

# ECOM'19: The SIGIR 2019 Workshop on eCommerce

Jon Degenhardt  
eBay Inc.  
jdegenhardt@ebay.com

Utkarsh Porwal  
eBay Inc.  
uporwal@ebay.com

Surya Kallumadi  
Home Depot  
surya@ksu.edu

Andrew Trotman  
University of Otago  
andrew@cs.otago.ac.nz

## ABSTRACT

eCommerce Information Retrieval is receiving increasing attention in the academic literature, and is an essential component of some of the largest web sites (such as eBay, Amazon, Airbnb, Alibaba, Taobao, Target, Facebook, Home Depot, and others). These kinds of organisations clearly value the importance of research into Information Retrieval. The purpose of this workshop is to bring together researchers and practitioners of eCommerce IR to discuss topics unique to it, to set a research agenda, to examine how to build data sets, and how evaluate algorithms for research into this fascinating topic.

eCommerce IR is ripe for research and has a unique set of problems. For example, in eCommerce search there may be no hypertext links between documents (products); there is a click stream, but more importantly, there is often a buy stream. eCommerce problems are wide in scope and range from user interaction modalities through to dynamic updates of a rapidly changing collection on auction sites, and the experientiveness of some products (such as Airbnb bookings).

This workshop is a follow up to very successful workshops held at SIGIR 2017 and SIGIR 2018. This year we will be running a data challenge (sponsored by eBay) which will allow us to follow up on multiple aspects that were discussed in the previous workshops (in particular, deterministic rank orders and how to evaluate these).

## 1 MOTIVATION

In this workshop we will explore approaches for search and recommendations of products, as well as other aspects of eCommerce Information Retrieval. Although the task is the same as web-page search (fulfill a user's information need), the way in which this is achieved is different. On product sites, the traditional web-page ranking features are either not present or are present in a different form.

The entities that need to be discovered (the information that fulfills the need) might be unstructured, associated with structure, semi-structured, or have facets such as: price, ratings, title, description, and seller location. Domains with such facets raise interesting challenges such as a) relevance and ranking functions that take into

account the tradeoffs across various facets b) recommendations based on entity similarity c) recommendations based on user location (e.g. shipping cost). In the case of eCommerce search and recommendation these challenges require an understanding of product attributes, user behavior, and query context. Product sites are also characterized by the presence of a dynamic inventory with a high rate of change and turnover, and a long tail of query distribution.

The features available for machine learning and for building click models are different in eCommerce than in web search. As well as queries, hover time, clicks, and browse times, eCommerce sites also have add-to-basket, purchase, remove-from-basket, return faulty goods, and so on. When incorporating online (and offline) promotions, personalization (such as individual pricing) the click models are vastly more complex than seen for web search.

Outside of search but still within IR, the same feature in different domains can have a different meaning. For example, in email filtering the presence of a product name along with a price is a strong indication of spam, but within an auction setting this likely indicates a valid product. Natural language translation also differs from other domains – company names, product names, and even product descriptions do not translate well with existing tools. Similar problems exist with knowledge graphs that are not customised to match the product domain.

This workshop will bring together researchers and practitioners in order to identify and discuss core research problems in eCommerce search and recommendation. This serves many purposes. First, collaboration: it will bring the community together in a way that rarely happens. Second, funds: it will attract research funding to this increasingly important domain. Third, research: it will attract researchers to eCommerce and product search. Finally, it will help broaden the definition of Information Retrieval at conferences such as SIGIR.

The workshop will continue to examine data availability. ECOM 2018 secured the release of data from Rakuten, who ran a classification data challenge. ECOM19 is working with eBay on the release data for a search task. As the purpose of a product site is to make entity data publicly available, the same security concerns that plague other search domains do not exist. However sales and seller information is private and proprietary and likely to be unavailable.

## 2 RELATED WORKSHOPS

ECOM19 builds on previous workshops at SIGIR 2017<sup>1</sup> and SIGIR 2018.<sup>2</sup> by putting some of the work presented there into practice

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

SIGIR '19, July 21–25, 2019, Paris, France

© 2019 Copyright held by the owner/author(s).

ACM ISBN 978-1-4503-6172-9/19/07.

<https://doi.org/10.1145/3331184.3331643>

<sup>1</sup><http://sigir-ecom.weebly.com/>

<sup>2</sup><https://sigir-ecom.github.io>

with the data challenge – in particular the work on deterministic sort orders (such as price low to high).

The SIGIR 2018 tutorial “Information Discovery in E-commerce”<sup>3</sup> provides much of the necessary background to delegates who might need to understand state of the art before attending the workshop. The IJCAI-17 Workshop on AI Applications in E-Commerce (WAAE 2017) dealt with some of the research areas that our workshop will cover, however the scope of WAAE was much broader. Product search and eCommerce IR was a small aspect in a wide range of topics chosen to be addressed by WAAE. The Web Conference ran the “First Workshop on e-Commerce and NLP”,<sup>4</sup> covering a complementary field, NLP.

Prior data challenges include the Rakuten challenge at our SIGIR 2018 workshop<sup>5</sup> and the “CIKM Cup 2016 Track 2: Personalized E-Commerce Search Challenge”.<sup>6</sup>

The ACM SIGecom runs an annual conference that includes search as well as recommendation in the cfp. However, for many years search has not been well represented at this conference.

This workshop is related to the ESAIR series of workshops that explored semantic annotation – product annotation is commonly seen in eCommerce sites.

### 3 THEME AND PURPOSE

The primary theme is eCommerce search and recommendation. The purpose of the workshop is to provide a venue for publication and discussion of Information Retrieval research and ideas as they pertain to products and eCommerce. We will be bringing together practitioners and researchers from academia and industry to discuss the challenges and approaches to search and recommendation.

The goal is to foster collaboration and discussion with the broader IR community. We are happy to state that we have an agenda to raise awareness within the academic community of the problems faced by this domain.

#### 3.1 Scope

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

- Machine learning techniques for eCommerce applications
- Semantic representation for users, products, services & queries
- Structured data and faceted search
- Query intent, suggestion, and auto-completion
- Temporal dynamics for Search and Recommendation
- Models for relevance and ranking for multi-faceted entities
- Deterministic sorting of results lists (e.g. price low to high)
- Click models for eCommerce domain
- Session aware, and session oriented search
- Knowledge graphs and ontologies for eCommerce
- Personalization & contextualization
- Indexing and search in a rapidly changing environment
- Diversity in product search and recommendations
- Strategies for resolving extremely low (or no) recall queries

- The use of external features such as reviews and ratings
- User interfaces
- Reviews and sentiment analysis
- The use of social signals in ranking and beyond
- The balance between revenue and relevance
- Trust and security
- Live experimentation
- Questions and answering, chat bots for eCommerce
- Cross-Lingual search and machine translation
- Resources and data sets

#### 3.2 Data Challenge

The ECOM18 workshop and Rakuten run the data challenge “Taxonomy Classification for eCommerce-scale Product Catalogs” in which participants addressed the problem of taking a product listing and choosing which category (from a taxonomy) that listing should belong. For ECOM19 we are working with eBay to secure a release of item data as well as queries and assessments for a search task. In particular we are examining relevance of for deterministic sort orders (e.g. price low to high). The task is to take the query, search the document collection, and to rank the results on price. This might appear to be a trivial problem, but it raises questions on assessment (how can we know that the assessments contain the lowest price relevant item?). It raises questions on evaluation (how should we score a run missing the lowest price item?). It raises questions on query semantics (does a user searching for “apple” want an apple or an Apple?). It raises questions about relevance (e.g. is a smartphone case relevant for a smartphone query?). Relevance (or otherwise) of these accessory items is especially problematic for deterministic sort orders as such items are often inexpensive and consequently can appear at the top of the SERP. Indeed, publications at ECOM18 discussed some of these problems – and we will put some of that work into practice.

### 4 WORKSHOP FORMAT

This workshop is a full day workshop. It will start with an invited talk from a well respected practitioner (or academic) who is tackling eCommerce search problems. Next will be presentation of selected papers. We will then break-out into small groups to identify key areas for future research in eCommerce search – with special emphasis on future data challenges. After lunch there will be a second invited talk, then a poster session. We will finish with a panel session. The workshop schedule and activities are structured to contain substantial time for discussion and engagement by all participants.

### 5 WORKSHOP OUTCOMES

The most important outcome of the workshop is the discussion between delegates. It is these discussions that lead to collaboration and future research – unarguably the ultimate goal of any workshop. We will capture what we can in the form of a SIGIR Forum workshop report.

An additional goal is to raise awareness of the fascinating problems that litter the eCommerce search battlefield. We hope that through discussion at the workshop, and SIGIR 2019, we can help steer the research community towards these problems and in doing so find solutions to some very difficult problems.

<sup>3</sup><https://sites.google.com/view/sigir2018-info-ec/home>

<sup>4</sup><https://sites.google.com/view/ecnlp>

<sup>5</sup><https://sigir-ecom.github.io/data-task.html>

<sup>6</sup><https://competitions.codalab.org/competitions/11161>