acm.org/doi/proceedings/10.1145/3503516>

²ADCS 2021 presentations: <https://bit.ly/ADCS2021Videos>

3.1 Organization

In 2021, the ADCS organizing committee was made up of five members:

- **General Chairs:** Joel Mackenzie (The University of Melbourne in 2021; now at the University of Queensland) and Damiano Spina (RMIT University);
- **Proceedings Chair:** Lida Rashidi (The University of Melbourne);
- **Web & Social Media Chair:** Valeriia Blinova (RMIT University); and
- **Best Papers Chair:** David Hawking (The Australian National University).

The steering committee, made up of the program/general chairs from the previous six iterations of ADCS, were instrumental in supporting the organization of ADCS 2021. ADCS was also supported by a total of 37 program committee members who helped review the submissions. In 2021, we endeavored to expand and diversify the program committee by inviting members of the community who were now in industry positions, or who had relocated outside of Australasia.

3.2 Registrants and Attendees

Since registration was free, as is the tradition for ADCS, not all registrants attended a session. However, we believe that the majority of registrants attended at least one session, with some members attending all sessions. An overview the registrants is as follows:

- In total, we had 110 registrants from 16 countries.
- Many registrants (78%) were from Australia or New Zealand, with the remaining 22% located further afield.
- The majority of attendees (66%) classified themselves as working for academic institutions. The remainder were either from industry, non-academic research organizations, or a combination thereof.
- There were a total of 45 unique institutions represented from across the world.
- More than half (62%) of attendees had never attended ADCS prior to 2021.
- Students made up 32% of registrants.

We hope that we can continue to welcome a more global participation in ADCS going forward.

4 Final Remarks

ADCS 2021 was the first virtual edition of ADCS (Figure 1). A novel approach to minimize virtual fatigue was implemented by running the conference as a virtual seminar series from December 2021 through to February 2022. After the final session we asked participants to fill out a short feedback form. We obtained a total of 16 responses, all from Australia. Of those, four respondents (25%) were attending ADCS for the first time. According to the feedback shared by the respondents, the highlights of ADCS '21 were both the content (papers and keynotes) and the format of spreading the virtual conference into weekly seminars. Almost all respondents (15 out of 16) indicated preference towards the virtual seminar format against a two-day virtual conference. As one would expect, the most negative aspect was not being able to catch up in person. We are aiming to

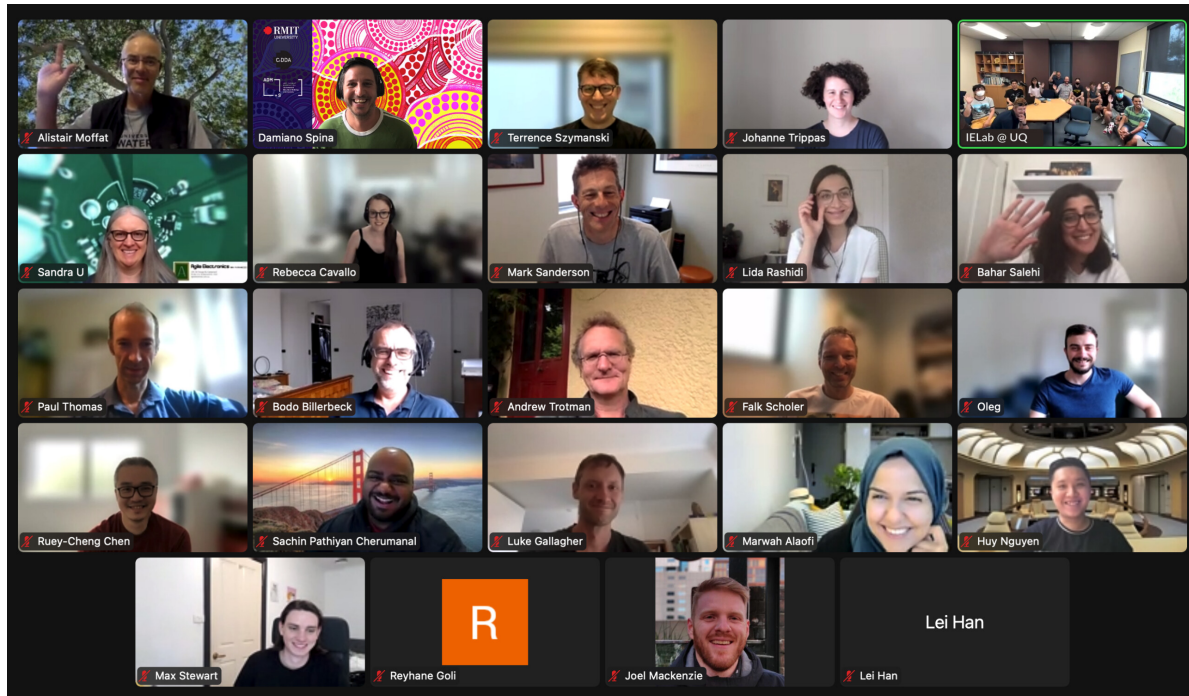


Figure 1: Happy faces at the ADCS 2021 Virtual Seminar on February 24, 2022. Shared with permission.

provide hybrid arrangements for 2022, and look forward to welcoming those who wish to attend ADCS 2022 to Adelaide in December!

Acknowledgments

ADCS 2021 was a success thanks to the support of the community and volunteers who contributed in various capacities. We would like to thank ACM SIGIR, RMIT University, Seek Limited, and the University of Melbourne for their generous sponsorship. We are grateful to our keynote speakers, Professor Min Zhang and Dr Terrence Szymanski, for sharing their experience with the community. We would also like to acknowledge the organizing committee, steering committee, and program committee for their assistance in making ADCS 2021 a success. Lastly, we would like to express our gratitude to all attendees of ADCS.

References

Joel Mackenzie and Damiano Spina, editors. *ADCS '21: Proceedings of the 25th Australasian Document Computing Symposium*, Virtual Event, Australia, 2021. Association for Computing Machinery, New York, NY, USA. ISBN 978-1-4503-9599-1. URL <https://doi.org/10.1145/3503516>.