



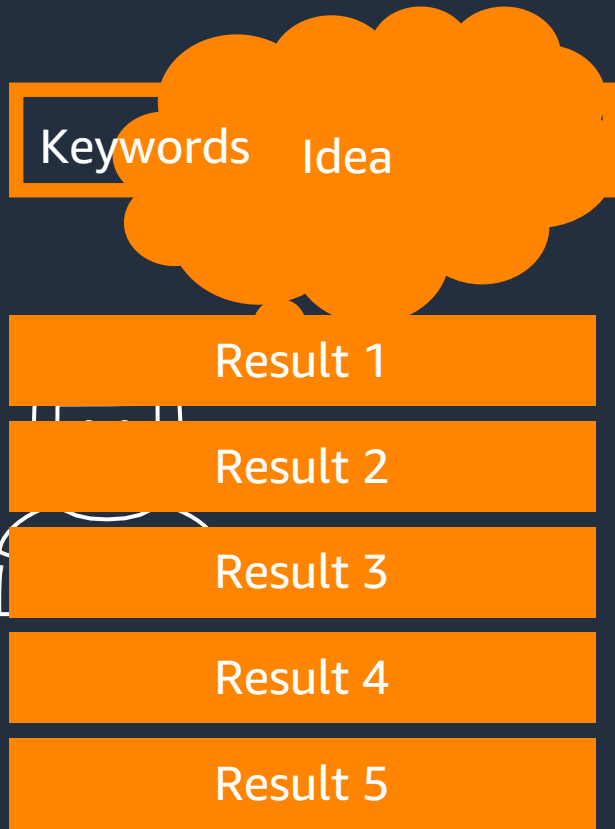
OPERATIONAL ANALYTICS

Analytics in 15

Improve Search Relevance with
Amazon OpenSearch Service

Gene Alpert

Sr. Analytics Specialist
Amazon Web Services



What is the problem?

Keyword quality?

Search algorithm?

Indexing quality?

Thought transcription?

Document quality?

Multiple complementary approaches

Apache Lucene

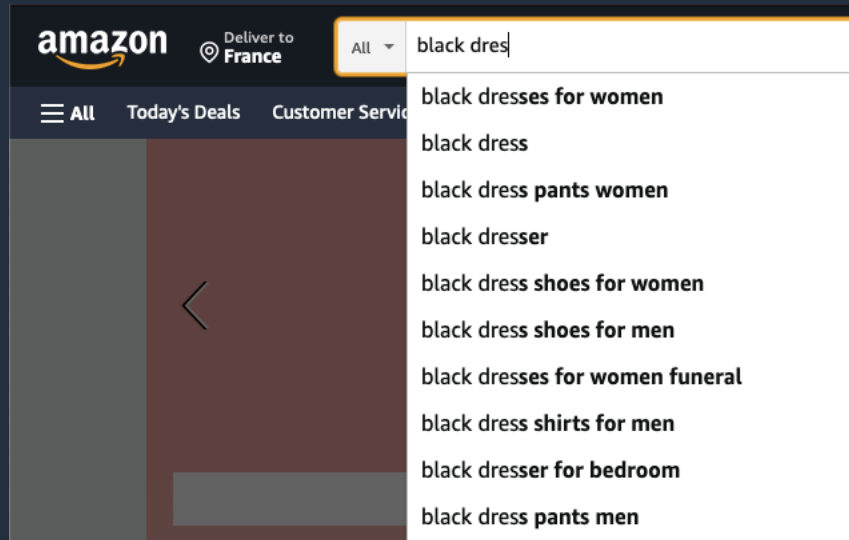
Analyzer:

- Character filter {0..n}
- Tokenizer {1}
- Token filter {0..n}

Default, language specific, domain specific analyzers

Score is unbounded

Suggesters



“type”: “search_as_you_type”

Or

Use n-grams directly, like edge n-grams

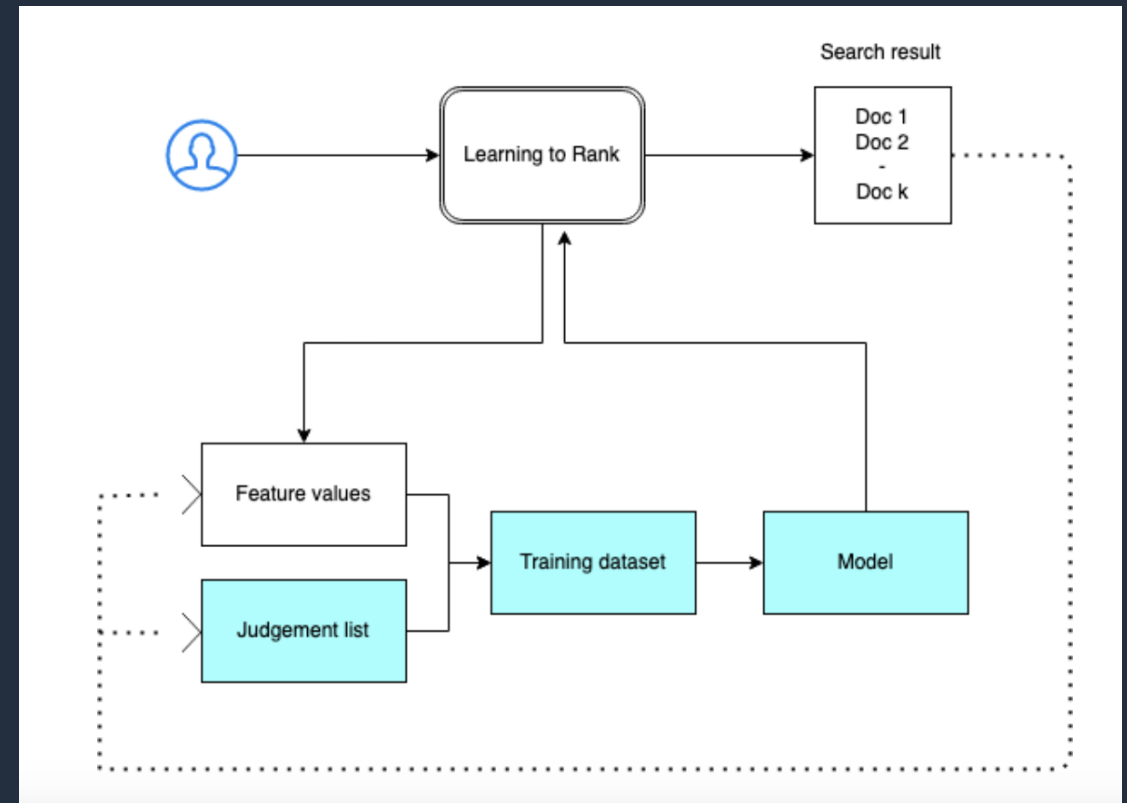
In-memory data structure (very fast)

ML-driven results re-ranking : Leverage Learning to Rank plugin



OpenSearch Service plugin for Learning to Rank

- Leverage ML techniques to re-rank results by manual grading
- Create the Judgement List , unique to catalog / source data
- Find features , this is OpenSearch query score ("sltr" filters)
- Find *feature set* is a set of features
- Update Judgement list
- Generate model with XGboost (SM) or LambdaMART
- Deploy Model to OpenSearch LTR plugin
- Search with LTR model



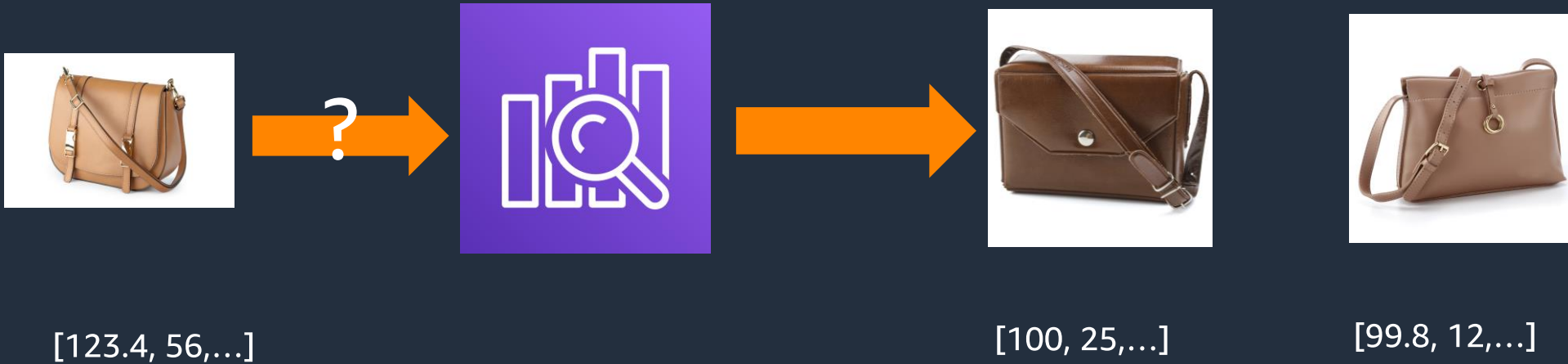
ML-driven Search : Improve Search User Experience



K-NN – Vector Search | Find the Nearest Neighbor

- Based on relationships between vectors and/or documents
- k nearest neighbors (k-NN) is a non-parametric
- KNN is data driven, supervised machine learning model
- KNN is lazy learning since we need to train based changes to data
- KNN for Amazon OpenSearch lets you search for points in a vector space
- “nearest neighbors” for those points by Euclidean distance or cosine similarity.

K – Nearest Neighbors



- Data represented as set of numbers (vectors)
- Euclidean distance or cosine similarity
- Product recommendations, Semantic matching, fraud detection

ML Commons – Bring Your own Model

- OpenSearch now supports BYOM
- Integrates with Open Source ML Commons Plugin
- Support for algorithm:
 - **K-means** | Unsupervised clustering
 - **Linear regression** | maps the linear relationship between inputs and outputs
 - **RCF** | unsupervised anomaly detection
 - **RCFSummarize** | hierarchical clustering technique
 - **Localization** | data exploration or root cause analysis
 - **Logistic Regression** | probability of a discrete outcome given an input variable

Semantic Search User Experience

“Do I need to buy anything extra to use with xbox one s controller?”

Semantic Search

Out of the box with his work for Xbox one?

it shows xbox 360. Does it work for ps3 as well?

Do I need to buy anything extra to used in xbox one s controller?

What adapter do I need for Xbox one

Do I need the Xbox one adapter to use the headset for the Xbox one?

do I need an adapter to use these with xbox 360?

How do i hook it up to xbox one? I have adapter

is this headset compatible with xbox 1S or above?

How do these headphones connect to the Xbox360 controller?

DOes it work with Xbox 360 or do you need the mixamp or chat cord?

Keyword Search

Will this **work** without the adaptor on the second gen **Xbox** one controllers?

Does it **work** for ps4

Does this **work** for a switch ?

Does it **work** for android?

does it **work** on ps4

Does it **work** for ps4?

Does this **work** with ps4

does this **work** with pc

Does this **work** for the LG V10

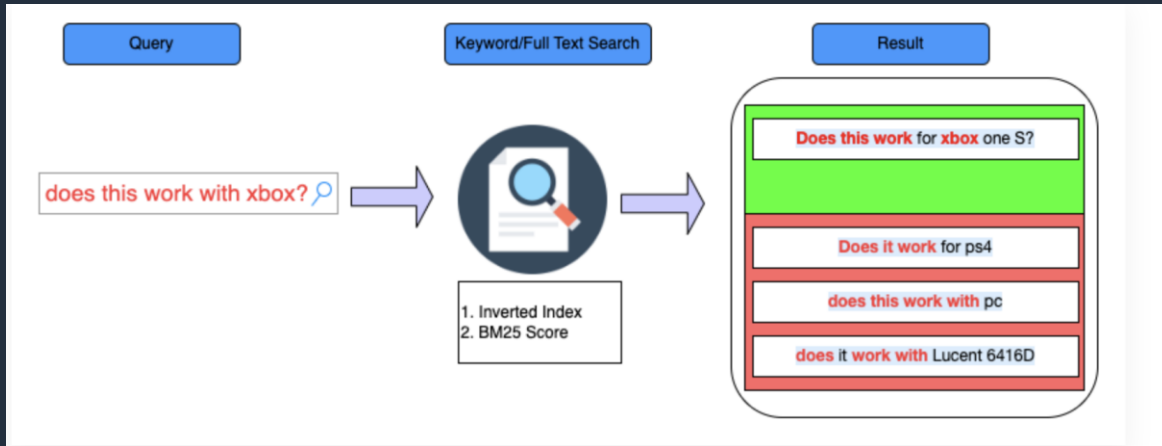
Does it **work** with galaxy s7?

DIY Workshop!



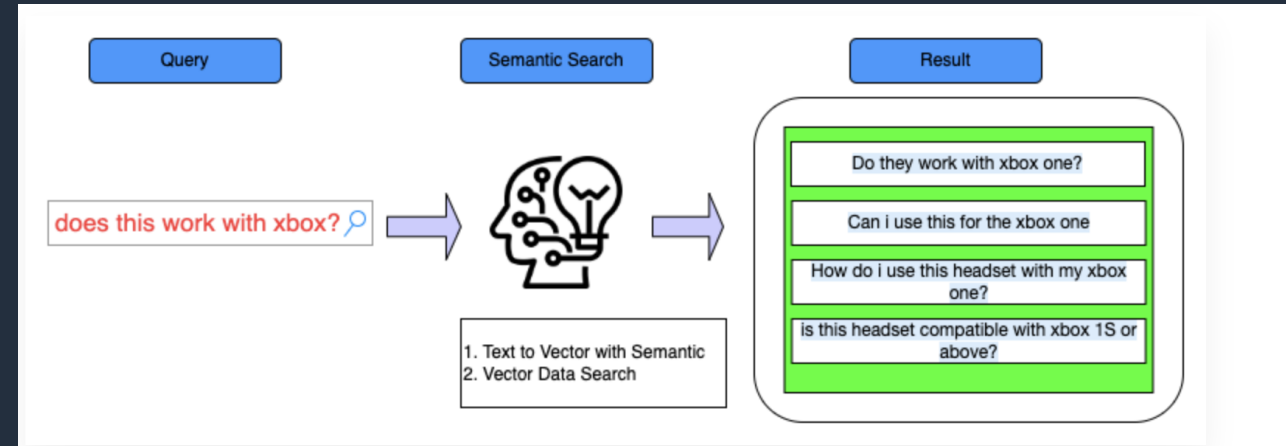
Semantic Search

Keyword Search



- OKAPI BM25 based ranking
- Keyword match
- Does not take context into consideration
- Text Analytics

Semantic Search



- Machine Language based ranking
- Conceptual & Contextual Search
- NLP & Text Data vectorization

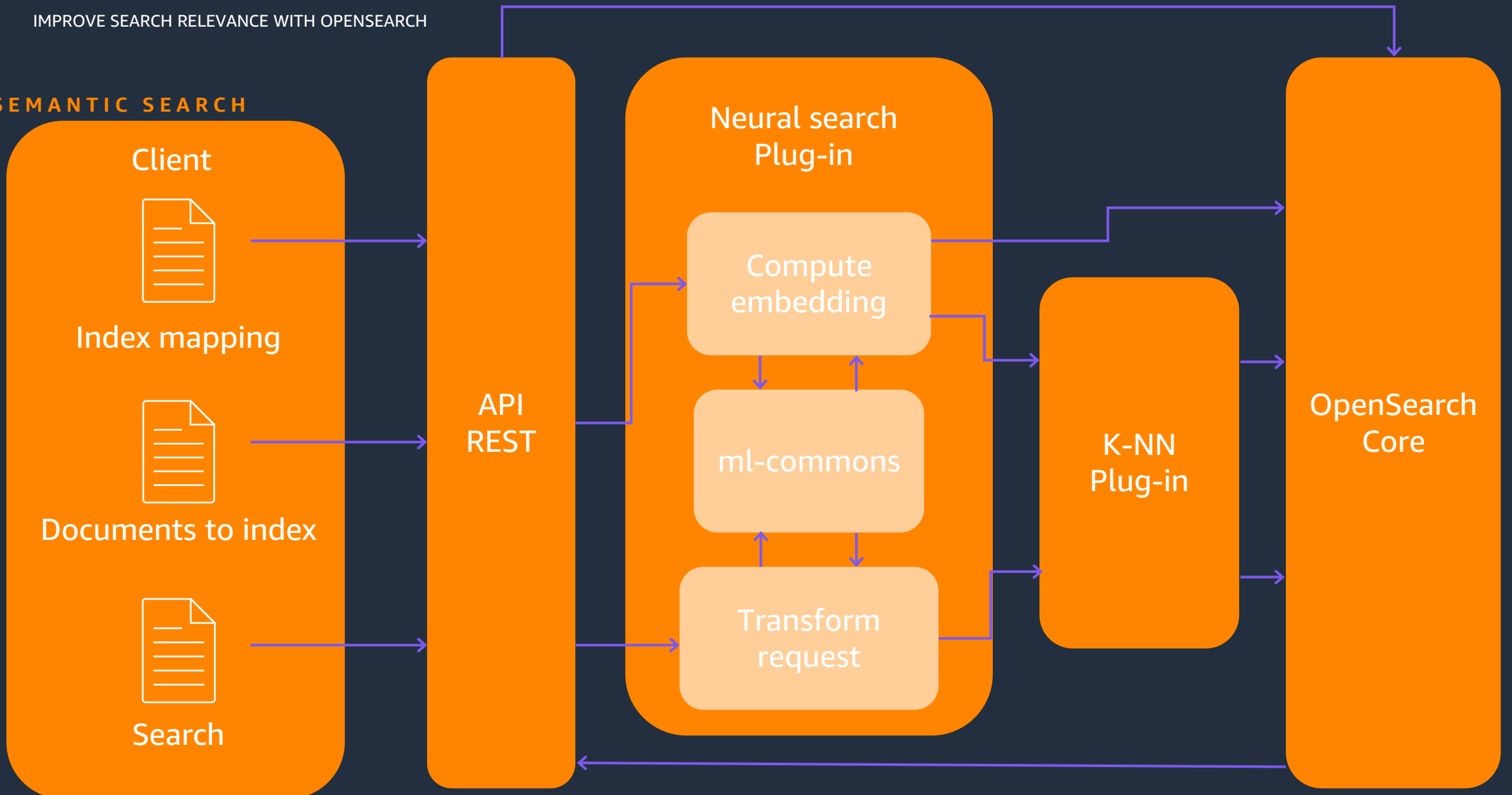
OpenSearch Neural Plugin

- Newly introduced with OpenSearch 2.5
- Still in beta
- Enables ML models into Search workload
- During Ingestion, Neural plugin transforms text to vectors
- During Search, Neural plugin does vector search

Contribute!



SEMANTIC SEARCH





Thank you!

Gene Alpert
Sr. Analytics Specialist
Amazon Web Services



To learn more, please visit:

<https://aws.amazon.com/big-data/datalakes-and-analytics/>