zalando

THE MODERN ARCHITECTURE OF SEARCH

Alaa Elhadba Mikio Braun

zalando

1500+ Brands

20 Million Customers

15 European Countries

Alaa Elhadba

Search Engineer

alaa.elhadba@zalando.de

@alaa_elhadba

Mikio Braun

Delivery Lead

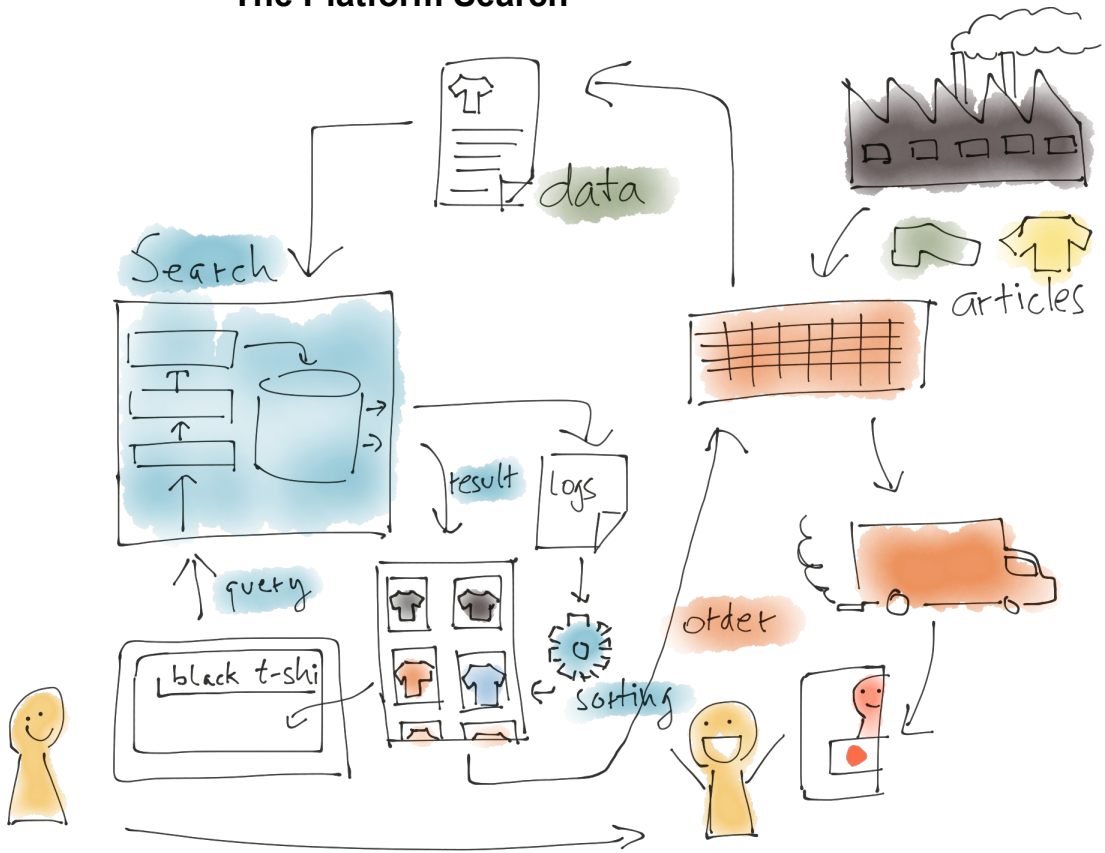
mikio.braun@zalando.de

@mikiobraun

zalando

The Platform Search

The platform search is a gateway offering search as a service for merchants and retailers. Its current focus is integration with its first and main stakeholder the fashion store.



zalando

OLD SKOOL SEARCH

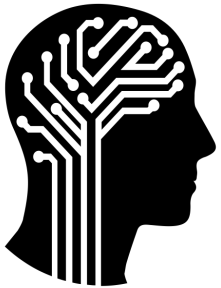
zalando

STRING MATCHING

White adidas performance sneakers

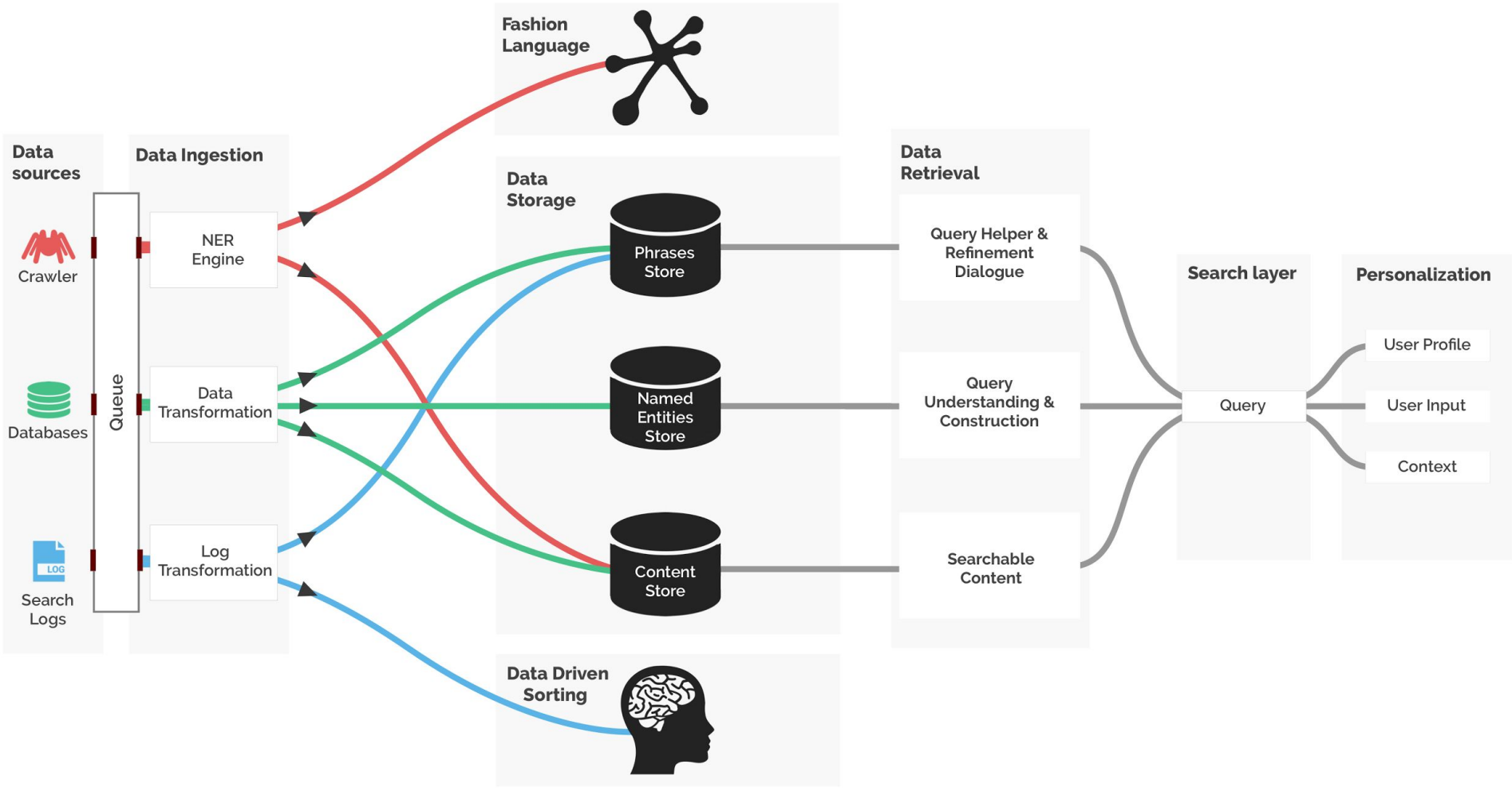
```
{
    "name": "STAN SMITH - Sneaker low - running white/collegiate red",
    "color": "white",
    "silhouette_code": "sneaker",
    "sku": "AD115B01K-A13",
    "url_key": "adidas-originals-stan-smith-sneaker-low-ad115b01k-a13",
    "brand": {
      "name": "adidas Originals",
      "code": "AD1",
      "url": "adidas-originals",
      "image_url": "86ba648f-737c-4b71-b13e-1354db4bbe50.jpg"
    },
    "categories": [
        "Katalog",
        "Damen",
        "Schuhe",
        "Sneaker"
    ],
    "in_stock": true
}
```


zalando



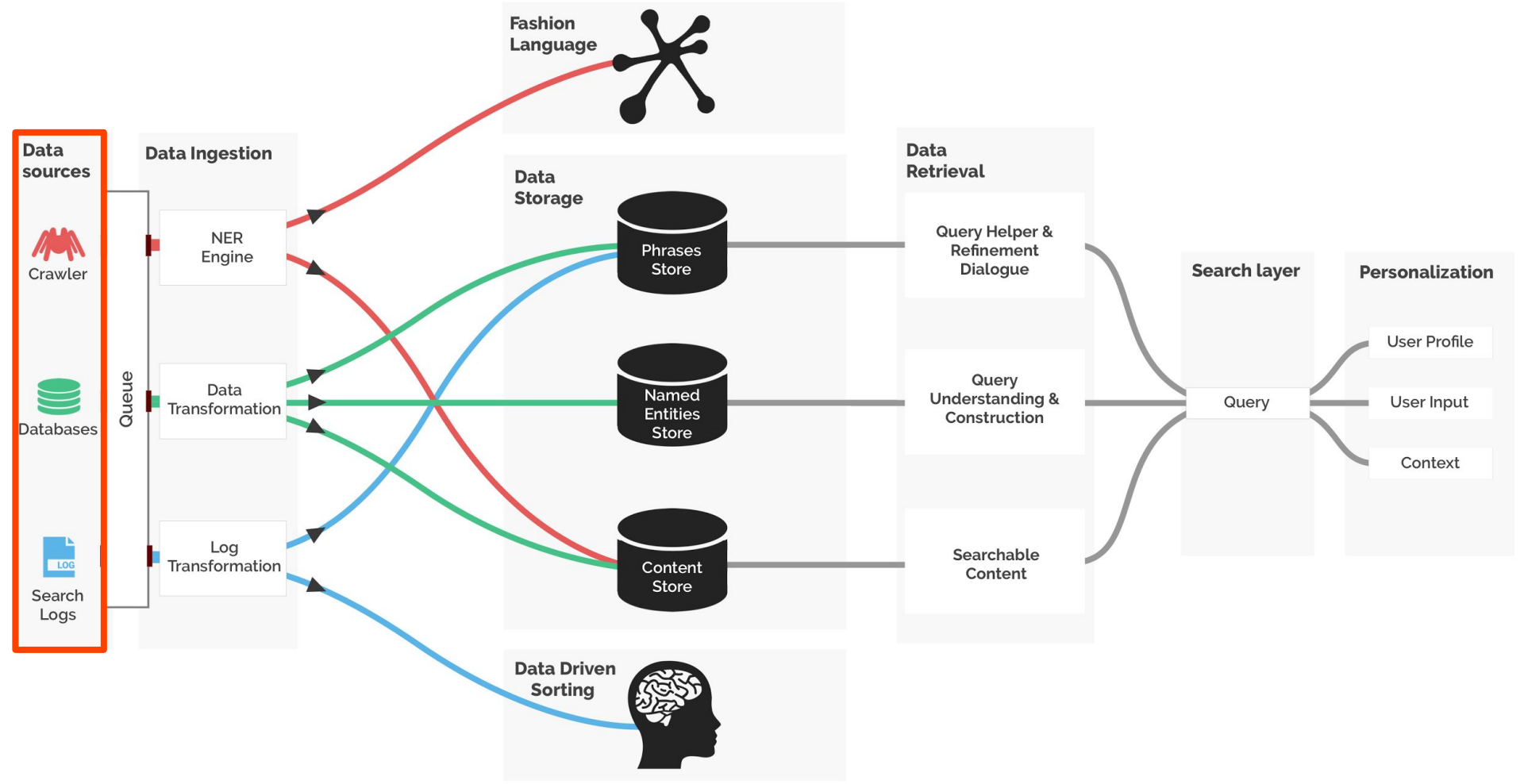
A DATA-DRIVEN WORLD

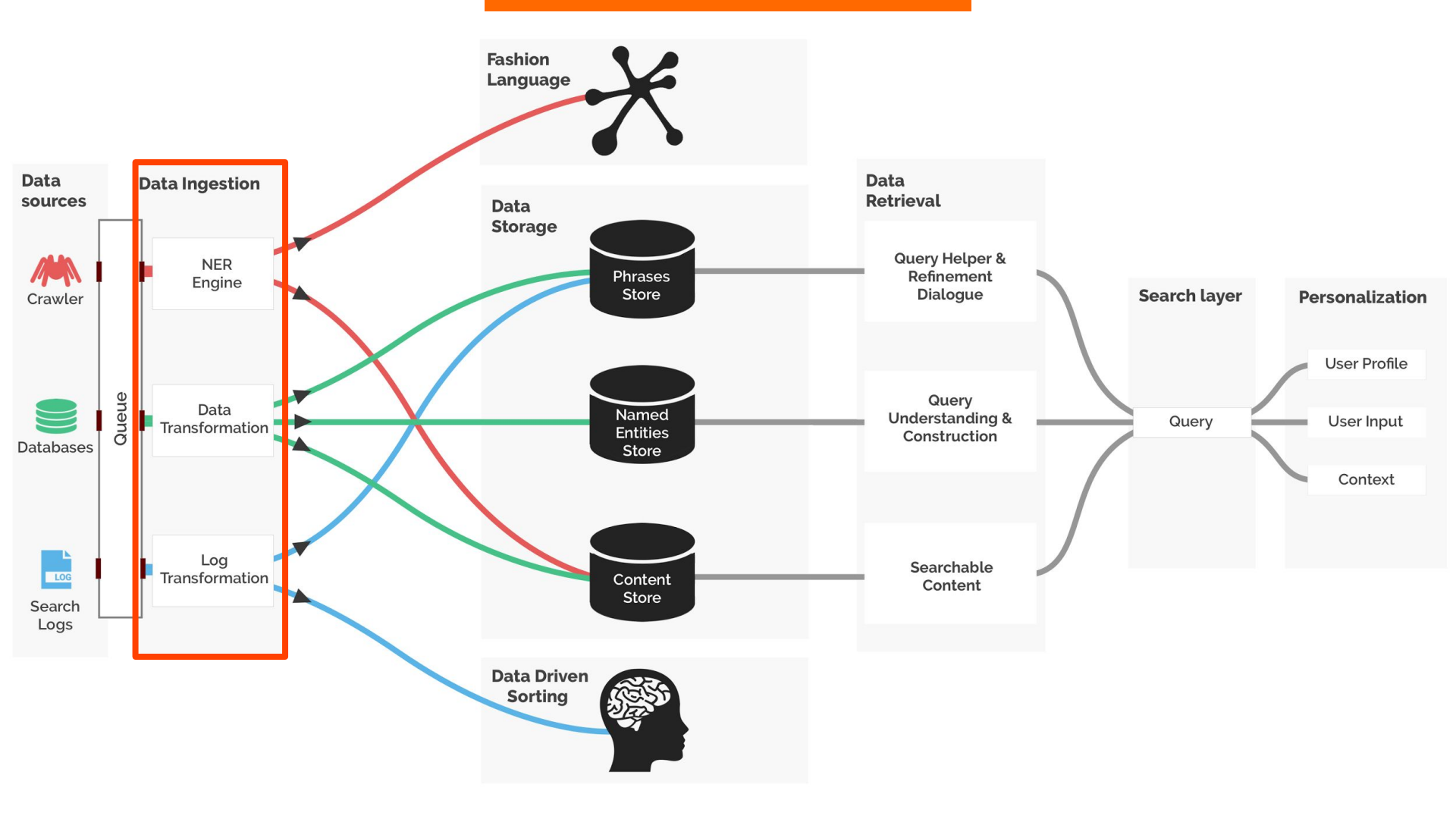
zalando

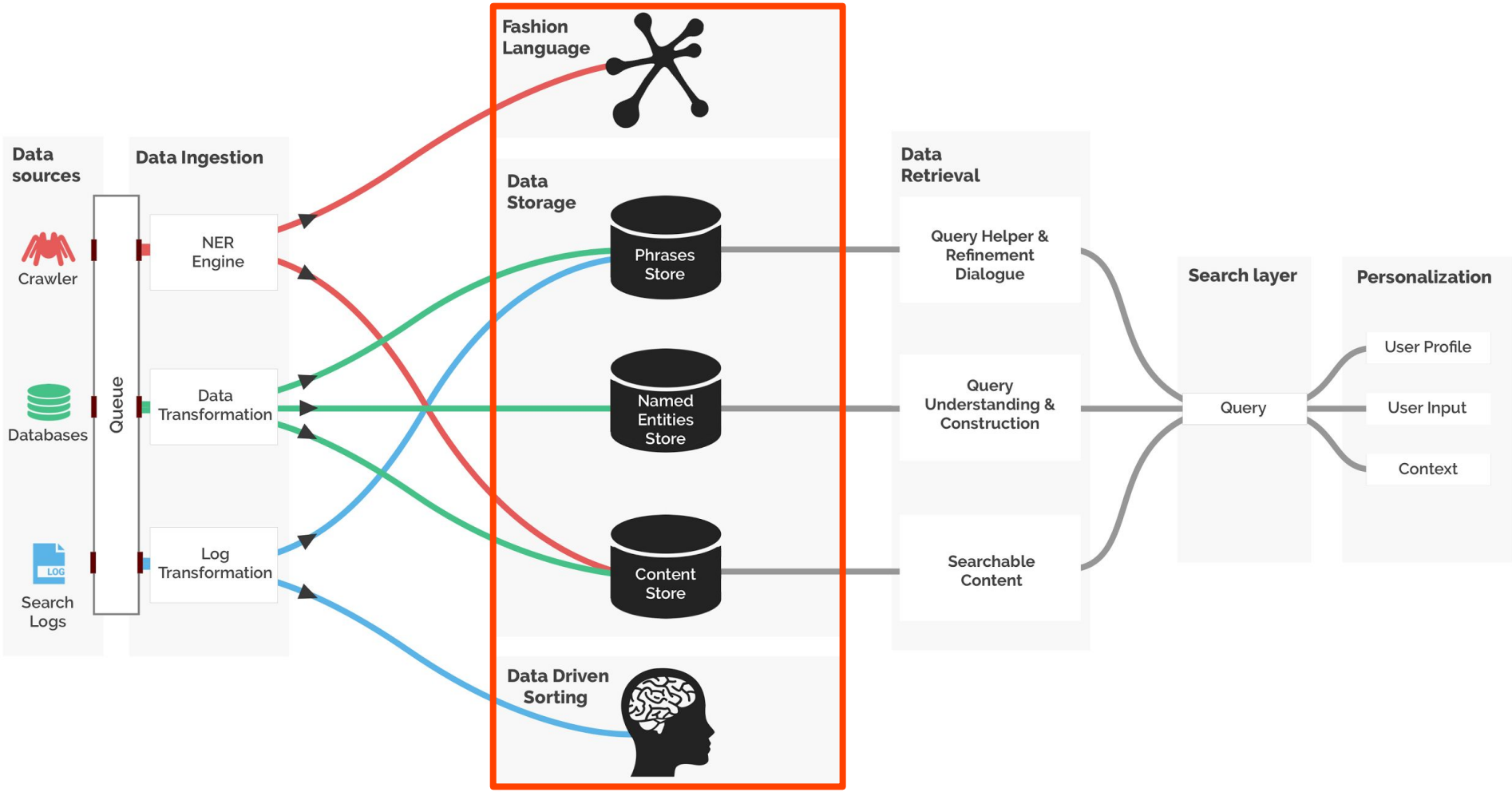

Fashion Language
Data sources
Data Ingestion
Crawler
Databases
Search Logs
Queue
NER Engine
Data Transformation
Log Transformation
Data Storage
Phrases Store
Named Entities Store
Content Store
Data Retrieval
Query Helper & Refinement Dialogue
Query Understanding & Construction
Searchable Content
Search layer
Query
Personalization
User Profile
User Input
Context
Data Driven Sorting

DATA INGESTION

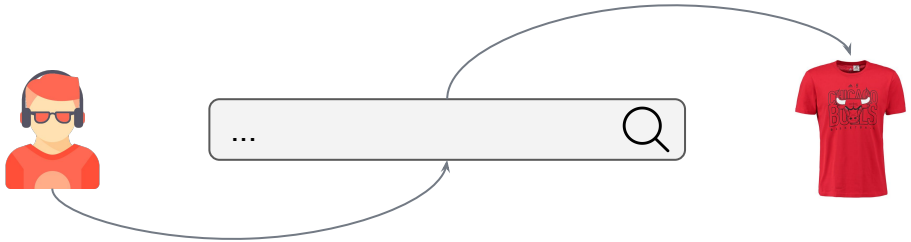
zalando


Fashion Language
Data sources
Crawler
Databases
Search Logs
Queue
Data Ingestion
NER Engine
Data Transformation
Log Transformation
Data Storage
Phrases Store
Named Entities Store
Content Store
Data Driven Sorting
Data Retrieval
Query Helper & Refinement Dialogue
Query Understanding & Construction
Searchable Content
Search layer
Query
Personalization
User Profile
User Input
Context

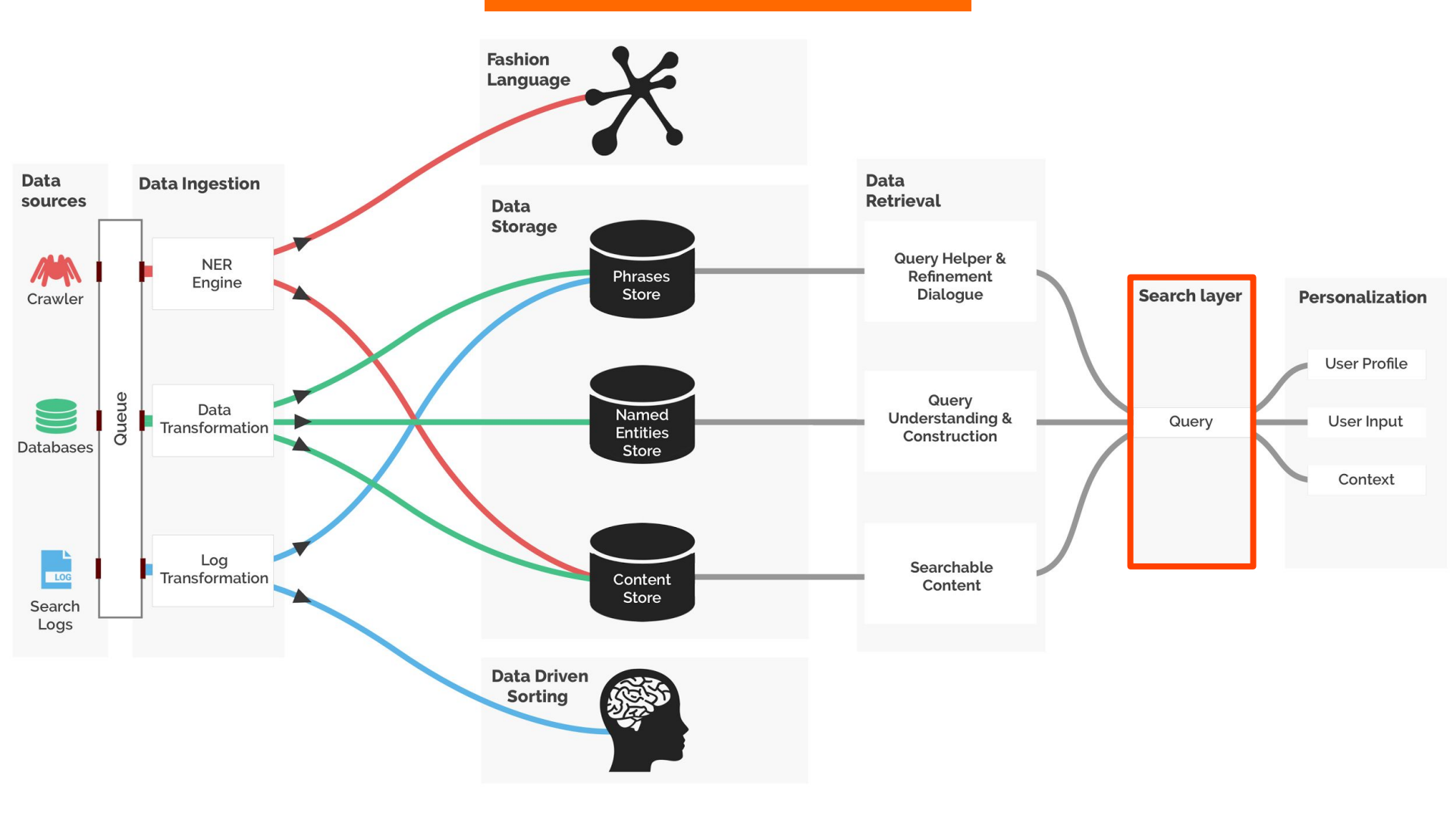

Fashion Language
Data sources
Data Ingestion
Crawler
Databases
Search Logs
Queue
NER Engine
Data Transformation
Log Transformation
Data Storage
Phrases Store
Named Entities Store
Content Store
Data Retrieval
Query Helper & Refinement Dialogue
Query Understanding & Construction
Searchable Content
Search layer
Query
Personalization
User Profile
User Input
Context
Data Driven Sorting

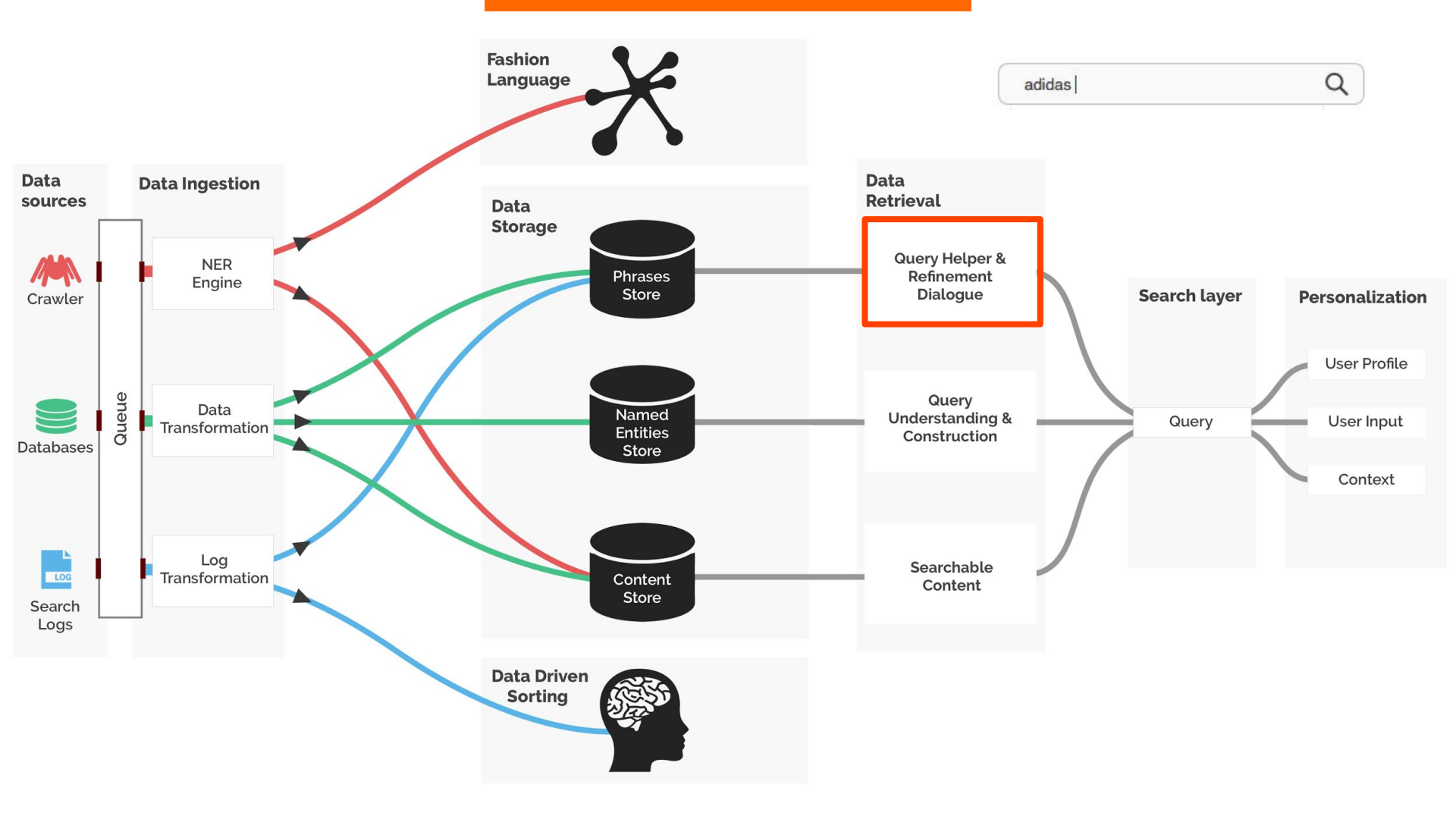

Data sources
Crawler
Databases
Search Logs
Queue
Data Ingestion
NER Engine
Data Transformation
Log Transformation
Fashion Language
Data Storage
Phrases Store
Named Entities Store
Content Store
Data Driven Sorting
Data Retrieval
Query Helper & Refinement Dialogue
Query Understanding & Construction
Searchable Content
Search layer
Query
Personalization
User Profile
User Input
Context

THE STORY OF A QUERY

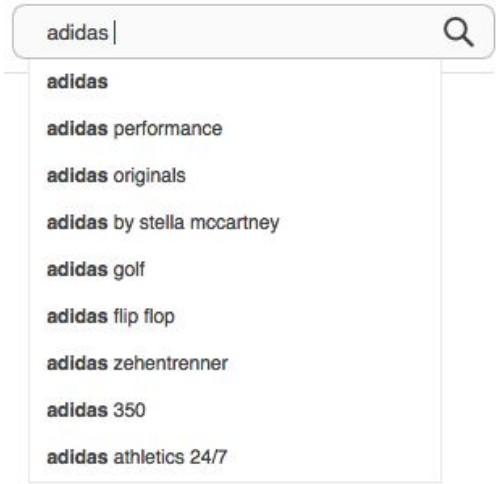


zalando


Fashion Language
Data sources
Data Ingestion
Crawler
Databases
Search Logs
Queue
NER Engine
Data Transformation
Log Transformation
Data Storage
Phrases Store
Named Entities Store
Content Store
Data Retrieval
Query Helper & Refinement Dialogue
Query Understanding & Construction
Searchable Content
Search layer
Query
Personalization
User Profile
User Input
Context
Data Driven Sorting

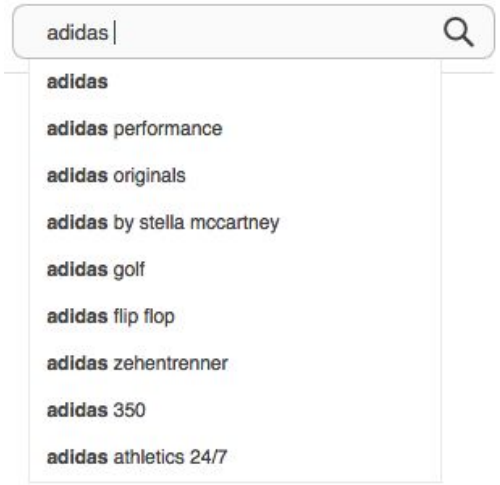

Fashion Language
adidas
Data sources
Data Ingestion
Data Storage
Data Retrieval
Crawler
Databases
Search Logs
Queue
NER Engine
Data Transformation
Log Transformation
Phrases Store
Named Entities Store
Content Store
Query Helper & Refinement Dialogue
Query Understanding & Construction
Searchable Content
Search layer
Query
Personalization
User Profile
User Input
Context
Data Driven Sorting

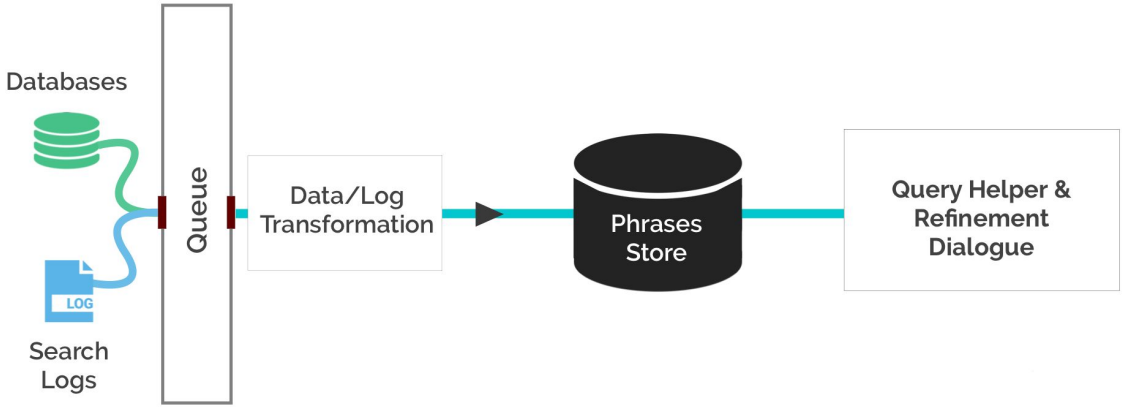
AUTO-COMPLETION



- Correct spell-error
- Predict the next word
- Understand synonym/acronym (`"Michael Kors" <-> "MK"`)
- Disambiguate (`"jack" -> "Jack & Jones"` or `"Jacke"`)
- Be personalized (sensitive to the context)
- Learn from user-feedback
- Have diverse results

GENERATING PHRASES



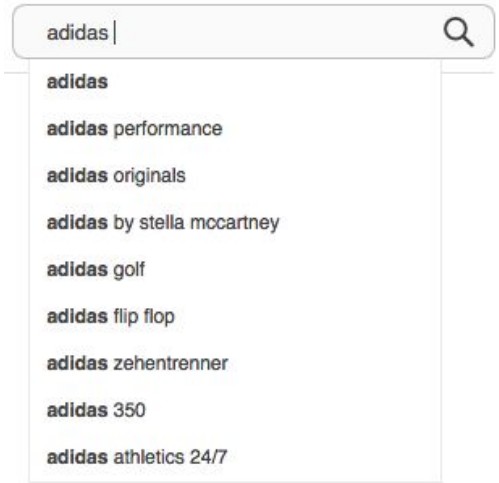


Brand + Color + Material + Category + Name + ... =

adidas schwarze schuhe

zalando

RANKING PHRASES



Score:

- Hit position
- Term frequency
- Entity preference (Entity bias)
- Brevity (Penalizing long queries that match multiple entities)
- User click (click count and recency bias)

AUTO-SUGGEST

lange kleide

Autosuggest:

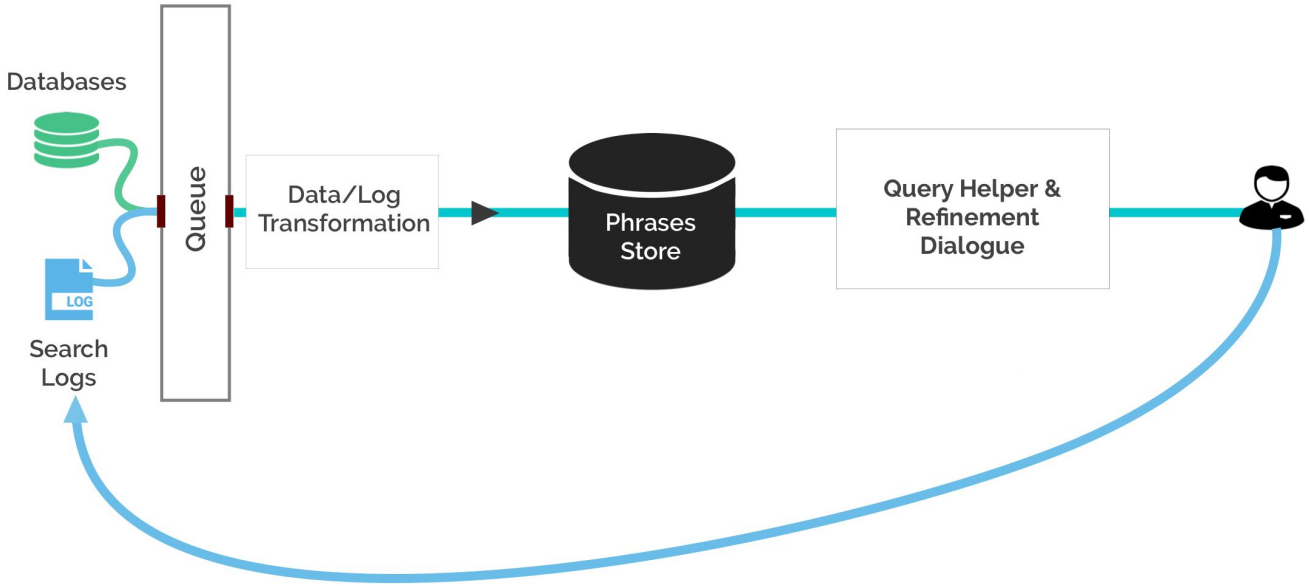
- Did you mean ? maxikleid

semantic space instead of text matching

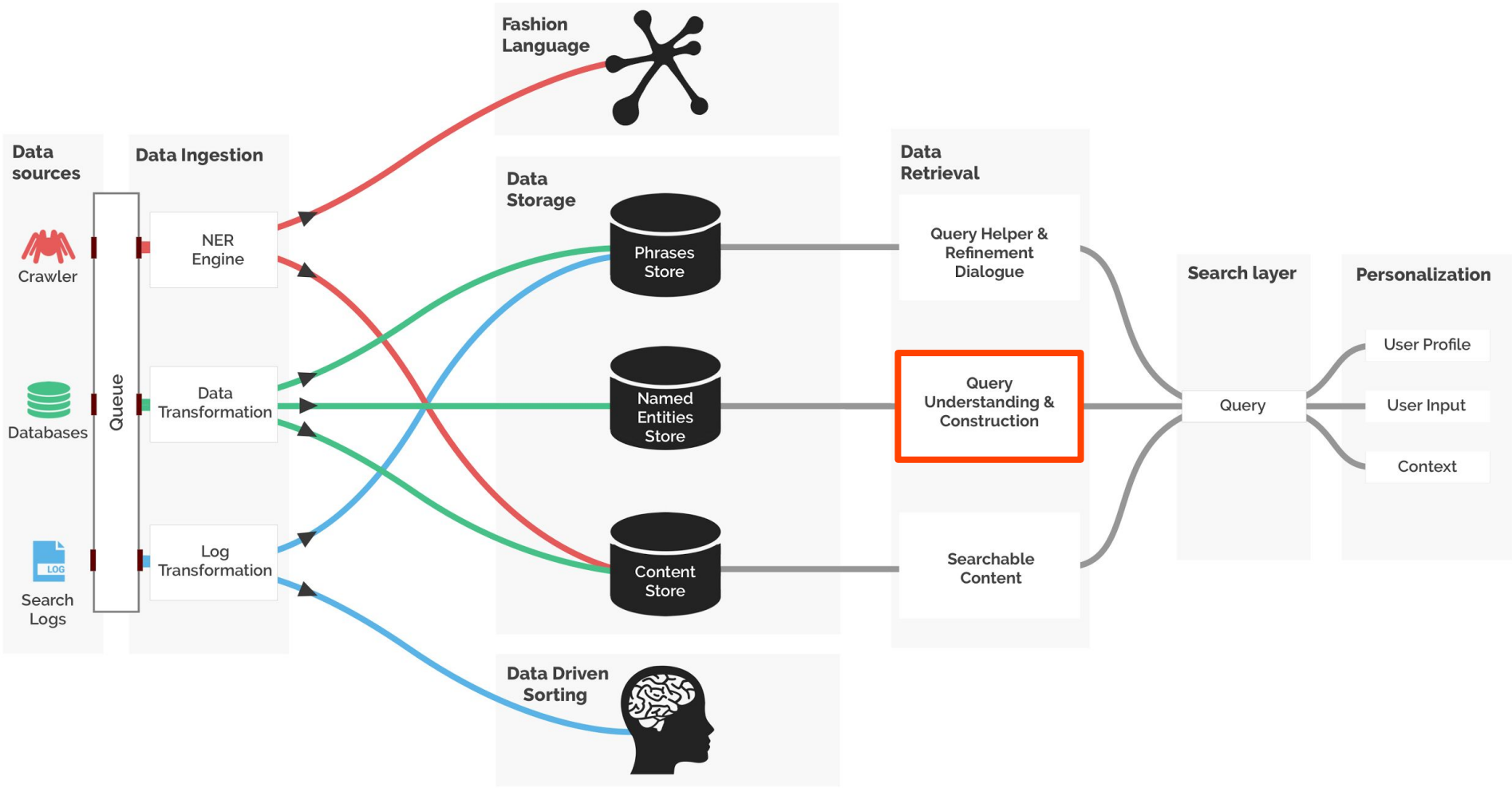
maxikleid

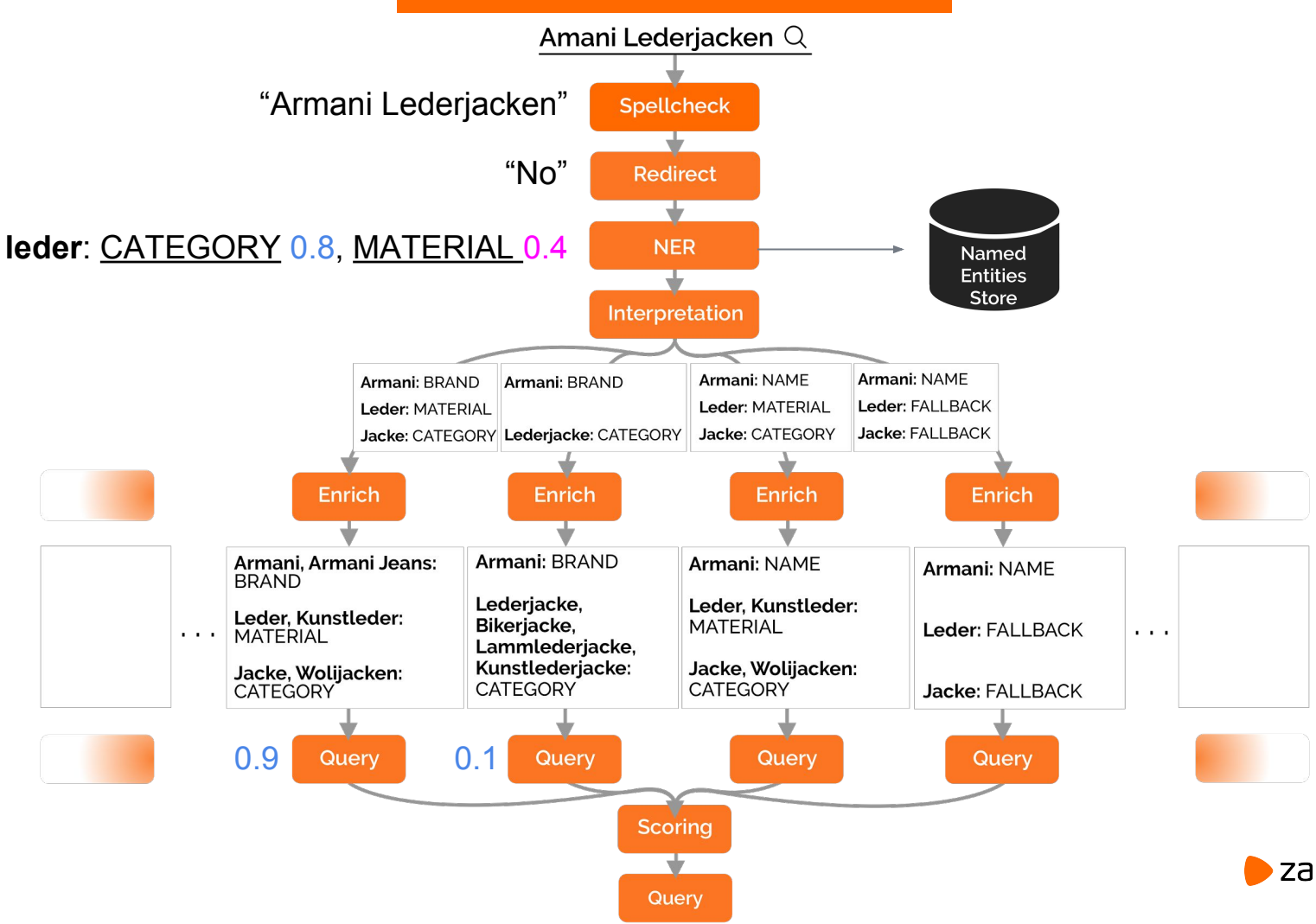
zalando

PERSONALISATION



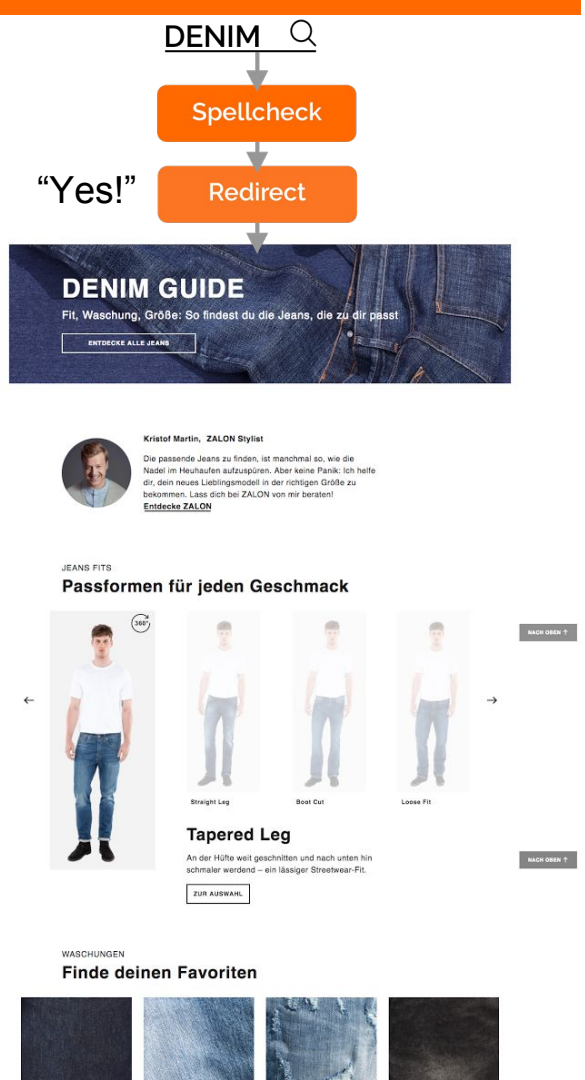
zalando


Fashion Language
Data sources
Data Ingestion
Data Storage
Data Retrieval
Crawler
Databases
Search Logs
Queue
NER Engine
Data Transformation
Log Transformation
Phrases Store
Named Entities Store
Content Store
Query Helper & Refinement Dialogue
Query Understanding & Construction
Searchable Content
Search layer
Query
Personalization
User Profile
User Input
Context
Data Driven Sorting

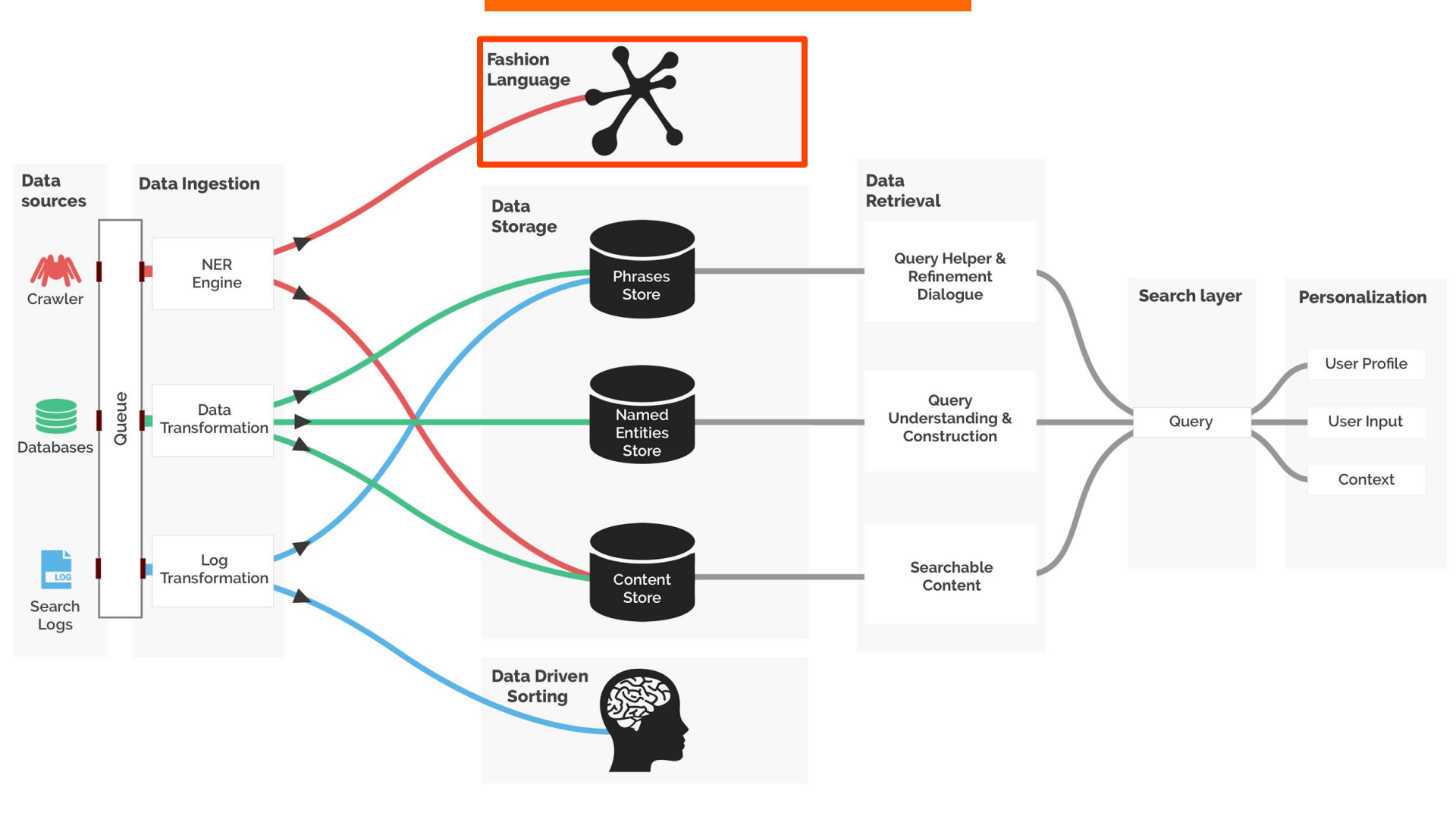

Amani Lederjacken
"Armani Lederjacken"
Spellcheck
"No"
Redirect
leder: CATEGORY 0.8, MATERIAL 0.4
NER
Named Entities Store
Interpretation
Armani: BRAND
Leder: MATERIAL
Jacke: CATEGORY
Armani: BRAND
Lederjacke: CATEGORY
Armani: NAME
Leder: MATERIAL
Jacke: CATEGORY
Armani: NAME
Leder: FALLBACK
Jacke: FALLBACK
Enrich
Enrich
Enrich
Enrich
Armani, Armani Jeans: BRAND
Leder, Kunstleder: MATERIAL
Jacke, Wolijacken: CATEGORY
Armani: BRAND
Lederjacke, Bikerjacke, Lammlederjacke, Kunstlederjacke: CATEGORY
Armani: NAME
Leder, Kunstleder: MATERIAL
Jacke, Wolijacken: CATEGORY
Armani: NAME
Leder: FALLBACK
Jacke: FALLBACK
0.9
Query
0.1
Query
Query
Query
Scoring
Query

zalando

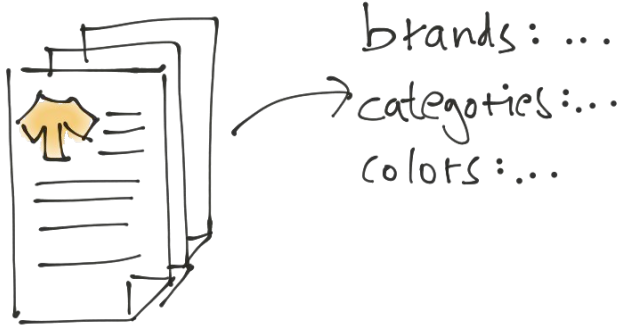
GUIDED SEARCH



zalando


Fashion Language
Data sources
Data Ingestion
Crawler
Databases
Search Logs
Queue
NER Engine
Data Transformation
Log Transformation
Data Storage
Phrases Store
Named Entities Store
Content Store
Data Driven Sorting
Data Retrieval
Query Helper & Refinement Dialogue
Query Understanding & Construction
Searchable Content
Search layer
Query
Personalization
User Profile
User Input
Context

FASHION KNOWLEDGE BEYOND ARTICLE DATA



Article data already produces a ton of data.

But what about…?

New trends

"boyfriend jeans"

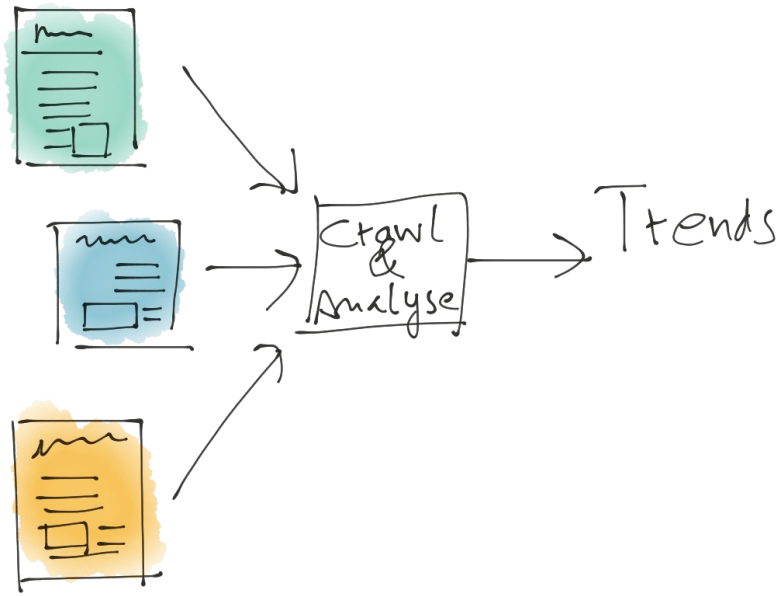
"Hobo" "80s"

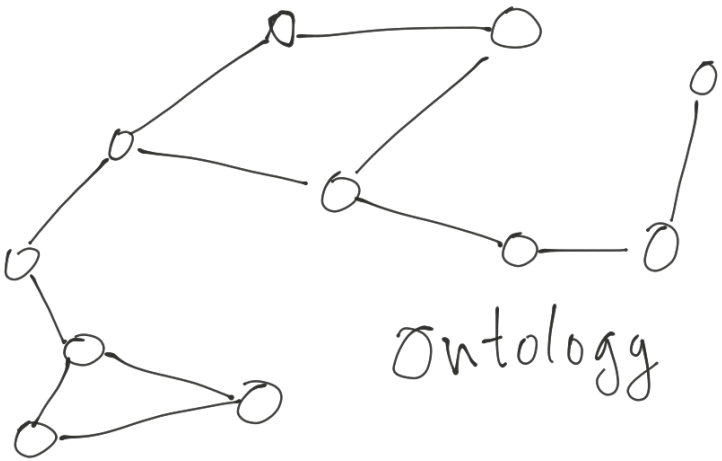
Occasions

"Wedding"

"Party"

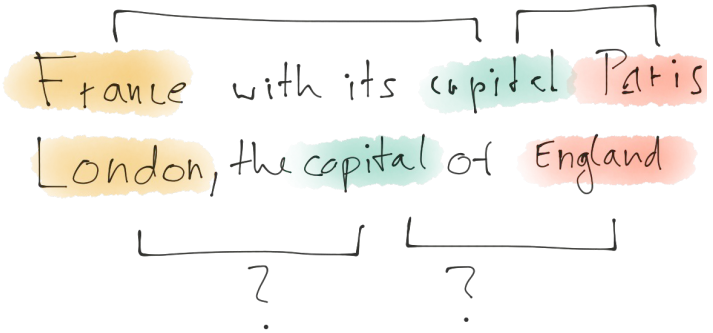
FASHION KNOWLEDGE BEYOND ARTICLE DATA

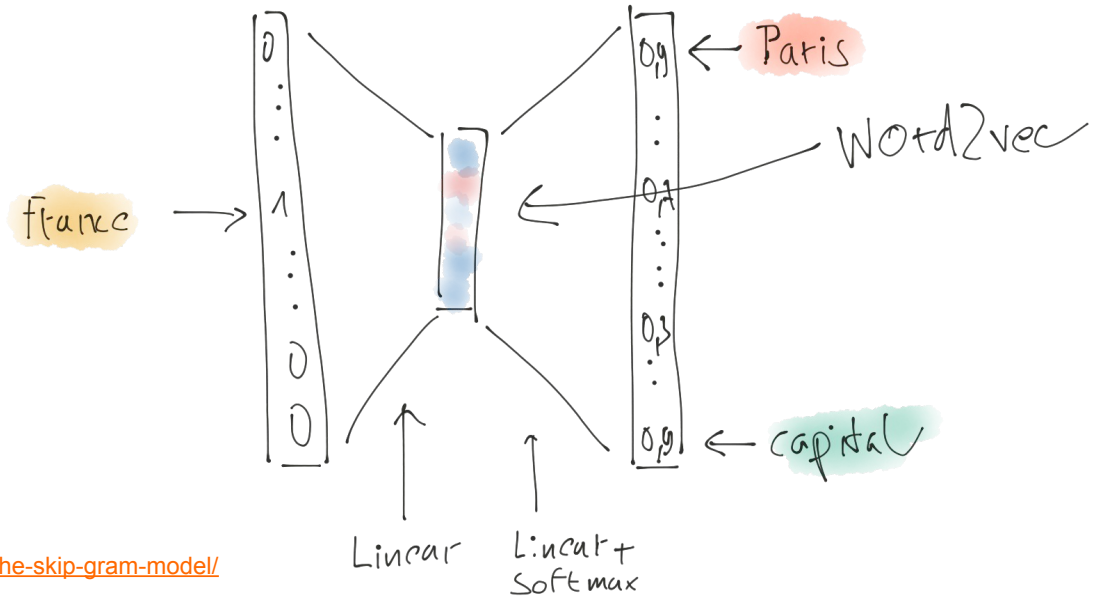




zalando

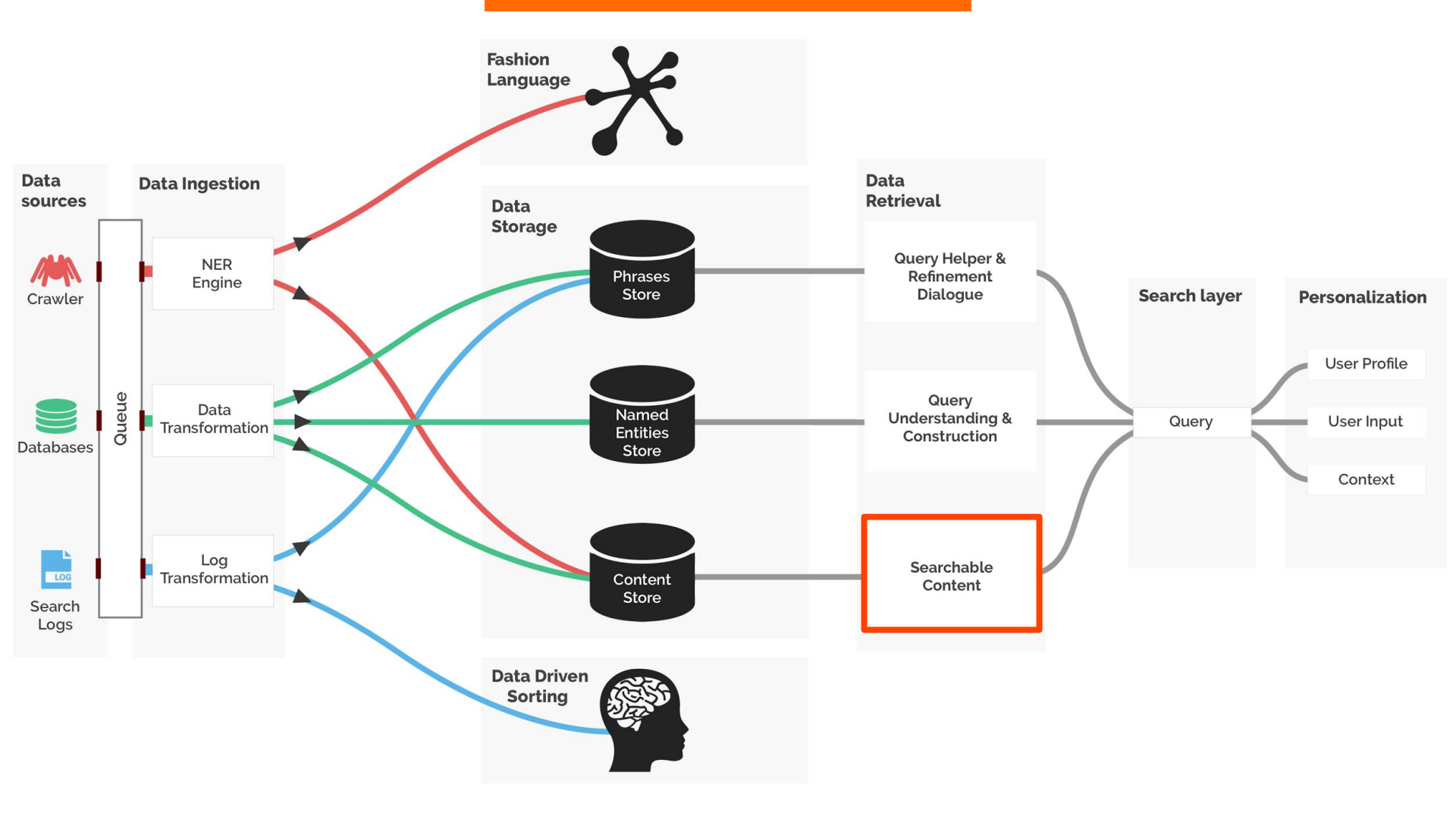
DATA DRIVEN SYNONYMS - WORD2VEC

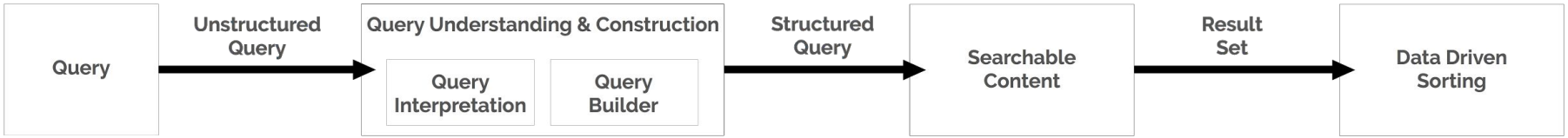




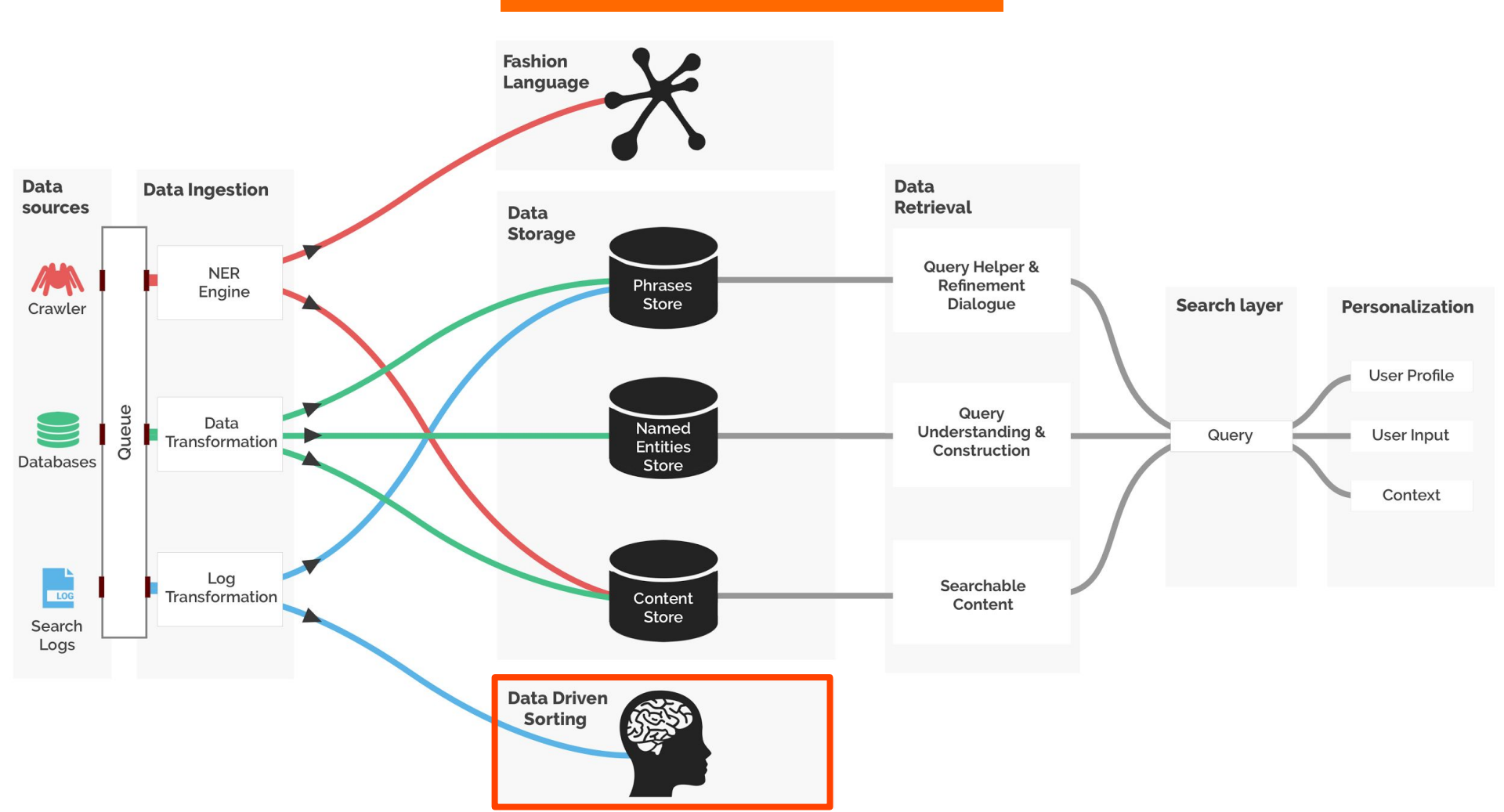
Excellent post: http://mccormickml.com/2016/04/19/word2vec-tutorial-the-skip-gram-model/

zalando

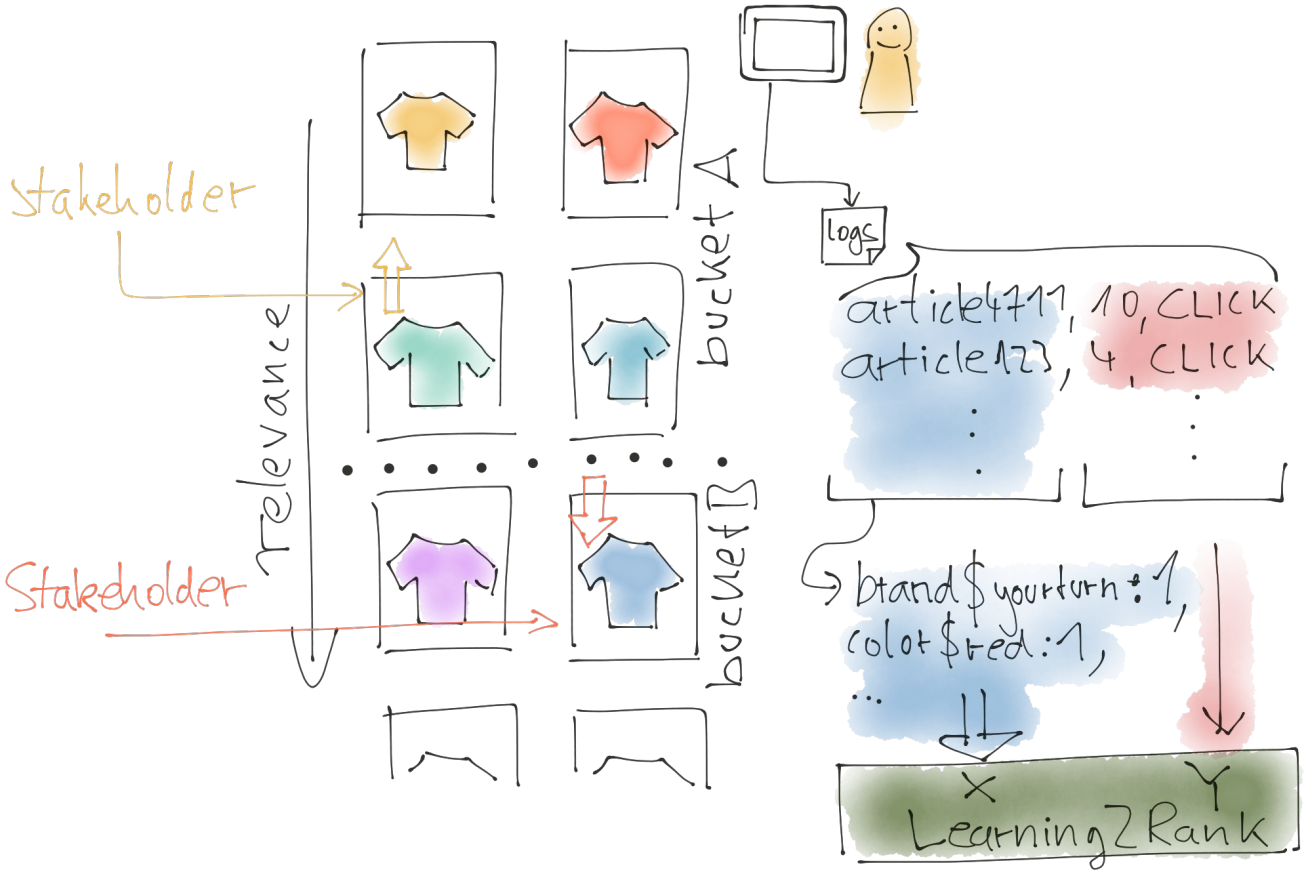

Fashion Language
Data sources
Data Ingestion
Crawler
Databases
Search Logs
Queue
NER Engine
Data Transformation
Log Transformation
Data Storage
Phrases Store
Named Entities Store
Content Store
Data Retrieval
Query Helper & Refinement Dialogue
Query Understanding & Construction
Searchable Content
Search layer
Query
Personalization
User Profile
User Input
Context
Data Driven Sorting


Query
Unstructured Query
Query Understanding & Construction
Query Interpretation
Query Builder
Structured Query
Searchable Content
Result Set
Data Driven Sorting

zalando


Fashion Language
Data sources
Data Ingestion
Data Storage
Data Retrieval
Crawler
Databases
Search Logs
Queue
NER Engine
Data Transformation
Log Transformation
Phrases Store
Named Entities Store
Content Store
Query Helper & Refinement Dialogue
Query Understanding & Construction
Searchable Content
Search layer
Query
Personalization
User Profile
User Input
Context
Data Driven Sorting

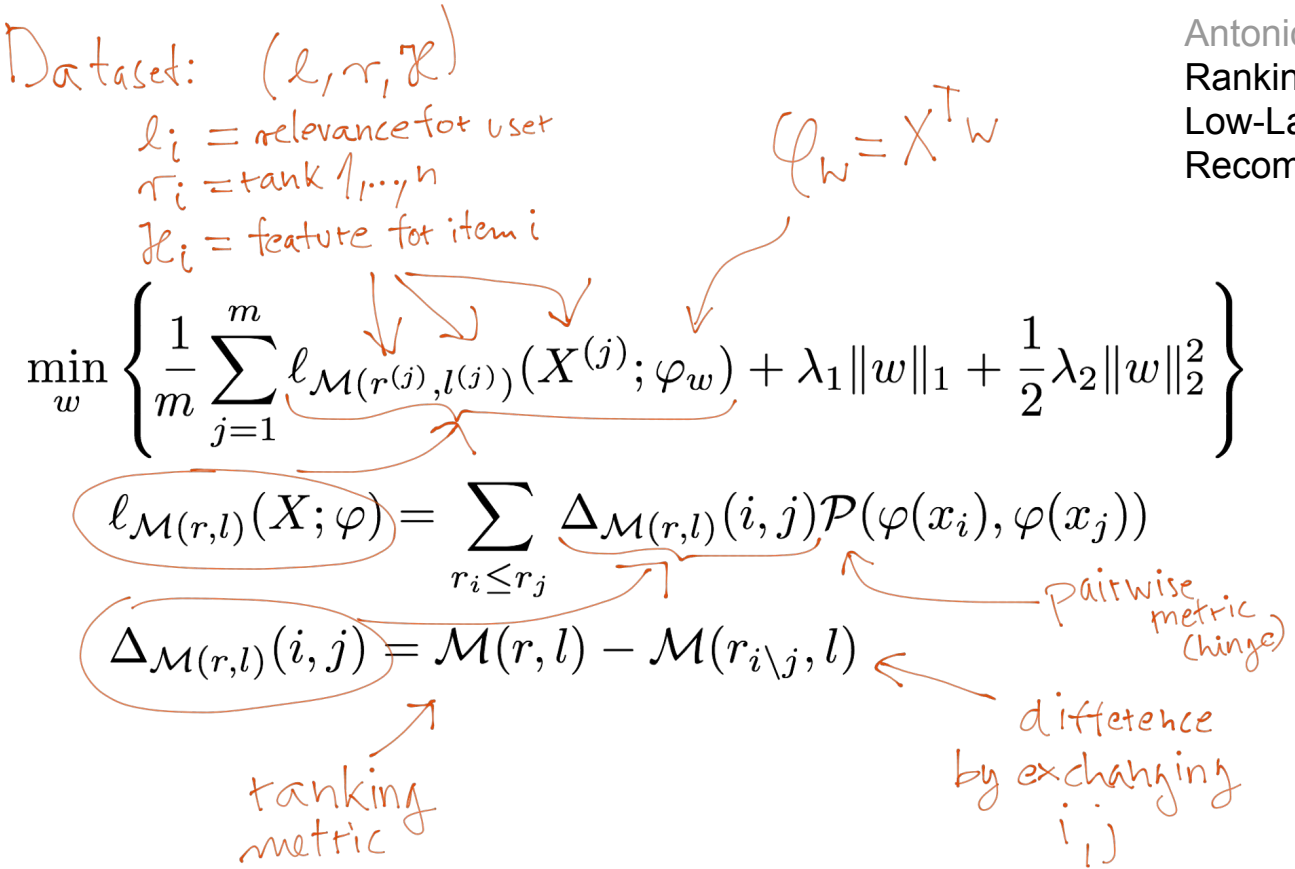
DATA DRIVEN SORTING



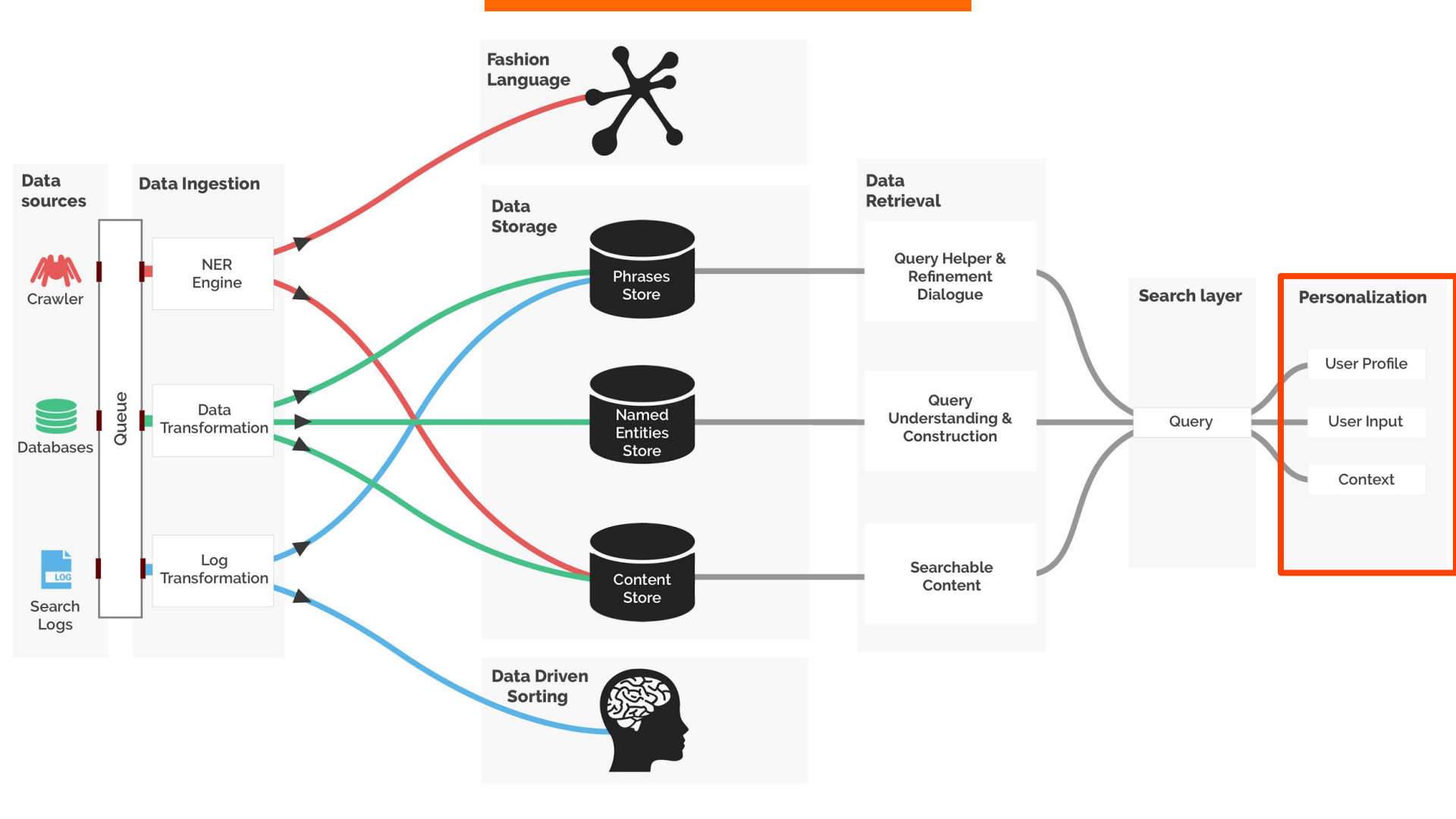
zalando

DATA DRIVEN SORTING: LEARNING TO RANK

Antonio Freno et al. "One-Pass Ranking Models for Low-Latency Product Recommendations", KDD 2015



zalando


Fashion Language
Data sources
Data Ingestion
Crawler
Databases
Search Logs
Queue
NER Engine
Data Transformation
Log Transformation
Data Storage
Phrases Store
Named Entities Store
Content Store
Data Retrieval
Query Helper & Refinement Dialogue
Query Understanding & Construction
Searchable Content
Search layer
Query
Personalization
User Profile
User Input
Context
Data Driven Sorting

PERSONALISATION

- Goal: Find the most relevant articles for each user.
- Give the customer the feeling, search is understanding what their needs are.
- Essentially part of every step: Query Helper, Query Analysis, Personalization.
- From simple methods (history, last orders) to full blown complex models (personalized ranking models)

THE END