

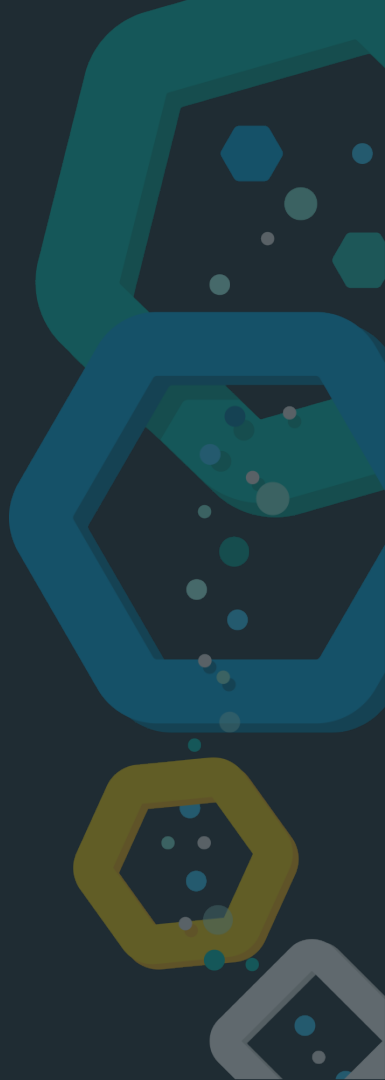


Finding e-commerce products using Elasticsearch

Alexander Reelsen | @spinscale
alex@elastic.co



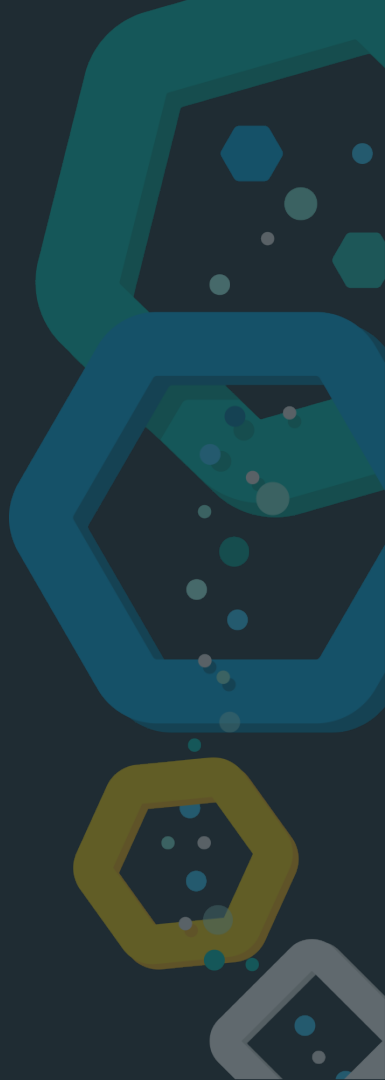
search is hard!



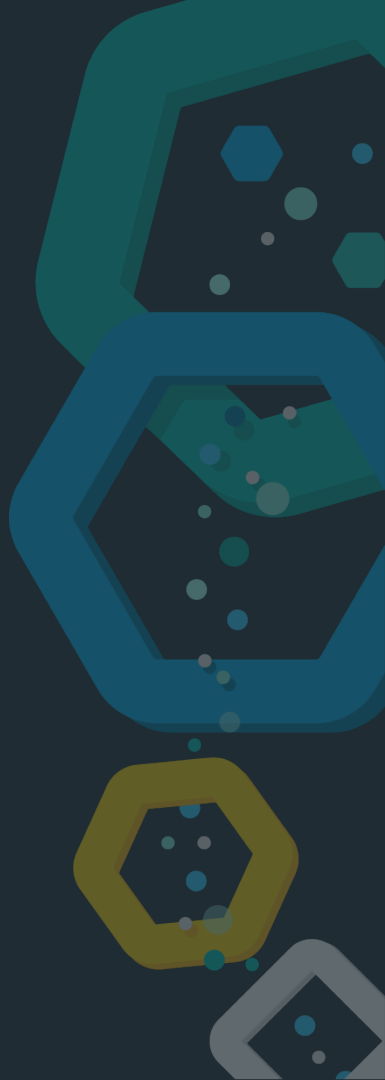
**search in
ecommerce is harder**



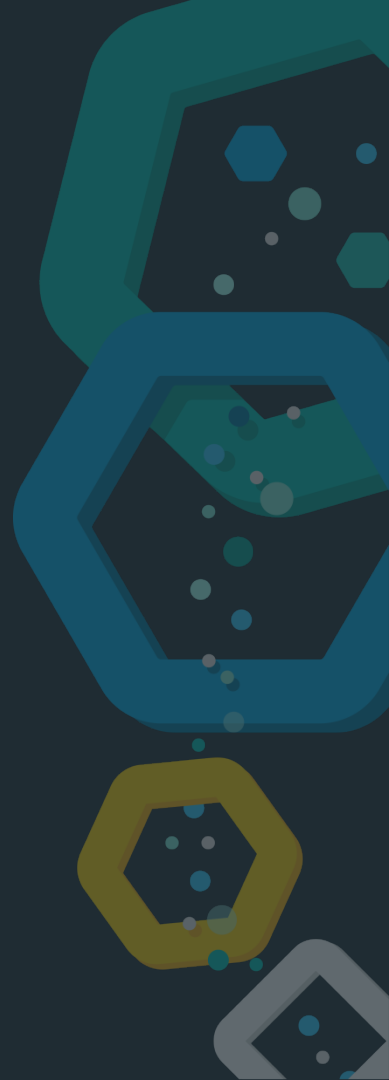
good data & good searches



bad data & smart searches



good data & worst searches



Agenda

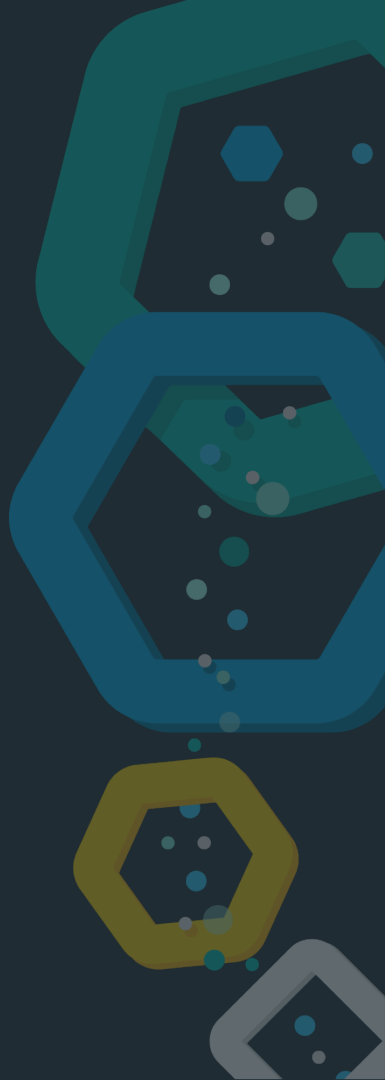
Agenda

facetted navigation

Agenda

facetted navigation

search bar

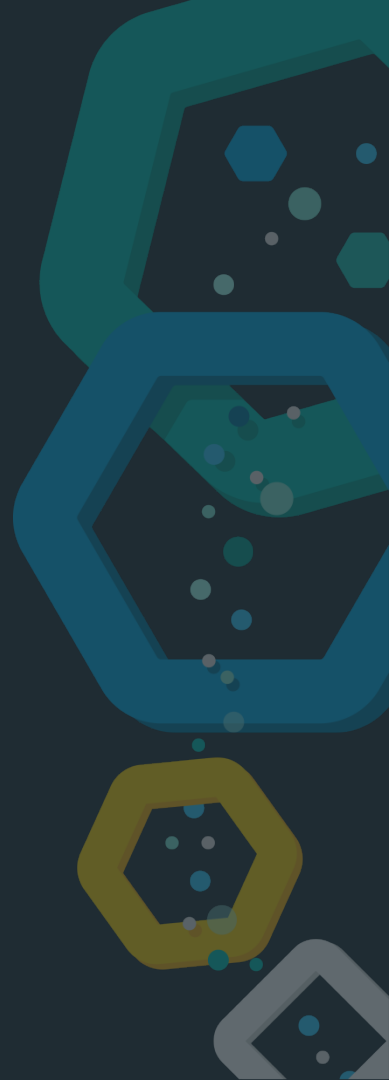


Agenda

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

faceted navigation

relevancy

synonyms

UOM

decompounding

smart searches

search bar

Agenda

clean data

faceted navigation

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

facetted navigation

variants

deduplication

analytics

synonyms

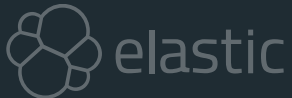
UOM

relevancy

decompounding

smart searches

search bar



search as you type



Agenda

analytics

clean data

faceted navigation

variants

deduplication

data quality

synonyms

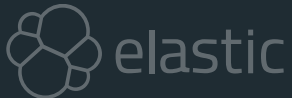
UOM

relevancy

decompounding

smart searches

search bar



search as you type



Agenda

analytics

clean data

faceted navigation

variants

deduplication

mobile

synonyms

data quality

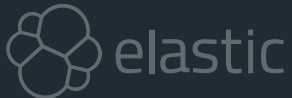
UOM

relevancy

decompounding

smart searches

search bar



search as you type



Agenda

product detail page

analytics

clean data

facetted 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

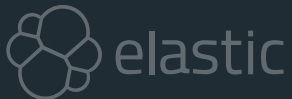
data quality

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

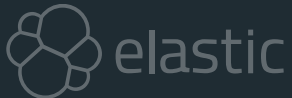
relevancy

decompounding

smart searches

search bar

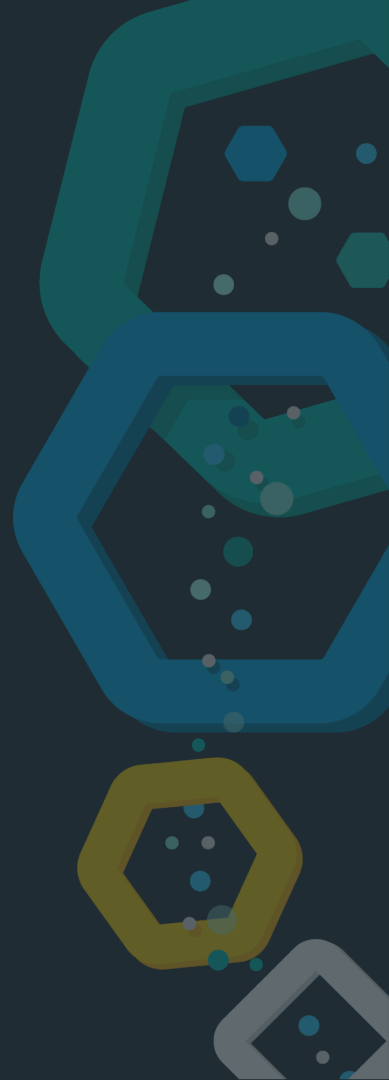
LTR



search as you type



demo



search bar

search bar

The screenshot shows the OTTO website's search bar and navigation menu. The search bar is located at the top left, with the OTTO logo to its left. The search bar contains the placeholder text "Suchbegriff / Artikelnr. eingeben" and a magnifying glass icon. To the right of the search bar are icons for "Service", "Mein Konto", "Merkzettel", and "Warenkorb". Below the search bar is a navigation menu with links for "Inspiration", "Damen", "Herren", "Kinder", "Wäsche/Bademode", "Sport", "Schuhe", "Große Größen", "Multimedia", "Haushalt", "Küche", "Heimtextilien", "Möbel", "Baumarkt", "Spielzeug", "Marken", and "%Sale%". Below the navigation menu are two promotional banners: "20€ für Neukunden in der App" and "0€ 100 Tage Zahlpause". Below these banners is a "FIFA19 APP Aktion!" banner featuring a PlayStation logo and a FIFA 19 game cover. The banner text reads "20€ auf FIFA 19 sparen! Nur in der App - mit dem Code 12952 bis 02.08.2019." and includes a chat icon in the bottom right corner.

The screenshot shows the Zalando website's search bar and navigation menu. The search bar is located at the top left, with the Zalando logo to its right. The search bar contains the placeholder text "Suchbegriff / Artikelnr. eingeben" and a magnifying glass icon. To the right of the search bar are icons for "Service", "Mein Konto", "Merkzettel", and "Warenkorb". Below the search bar is a navigation menu with links for "HILFE & KONTAKT", "KOSTENLOSER VERSAND & RÜCKVERSAND", "100 TAGE RÜCKGABERECHT", "DAMEN", "HERREN", "KINDER", "zalando", "DE | EN", "Anmelden", "Wunschzettel", and "Warenkorb". Below the navigation menu are two promotional banners: "20€ für Neukunden in der App" and "0€ 100 Tage Zahlpause". Below these banners is a "FIFA19 APP Aktion!" banner featuring a PlayStation logo and a FIFA 19 game cover. The banner text reads "20€ auf FIFA 19 sparen! Nur in der App - mit dem Code 12952 bis 02.08.2019." and includes a chat icon in the bottom right corner.

The banner features a photograph of three children (two boys and one girl) standing in a hallway, looking at each other. The text on the banner reads "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 text are three buttons labeled "DAMEN", "HERREN", and "KINDER".

Top Marken

smart searches

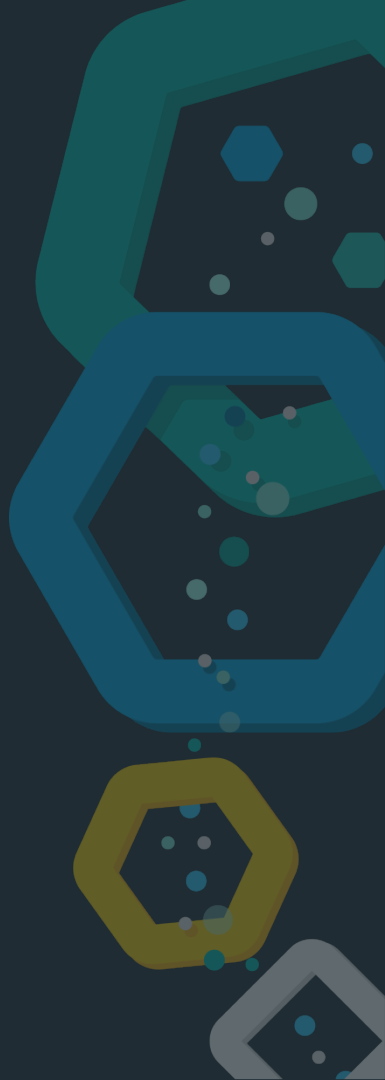
smart searches

nike running hoodie xl



smart searches

nike running hoodie xl



smart searches

nike running hoodie **xl**

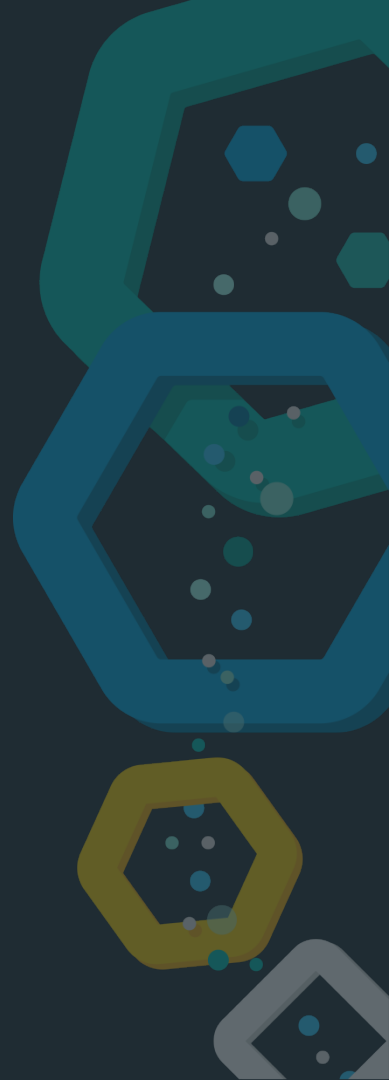


brand



size

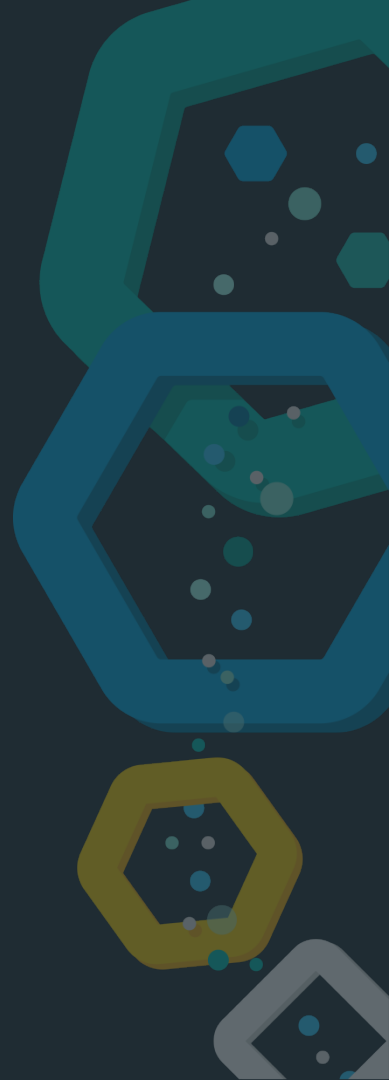
clean data



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?

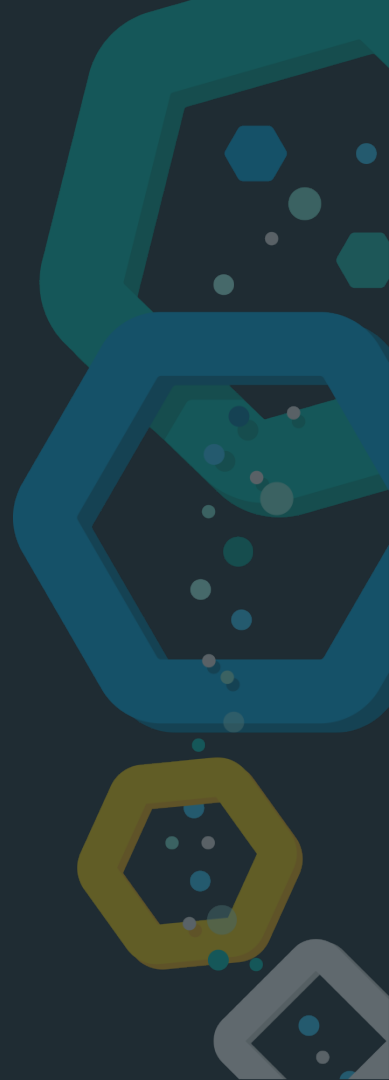
synonyms



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)

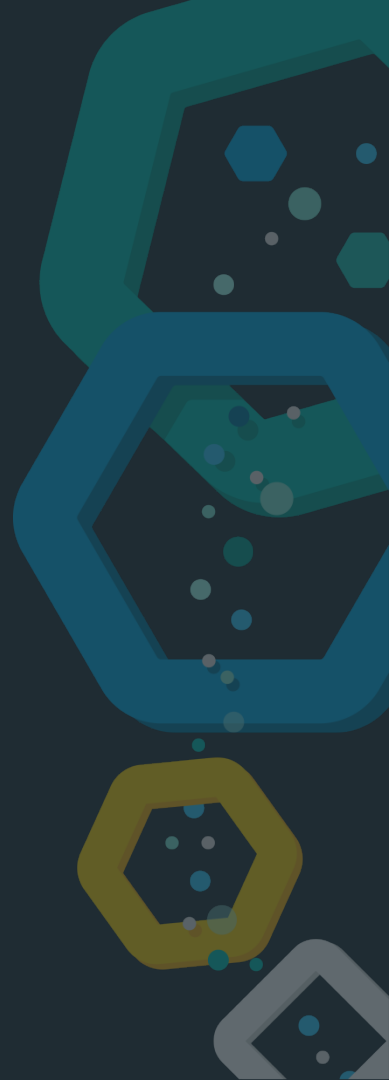
UOM



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 interface. At the top, the OTTO logo is on the left, followed by a search bar with the placeholder text "Suchbegriff / Artikelnr. eingeben". To the right of the search bar are navigation icons 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. The banner has a blue background with a PlayStation logo and the text "PlayStation." at the top. In the center, a FIFA 19 PS4 game case is displayed on a white pedestal. At the bottom of the banner, the text reads "20€ auf FIFA 19 sparen! Nur in der App - mit dem Code 12952 bis 02.08.2019." A yellow speech bubble icon is located in the bottom right corner of the page.

OTTO - Mode, Möbel & Technik x +

otto.de Incognito

OTTO Suchbegriff / Artikelnr. eingeben Service Mein Konto Merktzettel Warenkorb

Inspiration · Damen · Herren · Kinder · Wäsche/Bademode · Sport · Schuhe · Große Größen
Multimedia · Haushalt · Küche · Heimtextilien · Möbel · Baumarkt · Spielzeug · Marken · %Sale%

20€ für Neukunden in der App > 0€ 100 Tage Zahlpause >

FIFA19 APP Aktion!

PlayStation.

20€ auf FIFA 19 sparen! Nur in der App - mit dem Code 12952 bis 02.08.2019.

decompounding

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

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

decompounding

PUT example

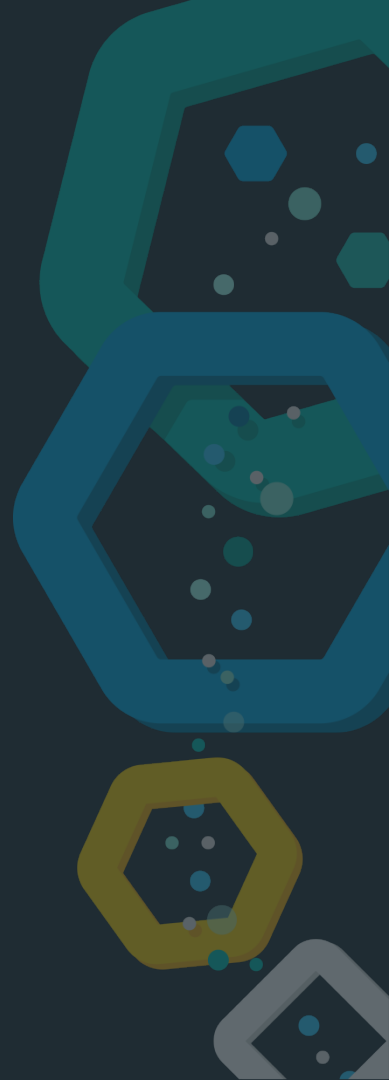
```
{
  "settings": {
    "index": {
      "analysis": {
        "analyzer": {
          "decompound_analyzer": {
            "type": "custom",
            "tokenizer": "standard",
            "filter": [ "dictionary_decompounder" ]
          }
        },
        "filter": {
          "dictionary_decompounder":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



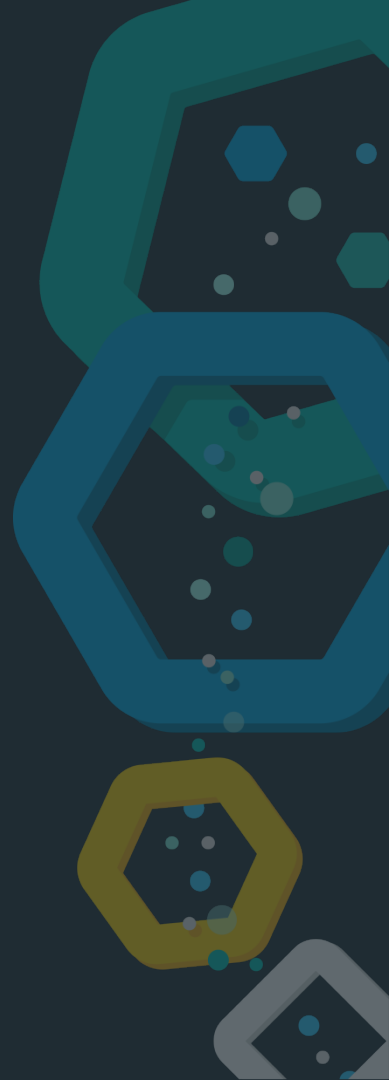
This functionality is experimental and may be changed or removed completely in a future release. Elastic will take a best effort approach to fix any issues, but experimental features are not subject to the support SLA of official GA features.

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. At the top, there is a navigation bar with links for 'DAMEN', 'HERREN', and 'KINDER'. The Zalando logo is prominently displayed in the center. To the right, there are icons for 'Anmelden', 'Wunschzettel', and 'Warenkorb'. A search bar with the text 'Suche' and a magnifying glass icon is located below the navigation. The main banner features a photograph of two children and the text 'Europas größte Auswahl an Fashion & Trends'. Below the banner, there are three buttons labeled 'DAMEN', 'HERREN', and 'KINDER'. Underneath the banner is a section titled 'Top Marken' with a grid of brand names: Abercrombie & Fitch, adidas, American Apparel, ASICS, Benetton, Bugatti, Calvin Klein, Converse, Esprit, Fossil, G-Star, and Hollister Co.

Schuhe, Mode und Accessoires x +

zalando.de

Incognito

DAMEN | HERREN | KINDER

zalando

Anmelden Wunschzettel Warenkorb

Menü Suche

Europas größte Auswahl an Fashion & Trends

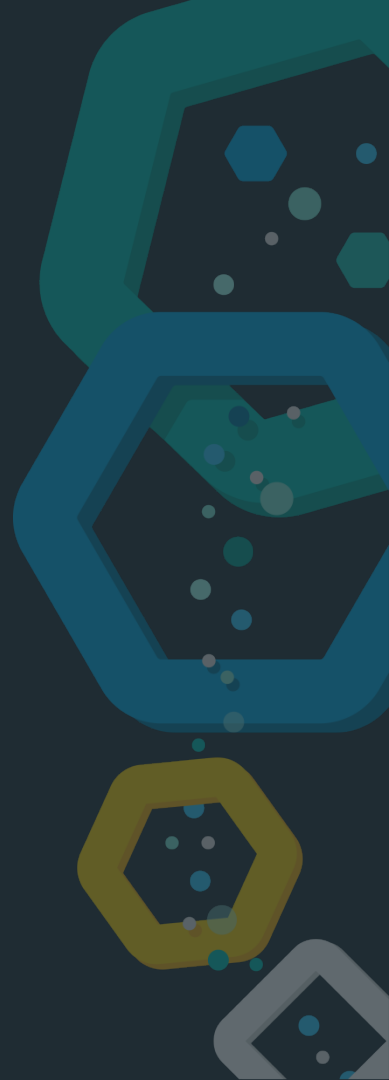
DAMEN HERREN KINDER

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



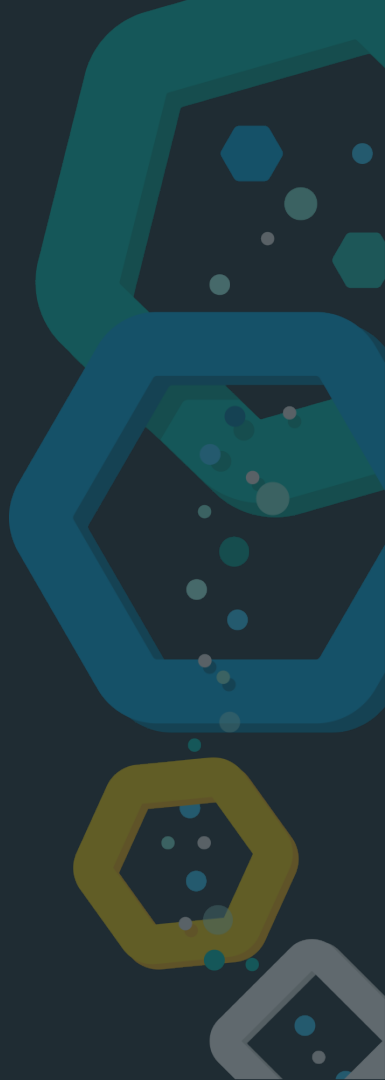
The `join` datatype is a special field that creates parent/child relation within documents of the same index. The `relations` section defines a set of possible relations within the documents, each relation being a parent name and a child name. A parent/child relation can be defined as follows:

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

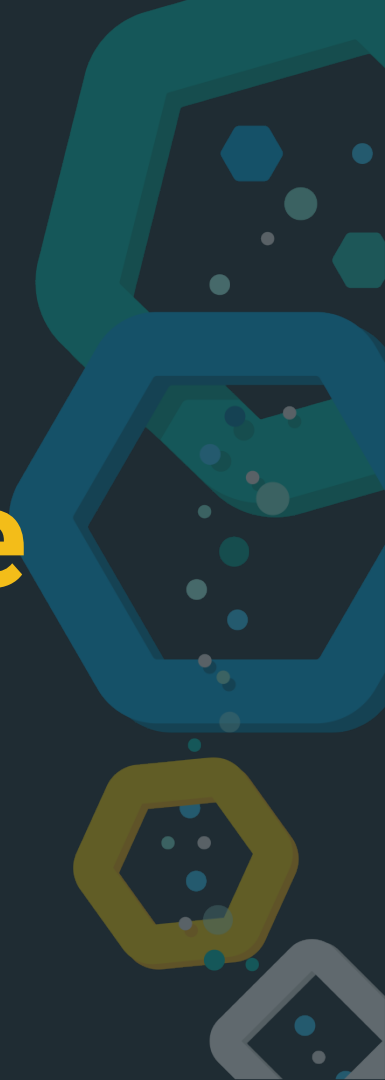


Allows to collapse search results based on field values. The collapsing is done by selecting only the top sorted document per collapse key. For instance the query below retrieves the best tweet for each user and sorts them by number of likes.

```
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

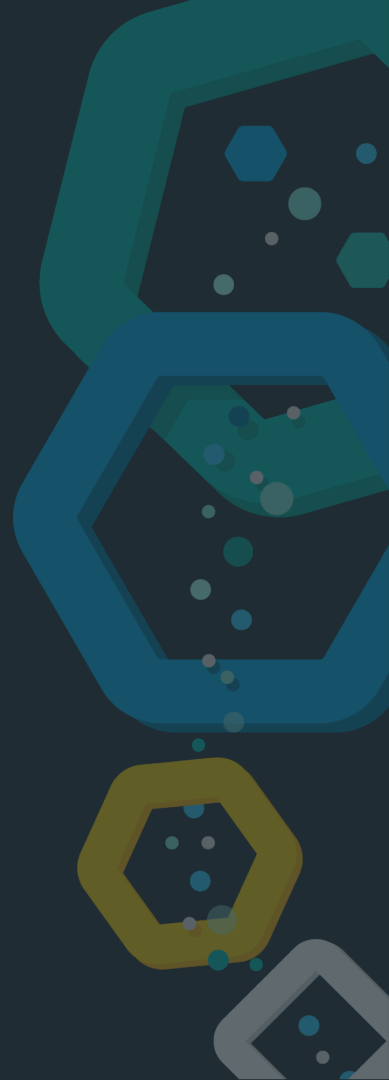
search as you type

Search-as-you-type datatype



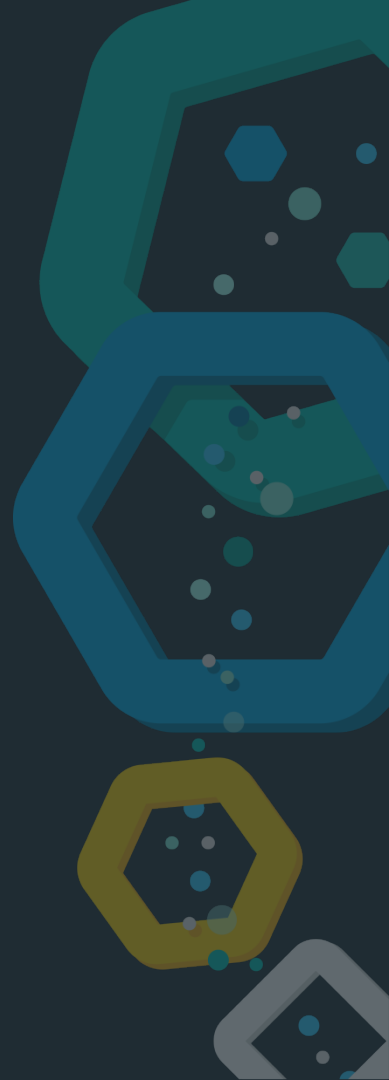
The `search_as_you_type` field type is a text-like field that is optimized to provide out-of-the-box support for queries that serve an as-you-type completion use case. It creates a series of subfields that are analyzed to index terms that can be efficiently matched by a query that partially matches the 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



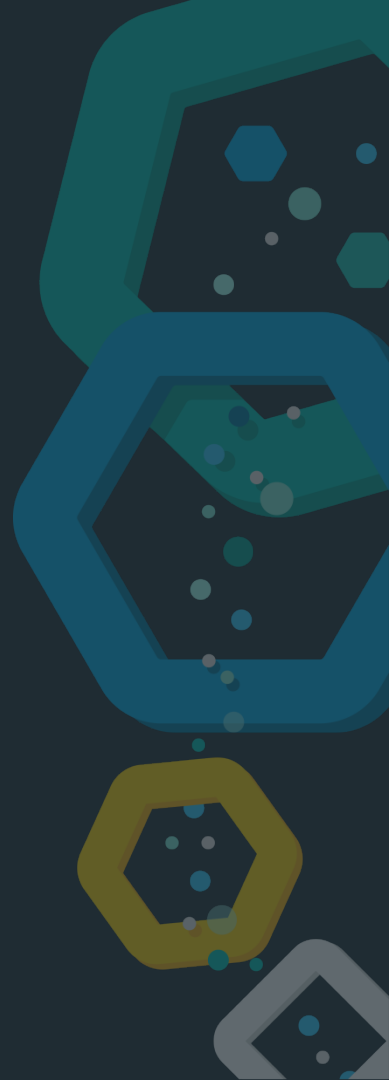
product detail page

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)

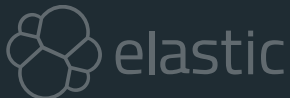


Emmanuel Ameisen @mlpowered · Jul 16

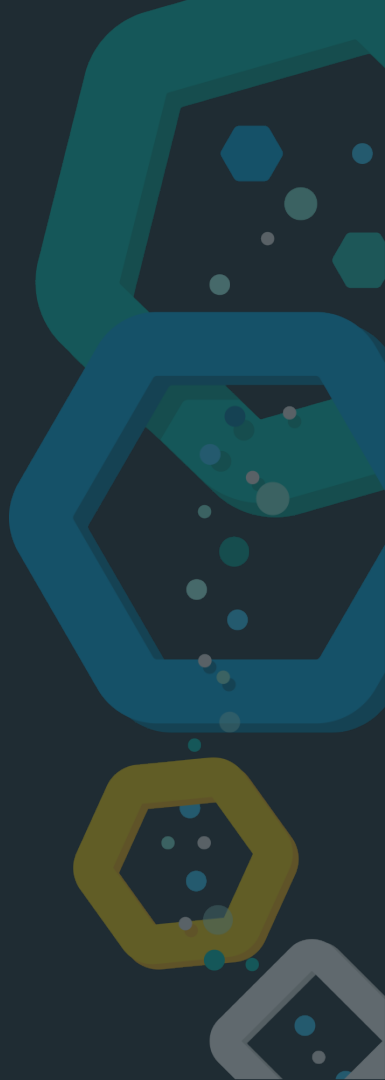
How to ship ML in practice:

- 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](#)



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)



search ui

elastic / search-ui

Used by 7 Watch 210 Star 455

Code Issues 66 Pull requests 2 Projects 1 Wiki Security Insights

Libraries for the fast development of modern, engaging search experiences. <https://swiftype.com/search-ui>

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

mountain

Search

Sort by: Relevance

Showing 1 - 20 out of 25 for: mountain

STATES

Filter states

- Alaska 6
- California 4
- Washington 3
- Colorado 2
- Montana 2

+ More

WORLD HERITAGE SITES

- false 16
- true 7

VISITORS

- 0 - 10000 9
- 10001 - 100000 5
- 100001 - 500000 2
- 500001 - 1000000 4
- 1000001 - 5000000 11

DATE ESTABLISHED

- More than 100 years ago 2
- 50 - 100 years ago 9
- Within the last 50 years 9

DISTANCE

- Nearby 6
- A longer drive 9
- Perhaps fly? 10

ACRES

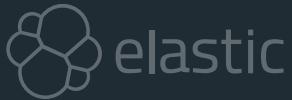
Large

Rocky Mountain

["page_link": "https://www.nps.gov/home/index.htm"](https://www.nps.gov/home/index.htm)
["title": "Rocky Mountains"](#)
["date_established": "1915-01-26T06:00:00+00:00"](#)
["world_heritage_site": false](#)
["states": "Colorado"](#)
["description": "Including mule deer, bighorn sheep, black bears, and cougars inhabit its igneous mountains"](#)
["visitors": 4373785](#)
["id": "park_rocky-mountain"](#)
["location": "40.4,-105.58"](#)
["square_m": 1079.6](#)
["acres": 268079.2](#)

Great Smoky Mountains

["page_link": "https://www.nps.gov/grosm/index.htm"](https://www.nps.gov/grosm/index.htm)
["title": "Great Smoky Mountains"](#)
["date_established": "1934-06-15T05:00:00+00:00"](#)
["world_heritage_site": true](#)
["states": "Tennessee,North Carolina"](#)
["description": "The Great Smoky Mountains, part of the Appalachian Mountains, span a wide range of elevations"](#)
["visitors": 9332796](#)
["id": "park_great-smoky-mountains"](#)
["location": "35.66,-83.53"](#)
["square_m": 2194.2](#)
["acres": 622426.68](#)



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

The screenshot shows the 'Relevance Tuning' interface for a 'marketing-website' index. The left sidebar contains navigation options: Overview, Analytics, Query Tester, and a 'MANAGE' section with Documents, Schema, and API Logs. Below that is 'SEARCH SETTINGS' with Synonyms, Curations, and Relevance Tuning (selected), and an 'ACCESS' section with Credentials. The main content area is titled 'Relevance Tuning' with the subtitle 'Set field weights and customize boosting.' It features a 'Manage Fields' section with a search bar for 22 fields. The 'title' field is selected, showing a weight of 1.9. Below this, there are sections for 'TEXT SEARCH' (enabled for queries), 'WEIGHT' (a slider set to 1.9), and 'BOOSTS' (an 'Add Boost' button). At the bottom, there are sections for 'sections' (weight 0) and 'body' (weight 1). On the right, a 'Preview' section shows search results with fields like title, body, sections, guide_identifier, and url, along with their respective scores (72.7, 72.7, 69.09).

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.

Standard

For small projects

\$49/month

BILLED MONTHLY

[Start free trial](#)

Features include:

- Basic analytics data
- Customizable relevance model
- Basic API logs

Pro

Ready for growth

\$199/month

BILLED MONTHLY

[Start free trial](#)

Features include:

- Multilingual support
- Additional search customizations
- Clickthrough Analytics
- Analytics API
- Extended API log retention

Premium

Mission-critical search

Custom pricing

[Contact us](#)

Features include:

- Engine-level user scoping
- Granular access control
- Dedicated hardware
- Dedicated point of contact
- Expert implementation team
- Enterprise SLA

Elastic App Search

Download Elastic App Search

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

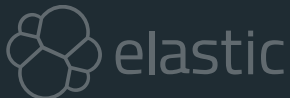
Version: 7.2.1

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

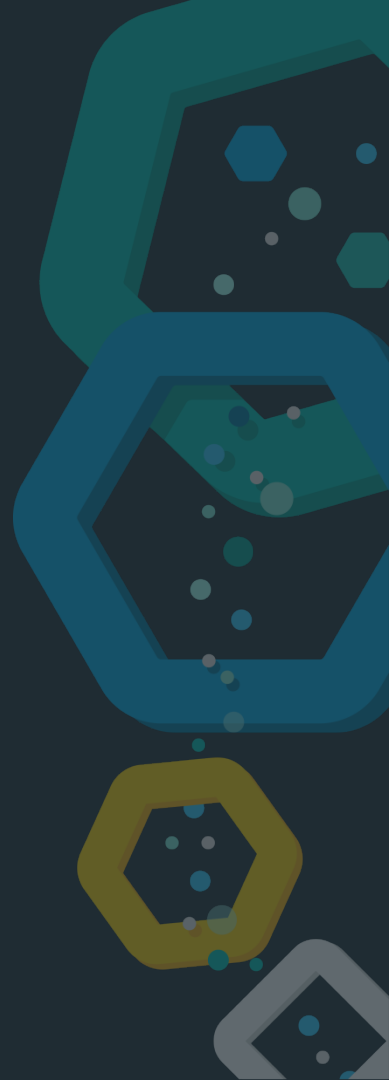
Relevant SEARCH


Doug Turnbull
John Berryman

FOREWORD BY Trey Grainger



 MANNING



 elastic

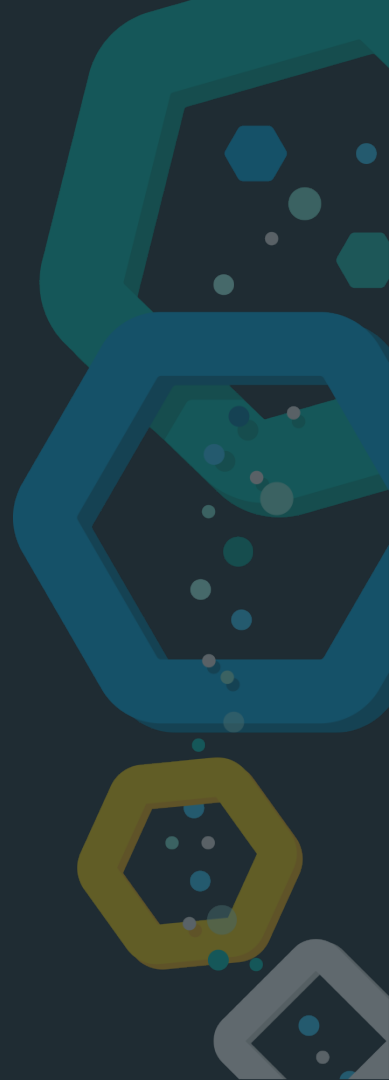
books



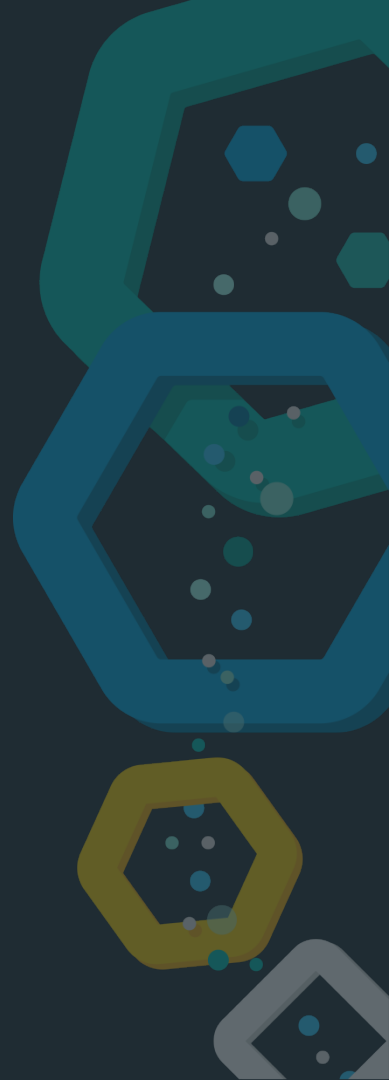
DEEP LEARNING for Search

Tommaso Teofili
Foreword by Chris Mattmann

 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

