UNIVERSITY OF AMSTERDAM

UvA-DARE (Digital Academic Repository)

ECOM'21: The SIGIR 2021 Workshop on eCommerce

Kallumadi, S.; King, T.H.; Malmasi, S.; de Rijke, M.

DOI
10.1145/3404835.3462820

Publication date
2021

Document Version
Final published version

Published in
SIGIR '21

License
CC BY

Link to publication

Citation for published version (APA):
Kallumadi, S., King, T. H., Malmasi, S., & de Rijke, M. (2021). ECOM'21: The SIGIR 2021 Workshop on eCommerce. In *SIGIR '21: proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval : July 11-15, 2021, virtual event, Canada* (pp. 2685-2688). Association for Computing Machinery. https://doi.org/10.1145/3404835.3462820

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: https://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, P.O. Box 19185, 1000 GD Amsterdam, The Netherlands. You will be contacted as soon as possible.

UvA-DARE is a service provided by the library of the University of Amsterdam (https://dare.uva.nl)

Download date:12 Oct 2026

ECOM'21: The SIGIR 2021 Workshop on eCommerce

Surya Kallumadi
Lowe's Companies, Inc.
surya@ksu.edu

Tracy Holloway King
Adobe
tking@adobe.com

Shervin Malmasi
Amazon.com, Inc.
malmasi@amazon.com

Maarten de Rijke
University of Amsterdam & Ahold Delhaize
m.derijke@uva.nl

ABSTRACT

eCommerce Information Retrieval (IR) is receiving increasing attention in the academic literature and is an essential component of some of the world's largest web sites (e.g. Airbnb, Alibaba, Amazon, eBay, Facebook, Flipkart, Lowe's, Taobao, and Target). SIGIR has for several years seen sponsorship from eCommerce organisations, reflecting the importance of IR research to them. The purpose of this workshop is (1) to bring together researchers and practitioners of eCommerce IR to discuss topics unique to it, (2) to determine how to use eCommerce's unique combination of free text, structured data, and customer behavioral data to improve search relevance, and (3) to examine how to build datasets and evaluate algorithms in this domain. Since eCommerce customers often do not know exactly what they want to buy (i.e. navigational and spearfishing queries are rare), recommendations are valuable for inspiration and serendipitous discovery as well as basket building.

The theme of this year's eCommerce IR workshop is **ensuring fairness in search and recommendations for eCommerce**. The workshop includes papers on this topic as well as a panel focused on this area (see Section 4). In addition, Coveo is sponsoring an eCommerce data challenge on session-based prediction for predicting the next action with a special subtask on cart abandonment. The data challenge reflects themes from prior SIGIR workshops in 2017, 2018, 2019, and 2020.

CCS CONCEPTS

• **Information systems** → **Online shopping**; **Specialized information retrieval**;

KEYWORDS

eCommerce information retrieval, product search, recommender systems, fairness, bias

ACM Reference Format:
Surya Kallumadi, Tracy Holloway King, Shervin Malmasi, and Maarten de Rijke. 2021. ECOM'21: The SIGIR 2021 Workshop on eCommerce. In *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '21), July 11–15, 2021, Virtual Event, Canada.* ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/3404835.3462820

This work is licensed under a Creative Commons Attribution International 4.0 License.

SIGIR '21, July 11–15, 2021, Virtual Event, Canada.
© 2021 Copyright held by the owner/author(s).
ACM ISBN 978-1-4503-8037-9/21/07.
https://doi.org/10.1145/3404835.3462820

1 MOTIVATION

Search and recommendation have applications ranging from traditional web search to document collections to vertical search systems. In this workshop, we explore approaches to search and recommendation of products in eCommerce information retrieval (IR). Although the basic search task (i.e. fulfill a user's information need) is the same as web search, the way in which this is achieved is different. On eCommerce sites (e.g. Alibaba, Amazon, eBay, Etsy, Flipkart, Walmart), the data available for retrieval and ranking are different as are the signals of success (e.g. adding items to a cart, purchasing) [5].

The entities that need to be discovered are combinations of unstructured text (e.g. titles, descriptions, reviews), images, and structured data (e.g. price, brand, ratings, popularity, revenue). This complex combination of data raises interesting research challenges including recall (also known as matching) and ranking functions that take into account the trade-offs across facets with respect to the input query. The features available for building click models used in ranking are different, and often stronger, in eCommerce than in web search. As well as queries, hover time, clicks, and browse time, eCommerce sites also have signals from add-to-cart, purchase, side-by-side comparison, remove-from-cart, return of goods, etc. Furthermore, when incorporating promotions and personalization such as individual pricing, the click models are more complex than in web search. eCommerce is also characterized by a dynamic inventory with a high rate of change and turnover, and a very long tailed query distribution.

This workshop brings together researchers and practitioners from academia and industry to identify and discuss core research problems in eCommerce search and recommendation. The workshop aims to foster collaboration by bringing the community together in a way that rarely happens outside of focused workshops like this one, to attract research funding to this increasing important domain, and to introduce IR researchers and postgraduate students to eCommerce and product discovery. Finally, it will help broaden the definition of IR at research venues such as SIGIR. To support these goals, the workshop features a **special theme** as well as a **data challenge**.

Special Theme: Fairness in eCommerce. During 2020 and 2021, the economic and societal importance of eCommerce grew tremendously. Now, more than ever, customers and suppliers rely on eCommerce. In line with this growing importance in people's lives, the

theme of this year's eCommerce IR workshop is ensuring fairness in search and recommendations for eCommerce. Here, fairness refers to three types of stakeholders: customers who require access to online goods, suppliers who are competing for customers, and the broader society which is concerned with fairness and sustainability. See Section 3.2 for details.

Data Challenge. The workshop will continue to support data availability for eCommerce IR research. In 2018, 2019, and 2020 we released data from Rakuten and eBay. This year Coveo will release data for a large-scale session-based interaction prediction task. The data from the challenge will be fully anonymized and reviewed per the special theme of trust and fairness. See Section 3.3 for details.

2 RELATED MATERIAL AND WORKSHOPS

Two eCommerce search overviews [5], of which three of the organizers are authors, and the SIGIR 2018 tutorial "Information Discovery in E-commerce"[1]) provide background to potential participants of ECOM21 who want to understand the state of the art before attending the workshop.

The interest in this workshop by the SIGIR community has continuously increased over the years, as reflected by the growth in participation and paper submissions. Prior workshops that this workshop builds on include SIGIR 2017 (~60 delegates),[2] SIGIR 2018 (~90 delegates),[3] SIGIR 2019 (~130 delegates),[4] and SIGIR 2020 (~220 delegates).[5] We believe a significant number of SIGIR participants will be interested in this year's workshop theme of fairness in eCommerce and the panel focused on this topic reflecting industry and academic perspectives.

Prior eCommerce search and recommendation data challenges include: the Rakuten challenge at the SIGIR 2018 workshop (28 participating groups);[6] the "CIKM Cup 2016 Track 2: Personalized E-Commerce Search Challenge" (19 participating groups);[7] the eBay Inc. challenge at the SIGIR 2019 workshop (10 participating groups);[8] and the Rakuten France challenge at the SIGIR 2020 workshop (16 participating groups). We expect similar numbers to participate in this year's challenge on session-based prediction of next action using real-world data provided by Coveo.

The IJCAI-17 Workshop on AI Applications in E-Commerce (WAAE 2017)[9] dealt with some of the research areas that our workshop will cover; however the scope of WAAE was much broader covering all of AI as opposed to search and recommendations. In 2019 The Web Conference held the "First Workshop on e-Commerce and NLP"[10] which covers a complementary field, Natural Language Processing, and was followed up by two workshops in 2020 at WebConf[11] and ACL.[12] The ACM Special Interest Group on E-commerce (SIGecom) runs an annual conference that includes search as well as recommendation in the call for papers. However, for many years, search and IR has not been well represented at this conference. The SIGIR ECOM21 workshop addresses this shortfall by examining eCommerce from the perspective of search and recommendation.

[1] https://sites.google.com/view/sigir2018-info-ec/home
[2] http://sigir-ecom.weebly.com/
[3] https://sigir-ecom.github.io/ecom2018
[4] https://sigir-ecom.github.io/ecom2019
[5] https://sigir-ecom.github.io/ecom2020
[6] https://sigir-ecom.github.io/ecom2018/data-task.html
[7] https://competitions.codalab.org/competitions/11161
[8] https://sigir-ecom.github.io/ecom2019/data-task.html
[9] No longer online
[10] https://sites.google.com/view/ecnlp/past-workshops/ecnlp-2019
[11] https://sites.google.com/view/ecnlp/past-workshops/www-2020
[12] https://sites.google.com/view/ecnlp/past-workshops/acl-2020

3 THEME AND PURPOSE

The primary theme of the workshop is eCommerce search and recommendation. The purpose of the workshop is to provide a venue for discussion and publication of IR research as it pertains to eCommerce. The workshop brings together practitioners, researchers, and applied researchers from academia and industry to discuss the challenges and approaches to eCommerce search and recommendation. It aims to foster collaboration and discussion with the broader IR community and to raise awareness within the academic community of the unique challenges faced by the eCommerce domain.

3.1 Scope

The workshop relates to all aspects of eCommerce search and recommendation. Research topics and challenges that are frequently encountered in this domain include:

- Ranking and Whole Page Relevance
  - Diversity in product search and recommendations
  - Relevance models for multi-faceted entities
  - Relevance vs. revenue
  - Ranking features and learning mechanisms
  - Deterministic sorts (e.g. price low to high)
  - Temporal dynamics and seasonality
- Query Understanding
  - Query intent, query suggestions, and auto-completion
  - Strategies for resolving low or zero recall queries
  - Converting across modalities (e.g. text, structured data, images)
- Document Understanding
  - Categorization and facets
  - Reviews and sentiment analysis
- Recommendation and Personalization
  - Personalization & contextualization, including the use of personal facets such as age, gender, location
  - Blending recommendations and search results
- Representations and Data
  - Data engineering and provenance for eCommerce
  - Semantic representation of products, queries, and customers
  - Construction and use of knowledge graphs for eCommerce
- IR Fundamentals for eCommerce
  - Cross-lingual search and machine translation
  - Machine learning techniques for eCommerce applications
  - Indexing and search in rapidly changing environments (e.g. auction sites)
  - Experimentation techniques including AB testing and Multi-armed bandits
- Other challenges
  - Trust and fairness in eCommerce
  - UX for eCommerce
  - The role of search in trust and security for marketplaces
  - Question answering and chat bots for eCommerce

3.2 Special Theme: Fairness in eCommerce

During 2020 and 2021, the economic and societal importance of eCommerce grew tremendously. Now, more than ever, customers and suppliers rely on eCommerce. In line with this growing importance, the theme of this year's eCommerce IR workshop is ensuring fairness in search and recommendation for eCommerce. Here, fairness refers to three types of stakeholders.

First, customers of eCommerce require equal access to online goods regardless of their demographics, including socioeconomic status, location and special needs. In addition, eCommerce sites must respect the privacy of their customers' online actions. This is an increasingly pressing issue given the desire for personalization of search and recommendation results and the reliance of eCommerce systems on individual and aggregated customer behavioral data.

Second, for suppliers who connect to customers through eCommerce platforms such as Amazon, Bol.com, and eBay, fairness includes notions such as fairness of exposure, especially in the organic search and recommendation results where all the suppliers are competing for the top slots. This is further complicated when the platform is also a supplier, e.g. in the case of Amazon and Bol.com. Performance advertising, i.e. sponsored products, is increasingly used by platforms to help with exposure, especially for new suppliers and products. This introduces additional trust issues where customers may confuse sponsored placement with organic search results.

Third, fairness has broader societal implications around social responsibility. Increasingly eCommerce sites allow users to search and refine by ways to reduce packaging, reduce transportation, reduce waste (e.g. of perishable products), purchase ethically sourced foods and manufactured goods, etc.

3.3 Data Challenge

At the 2018 workshop, we collaborated with Rakuten to run a data challenge "Taxonomy Classification for eCommerce-scale Product Catalogs"[13] which addressed the problem of taking a product listing and predicting the category (from a taxonomy) that listing should belong under. Academic, industrial, and independent participation was high, with 28 teams submitting runs. We worked with Rakuten to release the data from the challenge so that others can continue research on this topic.

For the 2019 workshop, eBay Inc. released item data as well as queries and assessments for a "High Accuracy Recall Task."[14] This task examined relevance of the recall set for deterministic sort orders (e.g. price low to high) and a task not seen in web search.

The 2020 workshop hosted a data challenge targeting taxonomy classification for eCommerce-scale multi-modal product catalogs. Rakuten France provided data including several million product titles, images and descriptions from their French catalog, featuring thousands of taxonomy-structured labels. The cataloging of product listings through taxonomy categorization is a fundamental problem in eCommerce, with applications ranging from query understanding to personalized search and recommendations.

For this year's workshop, Coveo[15] is providing a large dataset of customer behavior sessions from a mid-size eCommerce domain. The task is to predict customer interactions. Advances in this research area have been limited due to the lack of real data from eCommerce companies. Making the data available will benefit researchers and practitioners who have not had the opportunity to contribute their ideas due to the previous lack of data.

The Coveo dataset presents product level data from a mid-size eCommerce company, coupled with search-based interactions and content-based information. As such, this data challenge is a representative example of thousands of shops in the industry, with realistic traffic, conversion rate, abandonment ratio, etc.

The data challenge objectives are to:

- advance the field with the release of a new, fine-grained, session-based dataset, and
- propose inference tasks that are interesting for retailers of all sizes.

3.4 Workshop Outcomes

We believe that the most important outcome of the workshop is the discussion between individual participants at the workshop. It is these discussions that lead to collaboration across institutions, including across academia and industry, and future research. Insights will be captured in the form of a SIGIR Forum workshop report. As with previous workshops, the proceedings of the workshop will be published through CEUR Workshop Proceedings to ensure they are appropriately archived.

This year's workshop theme of fairness in eCommerce and the data challenge in session-based personalization are particularly timely given the 2021 zeitgeist with the increasing importance of eCommerce and the awareness of the dangers of biased ML and AI algorithms, including those used in IR, to society. We hope that through discussion at the workshop and more generally at SIGIR 2021, this workshop can help steer the research community towards these problems and in doing so work towards solutions.

4 WORKSHOP FORMAT

ECOM20 was held entirely virtually with a format to encourage maximum participation and discussion: significant time was allotted as part of each session for discussion and the chairs and workshop organizers acted as facilitators. ECOM21 will also be entirely virtual.

REFERENCES

[1] Jon Degenhardt, Giuseppe Di Fabbrizio, Surya Kallumadi, Mohit Kumar, Andrew Trotman, Yiu-Chang Lin, and Huasha Zhao (Eds.). 2018. *Proceedings of the SIGIR 2018 Workshop on eCommerce.* CEUR Workshop Proceedings. http://ceur-ws.org/Vol-2319/ Co-located with the 41st International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR 2018).

[2] Jon Degenhardt, Surya Kallumadi, Maarten de Rijke, Luo Si, Andrew Trotman, and Yinghui Xu (Eds.). 2017. *Proceedings of the SIGIR 2017 Workshop on eCommerce.* CEUR Workshop Proceedings. http://ceur-ws.org/Vol-2311/ Co-located with the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR 2019).

[3] Jon Degenhardt, Surya Kallumadi, Utkarsh Porwal, and Andrew Trotman (Eds.). 2019. *Proceedings of the SIGIR 2019 Workshop on eCommerce.* CEUR Workshop Proceedings. http://ceur-ws.org/Vol-2410/ Co-located with the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR 2019).

[13] https://sigir-ecom.github.io/ecom2018/data-task.html
[14] https://sigir-ecom.github.io/ecom2020/data-task.html
[15] https://www.coveo.com

[4] Dietmar Jannach, Surya Kallumadi, Tracy Holloway King, Weihua Luo, and Shervin Malmasi (Eds.). 2020. *Proceedings of the SIGIR 2020 Workshop on eCommerce*. CEUR Workshop Proceedings. Co-located with the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR 2020).

[5] Manos Tsagkias, Tracy Holloway King, Surya Kallumadi, Vanessa Murdock, and Maarten de Rijke. 2020. Challenges and Research Opportunities in eCommerce Search and Recommendations. *SIGIR Forum* 54, 1 (June 2020), 1–23.