



Finding e-commerce products using Elasticsearch

Alexander Reelsen | @spinscale
alex@elastic.co



search is hard!



**search in
ecommerce is harder**



good data & good searches



bad data & smart searches



good data & worst searches



Agenda



Agenda

facetted navigation

Agenda

facetted navigation

search bar



Agenda

facetted navigation

clean data

search bar



Agenda

faceted navigation

clean data

smart searches

search bar

Agenda

clean data

faceted navigation

synonyms

smart searches

search bar



Agenda

faceted navigation

clean data

UOM

synonyms

smart searches

search bar

Agenda

clean data

faceted navigation

decompounding

synonyms

UOM

smart searches

search bar

Agenda

clean data

facetted navigation

relevancy

synonyms

UOM

decompounding

smart searches

search bar

Agenda

faceted navigation

clean data

variants

synonyms

UOM

relevancy

decompounding

smart searches

search bar

Agenda

faceted navigation

clean data

variants

deduplication

synonyms

UOM

relevancy

decompounding

smart searches

search bar

Agenda

clean data

faceted navigation

variants

deduplication

search as you type

synonyms

UOM

relevancy

decompounding

smart searches

search bar

Agenda

clean data

faceted navigation

variants

deduplication

analytics

synonyms

UOM

relevancy

decompounding

smart searches

search bar



search as you type



Agenda

analytics

clean data

faceted navigation

variants

deduplication

data quality

synonyms

UOM

relevancy

decompounding

smart searches

search bar



search as you type



Agenda

analytics

clean data

faceted navigation

variants

deduplication

mobile

synonyms

data quality

UOM

relevancy

decompounding

smart searches

search bar



search as you type



Agenda

product detail page

analytics

clean data

faceted navigation

variants

deduplication

synonyms

data quality

UOM mobile

relevancy

decompounding

smart searches

search bar

 elastic

search as you type



Agenda

analytics

product detail page

clean data

facetted navigation

variants

deduplication

LTR

synonyms

data quality

UOM mobile relevancy

decompounding

smart searches

search bar



search as you type

Agenda

multi language

analytics

product detail page

clean data

faceted navigation

variants

deduplication

synonyms

data quality

UOM mobile

relevancy

decompounding

smart searches

search bar

LTR

 elastic

search as you type



Agenda

analytics

product detail page

clean data

faceted navigation

variants

deduplication

ETIME

synonyms

multi language

data quality

UOM mobile

relevancy

decompounding

smart searches

search bar

LTR

 elastic

search as you type



demo



search bar



search bar

The screenshot shows the OTTO website interface. At the top left is the OTTO logo. To its right is a search bar with the placeholder text "Suchbegriff / Artikelnr. eingeben" and a magnifying glass icon. Further right are icons for "Service", "Mein Konto", "Merktzettel", and "Warenkorb". Below the search bar is a horizontal navigation menu with categories: "Inspiration", "Damen", "Herren", "Kinder", "Wäsche/Bademode", "Sport", "Schuhe", "Große Größen", "Multimedia", "Haushalt", "Küche", "Heimtextilien", "Möbel", "Baumarkt", "Spielzeug", "Marken", and "%Sale%". Below the menu are two promotional banners: "20€ für Neukunden in der App" and "0€ 100 Tage Zahlpause". The main content area features a "FIFA19 APP Aktion!" banner with a PlayStation logo, a game cover, and the text "20€ auf FIFA 19 sparen! Nur in der App - mit dem Code 12952 bis 02.08.2019." A small chat icon is in the bottom right corner.

The screenshot shows the Zalando website interface. At the top is a cookie consent banner. Below it is a navigation menu with "HILFE & KONTAKT", "KOSTENLOSER VERSAND & RÜCKVERSAND", and "100 TAGE RÜCKGABERECHT". The Zalando logo is centered, with "DE | EN | Anmelden | Wunschkorbzeitel | Warenkorb" on the right. Below the logo is a horizontal navigation menu with categories: "Get the Look", "NEU", "Bekleidung", "Schuhe", "Sport", "Accessoires", "Beauty", "Premium", "Marken", "Sale %", and "Suche" (with a magnifying glass icon). The main content area features a large banner with the text "Europas größte Auswahl an Fashion & Trends" and "Entdecke über 1900 Marken und lass dich modisch inspirieren. Bestelle einfach und bequem nach Hause - mit unserem kostenlosen Versand & Rückversand." Below the banner are three buttons: "DAMEN", "HERREN", and "KINDER".

smart searches

smart searches

nike running hoodie xl



smart searches

nike running hoodie xl



smart searches

nike running hoodie **xl**



brand



size



clean data



clean data

- » Hardest thing to do ever
- » Formats being accepted? JSON, XML, CSV, EDIFACT?
- » How to train merchants?
- » Another local cleansing step? Accountability on failure?
- » If you fail here, stop optimising your search!
- » indexing pipeline: applying synonyms?

synonyms



synonyms

- » topf => kochtopf
- » naik => nike
- » portmonee => geldbörse
- » who maintains this list?
- » who keeps it updated?
- » who matches this against your worst queries, that return 0 hits?
- » reloadable without index closing (since ES 7.3)

UOM



UOM

- » Unit of Measure (100cm vs. 1m)
- » Requires normalization: part of data cleansing
- » Dissecting into a base unit and a value in order to query
- » Who is doing this already?

- » JSR 385: Units of Measurement API 2.0
- » Could be done in an Ingest Processor



decompounding



decompounding

The screenshot shows the OTTO website homepage. At the top, there is a navigation bar with the OTTO logo, a search bar containing the text "Suchbegriff / Artikelnr. eingeben", and links for "Service", "Mein Konto", "Merktzettel", and "Warenkorb". Below the navigation bar, there are category links: "Inspiration", "Damen", "Herren", "Kinder", "Wäsche/Bademode", "Sport", "Schuhe", "Große Größen", "Multimedia", "Haushalt", "Küche", "Heimtextilien", "Möbel", "Baumarkt", "Spielzeug", "Marken", and "%Sale%". Two promotional banners are visible: "20€ für Neukunden in der App" and "0€ 100 Tage Zahlpause". The main content area features a "FIFA19 APP Aktion!" banner with a PlayStation logo and a PS4 game case for FIFA 19. The banner text reads: "20€ auf FIFA 19 sparen! Nur in der App - mit dem Code 12952 bis 02.08.2019." The OTTO logo is also present in the bottom right corner of the page.

decompounding

```
GET _analyze?filter_path=**.token
{
  "text": "Blumentopf Aero mit Leuchte"
}
```

```
{
  "tokens" : [
    {
      "token" : "blumentopf"
    },
    {
      "token" : "aero"
    },
    {
      "token" : "mit"
    },
    {
      "token" : "leuchte"
    }
  ]
}
```

decompounding

PUT example

```
{
  "settings": {
    "index": {
      "analysis": {
        "analyzer": {
          "decompound_analyzer": {
            "type": "custom",
            "tokenizer": "standard",
            "filter": [ "dictionary_decompounder" ]
          }
        },
        "filter": {
          "dictionary_decompounder": {
            "type": "dictionary_decompounder",
            "word_list": [ "topf" ]
          }
        }
      }
    }
  }
}
```

GET example/_analyze?filter_path=**.token

```
{
  "analyzer": "my_analyzer",
  "text": "Blumentopf Aero mit Leuchte"
}
```

```
{
  "tokens" : [
    {
      "token" : "Blumentopf"
    },
    {
      "token" : "topf"
    },
    {
      "token" : "Aero"
    },
    {
      "token" : "mit"
    },
    {
      "token" : "Leuchte"
    }
  ]
}
```

relevancy



relevancy

- » relevancy needs to be defined by the business owners (who rarely understand it)
- » BM25 is not the score you are looking for
- » need to incorporate business/product metrics
- » provision, item on stock, location, free shipping, last sale

relevancy

- » Search for 'bicycle'
- » Are 20 different bikes relevant results?
- » What about locks, lights, clothes? Maybe go with 10 bikes, 3 accessoires?
- » User bought a bike three months ago, maybe he is searching for equipment? Or a replacement tire?

relevancy

» are there certain products you always want to score higher?

```
GET /_search
{
  "query": {
    "pinned" : {
      "ids" : ["1", "4", "100"],
      "organic" : {
        "match":{
          "description": "iphone"
        }
      }
    }
  }
}
```

relevancy

Rank feature datatype

A `rank_feature` field can index numbers so that they can later be used to boost documents in queries with a `rank_feature` query.

Script score query

The `script_score` allows you to modify the score of documents that are retrieved by a query. This can be useful if, for example, a score function is computationally expensive and it is sufficient to compute the score on a filtered set of documents.

To use `script_score`, you have to define a query and a script - a function to be used to compute a new score for each document returned by the query. For more information on scripting see [scripting documentation](#).

Dense vector datatype

X-Pack



WARNING

This functionality is experimental and may be changed or removed completely in a future release. Elastic will take a best effort approach to fix any issues, but experimental features are not subject to the support SLA of official GA features.

A `dense_vector` field stores dense vectors of float values. The maximum number of dimensions that can be in a vector should not exceed 1024. A `dense_vector` field is a single-valued field.

These vectors can be used for [document scoring](#). For example, a document score can represent a distance between a given query vector and the indexed document vector.

You index a dense vector as an array of floats.

variants



variants

The screenshot shows the Zalando website homepage. The browser address bar displays 'zalando.de'. The navigation bar includes links for 'DAMEN', 'HERREN', and 'KINDER', the Zalando logo, and icons for 'Anmelden', 'Wunschzettel', and 'Warenkorb'. A search bar with the text 'Suche' is positioned on the right. The main banner features a photograph of two children with the text 'Europas größte Auswahl an Fashion & Trends' and three buttons labeled 'DAMEN', 'HERREN', and 'KINDER'. Below the banner is a 'Top Marken' section with a grid of brand names.

Top Marken

Abercrombie & Fitch	adidas	American Apparel	ASICS
Benetton	Bugatti	Calvin Klein	Converse
Esprit	Fossil	G-Star	Hollister Co.

variants

- » how to model variants and their differences?
- » just attributes? and price? product title and description?
- » search: across all variants or the main products?
- » display: variants as own results or group them?
- » display: what happens when one product is out of stock?



variants

Join datatype



The `join` datatype is a special field that creates parent/child relation within documents of the same index. The `relations` section defines a set of possible relations within the documents, each relation being a parent name and a child name. A parent/child relation can be defined as follows:

Inner hits



The `parent-join` and `nested` features allow the return of documents that have matches in a different scope. In the parent/child case, parent documents are returned based on matches in child documents or child documents are returned based on matches in parent documents. In the nested case, documents are returned based on matches in nested inner objects.

deduplication



deduplication

- » Safe: ISBN, ASIN
- » Unsafe: Product images, description, name, release date, size
- » query time or index time?



deduplication

Field Collapsing

edit

Allows to collapse search results based on field values. The collapsing is done by selecting only the top sorted document per collapse key. For instance the query below retrieves the best tweet for each user and sorts them by number of likes.

```
GET /twitter/_search
{
  "query": {
    "match": {
      "message": "elasticsearch"
    }
  },
  "collapse" : {
    "field" : "user" ①
  },
  "sort": ["likes"], ②
  "from": 10 ③
}
```

[COPY AS CURL](#) [VIEW IN CONSOLE](#) 

search as you type

search as you type

- » "The importance of search-as-you-type cannot be overstated"
- » Hint: make a user test first. There are users who do not look up when typing!
- » Speed is key
- » Rank your suggestions on your own criteria!
- » Ensure exact hits are scored up (brown fox vs. brown foxes)
- » Steer the user without showing any search results
- » Possibly an own index with reduced result set
- » Analyze searches and adapt to follow trends

search as you type

Search-as-you-type datatype



The `search_as_you_type` field type is a text-like field that is optimized to provide out-of-the-box support for queries that serve an as-you-type completion use case. It creates a series of subfields that are analyzed to index terms that can be efficiently matched by a query that partially matches the entire indexed text value. Both prefix completion (i.e. matching terms starting at the beginning of the input) and infix completion (i.e. matching terms at any position within the input) are supported.



analytics



analytics

- » conversion rate
- » search results with zero hits
- » "one search and out"
- » busiest hours (planning downtime)
- » recommendations



product detail page

product detail page

- » crucial to make a sale
- » what to display, if the product is out of stock
- » what to display, if the product is EOL?
- » dynamic price calculation



LTR



LTR



Doug Turnbull
@softwaredoug

Everytime someone contacts [@o19s](#) and asks for "learning to rank" or some other ML relevance thing often they haven't done step 1 (some basic relevance work)



Emmanuel Ameisen @mlpowered · Jul 16

How to ship ML in practice:

- 1/ Write a simple rule based solution to cover 80% of use cases
- 2/ Write a simple ML algorithm to cover 95% of cases
- 3/ Write a filtering algorithm to route inputs to the correct method
- 4/ Add monitoring
- 5/ Detect drift
- ...
- 24/ Deep Learning

[Show this thread](#)

8:11 PM · Jul 17, 2019 · [Twitter Web Client](#)



summary



summary

- » ecommerce search is complex
- » so many things to take into account...
- » untold: index strategies, updates, management
- » always have a middleware (UI, query injection, a/b testing, landing pages, redirects, query logging, business owner endpoint)



Elastic App Search



Products Learn Company Pricing

Contact

Try Free

Login



App Search

As a Service Pricing Demo Request Login



Advanced search made simple

The curated experience of Elastic App Search brings the focused power of Elasticsearch to a refined set of APIs and intuitive dashboards. Leverage the seamless scalability, tunable relevance controls, thorough documentation, well-maintained clients, and robust analytics to build a leading search experience with ease.

Enter your email

Start Trial

By submitting you agree to the [Elastic Cloud Standard Terms of Service](#) and to receive occasional emails from Elastic. Your personal data will be processed in accordance with Elastic's [privacy statement](#)

Download the latest version



Elastic App Search

App Search Service [Pricing](#) [Demo Request](#) [Login](#)

App Search Service pricing

Start with our 14-day trial. No credit card required. Cancel anytime.

<h3>Standard</h3> <p>For small projects</p> <p>\$49/month BILLED MONTHLY</p> <p>Start free trial</p> <p>Features include:</p> <ul style="list-style-type: none">Basic analytics dataCustomizable relevance modelBasic API logs	<h3>Pro</h3> <p>Ready for growth</p> <p>\$199/month BILLED MONTHLY</p> <p>Start free trial</p> <p>Features include:</p> <ul style="list-style-type: none">Multilingual supportAdditional search customizationsClickthrough AnalyticsAnalytics APIExtended API log retention	<h3>Premium</h3> <p>Mission-critical search</p> <p>Custom pricing</p> <p>Contact us</p> <p>Features include:</p> <ul style="list-style-type: none">Engine-level user scopingGranular access controlDedicated hardwareDedicated point of contactExpert implementation teamEnterprise SLA
--	---	--

 elastic

Elastic App Search

Download Elastic App Search

Learn more about Elastic App Search. [Release blog »](#)

Version: 7.2.1

Release date: July 30, 2019

License: **Elastic License**

Downloads: [📄](#) **MACOS/LINUX** sha

Notes: View the **documentation**.



<https://www.elastic.co/blog/elastic-app-search-7-2-0-released>

books



books

With applications for Solr and Elasticsearch

Relevant SEARCH

Doug Turnbull
John Berryman

Foreword by Trey Grainger



MANNING



elastic

books



DEEP LEARNING for Search

Tommaso Teofili
Foreword by Chris Mettman

 MANNING



links



links

» <https://project-a.github.io/on-site-search-design-patterns-for-e-commerce/>

Thank you for listening!

Alexander Reelsen
@spinscale
alex@elastic.co

