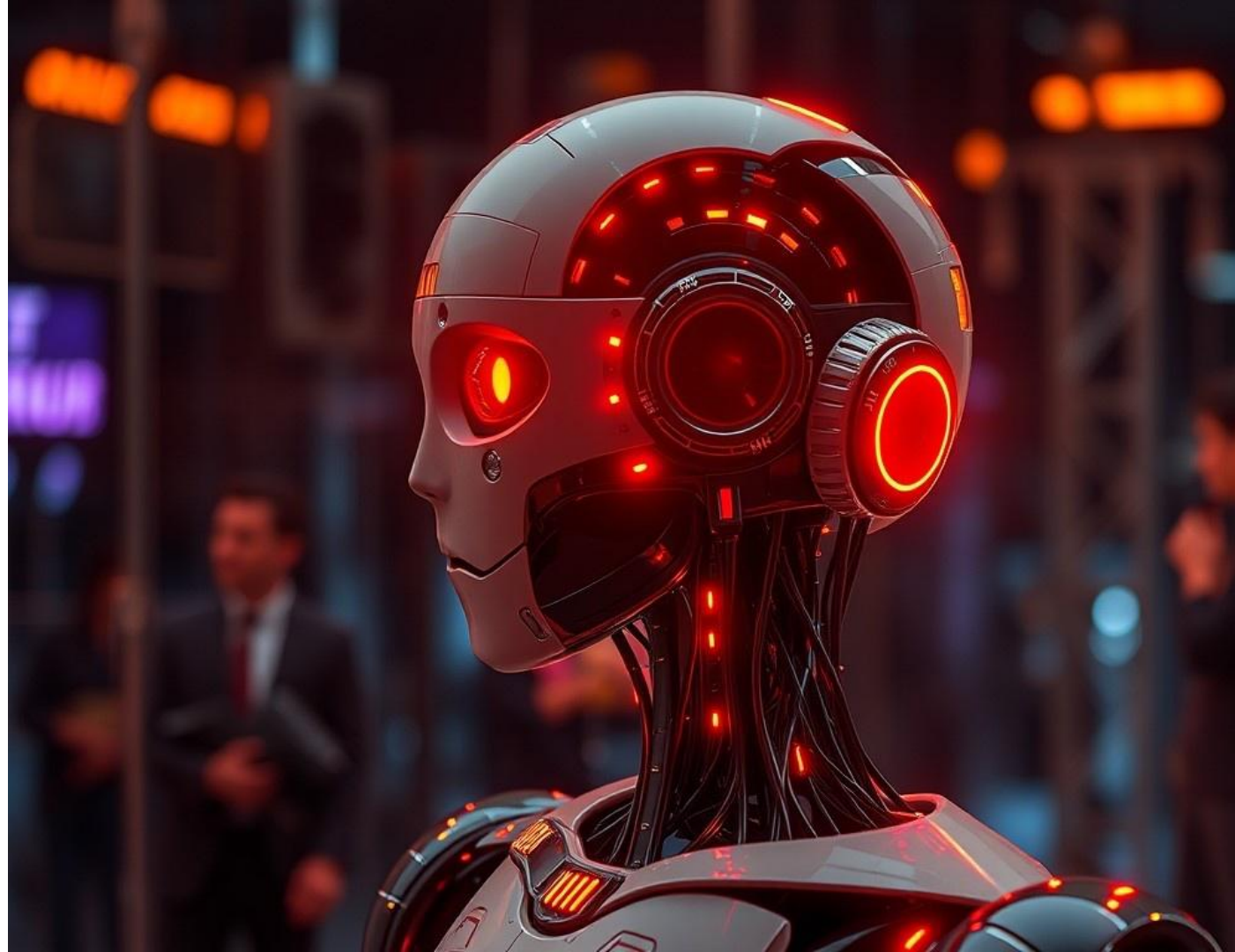


Solr's Place in an AI World



What spurred me on to give this talk...

This may get a bit personal.....

So how did we get
here?

Back in the dawn of time, when boolean ruled the planet!

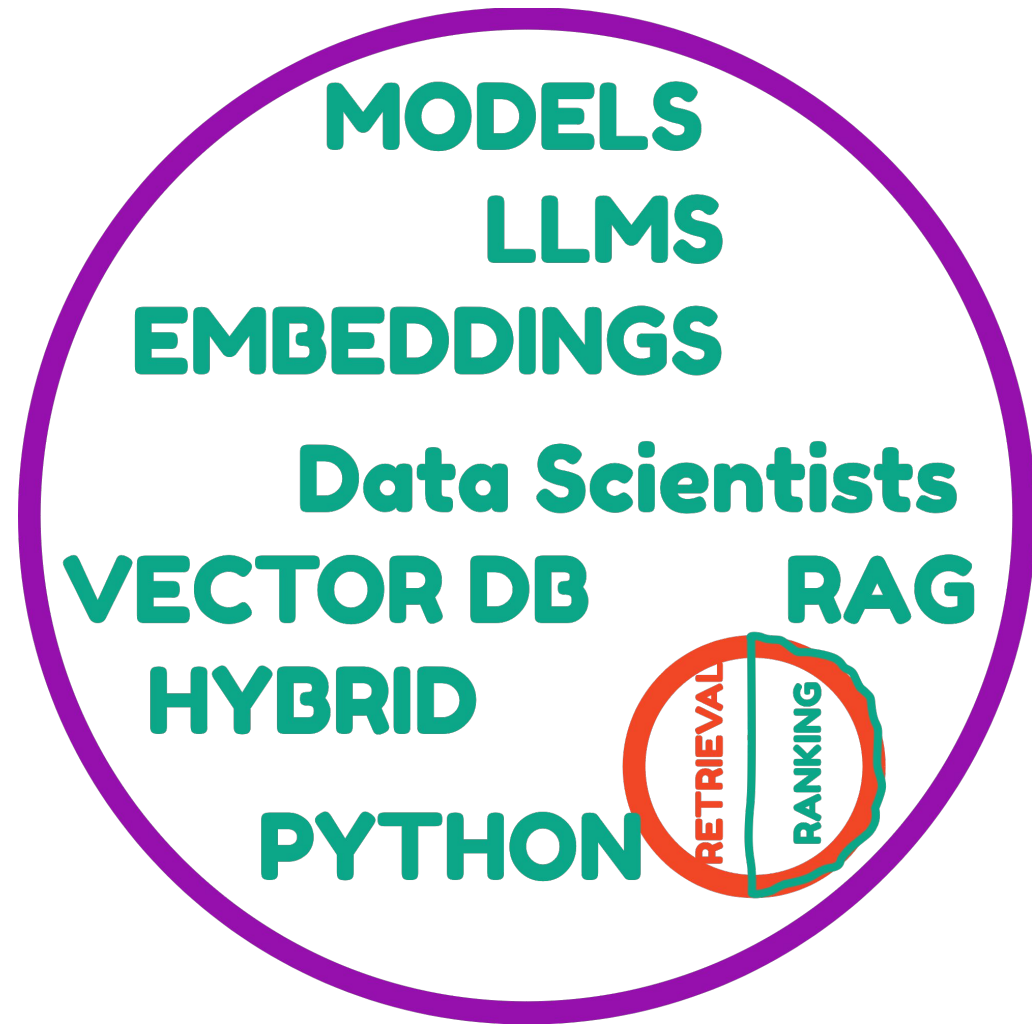


Then 10 blue links on a web page wasn't enough.....



Meanwhile.....

Everything
changed.....



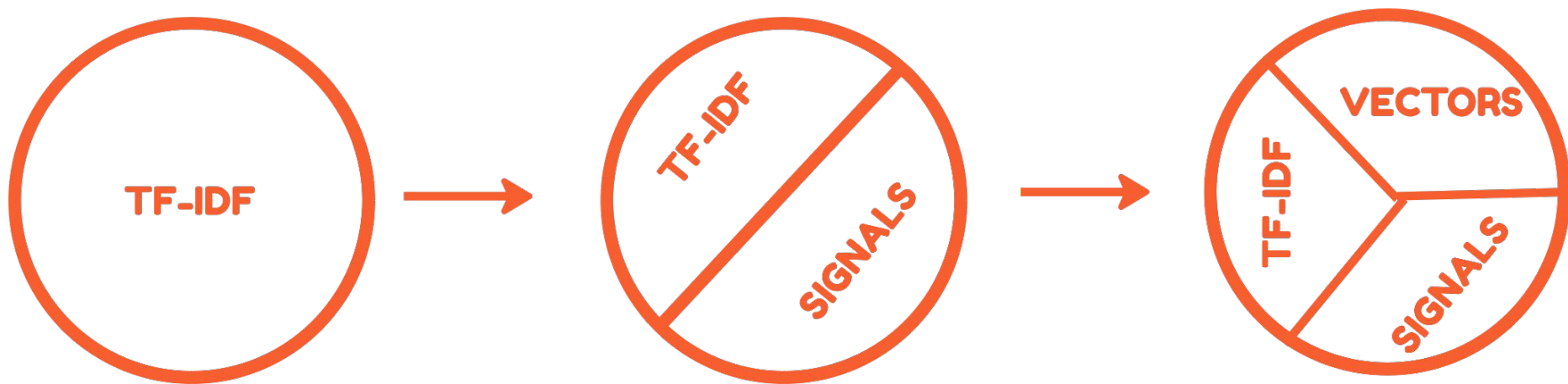
What is the scope of
Solr?

We as a community really focus on....



TF-IDF

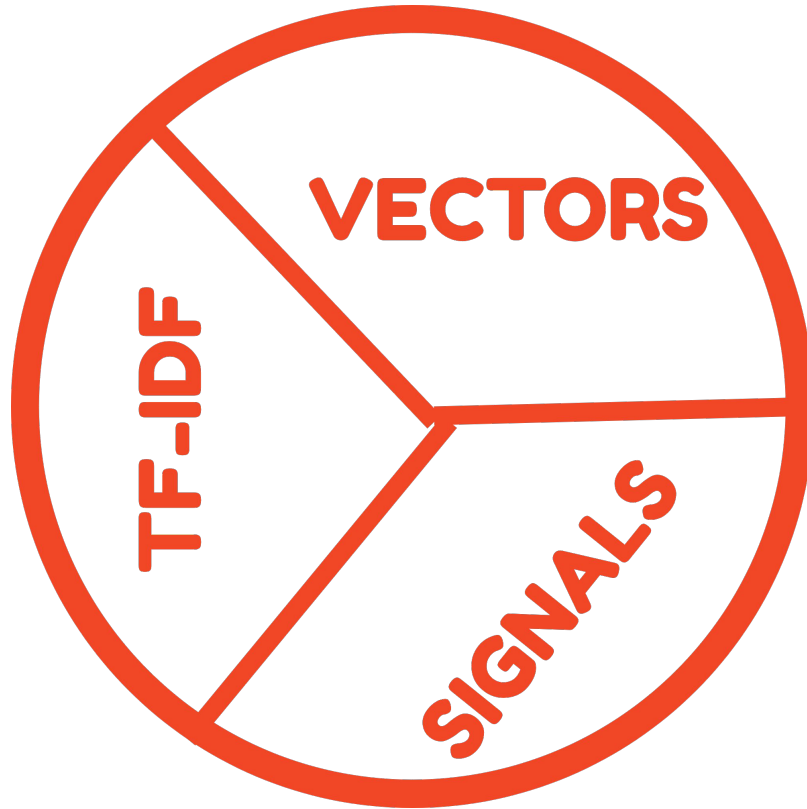
This is what a modern algorithm needs to do “the magic”



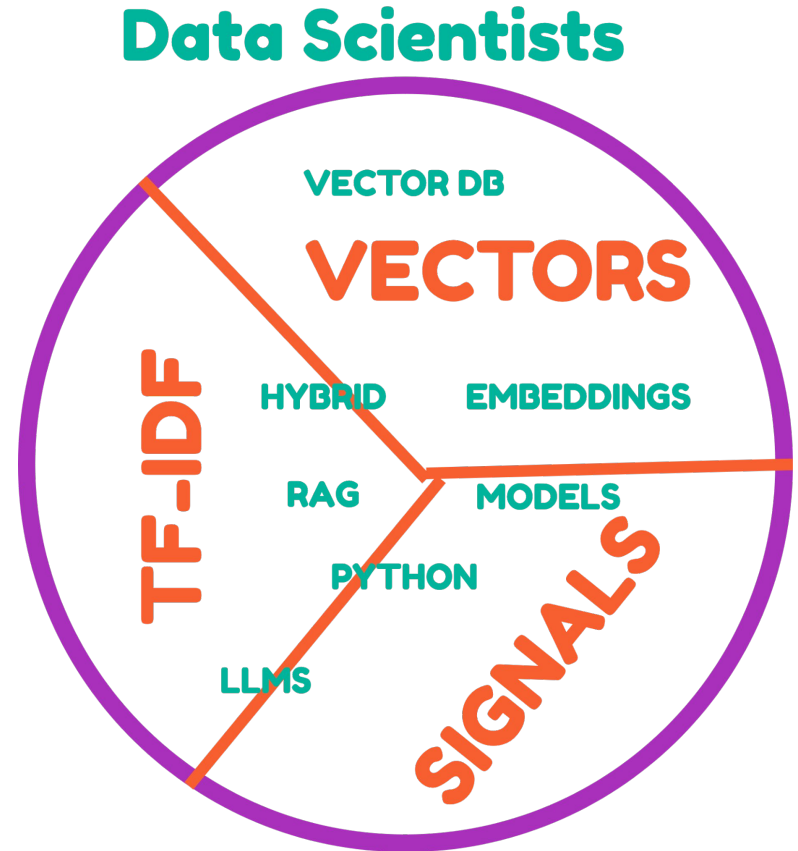
Maybe it's time to
thank TF/IDF for it's
service.....?

And dethrone it from
it's centrality in our
lives...

And think about focusing on what makes Solr the **best** platform for:

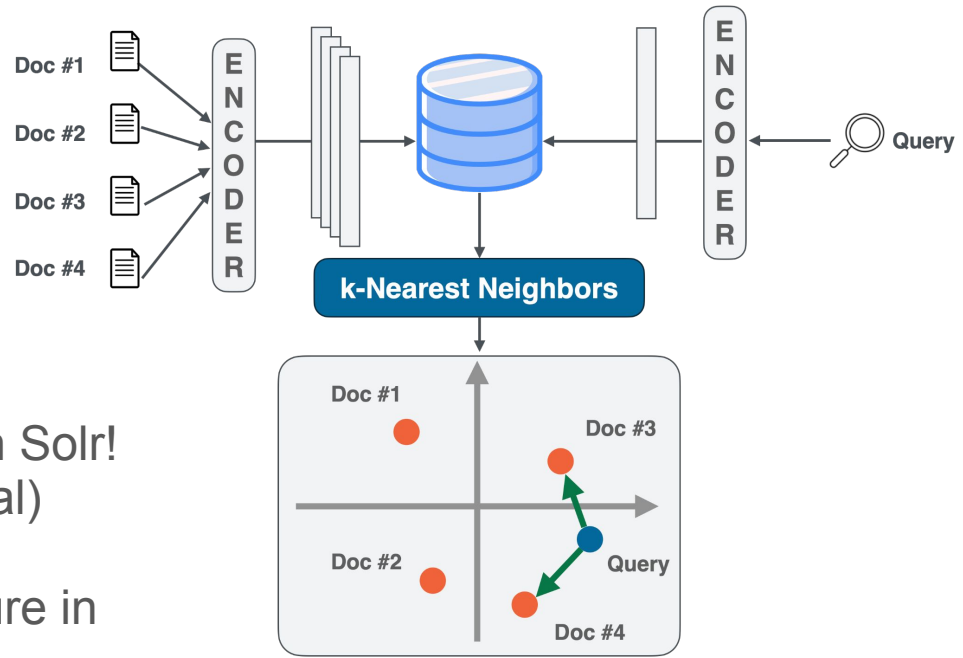


That attracts folks who care about:



We aren't starting from
zero....

KNN.. Vector Story...



- Lucene introduced native HNWS
- Idea was to introduce in Solr both
A field type and Query Parser
-> huge enabler for semantic search on Solr!
- Support for pre/post filtering (also lexical)
- Support for high dimensional vectors
- Vector similarity can be used as a feature in Learning To Rank
- [In Progress] Development of a Embedding Model resource store + Update Request Processor + Query parser to automatically infer

```
q={!knn f=vector topK=10}[1.0, 2.0, 3.0,  
4.0]
```

```
<fieldType name="knn_vector"  
class="solr.DenseVectorField"  
vectorDimension="4"  
similarityFunction="cosine"/>  
<field name="vector" type="knn_vector"  
indexed="true" stored="true"/>
```


Hybrid Search...

- **Combine traditional** keyword-based (lexical) search with vector-based (neural) search
- **Retrieval** of two sets of candidates:
 - one set of results coming from lexical matches with the query keywords
 - a set of results coming from the K-Nearest Neighbours search with the query vector
 - Union/Intersection of result sets
- **Ranking** of the candidates
 - Learning To Rank
 - Sum/Multiplication of normalised scores
- Reciprocal Rank Fusion
 - <https://github.com/apache/solr/pull/2489> work in progress - paused, waiting for funding/time

LLMS!


- Vocabulary mismatch problem -> LLMs are great in solving it
- From natural language -> structured Solr query
- We want to identify the ebay field for each entity in the query and use terms in the index that are close semantically
- Given this JSON structure (Solr document) ask the LLM to model the natural language query in the same way
- We need to match the index -> from the Solr index collect the term dictionaries for each field
- TO DO: Idea to build a new Query Parser that does this internally
- Some ref:

<https://sease.io/2024/08/from-natural-language-to-structured-solr-queries-using-llms.html>



Model Based Enrichment

```
curl --data-binary vocab.txt -X PUT  
http://localhost:8983/api/cluster/files/models/sentiment/ vocab.txt
```

apache/solr

#1999 **SOLR-17023: Use Modern NLP Models via ONNX and Apache...** 

22 comments 23 reviews 12 files +120 -4

 epugh · October 10, 2023 · 30 commits 

 github.com

SOLR-17023: Use Modern NLP Models via ONNX and Apache OpenNLP with Solr by epugh · Pull Request #1999 · apache/solr


<https://issues.apache.org/jira/browse/SOLR-17023> Description Code and BATS test that demonstrates downloading a sentiment classification model from Huggingface, converting it to Onnx model, uploadi...

Model Based Enrichment

```
curl --data-binary vocab.txt -X PUT  
http://localhost:8983/api/cluster/files/models/sentiment/vocab.txt
```

```
curl --data-binary model.txt -X PUT  
http://localhost:8983/api/cluster/files/models/sentiment/ model.onnx
```

apache/solr

#1999 **SOLR-17023: Use Modern NLP Models via ONNX and Apache...** 

22 comments 23 reviews 12 files +120 -4

epugh • October 10, 2023 30 commits

github.com

SOLR-17023: Use Modern NLP Models via ONNX and Apache OpenNLP with Solr by epugh · Pull Request #1999 · apache/solr

<https://issues.apache.org/jira/browse/SOLR-17023> Description Code and BATS test that demonstrates downloading a sentiment classification model from Huggingface, converting it to Onnx model, uploadi...

Model Based Enrichment

```
curl --data-binary vocab.txt -X PUT
http://localhost:8983/api/cluster/files/models/sentiment/vocab.txt
```

```
curl --data-binary vocab.txt -X PUT
http://localhost:8983/api/cluster/files/models/sentiment/model.onnx
```

```
curl -X POST -H 'Content-type:application/json' -d '{
  "add-updateprocessor": {
    "name": "sentimentClassifier",
    "class":
"solr.processor.DocumentCategorizerUpdateProcessorFactory",
    "modelFile": "models/sentiment/model.onnx",
    "vocabFile": "models/sentiment/vocab.txt",
    "source": "name",
    "dest": "name_sentiment"
  }
}' "http://localhost:8983/solr/techproducts/config"
```

apache/solr

#1999 **SOLR-17023: Use Modern NLP Models via ONNX and Apache...**



22 comments 23 reviews 12 files +120 -4

epugh · October 10, 2023 · 30 commits

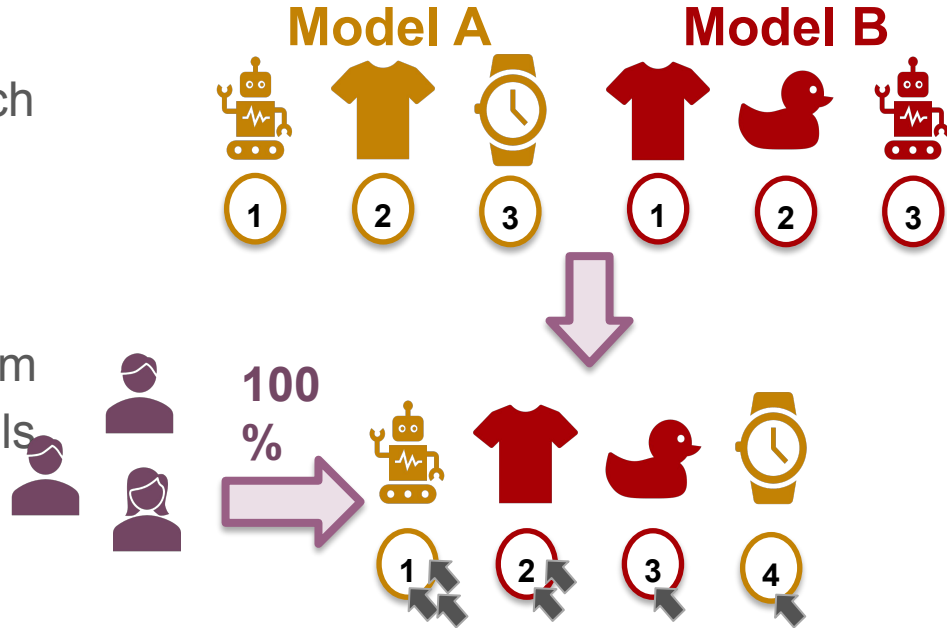
github.com

SOLR-17023: Use Modern NLP Models via ONNX and Apache OpenNLP with Solr by epugh · Pull Request #1999 · apache/solr

<https://issues.apache.org/jira/browse/SOLR-17023> Description Code and BATS test that demonstrates downloading a sentiment classification model from Huggingface, converting it to Onnx model, uploadi...

Interleaving Story

- Alternative to A/B testing
- Originated by internal Sease research in the field (<https://arxiv.org/abs/2303.10094>)
- Sponsored and then applied to a client project
- Allows to combine search results from two different Learning To Rank models
- Limitation: doesn't work distributed



`http://localhost:8983/solr/techproducts/query?q=test&rq={!ltr
model=myModelA model=myModelB reRankDocs=100}&fl=id,score`

User Behavior Insights

- Holistic view of the users search journey
- Standard schema for tracking queries and events for Search
- Calculate Implicit Judgements
- Optimize hybrid search??
- Track interleaving???
- Implementations for OS, ES, and Solr

o19s/**ubi**

User Behavior Insights standard schema



4 Contributors 2 Issues 1 Discussion 13 Stars 4 Forks

 github.com

GitHub - o19s/ubi: User Behavior Insights standard schema

User Behavior Insights standard schema. Contribute to o19s/ubi development by creating an account on GitHub.

User Behavior Insights

```
curl -X POST -d '{
  "query" : "ram OR memory",
  "filter": [
    "inStock:true"
  ],
  "limit":2,
  "params": {
    "ubi": "true",
    "query_id": "c3d22be7-6bdc-4250-91e1",
    "user_query": {
      "query": "RAM memory",
      "experiment": "supersecret",
      "page": 1,
      "filter": "productStatus:available"
    }
  }
}'
"http://localhost:8983/solr/techproducts/query"
```

apache/solr

#2452 User Behavior Insights implementation for Apache Solr



7 comments 0 reviews 11 files +1188 -4

epugh · May 9, 2024 · 35 commits

github.com

User Behavior Insights implementation for Apache Solr by epugh · Pull Request #2452 · apache/solr

Description I am working with other folks, especially Stavros Macrakis (macrakis@gmail.com), to come up with a solution for understanding what users are doing in response to search results. We ha...

Where do we go from
here?

- Neural Search is very different access patterns from Keyword Search --> lets deal with it!
- We need more AI and ML People as Committers --> We [said this in 2023 in Halifax](#), and it hasn't changed.
- We need to stick to Lucene's cutting edge like glue --> Lucene is key to vector story
- Let's hope Java continues to innovate to compete with Rust
- Support new workloads that aren't just TF/IDF

Or... We become just
a retrieval engine.

So let's make sure...



Solr 

The word 'Solr' is written in a dark blue, sans-serif font. To the right of the word is a red circular icon with white diagonal lines radiating from the center, resembling a sun or a fan.