UNIVERSITY OF AMSTERDAM

UvA-DARE (Digital Academic Repository)

eCom'22: The SIGIR 2022 Workshop on eCommerce

Kale, A.; Kallumadi, S.; King, T.H.; Malmasi, S.; de Rijke, M.; Tagliabue, J.

DOI
10.1145/3477495.3531701

Publication date
2022

Document Version
Final published version

Published in
SIGIR '22

License
CC BY

Link to publication

Citation for published version (APA):
Kale, A., Kallumadi, S., King, T. H., Malmasi, S., de Rijke, M., & Tagliabue, J. (2022). eCom'22: The SIGIR 2022 Workshop on eCommerce. In *SIGIR '22: proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval : July 11-15, 2022, Madrid, Spain* (pp. 3485-3487). The Association for Computing Machinery. https://doi.org/10.1145/3477495.3531701

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:07 Oct 2026

eCom'22: The SIGIR 2022 Workshop on eCommerce

Ajinkya Kale
Adobe
akale@adobe.com

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
m.derijke@uva.nl

Jacopo Tagliabue
Coveo
jtagliabue@coveo.com

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 seen sponsorship from eCommerce organisations for the past several years, 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 **Bridging IR Metrics and Business Metrics and Multi-objective Optimization**. The workshop includes papers on this topic as well as a panel focused on this area (see Section 3). In addition, Farfetch is sponsoring a recommendation challenge focused on outfit completion: as part of the event, Farfetch will release to the research community a novel, large dataset containing multi-modal information and extensive labels curated by fashion experts. The data challenge reflects themes from prior SIGIR workshops in 2017, 2018, 2019, 2020, 2021.

CCS CONCEPTS

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

KEYWORDS

eCommerce information retrieval, product search, recommender systems, metrics, evaluation

ACM Reference Format:
Ajinkya Kale, Surya Kallumadi, Tracy Holloway King, Shervin Malmasi, Maarten de Rijke, and Jacopo Tagliabue. 2022. eCom'22: The SIGIR 2022 Workshop on eCommerce. In *Proceedings of the 45th Int'l ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '22), July 11–15, 2022, Madrid, Spain.* ACM, New York, NY, USA, 3 pages. https://doi.org/10.1145/3477495.3531701

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

SIGIR '22, July 11–15, 2022, Madrid, Spain.
© 2022 Copyright held by the owner/author(s).
ACM ISBN 978-1-4503-8732-3/22/07.
https://doi.org/10.1145/3477495.3531701

1 MOTIVATION FOR THE WORKSHOP

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 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).

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 (matching) and ranking functions that take into account the trade-offs across facets with respect to the user's 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. 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 tail of 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, to attract research funding to this increasingly 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**.

1.1 Data Challenge

The workshop will continue to support data availability for eCommerce IR research. In 2018, 2019, 2020, 2021 we released data from Rakuten, eBay and Coveo. This year, Farfetch will release data and organize a data challenge in the fashion eCommerce domain. See Section 2.3 for details.

1.2 Appropriateness to SIGIR

The primary theme of the workshop is eCommerce search and recommendations (i.e. Information Retrieval). As such it is appropriate for SIGIR. The recent publication of a position paper on search and discovery in eCommerce in SIGIR Forum supports this [7]. Moreover, the special theme of this year (bridging IR metrics and business metrics and multi-objective optimization) aligns well with both longstanding and current interests of the SIGIR community.

2 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. We will bring together practitioners, researchers, and applied researchers from academia and industry to discuss the challenges and approaches to eCommerce search and recommendation. We aim 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.

2.1 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
  - Optimization for IR and business metrics (special theme for 2022)
  - 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 and contextualization, including the use of personal facets such as age, gender, location
  - Privacy, bias and ethics in eCommerce IR
  - 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, transparency, and fairness in eCommerce
  - UX for eCommerce
  - The role of search in trust and security for marketplaces
  - Question answering and chat bots for eCommerce

2.2 Special Theme: Bridging IR metrics and business metrics and multi-objective optimization

While search and recommendation models are trained on implicit feedback generated by user interactions, the impact of these models are measured in terms of one or more business metrics such as Gross Merchandise Value (GMV), Average Order Value (AOV), and Conversion. There exists a gulf between metrics used for optimizing these models and measuring the business impact of these models. In addition, there exists a delay in the feedback of customers interacting with the products surfaced by search and recommendations, and their eventual purchase; this is especially true for some verticals (e.g. DIY), where browsing online and buying offline is a common practice. In an increasingly "hybrid" or "multi-channel" shopping world, reconciling short-term, easy-to-attribute signals with long-term, noisy business rewards is a crucial challenge that opens up new research opportunities for the IR community.

In addition to having invited speakers who are experts in this area for eCom'22, we solicit papers from academia and industry focused on this topic.

2.3 Data Challenge

In order to facilitate community building in the eCommerce information retrieval domain, we have been organizing data challenges in conjunction with this workshop. For previous workshops, eCommerce companies such as Rakuten (2018, 2020), eBay (2019), and Coveo (2021) have supported this initiative by organizing data challenges. These data challenges have spanned a wide variety of topics including Taxonomy classification, Session-oriented personalization, and High accuracy recall. These data challenges have led to increased interest in the eCommerce domain with participation from diverse groups across industry and academia and have led to building a community of researchers interested in this domain. Often, even after the end of the challenge, the dataset made available to the community continues to be downloaded and frequently used for benchmarking and educational and product development purposes. For the fifth year in a row we have worked with industry researchers to create a data challenge in 2022, the *Farfetch Data Challenge*.

Farfetch is one of the largest fashion eCommerce companies in the world. The *Farfetch Data Challenge* revolves around outfit composition. Currently fashion experts are tasked with curating outfits according to a "style". To scale this manual process, we are asking participants to develop models for outfit creation. To solve this important use case, a new dataset will be released, including 300, 000 products (with multi-modal meta-data), and 240, 000 outfits curated by fashion experts. Participants will be asked to prepare

recommenders that, given an item, predict products that "complete that outfit".

The challenge sits at the intersection of problems and methods that are interesting to a broad set of practitioners, inside and outside of fashion. Since outfits cannot be modelled with traditional data-mining approaches, the challenge invites participants to borrow tools from computer vision, NLP and recent multi-modal techniques [1]. In particular, given the power-law nature of product interactions, developing content-based strategies that rely on product understanding and not just behavioral data is crucial for most, if not all, retailers.

To run the competition, we will leverage the same tools and processes that proved to be successful in the past: the organizers will maintain an online leaderboard and a slack channel for questions, clarifications, announcements etc. We will also invite paper submissions from all participants. The accepted contributions as well as the winners will prerecord a presentation on their submission and will participate in a session dedicated to the data challenge.

2.4 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 to future research. We will capture what we can in the form of a SIGIR Forum workshop report and, as with previous workshops, we will produce a proceedings of the workshop and work with CEUR Workshop Proceedings to ensure they are appropriately archived.

This year's workshop theme of metrics and measurement and the data challenge in the fashion domain by Farfetch are of increasing importance to eCommerce. We hope that through discussion at the workshop and, more generally, at SIGIR 2022, we can help steer the research community towards these problems and in doing so work towards solutions.

3 WORKSHOP FORMAT

eCom'21 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 [6]. eCom'22 will be held as a full day hybrid workshop to accommodate for diverse participation.

REFERENCES

[1] Patrick John Chia, Jacopo Tagliabue, Federico Bianchi, Ciro Greco, and Diogo Gonçalves. 2022. "Does It Come in Black" CLIP-like Models are Zero-shot Recommenders. *arXiv* abs/2204.02473 (2022).

[2] 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).

[3] 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).

[4] 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).

[5] 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).

[6] Surya Kallumadi, Tracy Holloway King, Shervin Malmasi, and Maarten de Rijke. 2021. eCom'21: The SIGIR 2021 Workshop on eCommerce. In *SIGIR 2021: 44th international ACM SIGIR Conference on Research and Development in Information Retrieval.* ACM, 2685–2688.

[7] 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.