User Behavior Insights

Originally titled Your Search Engine Needs a Memory

Eric Pugh Stavros Macrakis April 2024 - OpenSource Connections

- AWS OpenSearch



Track clicks using the Generalizable Signals Format #68



0

epugh opened this issue on Sep 29, 2021 · 3 comments



epugh commented on Sep 29, 2021

recommended by Trey, Max, Doug in Al Powered search ;-)

Membe

Haystack 2019 (External) >>



Stavros Macrakis (Σταῦρος Μακράκης) macrakis@alum.mit.edu via opensourceconnections.com to hello -

Mon, Aug 6, 2018, 4:04 PM

Any plans for a 2019 edition?

Thanks,

-s ← Reply ← Reply all → Forward

Search needs data





To determine what a "good" result is

Databases– unambiguous criteria

- Pure Boolean search
- SQL execution

Search – inherently fuzzy

- Usually too many results to show
- Precision are the results good?
- Recall did we show the good results?
 What is "good"?
- Show best results first
 - What is "best"?



To optimize for corpus and user

- Document search and recommendations (heavy on text)
 - Academic documents
 - News
 - Intranet, highly heterogenous, inconsistent formats
 - Knowledge management
 - Legal and regulatory, both high-precision (relevant regulation) and high-recall (discovery)
 - FAQs, call centers, troubleshooting
- Looking for items, not documents (heavy on structured data)
 - E-commerce search and recommendations: looking for a thing to buy
 - Anything from plumbing parts to real estate
 - Media movies, TV, ...
 - Job search; placements
 - Expert search
 - Hotel, restaurant, flight search, fractional jets real-time availability
 - Ride share real-time position
 - Photos by description or by similarity; SmugMug very different from ShutterStock
- Embedded
 - Company search within a financial services app
 - Person search within email
- e 2023, Amazon Web Services, Inc. or its affiliates. All rights reserved. Amazon Confidential and Trademark

Unifying concept

Relevance ranking

Search

Relevance ranking

User Behavior Insights

Goal: Collect all information needed for evaluating search

- What did the user search for?
- How did they use typeahead?
- How did they refine the search (facets)?
- What did they click on?
 - ... other actions... view details, add to basket, checkout basket



What should all search implementations do, but don't?

Berlin Buzzwords 2022 survey

- Log search actions
- Understand application data
- Evaluate quality of results
- Talk to users and sponsors
- Regression testing
- Performance testing
- Multilingual coverage
- Privacy

Customer gaps

- Search specialist engineers
 - Search is a *discipline*
- Logging
- Quality evaluation
- Incorporating multiple signals
- Tuning

The virtuous circle of search relevance

Search processing

- Pipeline processor
 - Modular query processing
 - Quergy rewriter
 - Modular result processing
 - Relevance integrations
 - Personalization integrations
- Integrated semantic search
- Combining semantic and lexical (BM25) ۲
- Sparse neural •

Search tuning

- Search configurations
- **Regression evaluation**
- Manual tuning, Bayesian optimization
- Semantic model fine-tuning
 - Learning to Rank

Measurement and analysis

- Online evaluation tools
- A/B testing
- Creation of judgement sets
- Offline evaluation tools
- Metrics

aws

Choosing search processing

- Lexical search
- Semantic vector search
- Customization, personalization
- Neural sparse retrieval (similar to SPLADE)
- Image
- Hybrid search: lexical, semantic, image



Measurement and analysis

User behavior insights

- Recording end-to-end search behavior
- Analysis dashboards



Search tuning

- Manual tuning
- Bayesian/grid parameter optimization
- Semantic model fine-tuning
- Combining neural and lexical rankings
- Learning to Rank



User Behavior Insights

Applications

- Dashboarding of KPIs
- Investigation by relevance engineers and data scientists
- Debugging
- Tuning of typeahead
- Real-time feedback to relevance calculation
- Input to machine learning models (Personalize, Learning to Rank, ...)



The User Behavior Insights project

- Discussion at Haystack 2022
- RFC 4619 (9/2022); RFC 12084 (1/2024, updated 4/2024)
- Community debated names
- All work done in the open on github
- Apache 2.0 license





We need your input! Just a few questions I promise!

Why am I excited about UBI?



Relevancy Framework for Search



What makes up UBI?

1. Code for collecting specific search related events by users (often JavaScript)

2. A shared schema for

- a. defining what the user is asking for (Query)
- b. Defining what documents/objects/records/answers the user is receiving back.
- c. Defining how we track and organize the resulting Events and make the causal connections between the Query and the follow on actions (click through, add to cart, etc)

3. An implementation in OpenSearch to facilitate this process

What makes up UBI?

Code for collecting energific energy related events by years (often laveScript)

- 2. A shared schema for
 - a. defining what the user is asking for (Query)
 - Defining what documents/objects/records/answers the user is receiving back.
 - Defining how we track and organize the resulting Events and make the causal connections between the Query and the follow on actions (click through, add to cart, etc)

3. An implementation in OpenSearch to facilitate this process





Show don't tell.....

```
GET _search
{
    "ubi": {
        "user_query": "the wind rises",
     },
     "query": {
        "query": {
            "query_string": {
               "query": "the wind AND (rises OR rising)"
        }
     }
}
```

Enabling UBI logging per query

Passing in our own explicit query_id.

```
POST /ubi events/doc
  "query id": "USER52-SESSION2-QUERY3",
  "application": "chorus",
  "action name": "add to cart",
  "user id": "1",
  "event attributes": {
   "object": {
     "object id": "RESULT C",
     . . . .
   "position": {...},
   "data": {...},
```

Events logging is just standard OS document to _index.

https://github.com/o19s/opensearch-ubi/blob/2.14.0/documentation/schemas.md#2ubi-events HOW STANDARDS PROLIFERATE: (SEE: A/C CHARGERS, CHARACTER ENCODINGS, INSTANT MESSAGING, ETC.)

SITUATION: THERE ARE 14 COMPETING STANDARDS.

14?! RIDICULOUS! WE NEED TO DEVELOP ONE UNIVERSAL STANDARD THAT COVERS EVERYONE'S USE CASES. YEAH!



SITUATION: THERE ARE 15 COMPETING STANDARDS.

https://xkcd.com/927/

HOW STANDARDS PROLIFERATE: (SEE: A/C CHARGERS, CHARACTER ENCODINGS, INSTANT MESSAGING, ETC.)

SITUATION: THERE ARE O COMPETING STANDARDS.

14?! RIDICULOUS! WE NEED TO DEVELOP ONE UNIVERSAL STANDARD THAT COVERS EVERYONE'S USE CASES. YEAH!

SOON:

SITUATION: THERE ARE 1 Cooperating STANDARDS.

https://xkcd.com/927/

Learn More



OpenSearch RFC for this work:

https://github.com/opensearch-project/OpenSearch/issues/12084

Docs for current code:

https://github.com/o19s/opensearch-ubi/blob/2.14.0/documentation/documentation/ .md

Schema:

https://github.com/o19s/opensearch-ubi/blob/2.14.0/documentation/schemas.md

Chorus w/ UBI: https://github.com/o19s/chorus-opensearch-edition/

Online Demo: <u>http://chorus-opensearch-edition.dev.o19s.com:4000/</u> and <u>http://chorus-opensearch-edition.dev.o19s.com:5601</u>



Track clicks using the Generalizable Signals Format



epugh opened this issue on Sep 29, 2021 · 3 comments

THANK YOU

ticket!



epugh commented on Sep 29, 2021

recommended by Trey, Max, Doug in AI Powered search ;-)

