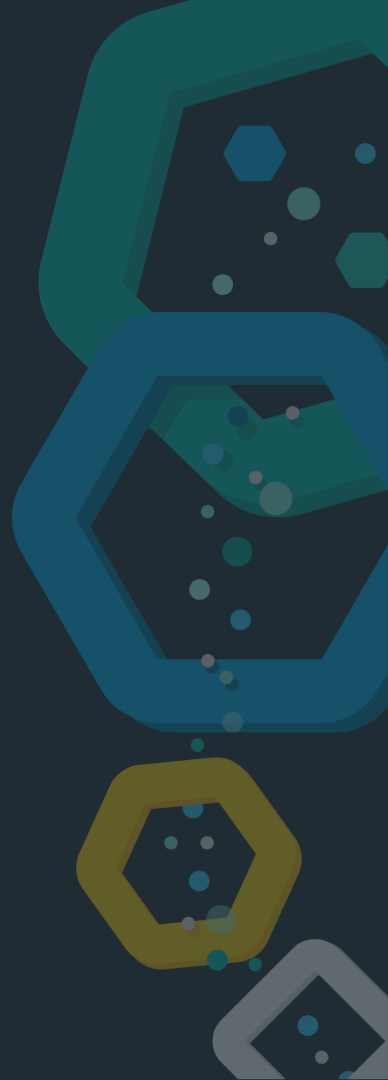
elastic

Finding e-commerce products
using Elasticsearch

Alexander Reelsen | @spinscale
alex@elastic.co

search is hard!

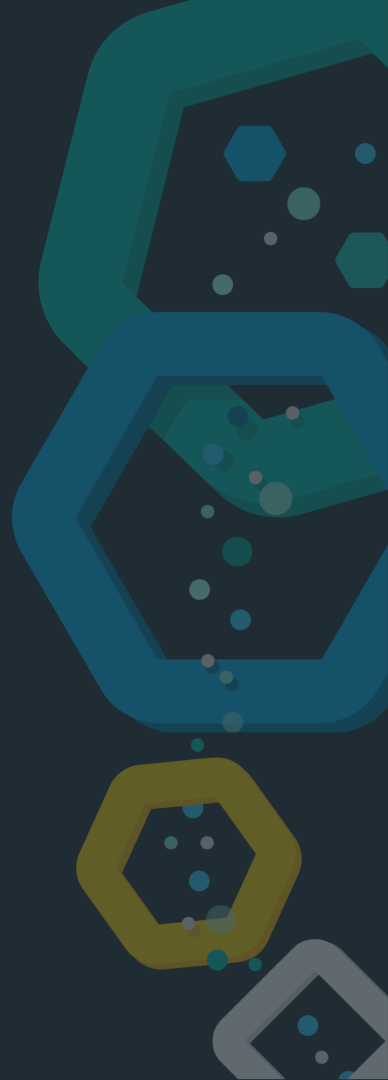


elastic

search in ecommerce is harder

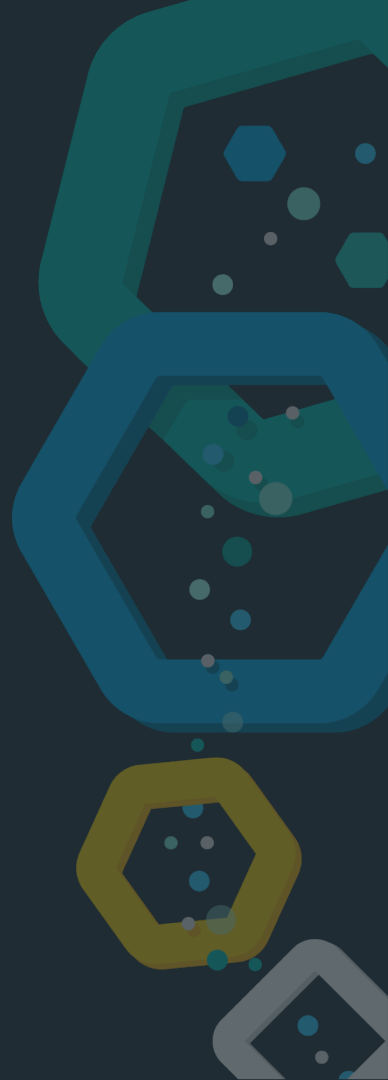
elastic

good data
&
good searches



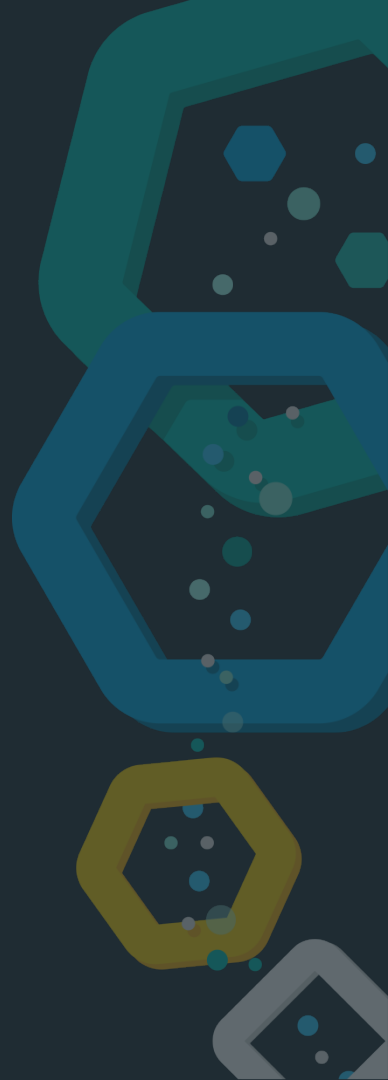
elastic

bad data
&
smart searches



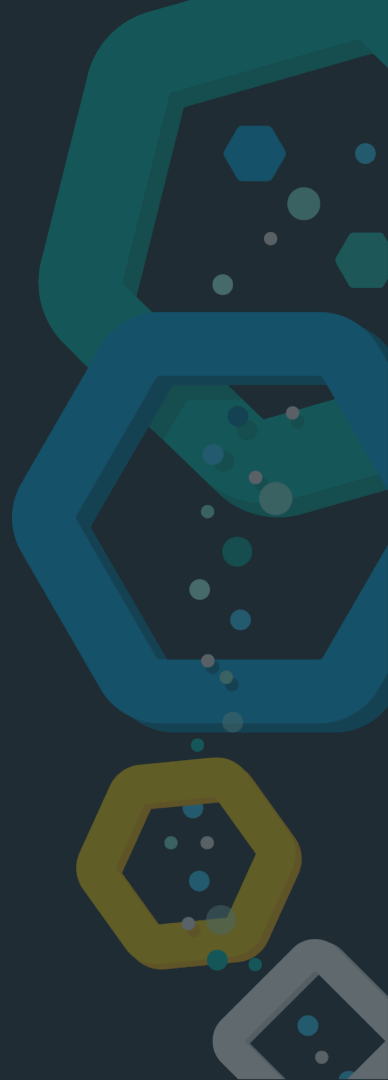
elastic

good data
&
worst searches



elastic

Agenda



elastic

Agenda

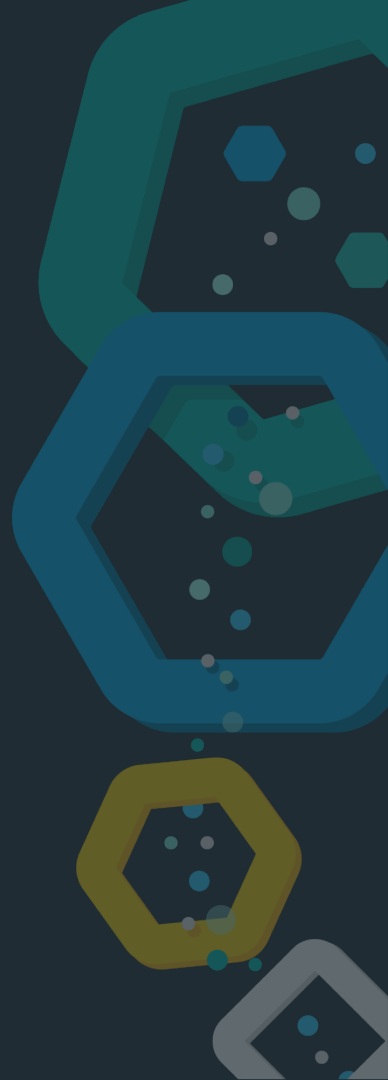
facetted navigation

elastic

Agenda

facetted navigation

search bar



elastic

Agenda

facetted navigation

clean data

search bar

elastic

Agenda

clean data

facetted navigation

smart searches

search bar

elastic

Agenda

clean data

facetted navigation

synonyms

smart searches

search bar

elastic

Agenda

clean data

facetted navigation

UOM

synonyms

smart searches

search bar

elastic

Agenda

clean data

facetted navigation

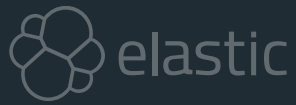
synonyms

decompounding

UOM

smart searches

search bar



Agenda

clean data

facetted navigation

relevancy

synonyms

UOM

decompounding

smart searches

search bar

elastic

Agenda

clean data

facetted navigation

synonyms

variants

UOM

relevancy

decompounding

smart searches

search bar

elastic

Agenda

clean data

facetted navigation

variants

deduplication

synonyms

UOM

relevancy

decompounding

smart searches

search bar

elastic

Agenda

clean data

facetted navigation

variants

deduplication

synonyms

search as you type

UOM

relevancy

decompounding

smart searches

search bar

elastic

Agenda

clean data

facetted navigation

variants

deduplication

analytics

synonyms

UOM

relevancy

decompounding

smart searches

search bar

search as you type

Agenda

analytics

clean data

facetted navigation

variants

deduplication

data quality

synonyms

UOM

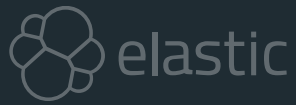
relevancy

decompounding

smart searches

search bar

search as you type



Agenda

analytics

clean data

facetted navigation

variants

deduplication

mobile

synonyms

data quality

UOM

relevancy

decompounding

smart searches

search bar

search as you type

Agenda

analytics

clean data

facetted navigation

variants

deduplication

product detail page

synonyms

data quality

UOM

mobile

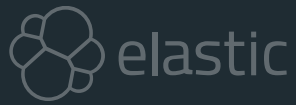
relevancy

decompounding

smart searches

search bar

search as you type



Agenda

product detail page

analytics

clean data

facetted navigation

variants

deduplication

LTR

synonyms

data quality

UOM

mobile

relevancy

decompounding

smart searches

search bar

search as you type

Agenda

product detail page

analytics

clean data

facetted navigation

variants

deduplication

multi language

synonyms

data quality

UOM

mobile

relevancy

decompounding

smart searches

search bar

LTR

search as you type

Agenda

product detail page

analytics

clean data

facetted navigation

variants

deduplication

ETIME

synonyms

multi language

data quality

mobile

UOM

relevancy

decompounding

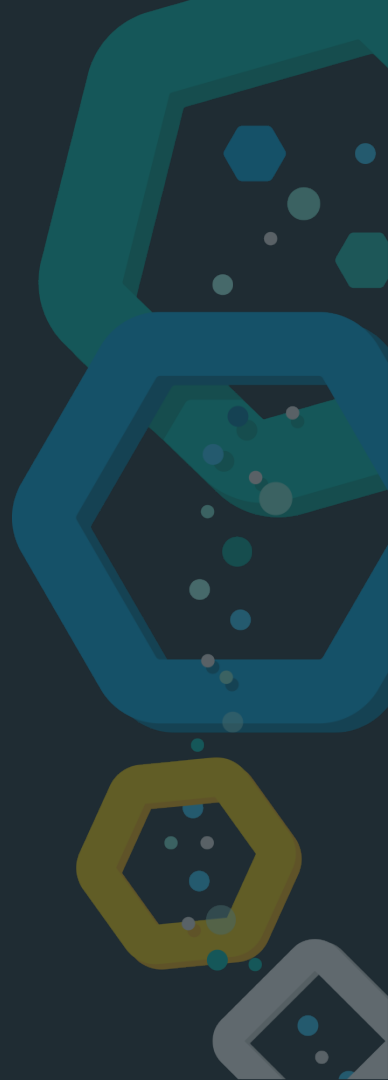
smart searches

search bar

LTR

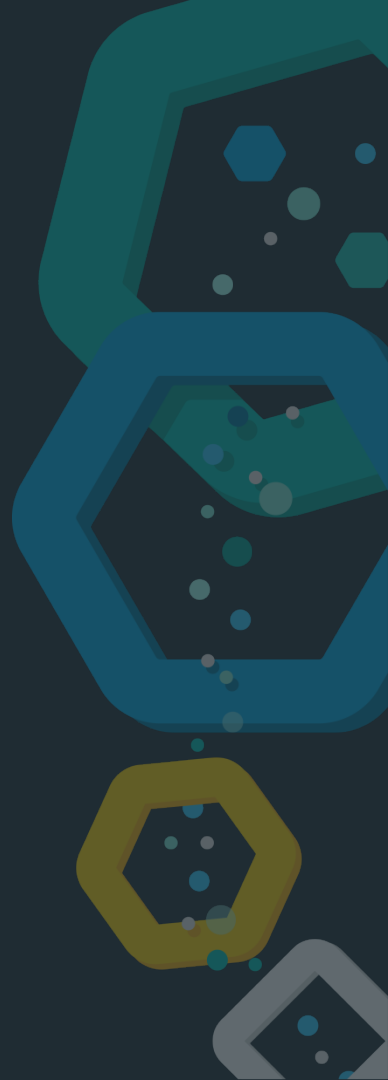
search as you type

demo



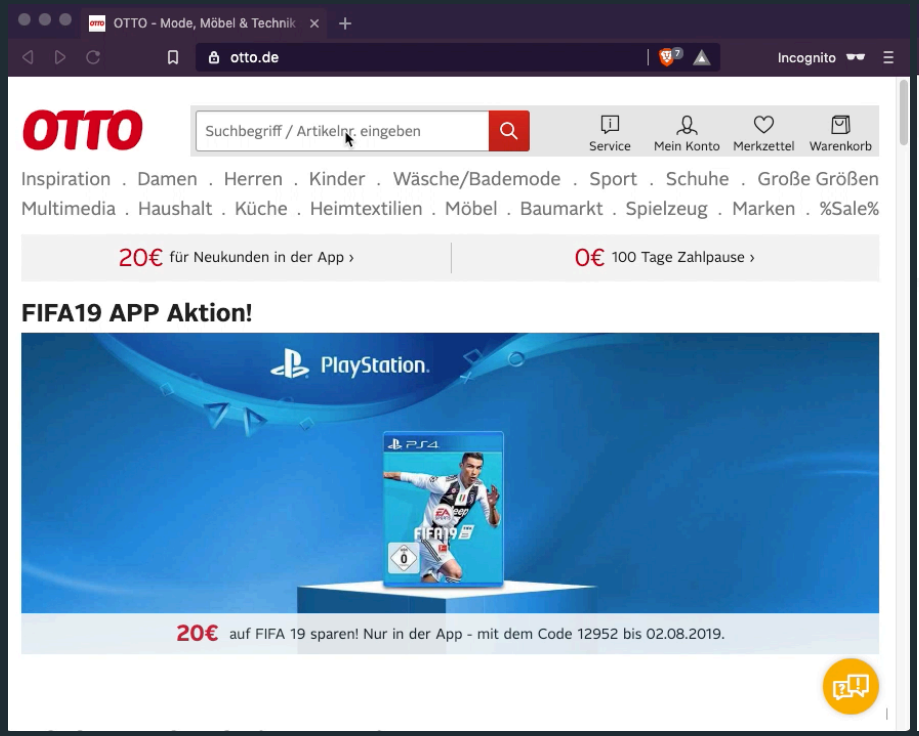
elastic

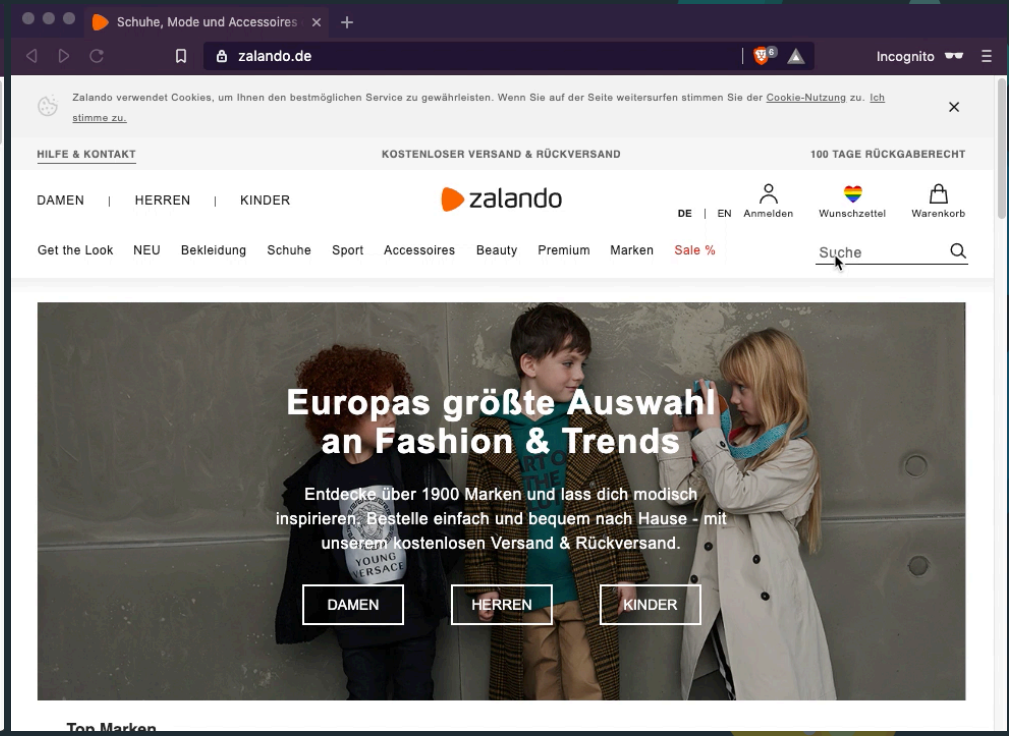
search bar



elastic

search bar





elastic

smart searches

elastic

smart searches

nike running hoodie xl

elastic

smart searches

nike running hoodie xl

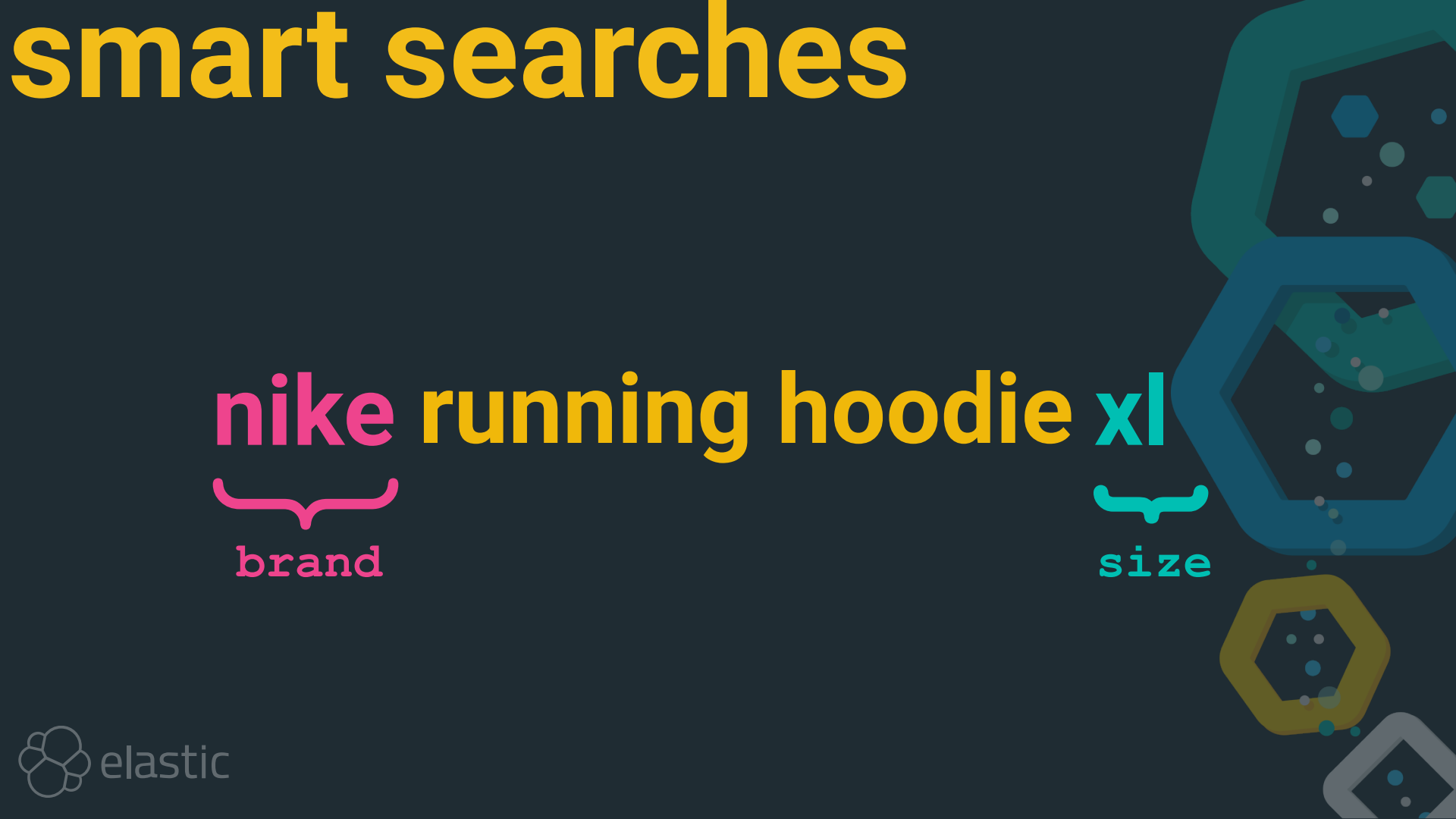
elastic

smart searches

nike running hoodie xl

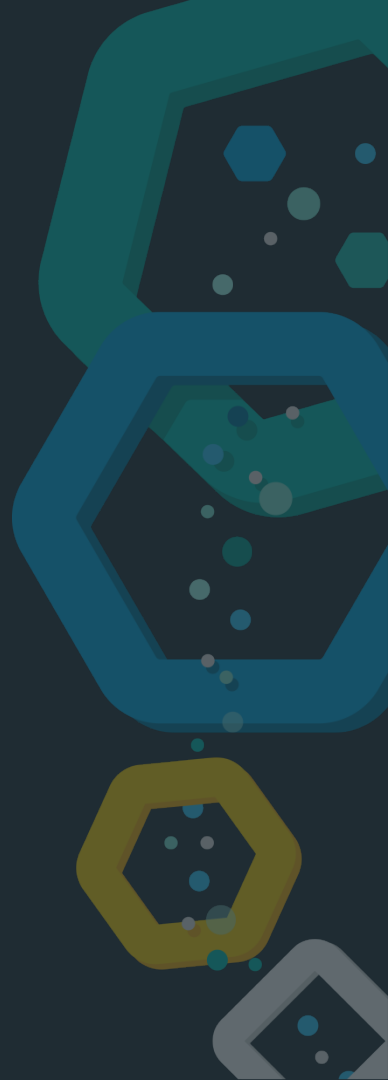
brand

size



elastic

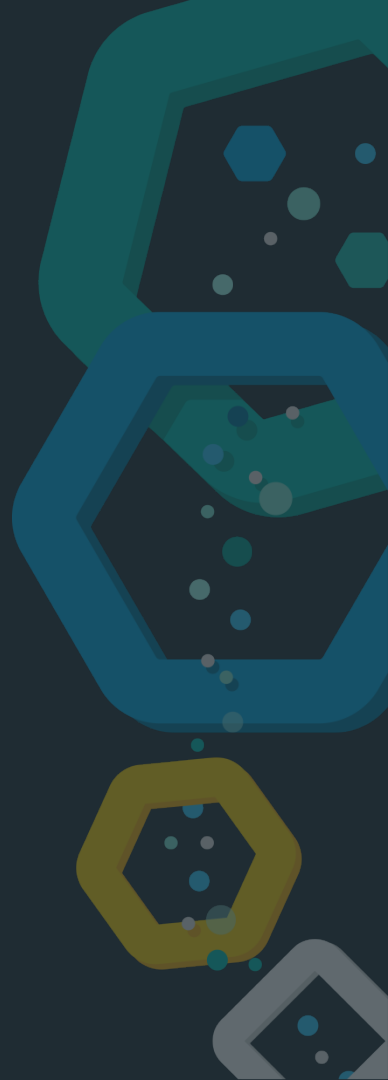
clean data



elastic

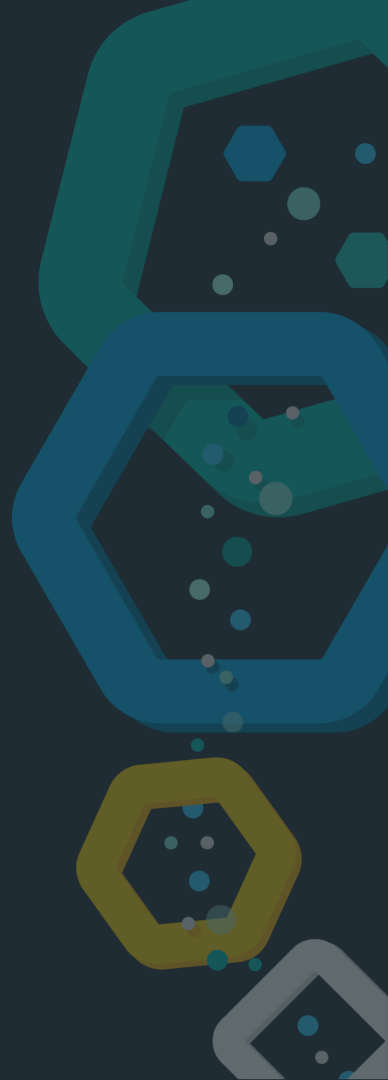
clean data

- 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?



elastic

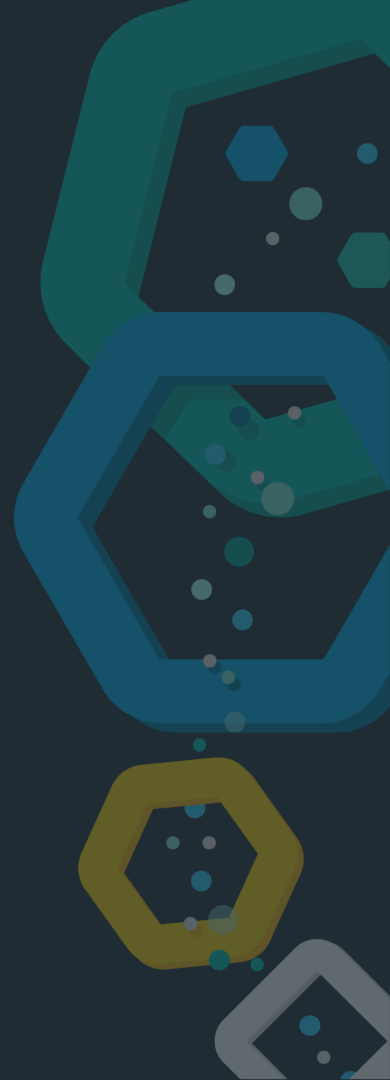
synonyms



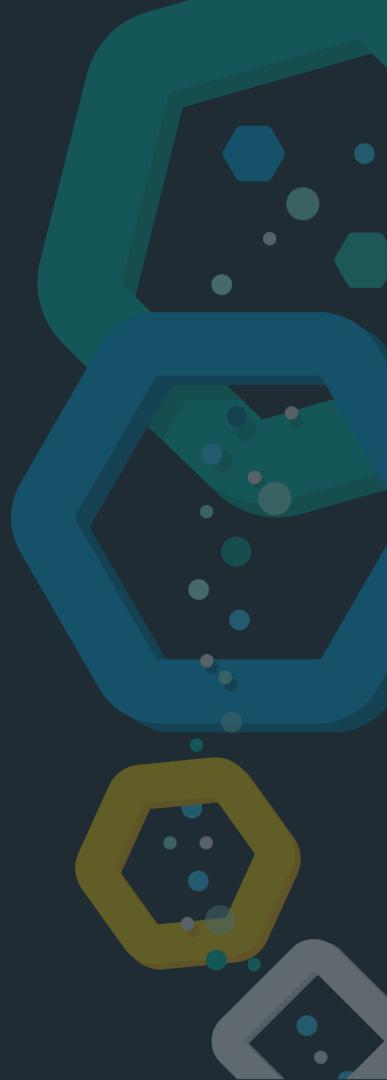
elastic

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

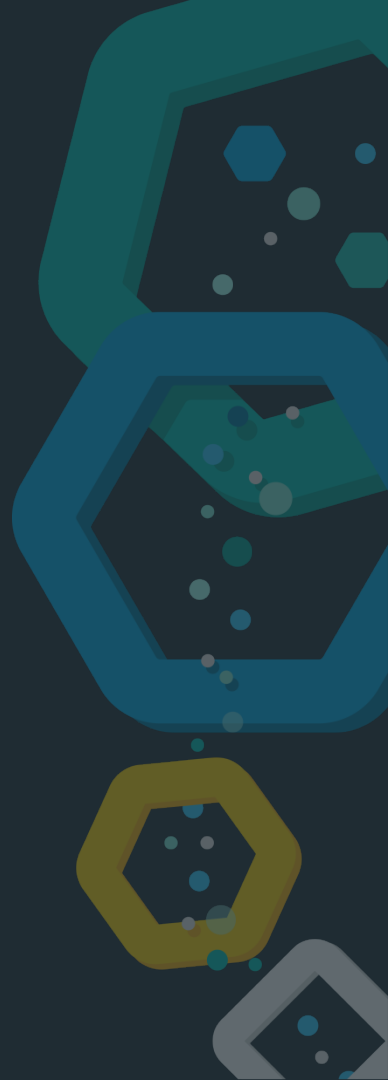


elastic

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

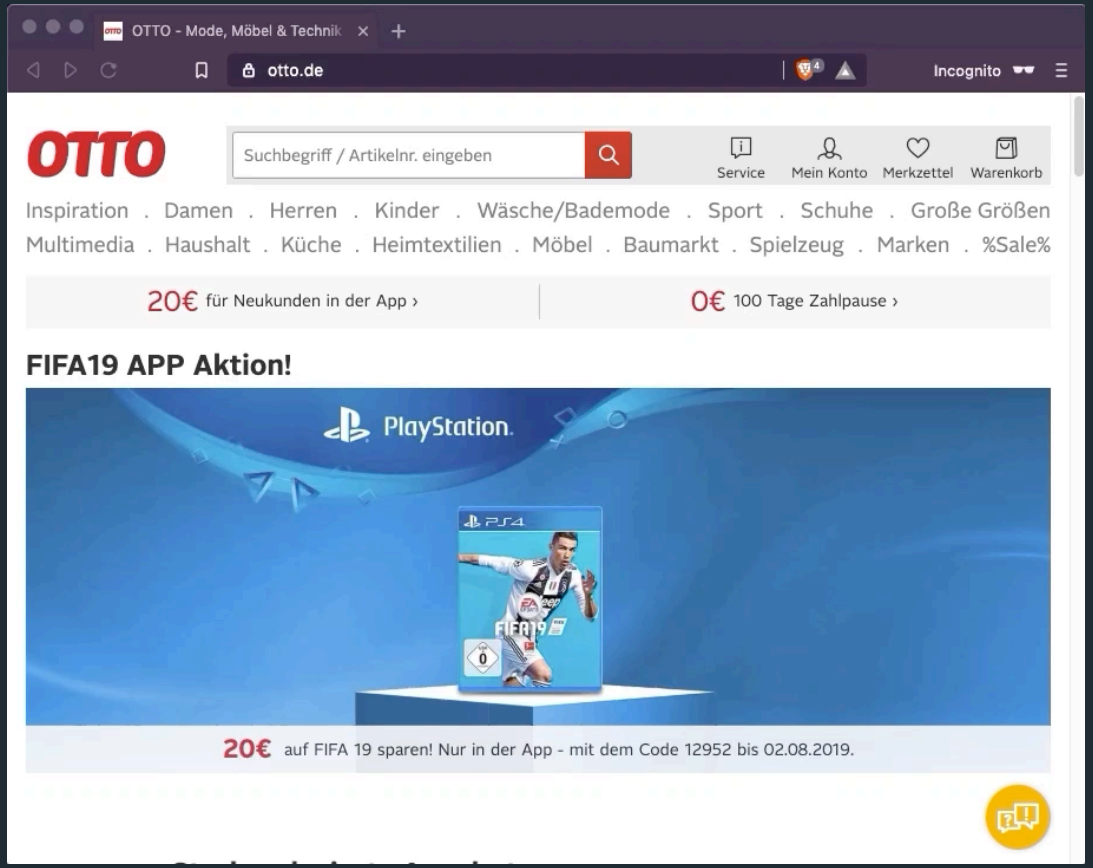


elastic

decompounding

elastic

decompounding



elastic

decompounding

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

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

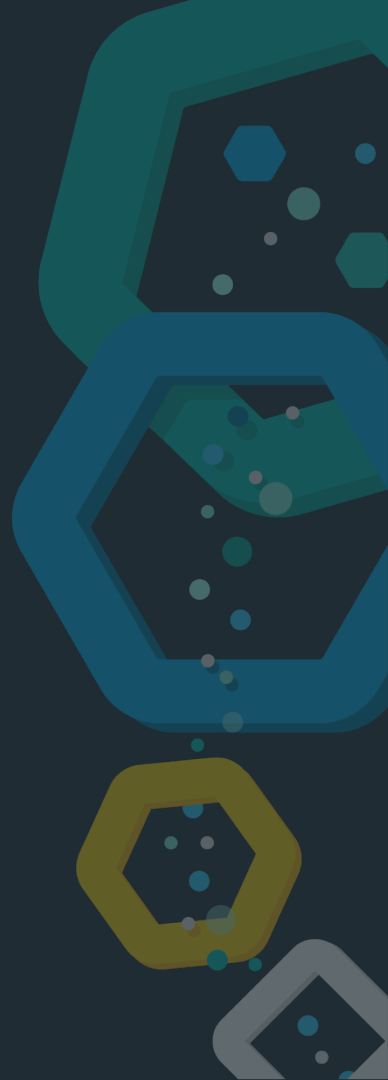
decompounding

```
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"
    }
```

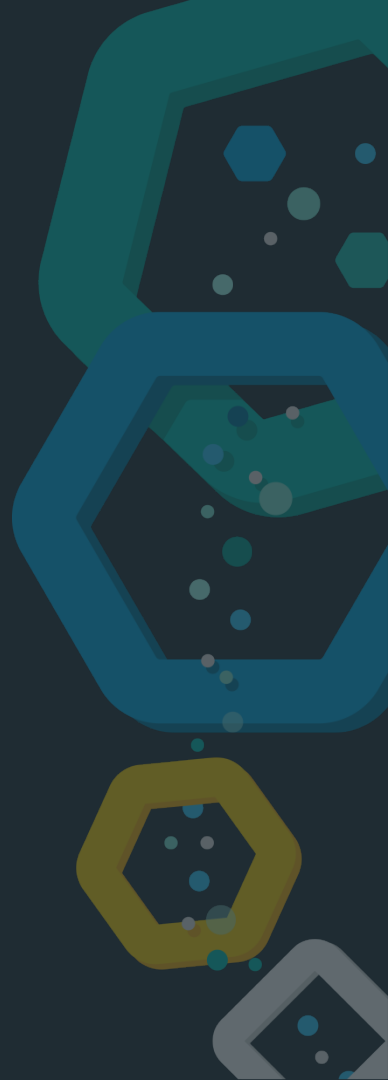
relevancy



elastic

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



elastic

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?

elastic

relevancy

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

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

elastic

relevancy

Rank feature datatype

edit

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

edit

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

edit

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.

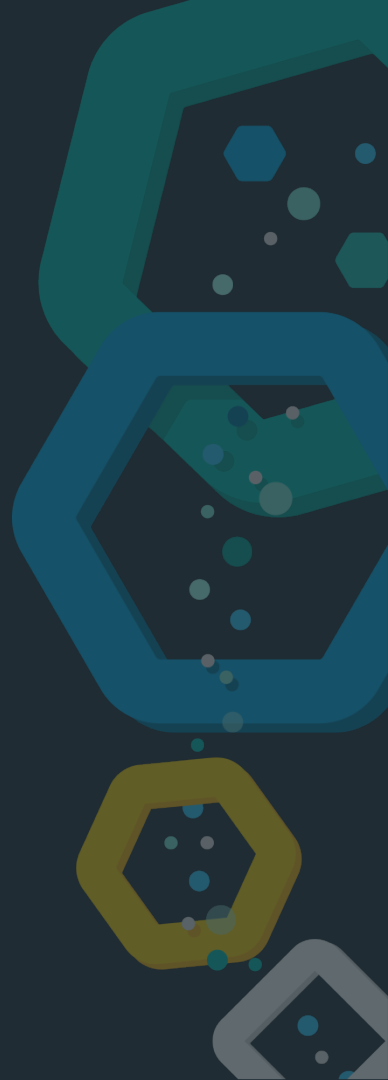
A `dense_vector` field stores dense vectors of float values. The maximum number of dimensions that can be in a vector should not exceed 1024. A `dense_vector` field is a single-valued field.

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.

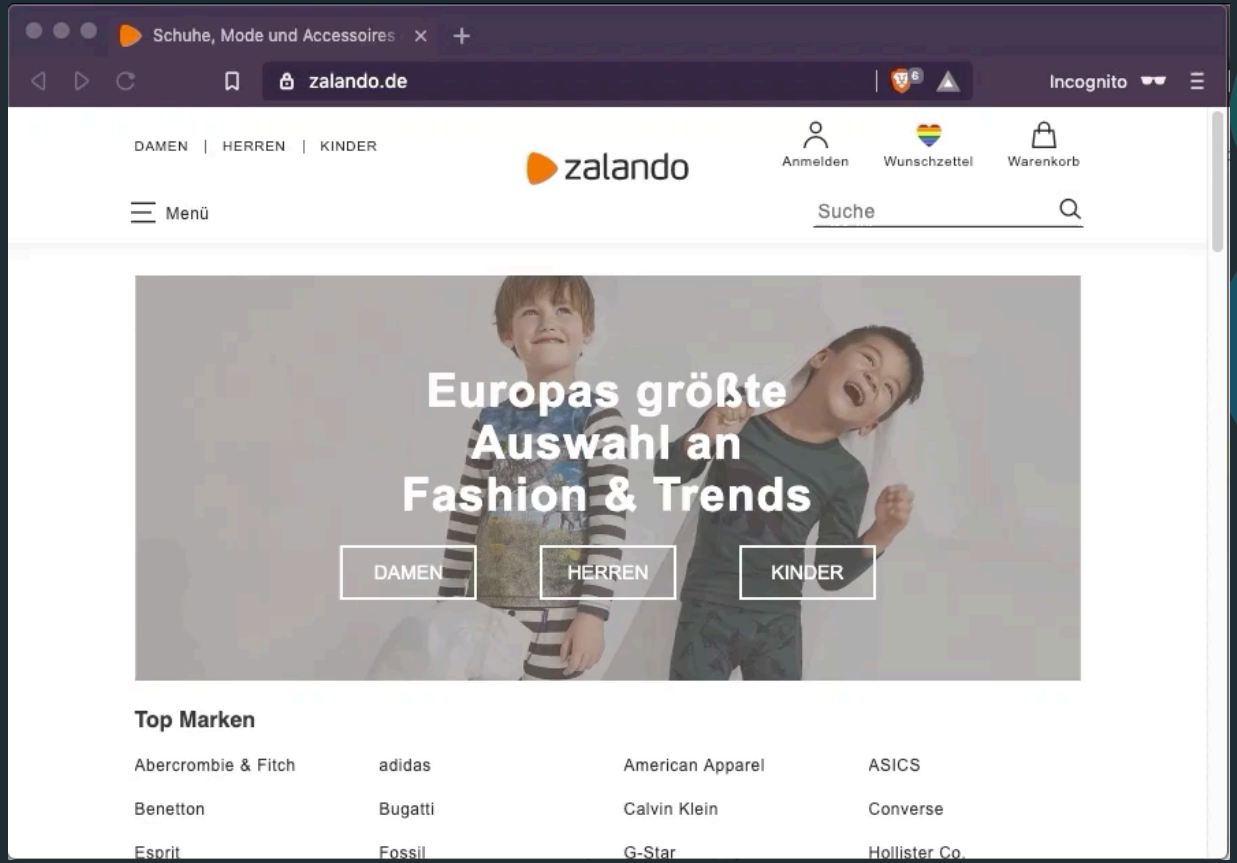
elastic

variants



elastic

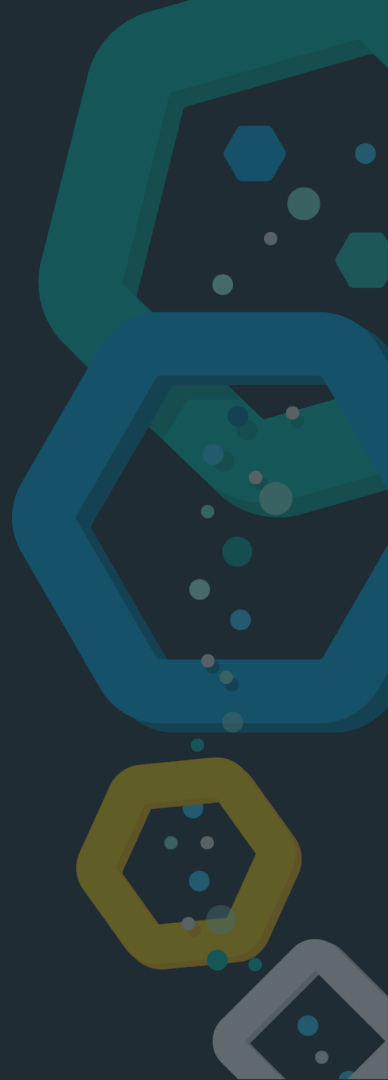
variants



elastic

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?



elastic

variants

Join datatype

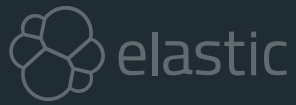
edit

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:

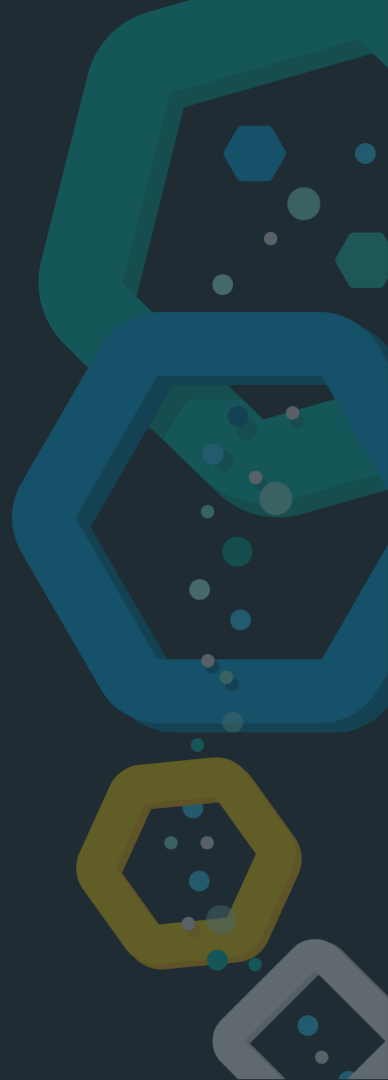
Inner hits

edit

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.



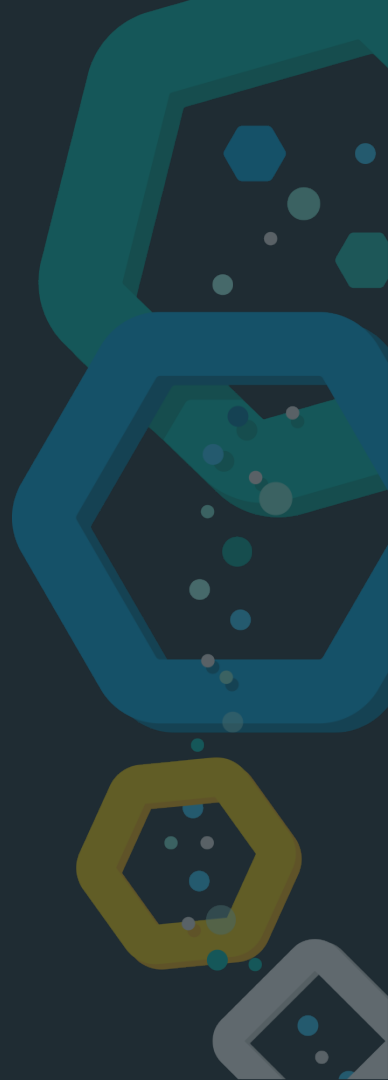
deduplication



elastic

deduplication

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



elastic

deduplication

Field Collapsing

edit

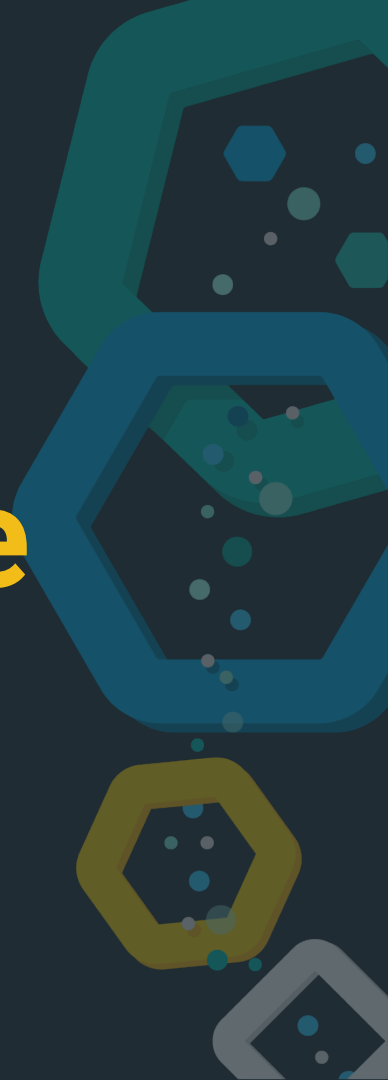
Allows to collapse search results based on field values. The collapsing is done by selecting only the top sorted document per collapse key. For instance the query below retrieves the best tweet for each user and sorts them by number of likes.

```
GET /twitter/_search
{
    "query": {
        "match": {
            "message": "elasticsearch"
        }
    },
    "collapse" : {
        "field" : "user" ❶
    },
    "sort": ["likes"], ❷
    "from": 10 ❸
}
```

COPY AS CURL VIEW IN CONSOLE

elastic

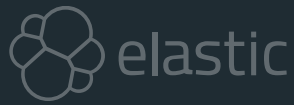
search as you type



elastic

search as you type

- "The importance of seach-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

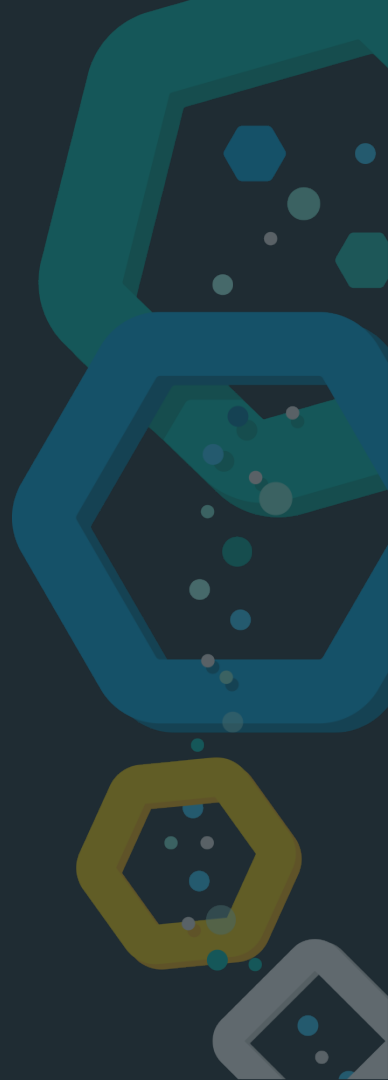


search as you type

Search-as-you-type datatype

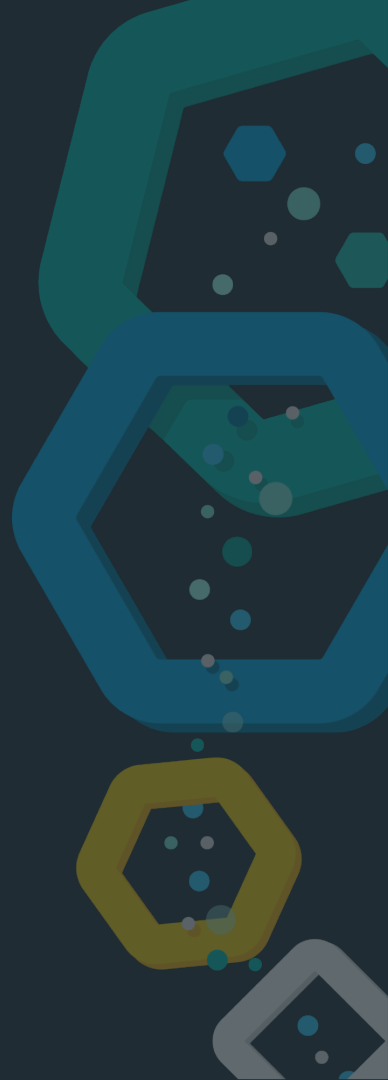
edit

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.



elastic

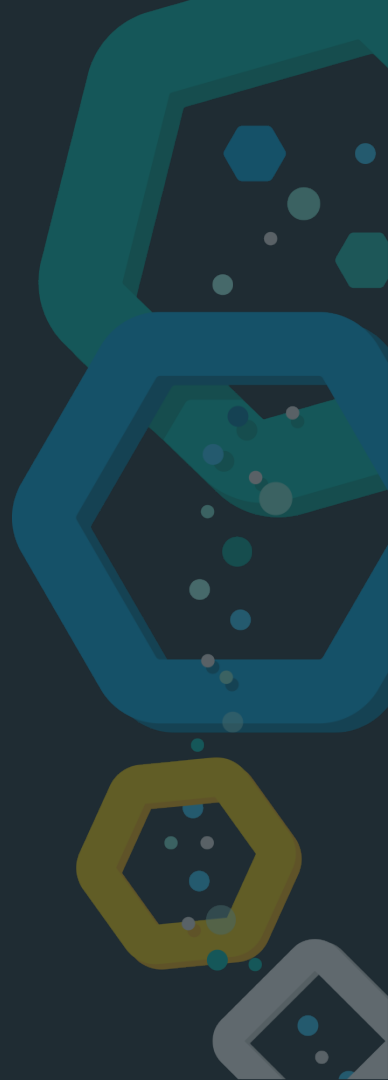
analytics



elastic

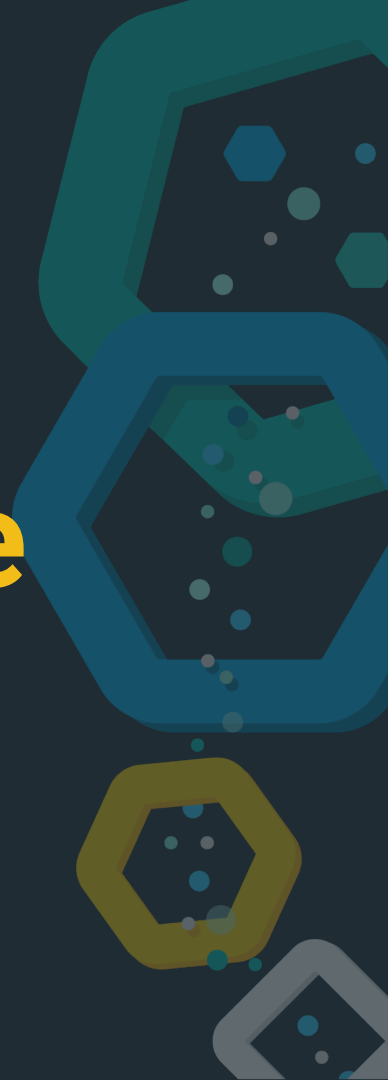
analytics

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



elastic

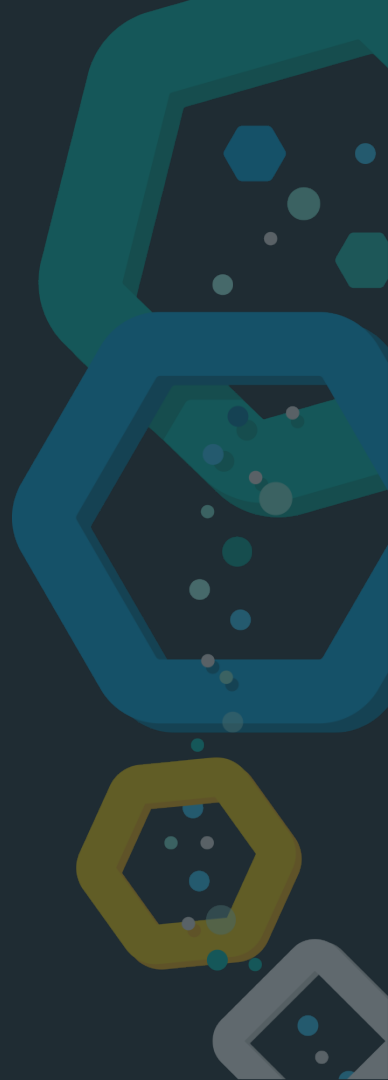
product detail page



elastic

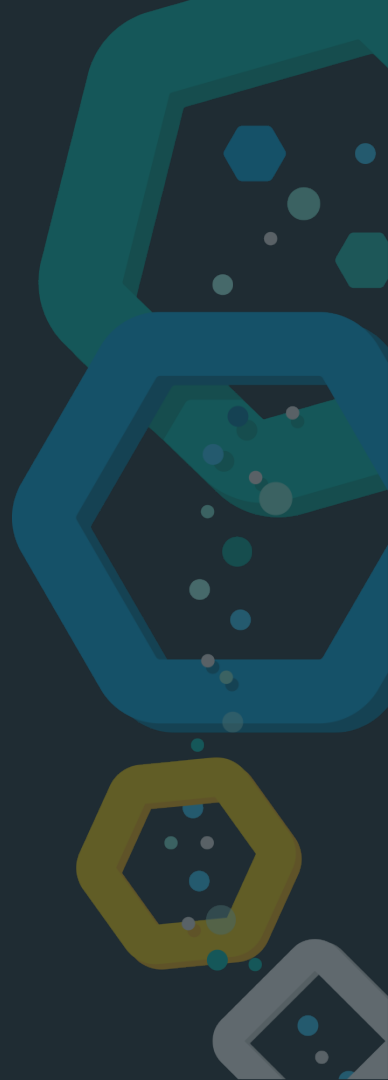
product detail page

- 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



elastic

LTR



elastic

LTR

Doug Turnbull
@softwaredoug

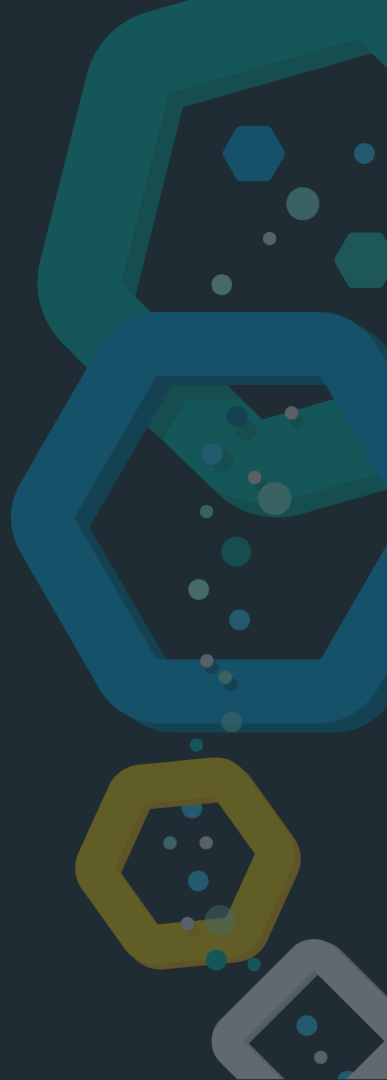
Everytime someone contacts @o19s and asks for "learning to rank" or some other ML relevance thing often they haven't done step 1 (some basic relevance work)

> 
> **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

elastic

summary



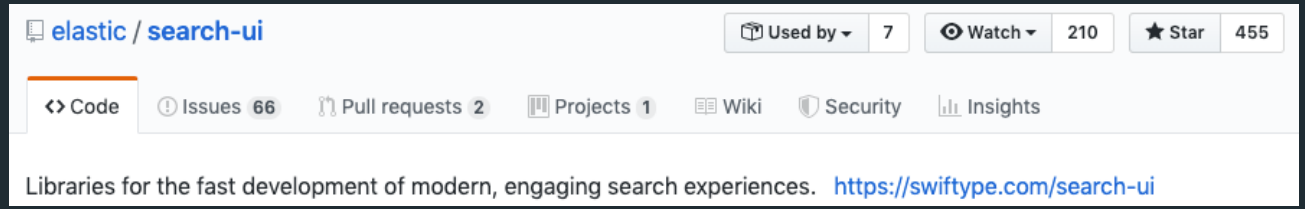
elastic

summary

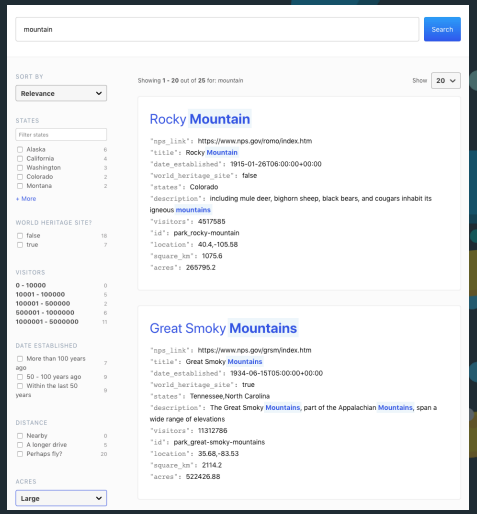
- ecommerce search is complex
- so many things to take into account...
- untold: index strategies, updates, management
- always have a middleware (UI, query injection, a/b testing, landing pages, redirects, query logging, business owner endpoint)

elastic

search ui

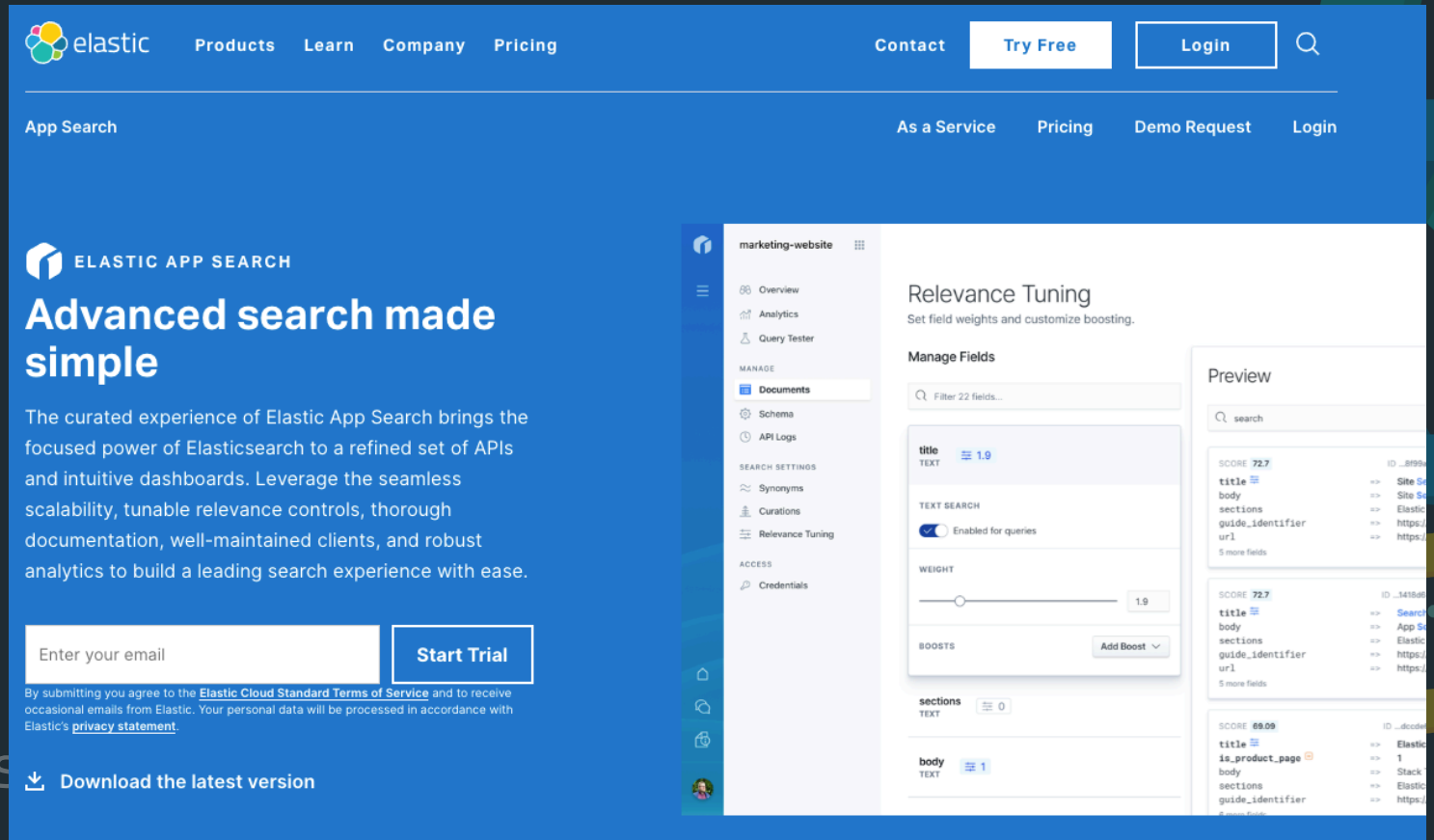


https://github.com/elastic/search-ui

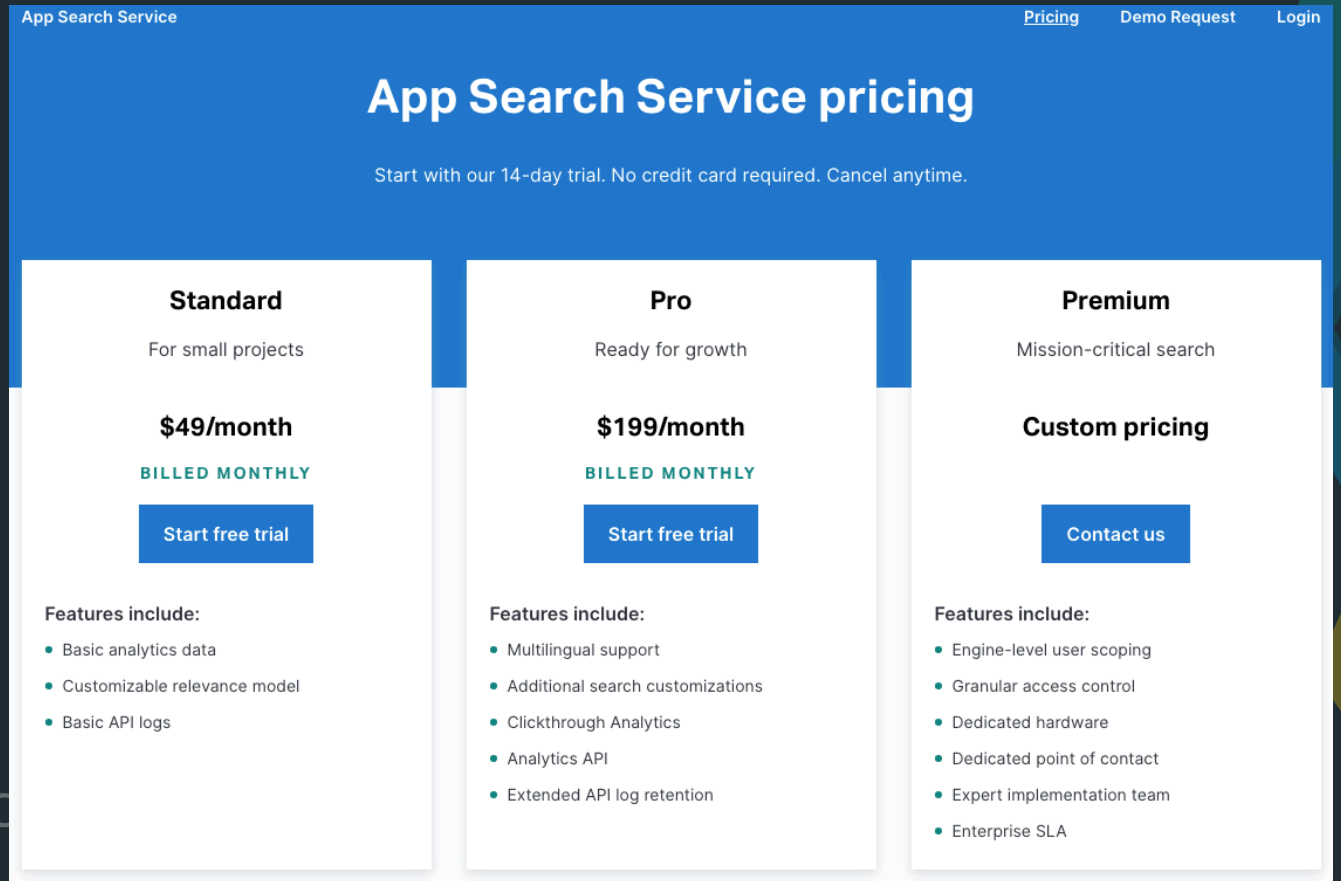


elastic

Elastic App Search

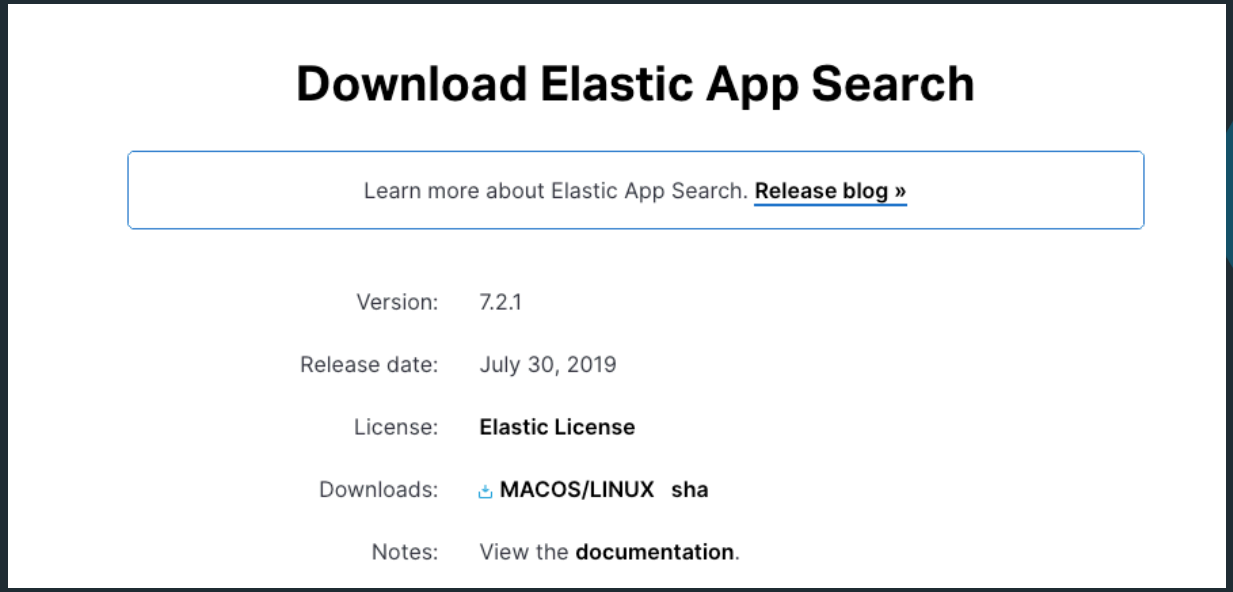


Elastic App Search



elastic

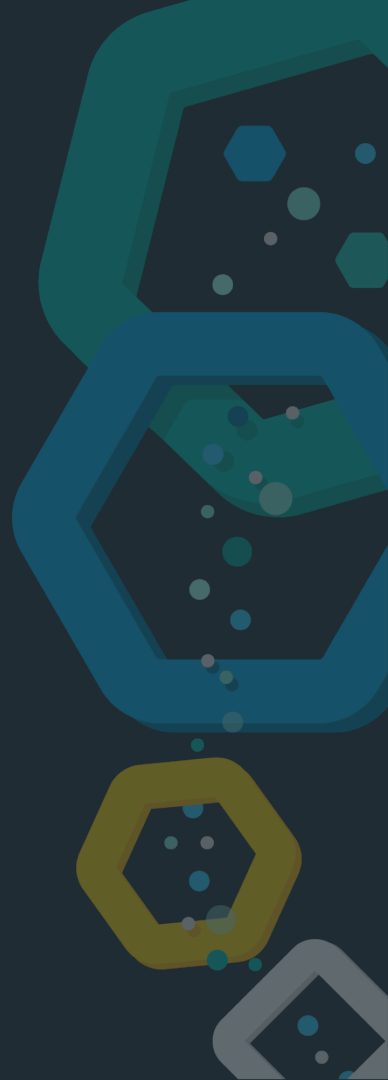
Elastic App Search



elastic

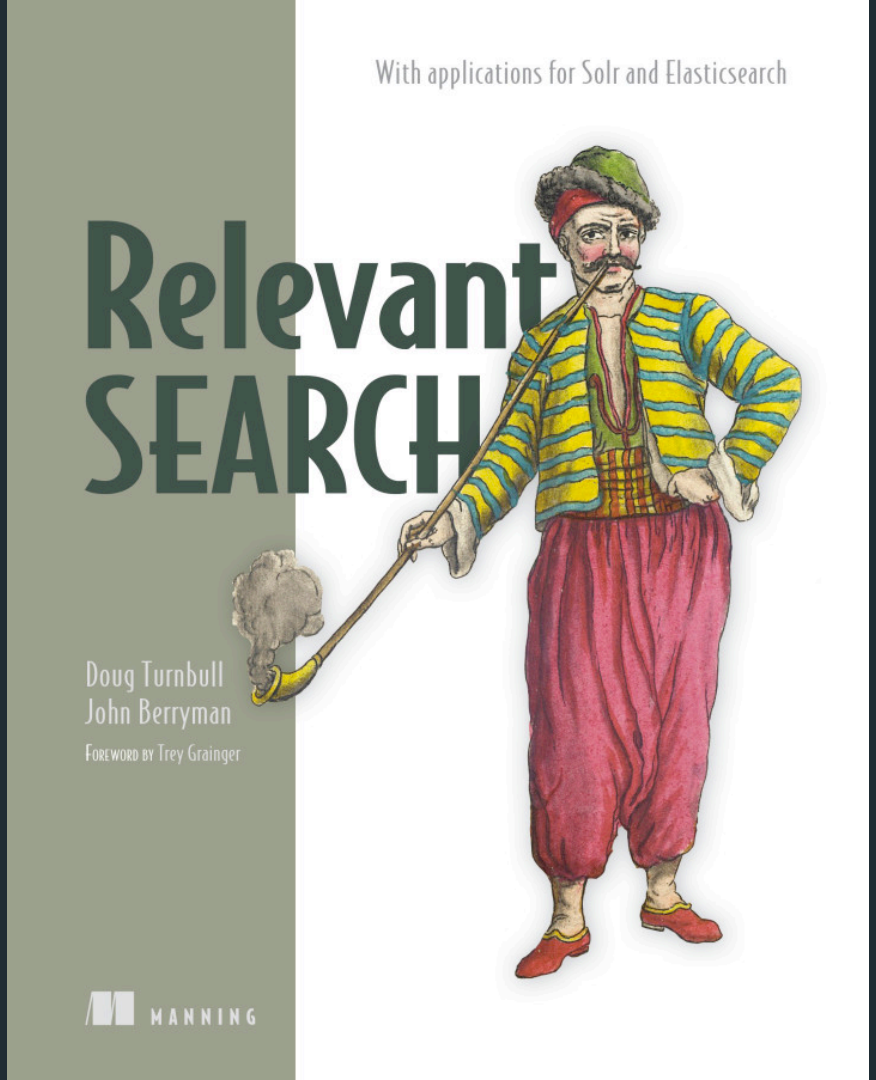
https://www.elastic.co/blog/elastic-app-search-7-2-0-released

books



elastic

books



elastic

books

DEEP LEARNING for Search

Tommaso Teofili

Foreword by Chris Mattmann

MANNING

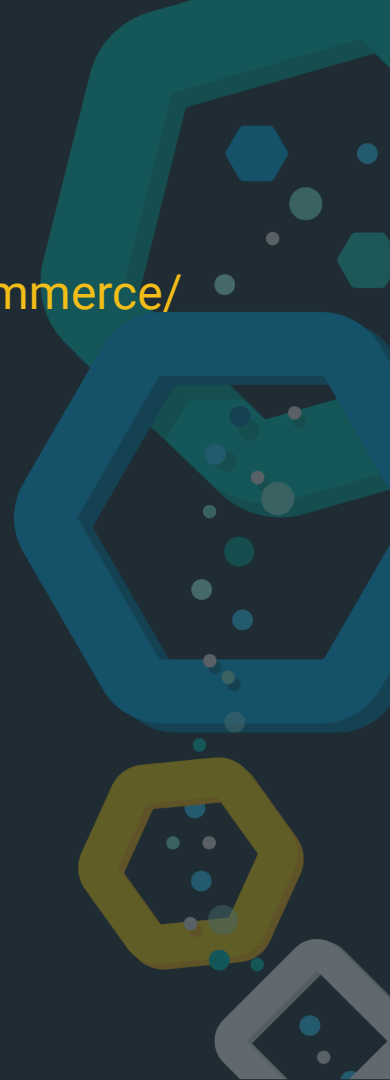
elastic

links

elastic

links

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



elastic

Thank you for listening!

Alexander Reelsen
@spinscale
alex@elastic.co

elastic