

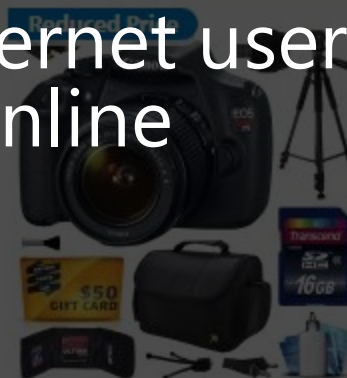
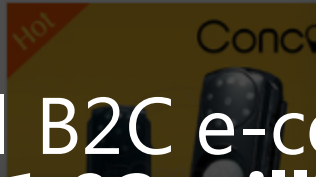
Personalized E-Commerce Search

CIKM Cup 2016 Track 2

Nik Spirin, **CIKM Cup 2016 Co-Chair**
University of Illinois at Urbana-Champaign



DIGINETICA
Retail Technology Company



- **40 percent** of worldwide internet users have bought products or goods online
- **Over 1 billion** online buyers
- Global B2C e-commerce sales are expected to reach **1.92 trillion** U.S. dollars by 2016

Problem description

- Input:
 - anonymized+obfuscated search and browsing logs
 - anonymized transactions
 - product catalog
 - a subset of queries (query-less / query-full)
- Output:
 - Predict relevance labels and re-rank products returned by the non-personalized ranker
- Evaluation metric:
 - NDCG

Example data: train-queries.csv

- queryId (serial)
- sessionId (serial)
- userId (serial)
- timestart (time since the first query in a session, in milliseconds)
- duration (page dwell time, in milliseconds)
- eventdate (calendar date)
- searchstring.tokens (comma separated hashed query tokens; empty if it is a query-less case)
- categoryId (product category ID; empty if it is a query-full session)
- items (productIDs returned by the default ranking algorithm on the SERP; this IDs must be re-ranked).
- is.test (TRUE/FALSE; TRUE if it is a test query)
- regionId (geographical region of a query; serial).

Example data: products.csv

- productID (serial)
- priceLog2 (log-transformed product price)
- product.name.tokens (comma separated hashed product name tokens)
- imageName (name of the corresponding product image)

Example data: product-categories.csv

- productCategoryID (serial)
- productID
- categoryID

Example data: train-purchases.csv

- sessionId (serial)
- timeframe (time since the first query in a session, in milliseconds)
- eventdate (calendar date)
- ordernumber (serial product orderID; groups all products purchased together ~ shopping cart; if a user bought several products, there are several records sharing the same ordernumber)
- itemId (purchased product)

Example data: train-item-views.csv

- sessionId
- userId
- itemId
- timeframe (time since the first query in a session, in milliseconds)
- eventdate (calendar date)

Dataset statistics

- The number of sessions: 573,935
- The number of products: 134,319,529
- The number of products viewed from search (including browsing after SERP): 2,451,565
- The average number of products viewed per search session (including browsing after SERP): 4.271
- The number of SERP clicks on products: 1,877,542
- The average number of SERP click per search session: 3.271
- The number of products purchased from search: 68,818
- The average number of products purchased from search session: 0.119

Relevance labels

- **0 (irrelevant) grade** corresponds to the products with no SERP clicks.
- **1 (somewhat relevant) grade** corresponds to the products, which were shown on the SERP and clicked by the user.
- **2 (relevant) grade** corresponds to the products, which were shown on the SERP, clicked by the user, and purchased.

Results

Previous			► Current		End
Phase 1: Validation Leaderboard			Phase 2: Testing Leaderboard		Competition Ends
Aug. 5, 2016, midnight UTC			Oct. 2, 2016, midnight UTC		Oct. 5, 2020, midnight UTC
	User	Team Name	FinalNDCG (weighted average)	SearchNDCG (query-full; textual queries)	CategoryNDCG (query-less; category facets)
1	minerva	Ali-Search	0.4262 (1)	0.5574 (1)	0.3935 (1)
2	Dmitrii_Nikitko		0.4149 (2)	0.5301 (2)	0.3861 (3)
3	tjy	red fruit yard	0.4015 (3)	0.4364 (5)	0.3928 (2)
4	wistuba		0.3769 (4)	0.4495 (4)	0.3588 (4)
5	joaopalotti		0.3712 (5)	0.4860 (3)	0.3425 (5)

- 88 registered participants
- Over 600 submissions in 2 months by top-5 teams
- The best result is 14.2% better than the baseline

Contacts

Nik Spirin

Email: spirinus@gmail.com

Skype: @spirinus

Twitter: @spirinus

Facebook: @spirinus

Instagram: @_spirinus_