

Index me baby!

An elastic introduction; Let's talk search engine!



Nerea Enrique
PHP Engineer @ ekino.
@ nerea.enrique@gmail.com

ekino.



**DRUPAL
DEVELOPER DAYS**
Burgas, Bulgaria - 2024

Hola!



ekino.



Who?

✌️ **Nerea Enrique**



@nep94



@curryed

Where?

🐘 **PHP Engineer at ekino.**

4 years

What?

@CurriedN

6 years

Content

01. Search Engines

02. Welcome to my library

03. Elastic math

04. Let's analyze everything

05. WDYT Drupal?



Search engines

Search engines

Full text search



Search engines

Full text search



Search engines

Tokenization

- Lexical analysis, lexing or tokenization
 - Process of converting a sequence of characters into a sequence of tokens
- Tokens
 - Strings/symbols with an assigned and thus identified meaning
- Break sentences into individual tokens
- Make it easier to understand and analyze it
- Easier to change tokens than sentences

Source: https://en.wikipedia.org/wiki/Lexical_analysis#Token
<https://beram-presentation.gitlab.io/php-static-analysis-101/#10>



Tokenization

The quick brown fox jumps over the lazy dog

Tokenization

The quick brown fox jumps over the lazy dog



The, quick, brown, fox, jumps, over, the, lazy, dog



Search engines

Inverted Index

- Map a token to documents
- Instead of a document to words

Inverted index

D1 = "The quick brown fox."
D2 = "The lazy dog sleeps."
D3 = "A quick nap is refreshing."



D1 = The, quick, brown, fox
D2 = The, lazy, dog, sleeps
D3 = A, quick, nap, is, refreshing

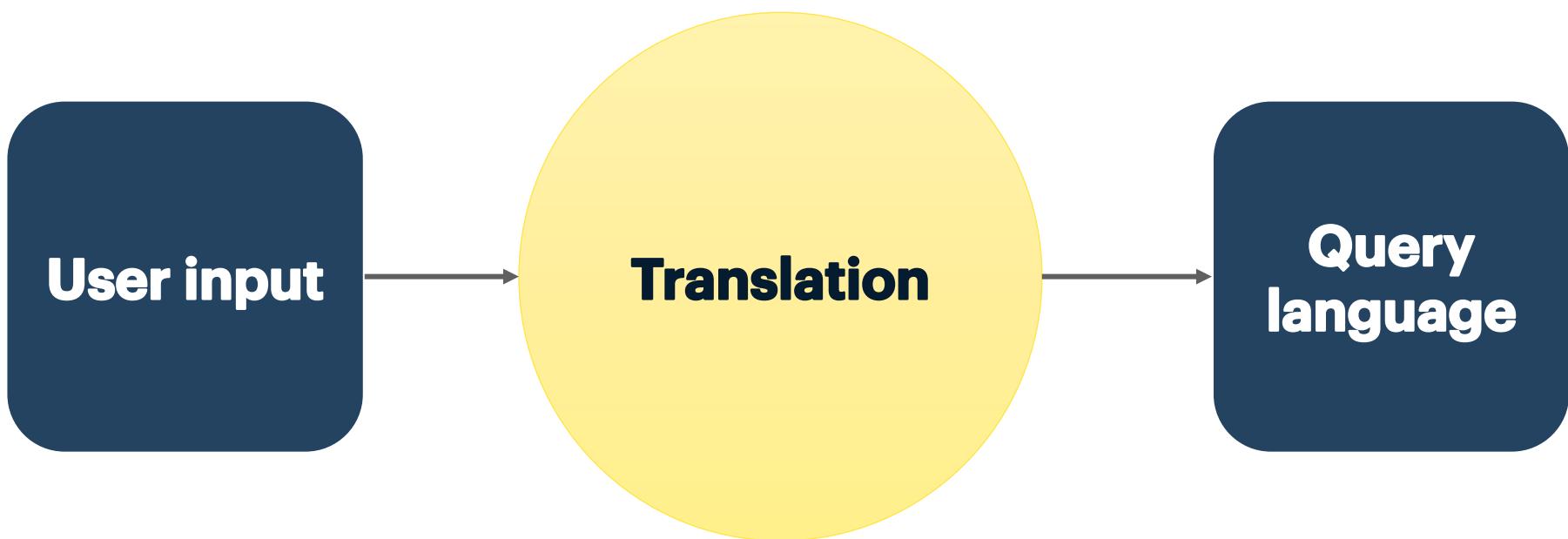
Inverted index

D1 = The, quick, brown, fox
D2 = The, lazy, dog, sleeps
D3 = A, quick, nap, is, refreshing

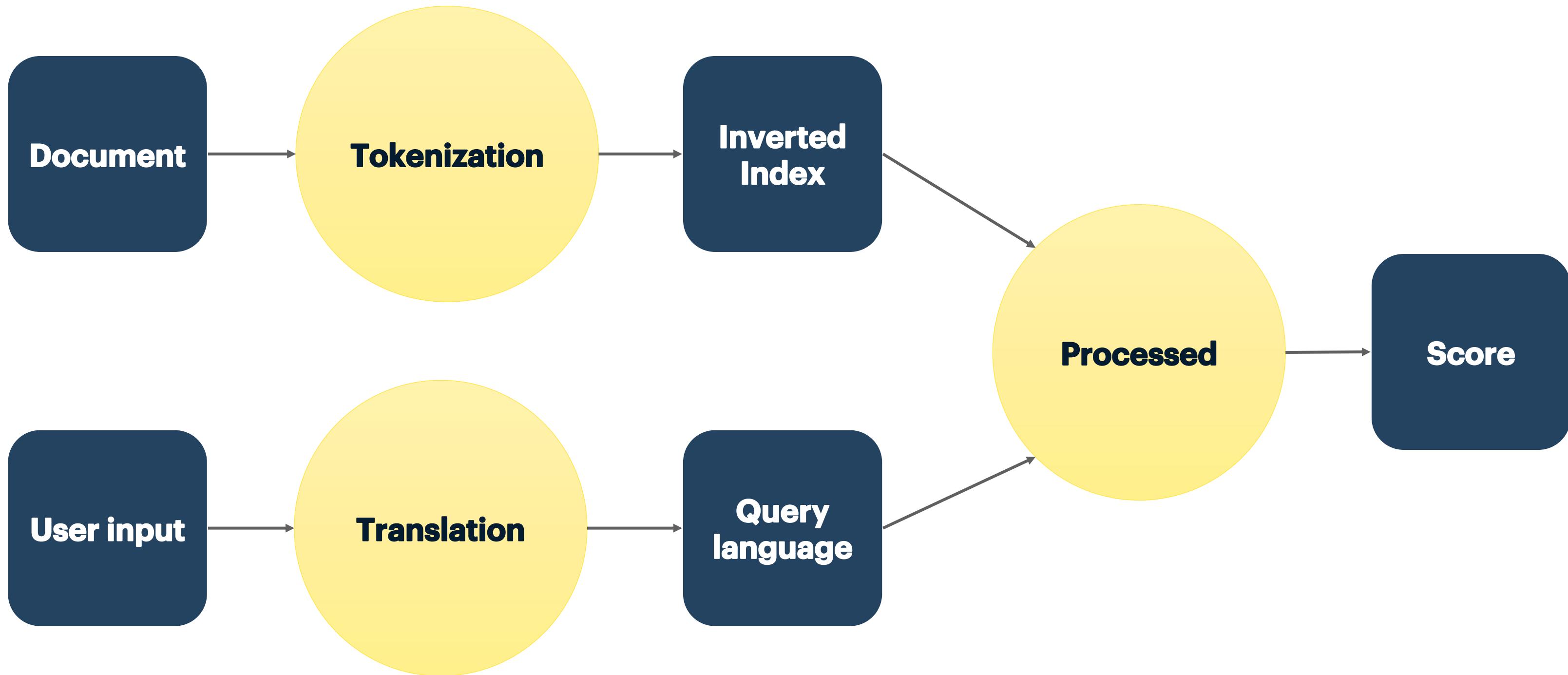


brown: [1]
dog: [2]
fox: [1]
is: [3]
lazy: [2]
nap: [3]
quick: [1, 3]
refreshing: [3]
sleeps: [2]
the: [1, 2]

Full text search



Full text search





Welcome to my library

The building not the library you might think of 😎

Library



Library

Title: Harry Potter and the Prisoner of Azkaban
Genre: Fantasy

Title: Harry Potter and the Deathly Hallows
Genre: Fantasy

Title: The Lord of the Rings
Genre: Fantasy

Title: The Adventures of Huckleberry Finn
Genre: Adventure fiction

Title: Harry Potter and the Half-Blood Prince
Genre: Fantasy

Title: The Silmarillion
Genre: Fantasy

Title: The Hobbit
Genre: Fantasy

Title: Harry Potter and the Philosopher's Stone
Genre: Fantasy

Title: The Grapes of Wrath
Genre: Historical fiction

Title: The Catcher in the Rye
Genre: Realistic fiction

Title: Harry Potter and the Goblet of Fire
Genre: Fantasy

Title: Harry Potter and the Order of the Phoenix
Genre: Fantasy

Title: Harry Potter and the Chamber of Secrets
Genre: Fantasy

Title: The Great Gatsby
Genre: Fiction, tragedy

Library



Library

Title: Harry Potter and the Prisoner of Azkaban
Genre: **Fantasy**

Title: Harry Potter and the Deathly Hallows
Genre: **Fantasy**

Title: Harry Potter and the Half-Blood Prince
Genre: **Fantasy**

Title: The Lord of the Rings
Genre: **Fantasy**

Title: The Adventures of Huckleberry Finn
Genre: Adventure fiction

Title: The Silmarillion
Genre: **Fantasy**

Title: The Hobbit
Genre: **Fantasy**

Title: Harry Potter and the Philosopher's Stone
Genre: **Fantasy**

Title: The Catcher in the Rye
Genre: Realistic fiction

Title: The Grapes of Wrath
Genre: Historical fiction

Title: Harry Potter and the Goblet of Fire
Genre: **Fantasy**

Title: Harry Potter and the Order of the Phoenix
Genre: **Fantasy**

Title: Harry Potter and the Chamber of Secrets
Genre: **Fantasy**

Title: The Great Gatsby
Genre: Fiction, tragedy

Fantasy

Library



Library

Title: Harry Potter and the Prisoner of Azkaban
Genre: Fantasy

Title: Harry Potter and the Deathly Hallows
Genre: Fantasy

Title: Harry Potter and the Half-Blood Prince
Genre: Fantasy

Title: The Silmarillion
Genre: Fantasy

Title: The Hobbit
Genre: Fantasy

Title: Harry Potter and the Philosopher's Stone
Genre: Fantasy

Title: The Catcher in the Rye
Genre: Realistic fiction

Title: The Grapes of Wrath
Genre: Historical fiction

Title: Harry Potter and the Goblet of Fire
Genre: Fantasy

Title: Harry Potter and the Order of the Phoenix
Genre: Fantasy

Title: Harry Potter and the Chamber of Secrets
Genre: Fantasy

Title: The Great Gatsby
Genre: Fiction, tragedy

Fantasy &
Harry &
Potter

Library



Library

Fantasy &
Harry &
Potter &
Phoenix

Title: Harry Potter and the Prisoner of Azkaban
Genre: Fantasy

Title: Harry Potter and the Deathly Hallows
Genre: Fantasy

Title: The Lord of the Rings
Genre: Fantasy

Title: The Adventures of Huckleberry Finn
Genre: Adventure fiction

Title: Harry Potter and the Half-Blood Prince
Genre: Fantasy

Title: The Silmarillion
Genre: Fantasy

Title: The Hobbit
Genre: Fantasy

Title: Harry Potter and the Philosopher's Stone
Genre: Fantasy

Title: The Catcher in the Rye
Genre: Realistic fiction

Title: The Grapes of Wrath
Genre: Historical fiction

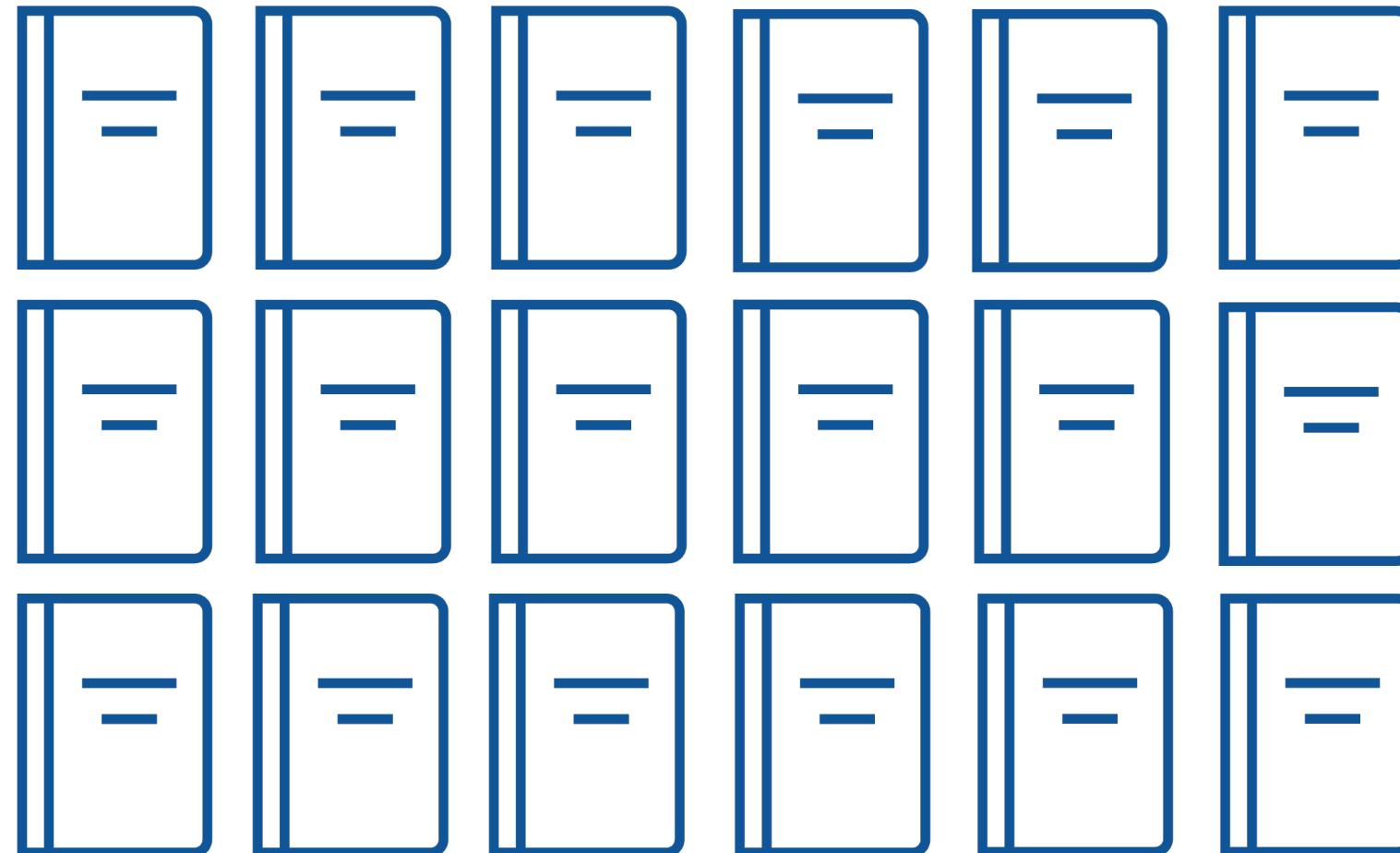
Title: Harry Potter and the Goblet of Fire
Genre: Fantasy

Title: Harry Potter and the Order of the Phoenix
Genre: Fantasy

Title: Harry Potter and the Chamber of Secrets
Genre: Fantasy

Title: The Great Gatsby
Genre: Fiction, tragedy

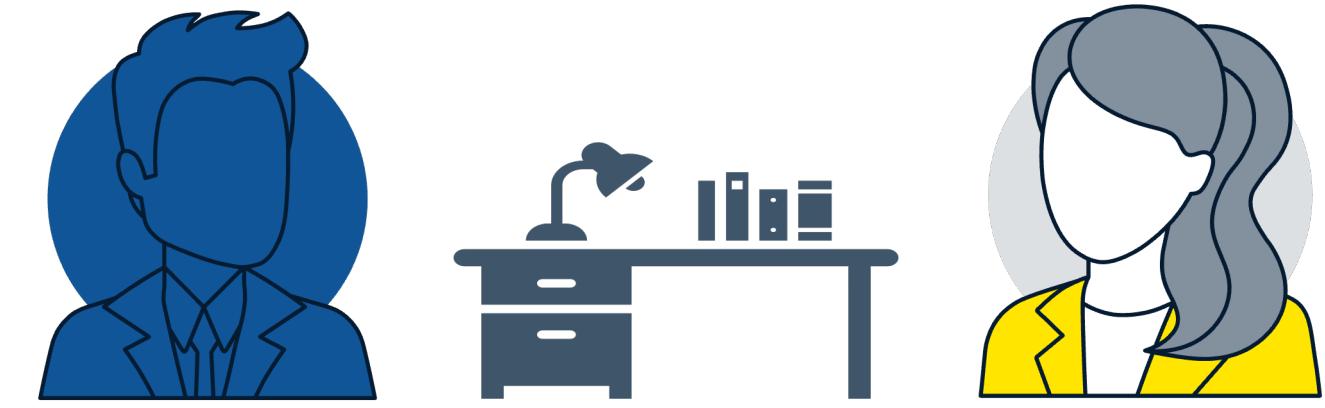
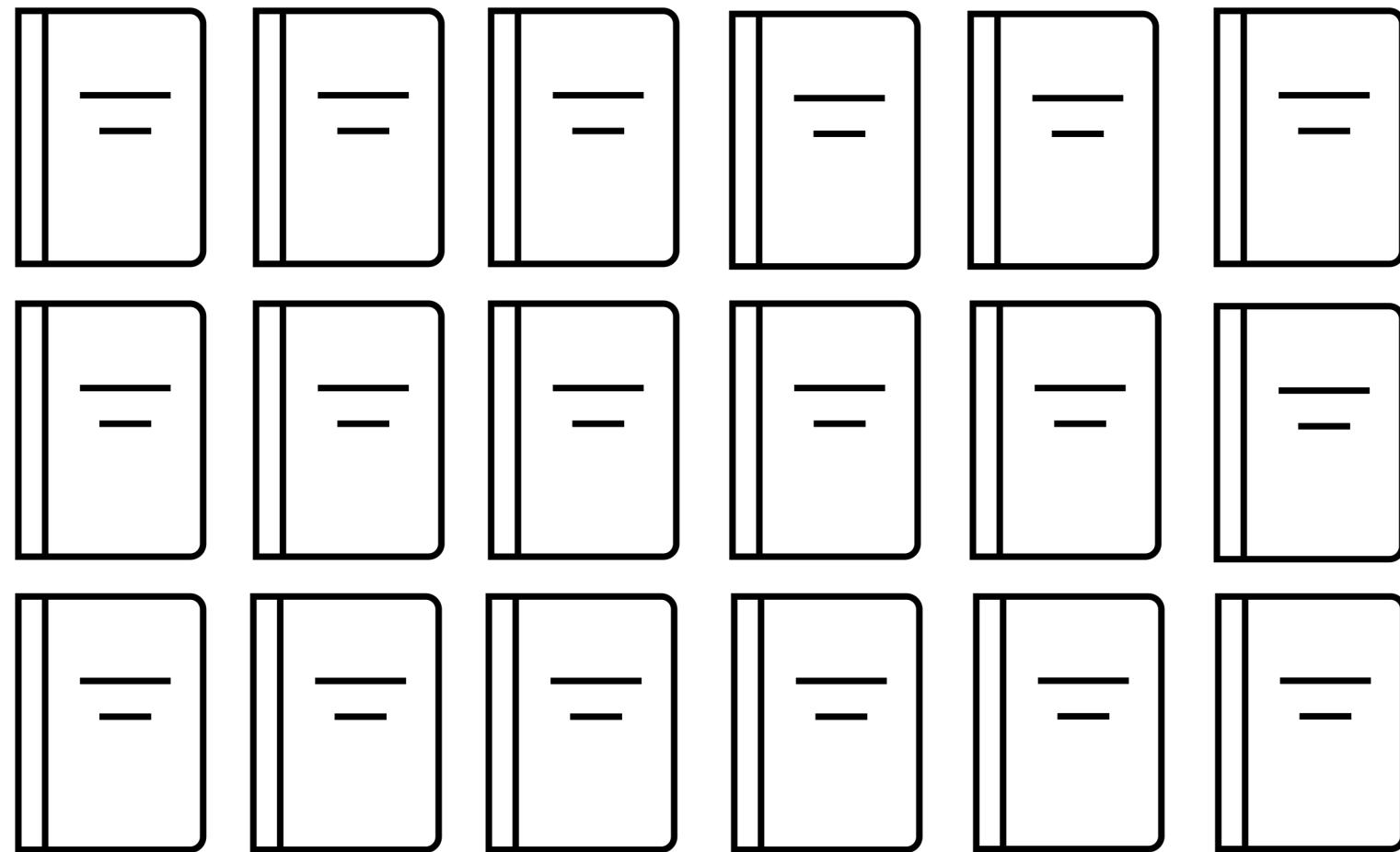
Library



Documents

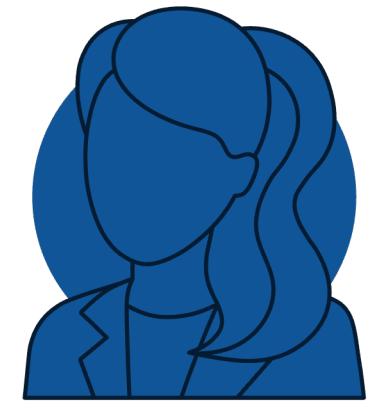
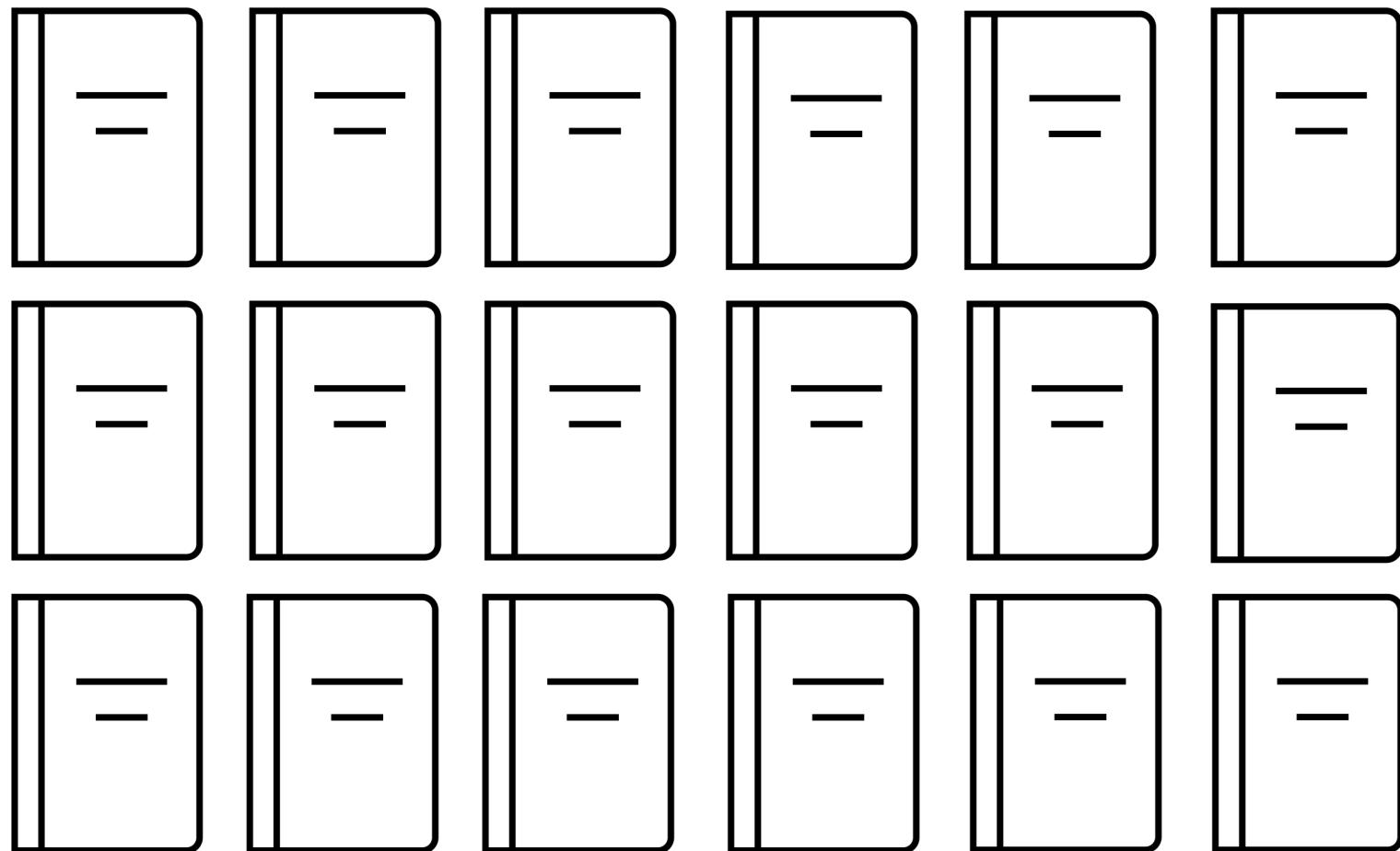


Library



Inverted
index

Library

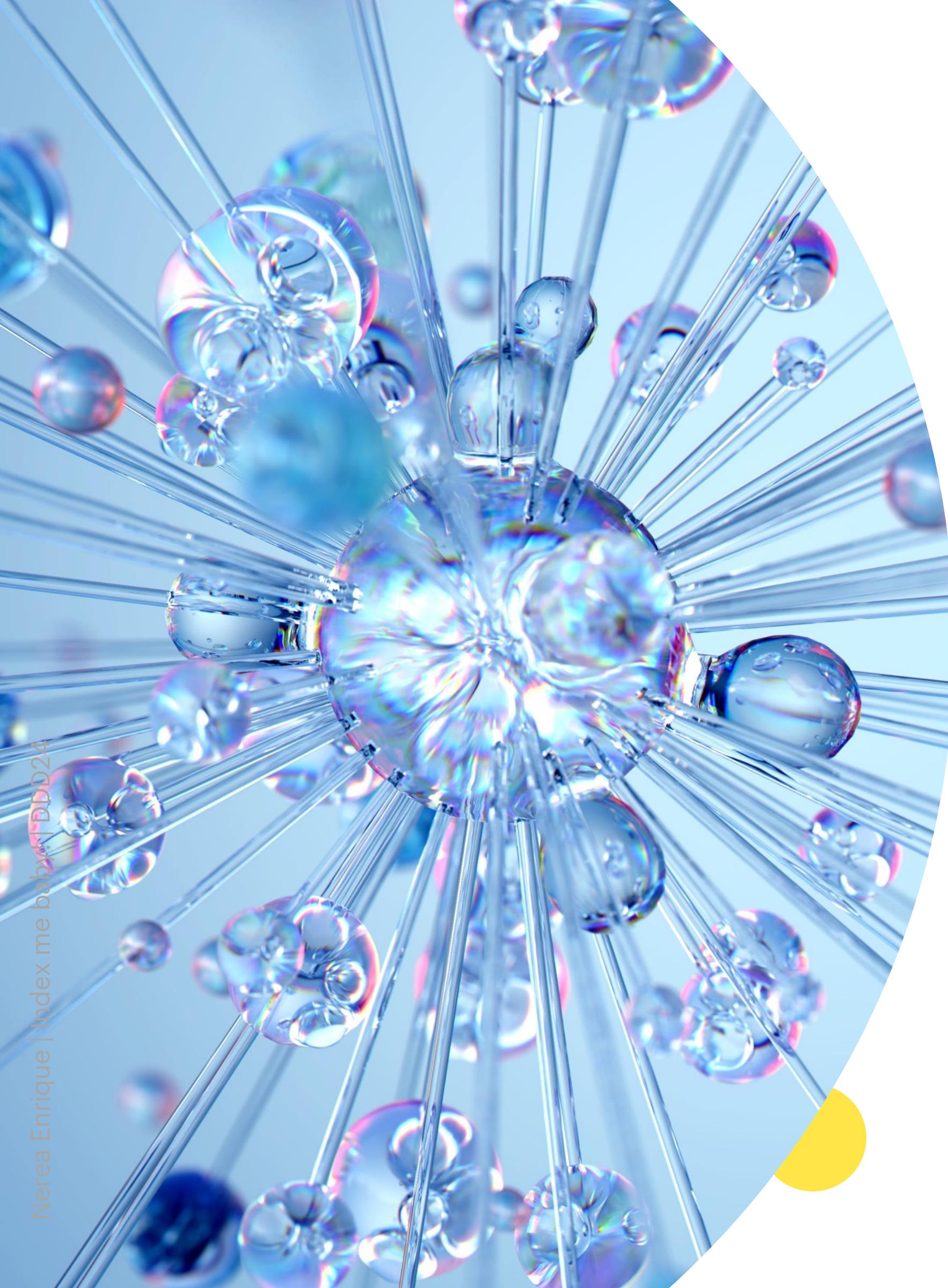


User



Elastic math





Elastic math

Some notions

- Inverse Document Frequency IDF
- Term Frequency TF
- Coordination factor
- Score

Elastic math

Inversed Document Frequency (IDF)

- Quantifies the importance of a term
- Calculated as

$$idf(D, tm) = \log \frac{totalDocuments}{documentsContaining(tm)}$$

Elastic math

Inversed Document Frequency (IDF)

Harry, Potter,
tragedy

- More frequently a term appears => less important it is => lower IDF it gets

31

Title: The Catcher in the Rye
Genre: Realistic fiction

Title: Well done if you can read this
Genre: Unfiction

Title: The Grapes of Wrath
Genre: Historical fiction

Title: Harry Potter and the Goblet of Fire
Genre: Fantasy

Title: Harry Potter and the Order of the Phoenix
Genre: Fantasy

Title: Harry Potter and the Chamber of Secrets
Genre: Fantasy

Title: The Great Gatsby
Genre: Fiction, tragedy

Source: <https://en.wikipedia.org/wiki/Tf%E2%80%93idf>

Term frequency (TF)

- Measures the frequency of a term within a document
- Calculated as

$$tf(D, tm) = \frac{\text{timesTermAppearsInDocument}(D, tm)}{\text{numberOfTermsWithinTheDocument}(D)}$$

Term frequency (TF)

D1 = "The cat chased the mouse. The mouse scaped."

D2 = "There is a mouse in the house."

Term frequency (TF)

D1 = "The cat chased the mouse. The mouse scaped."

D2 = "There is a mouse in the house."



$$\text{TF}(D1, \text{"mouse"}) = 2/8 = 0,25$$

$$\text{TF}(D2, \text{"mouse"}) = 1/7 = 0,14$$

Elastic math

Coordination factor

- Number of terms from the query matched in the document

$$cf = \frac{q(D)}{q}$$

Coordination factor

D1 = "The cat chased the mouse. The mouse scaped."

D2 = "There is a mouse in the house."



$$CF(D1, \text{"mouse"}) = 1/1 = 1$$

$$CF(D2, \text{"mouse"}) = 1/1 = 1$$

$$CF(D1, \text{"mouse, cat"}) = 2/2 = 1$$

$$CF(D2, \text{"mouse, cat"}) = 1/2 = 0,5$$

Elastic math

Score

- Computed value
- Based on IDF and TF
- Ranks documents

Score

Fantasy,
Harry
Potter,
Phoenix

- Higher the score => Higher the relevance

Title: The Catcher in the Rye
Genre: Realistic fiction

Title: Well done if you can read this
Genre: Unfiction

Title: The Grapes of Wrath
Genre: Historical fiction

Title: Harry Potter and the Goblet of Fire
Genre: Fantasy

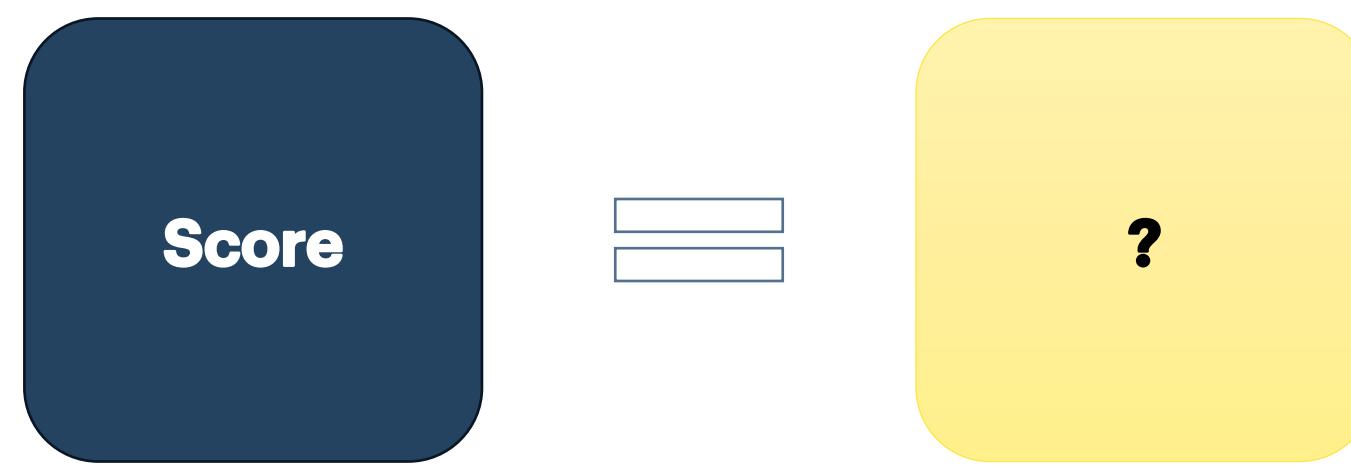
Title: Harry Potter and the Order of the Phoenix
Genre: Fantasy

Title: Harry Potter and the Chamber of Secrets
Genre: Fantasy

Title: The Great Gatsby
Genre: Fiction, tragedy

Elastic math

Score



Elastic math

Disclaimer

- Not the actual algorithm (simplified and changed for the presentation)
- Maths

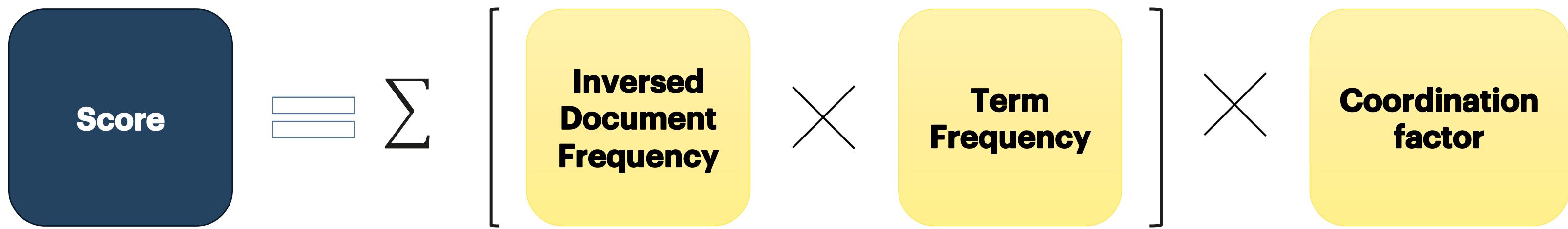


BM25

$$S(D, Q) = \sum_{i=1}^n IDF(q_i) \cdot \frac{f(q_i, D) \cdot (k_1 + 1)}{f(q_i, D) + k_1 \cdot (1 - b + b \cdot \frac{|D|}{avgdl})}$$

Elastic math

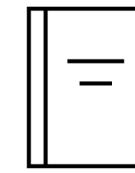
Score



Source: Mathias ARLAUD – Forum PHP 2022
<https://www.youtube.com/watch?v=UhONe6GSfGQ>

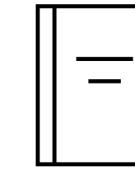
Elastic math

Example



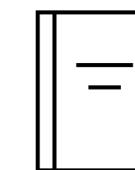
Genre : Fantasy, Title : Hunger Games

$$\text{Score} = [\text{IDF}(\text{fanta}) \cdot \text{TF}(\text{fanta})] \cdot \text{CF}$$



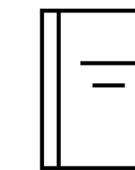
Genre : Fantasy, Title : Narnia

$$\text{Score} = [\text{IDF}(\text{fanta}) \cdot \text{TF}(\text{fanta})] \cdot \text{CF}$$



Genre : Fantasy, Title : Fantastic Beasts and Where to Find Them

$$\text{Score} = [\text{IDF}(\text{fanta}) \cdot \text{TF}(\text{fanta})] \cdot \text{CF}$$



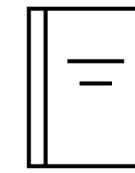
Genre : Horror, Title : IT

$$\text{Score} = [\text{IDF}(\text{fanta}) \cdot \text{TF}(\text{fanta})] \cdot \text{CF}$$



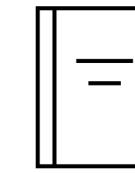
Elastic math

Example – TF



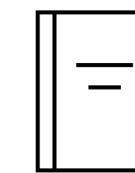
Genre : Fantasy, Title : Hunger Games

$Score = [IDF(fanta) . 1] . CF$



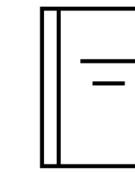
Genre : Fantasy, Title : Narnia

$Score = [IDF(fanta) . 1] . CF$



Genre : Fantasy, Title : Fantastic Beasts and Where to Find Them

$Score = [IDF(fanta) . 2] . CF$



Genre : Horror, Title : IT

$Score = [IDF(fanta) . 0] . CF$



Elastic math

Example - IDF



Genre : Fantasy, Title : Hunger Games

$$\text{Score} = [\log(4/3) \cdot 1] \cdot CF$$



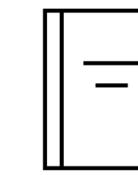
Genre : Fantasy, Title : Narnia

$$\text{Score} = [\log(4/3) \cdot 1] \cdot CF$$



Genre : Fantasy, Title : Fantastic Beasts and Where to Find Them

$$\text{Score} = [\log(4/3) \cdot 2] \cdot CF$$



Genre : Horror, Title : IT

$$\text{Score} = [\log(4/3) \cdot 0] \cdot CF$$



Reminder:

$$\text{Score} = [IDF(fanta) \cdot TF(fanta)] \cdot CF$$

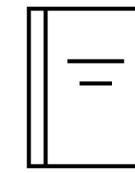
$TF = \text{timesTermAppearsInDocument} / \text{numberOfTermsWithinTheDocument}$

$IDF = \log(\text{totalDocuments} / \text{documentsContaining})$

$CF = Q(D) / Q$

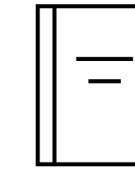
Elastic math

Example



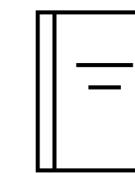
Genre : Fantasy, Title : Hunger Games

Score = [0,12 . 1]. CF



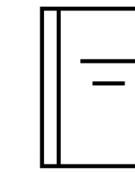
Genre : Fantasy, Title : Narnia

Score = [0,12 . 1]. CF



Genre : Fantasy, Title : Fantastic Beasts and Where to Find Them

Score = [0,12 . 2]. CF



Genre : Horror, Title : IT

Score = [0,12 . 0]. CF



Reminder:

Score = [IDF(fanta) . TF(fanta)]. CF

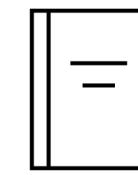
TF = timesTermAppearsInDocument / numberOfTermsWithinTheDocument

IDF = log(totalDocuments / documentsContaining)

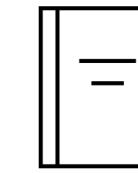
CF = Q(D) / Q

Elastic math

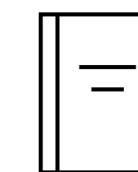
Example - CF



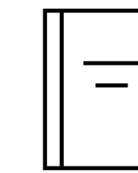
Genre : Fantasy, Title : Hunger Games

 $Score = [0,12 . 1]. 1$ 

Genre : Fantasy, Title : Narnia

 $Score = [0,12 . 1]. 1$ 

Genre : Fantasy, Title : Fantastic Beasts and Where to Find Them

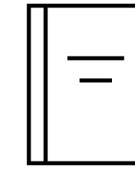
 $Score = [0,12 . 2]. 1$ 

Genre : Horror, Title : IT

 $Score = [0,12 . 0]. 0$ 

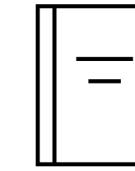
Elastic math

Example



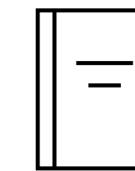
Genre : Fantasy, Title : Hunger Games

$$\text{Score} = [0,12 \cdot 1] \cdot 1 = 0,12$$



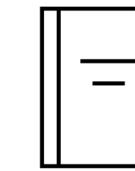
Genre : Fantasy, Title : Narnia

$$\text{Score} = [0,12 \cdot 1] \cdot 1 = 0,12$$



Genre : Fantasy, Title : Fantastic Beasts and Where to Find Them

$$\text{Score} = [0,12 \cdot 2] \cdot 1 = 0,24$$



Genre : Horror, Title : IT

$$\text{Score} = [0,12 \cdot 0] \cdot 0 = 0$$

*Reminder:*

$$\text{Score} = [\text{IDF}(fanta) \cdot \text{TF}(fanta)] \cdot \text{CF}$$

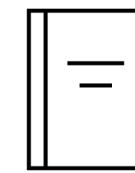
TF = timesTermAppearsInDocument / number_of_terms_within_the_document

$\text{IDF} = \log(\text{total_documents} / \text{documents_containing})$

$\text{CF} = Q(D) / Q$

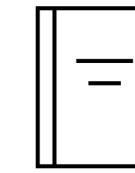
Elastic math

Example



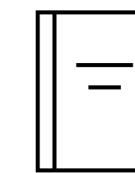
Genre : Fantasy, Title : Hunger Games

$$\text{Score} = [(0,12 \cdot 1) + (\text{IDF}(narnia) \cdot \text{TF}(narnia))]. 0,5 =$$



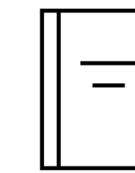
Genre : Fantasy, Title : Narnia

$$\text{Score} = [(0,12 \cdot 1) + (\text{IDF}(narnia) \cdot \text{TF}(narnia))]. 1 =$$



Genre : Fantasy, Title : Fantastic Beasts and Where to Find Them

$$\text{Score} = [(0,12 \cdot 2) + (\text{IDF}(narnia) \cdot \text{TF}(narnia))]. 0,5 =$$



Genre : Horror, Title : IT

$$\text{Score} = [(0,12 \cdot 0) + (\text{IDF}(narnia) \cdot \text{TF}(narnia))]. 0 =$$



Reminder:

$$\text{Score} = [(\text{IDF}(fanta) \cdot \text{TF}(fanta)) + (\text{IDF}(narnia) \cdot \text{TF}(narnia))]. \text{CF}$$

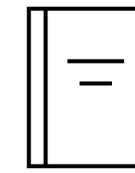
$\text{TF} = \text{timesTermAppearsInDocument} / \text{numberOfTermsWithinTheDocument}$

$\text{IDF} = \log(\text{totalDocuments} / \text{documentsContaining})$

$\text{CF} = Q(D) / Q$

Elastic math

Example



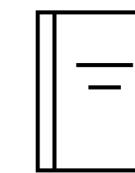
Genre : Fantasy, Title : Hunger Games

$$\text{Score} = [(0,12 \cdot 1) + (\log(4/1) \cdot 0)] \cdot 0,5 =$$



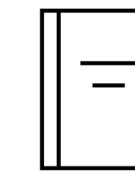
Genre : Fantasy, Title : Narnia

$$\text{Score} = [(0,12 \cdot 1) + (\log(4/1) \cdot 1)] \cdot 1 =$$



Genre : Fantasy, Title : Fantastic Beasts and Where to Find Them

$$\text{Score} = [(0,12 \cdot 2) + (\log(4/1) \cdot 0)] \cdot 0,5 =$$



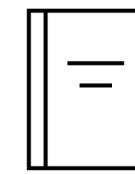
Genre : Horror, Title : IT

$$\text{Score} = [(0,12 \cdot 0) + (\log(4/1) \cdot 0)] \cdot 0 =$$



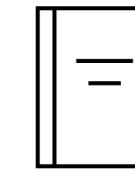
Elastic math

Example



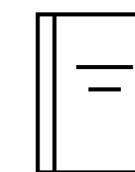
Genre : Fantasy, Title : Hunger Games

$$\text{Score} = [(0,12 \cdot 1) + (0,6 \cdot 0)] \cdot 0,5 =$$



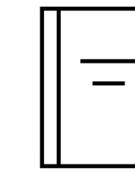
Genre : Fantasy, Title : Narnia

$$\text{Score} = [(0,12 \cdot 1) + (0,6 \cdot 1)] \cdot 1 =$$



Genre : Fantasy, Title : Fantastic Beasts and Where to Find Them

$$\text{Score} = [(0,12 \cdot 2) + (0,6 \cdot 0)] \cdot 0,5 =$$



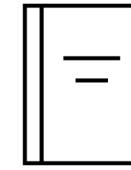
Genre : Horror, Title : IT

$$\text{Score} = [(0,12 \cdot 0) + (0,6 \cdot 0)] \cdot 0 =$$



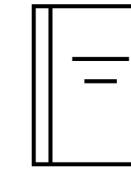
Elastic math

Example



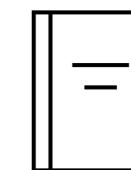
Genre : Fantasy, Title : Hunger Games

$$\text{Score} = [(0,12 \cdot 1) + (0,6 \cdot 0)] \cdot 0,5 = 0,06$$



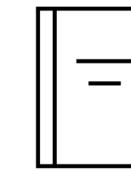
Genre : Fantasy, Title : Narnia

$$\text{Score} = [(0,12 \cdot 1) + (0,6 \cdot 1)] \cdot 1 = 0,72$$



Genre : Fantasy, Title : Fantastic Beasts and Where to Find Them

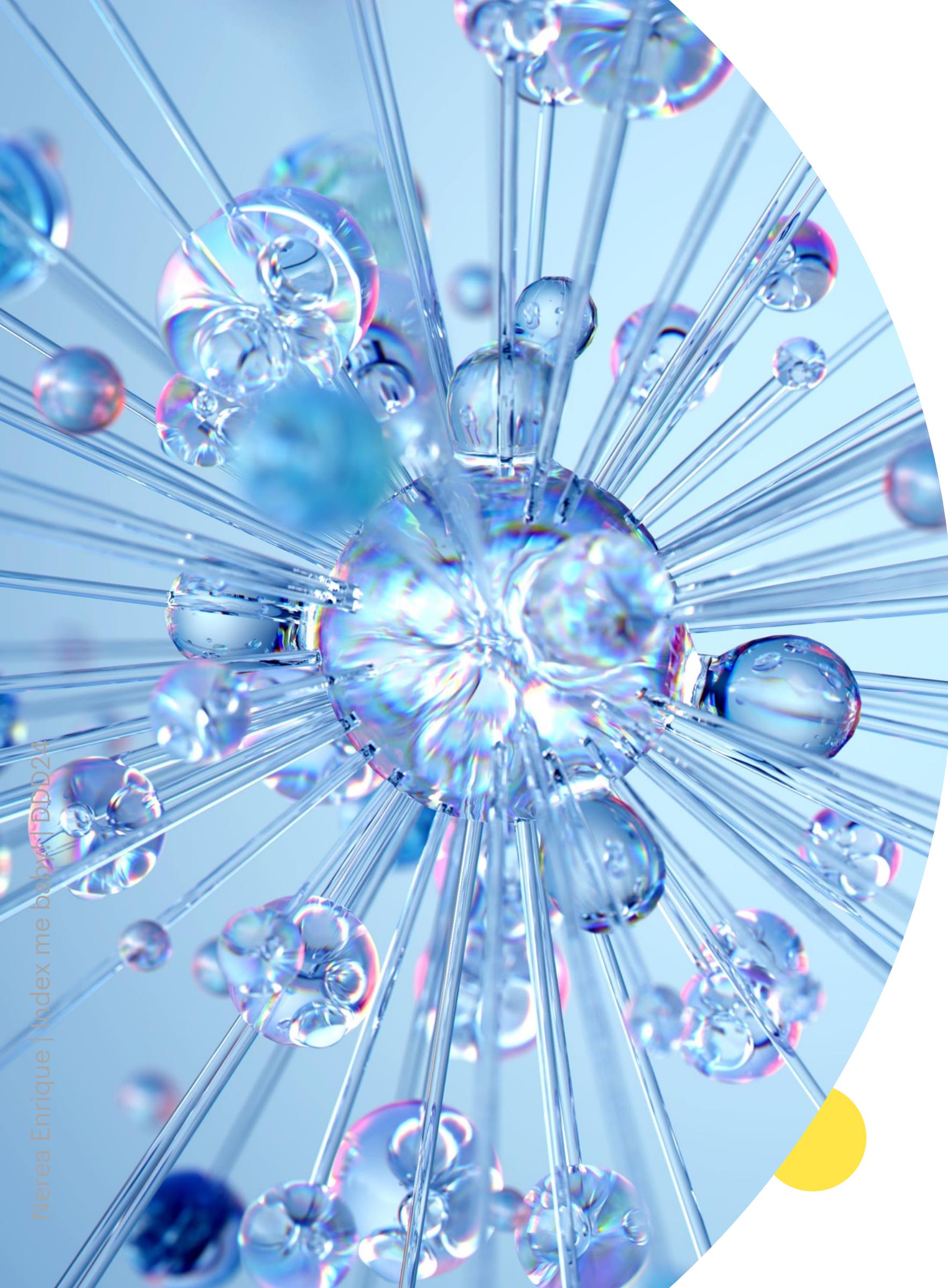
$$\text{Score} = [(0,12 \cdot 2) + (0,6 \cdot 0)] \cdot 0,5 = 0,12$$



Genre : Horror, Title : IT

$$\text{Score} = [(0,12 \cdot 0) + (0,6 \cdot 0)] \cdot 0 = 0$$





Elastic math

Some notions

- Score
- Inversed Document Frequency IDF
- Term Frequency TF
- Coordination factor

That's not all folks

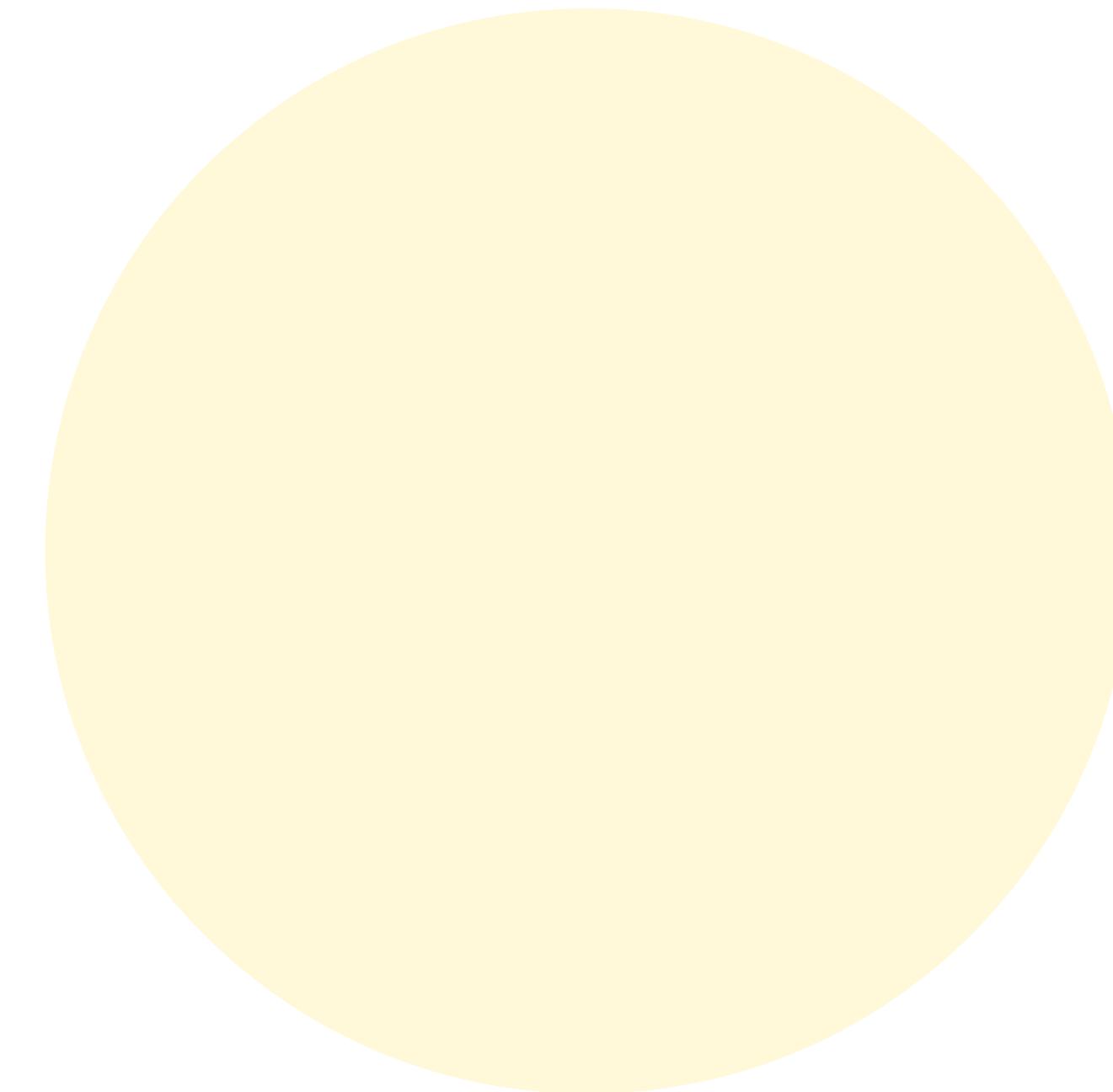
- Boosts
- Function score
- User input
- ...



Let's analyze everything



People talk with lots of words, how simplify a long, full phrase on some keywords that will be used for a search?

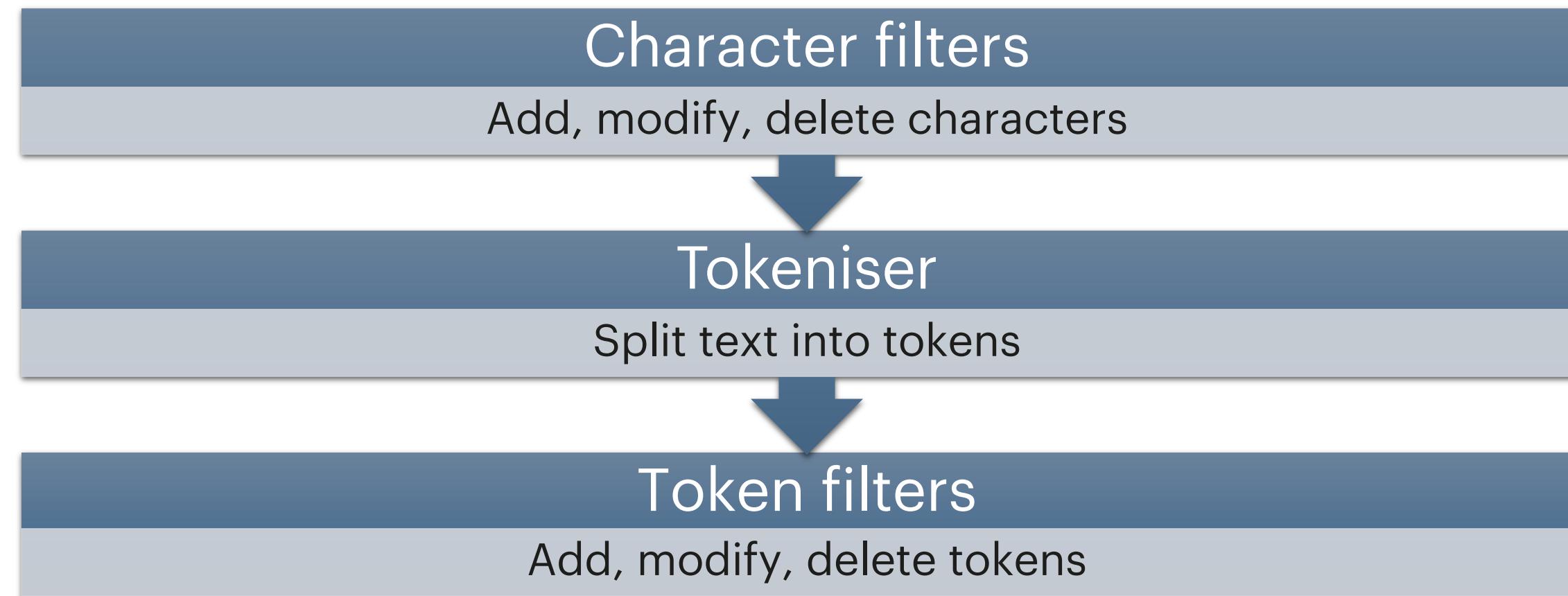


Let's analyse everything

What can they do?

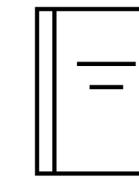
Let's analyze everything 

What can they do?

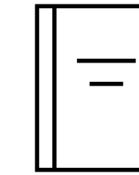


Let's analyze everything 

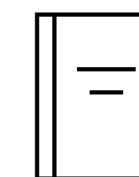
Example



Title: The Lord of the Rings

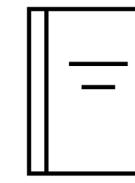


Title: The Great Gatsby



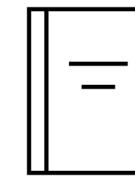
Title: The Adventures of Huckleberry Finn

Tokenisation



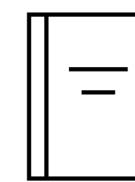
Title: The Lord of the Rings

Tokens = The, Lord, of, the, Rings



Title: The Great Gatsby

Tokens = The, Great, Gatsby

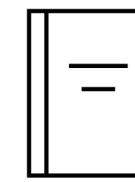


Title: The Adventures of Huckleberry Finn

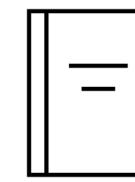
Tokens = The, Adventures, of, Huckleberry, Finn

Let's analyze everything 

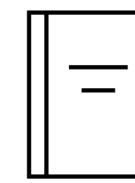
Token filter: lowercase



Title: The Lord of the Rings
Tokens = the, lord, of, the, rings

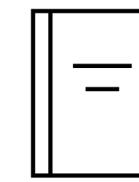


Title: The Great Gatsby
Tokens = the, great, gatsby

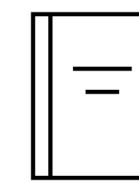


Title: The Adventures of Huckleberry Finn
Tokens = the, adventures, of, huckleberry, finn

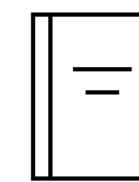
Token filter: stop



Title : The Lord of the Rings
Tokens = lord, rings



Title : The Great Gatsby
Tokens = great, gatsby



Title : The Adventures of Huckleberry Finn
Tokens = adventures, huckleberry, finn



Let's analyze everything

Disclaimer

- Each index contains the same 15 books
- Each index is configured distinctly

Let's analyze everything 

Query

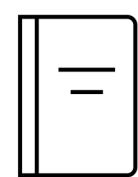
```
{  
  "query": {  
    "match": {  
      "description": "epic"  
    }  
  }  
}
```

Let's analyze everything 

Tokens - ngram



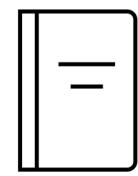
Total: 7



Title: The Lord of the Rings

Description: The Lord of the Rings is an [epic](#) high fantasy novel by the English author and scholar J. R. R. Tolkien.

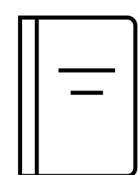
Score = 1.2574334



Title: The Silmarillion

Description: The Silmarillion is a collection of mythopoeic works by English writer J. R. R. Tolkien, edited and published posthumously by his son, Christopher Tolkien, in 1977, with assistance from Guy Gavriel Kay.

Score = 1.2202173



Title: The Hobbit

Description: The Hobbit, or There and Back Again is a children's fantasy novel by English author J. R. R. Tolkien.

Score = 0.70242006

...

The ngram tokenizer first breaks text down into words whenever it encounters one of a list of specified characters, then it emits N-grams of each word of the specified length.

N-grams are like a sliding window that moves across the word - a continuous sequence of characters of the specified length. They are useful for querying languages that don't use spaces or that have long compound words, like German.

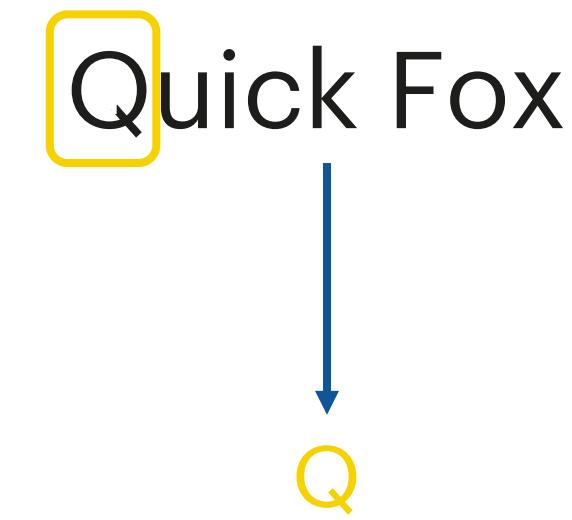
<https://www.elastic.co/guide/en/elasticsearch/reference/current/analysis-ngram-tokenizer.html>

Ngram

Quick Fox

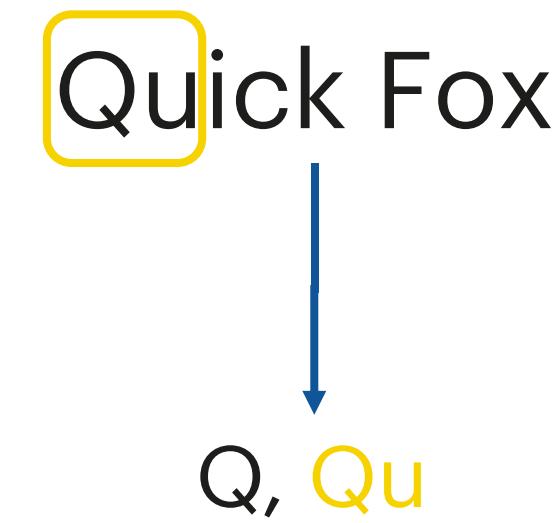
Ngram

Quick Fox



A diagram illustrating the concept of an N-gram. The word "Quick Fox" is written in black text. The letter "Q" is highlighted with a yellow square box. A blue arrow points from the yellow box down to the letter "Q" itself.

Ngram



Ngram

Quick Fox

Q, Qu, u

Ngram

Quick Fox

Q, Qu, u, ui

Ngram

Quick Fox



Q, Qu, u, ui, i, ic, c, ck, k, "k ", "", " F", F, Fo, o, ox, x

Ngram

epic



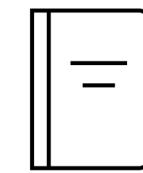
e, ep, p, pi, i, ic, c

Let's analyze everything 

Tokens - edge-gram



Total: 1



Title: The Lord of the Rings

Description: The Lord of the Rings is an [epic](#) high fantasy novel by the English author and scholar J. R. R. Tolkien.

Score = 6.6332088

The edge_ngram tokenizer first breaks text down into words whenever it encounters one of a list of specified characters, then it emits N-grams of each word where the start of the N-gram is anchored to the beginning of the word.

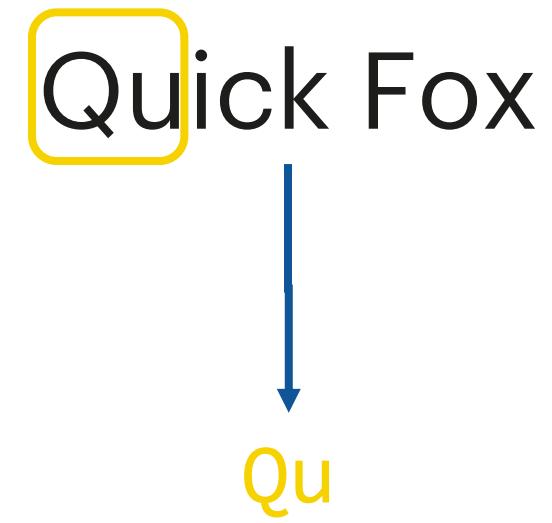
Edge N-Grams are useful for search-as-you-type queries.

<https://www.elastic.co/guide/en/elasticsearch/reference/current/analysis-edgengram-tokenizer.html>

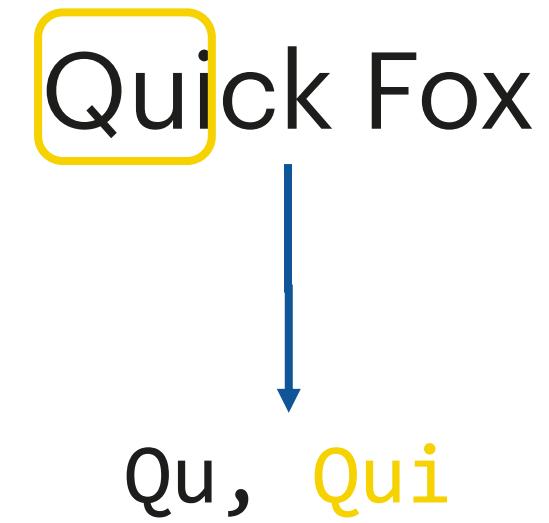
Edge-gram

Quick Fox

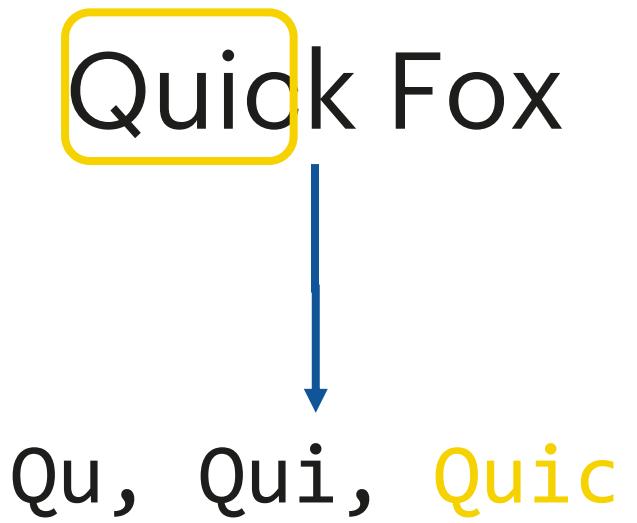
Edge-ngram



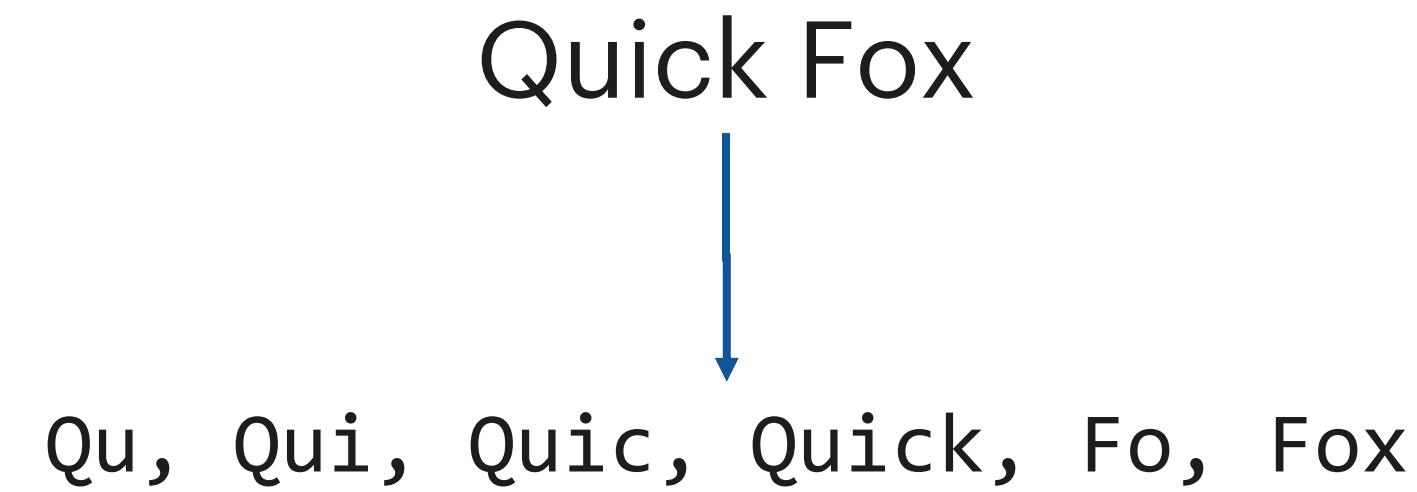
Edge-ngram



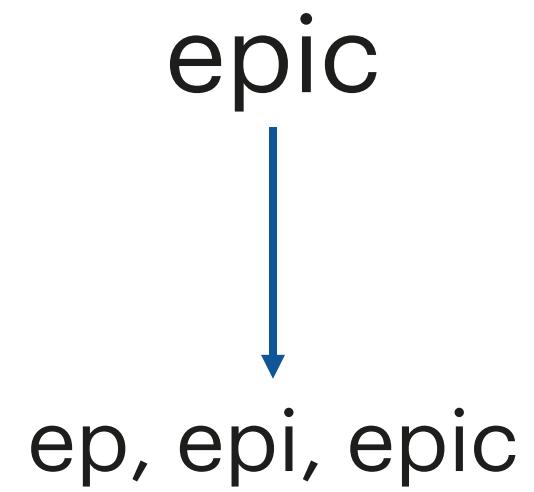
Edge-ngram



Edge-ngram



Edge-ngram

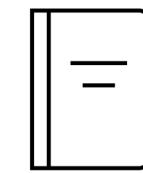


Let's analyze everything 

Filters – lowercase



Total: 1



Title: The Lord of the Rings

Description: The Lord of the Rings is an [epic](#) high fantasy novel by the English author and scholar J. R. R. Tolkien.

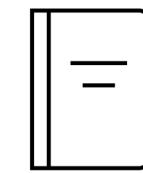
Score = 2.0468292

Let's analyze everything 

Filters – stemmer



Total: 1



Title: The Lord of the Rings

Description: The Lord of the Rings is an [epic](#) high fantasy novel by the English author and scholar J. R. R. Tolkien.

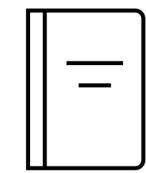
Score = 2.048967

Let's analyze everything 

Filters – stop



Total: 1



Title: The Lord of the Rings

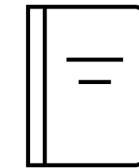
Description: The Lord of the Rings is an [epic](#) high fantasy novel by the English author and scholar J. R. R. Tolkien.

Score = 2.0478997

All filters with ngram



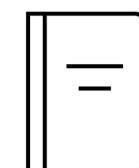
Total: 7



Title: The Grapes of Wrath

Description: The Grapes of Wrath is an American realist novel written by John Steinbeck and published in 1939. The book won the National Book Award and Pulitzer Prize for fiction, and it was cited prominently when Steinbeck was awarded the Nobel Prize in 1962.

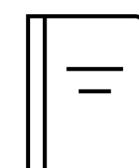
Score = 0.75382936



Title: The Silmarillion

Description: The Silmarillion is a collection of mythopoeic works by English writer J. R. R. Tolkien, edited and published posthumously by his son, Christopher Tolkien, in 1977, with assistance from Guy Gavriel Kay.

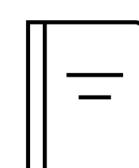
Score = 0.7357905



Title: The Catcher in the Rye

Description: The Catcher in the Rye is a novel by J. D. Salinger, partially published in serial form in 1945–1946 and as a novel in 1951. It was originally intended for adults but is read by adolescents for its themes of angst, alienation, and as a critique on superficiality in society.

Score = 0.70242006



Title: The Lord of the Rings

Description: The Lord of the Rings is an [epic](#) high fantasy novel by the English author and scholar J. R. R. Tolkien.

Score = 0.63124937

...

Let's analyze everything 

Books

Setting	Default settings	Ngram	Edge-ngram	Case sensitive	Stemmer	Stop	All
Score	2.048967	1.2574334	6.6332088	2.0468292	2.048967	2.0478997	0.63124937
Total	1	7	1	1	1	1	7

Let's analyze everything 

Books - EN

```
{
  "settings": {
    "analysis": {
      "filter": {
        "english_stop": {
          "type": "stop",
          "stopwords": "_english_"
        },
        "english_keywords": {
          "type": "keyword_marker",
          "keywords": ["example"]
        },
        "english_stemmer": {
          "type": "stemmer",
          "language": "english"
        },
        "english_possessive_stemmer": {
          "type": "stemmer",
          "language": "possessive_english"
        }
      },
      "analyzer": {
        "rebuilt_english": {
          "tokenizer": "standard",
          "filter": [
            "english_possessive_stemmer",
            "lowercase",
            "english_stop",
            "english_keywords",
            "english_stemmer"
          ]
        }
      }
    }
  }
}
```

Let's analyze everything 

Books - Bulgarian

```
{
  "settings": {
    "analysis": {
      "filter": {
        "bulgarian_stop": {
          "type": "stop",
          "stopwords": "_bulgarian_"
        },
        "bulgarian_keywords": {
          "type": "keyword_marker",
          "keywords": ["пример"]
        },
        "bulgarian_stemmer": {
          "type": "stemmer",
          "language": "bulgarian"
        }
      },
      "analyzer": {
        "rebuilt_bulgarian": {
          "tokenizer": "standard",
          "filter": [
            "lowercase",
            "bulgarian_stop",
            "bulgarian_keywords",
            "bulgarian_stemmer"
          ]
        }
      }
    }
  }
}
```



What about Drupal?

With ElasticSearch

What about Drupal?

Context

- Elastic search
- `search_api` module
- `elasticsearch_connector` module



What about Drupal?

What should we do?

- Before the index is prepared
- Access to the index
- Change its settings

What about Drupal?

Subscribe?

- `BuildIndexParamsEvent`
- `BuildSearchParamsEvent`
- `PrepareIndexEvent`
- `PrepareIndexMappingEvent`
- `PrepareMappingEvent`
- `PrepareSearchQueryEvent`

What about Drupal?

Subscribe!

```
class ChangeIndexEventSubscriber implements EventSubscriberInterface {  
  
    public static function getSubscribedEvents() {  
        return [  
            PrepareIndexEvent::PREPARE_INDEX => 'prepareIndex',  
        ];  
    }  
}  
"  
}
```

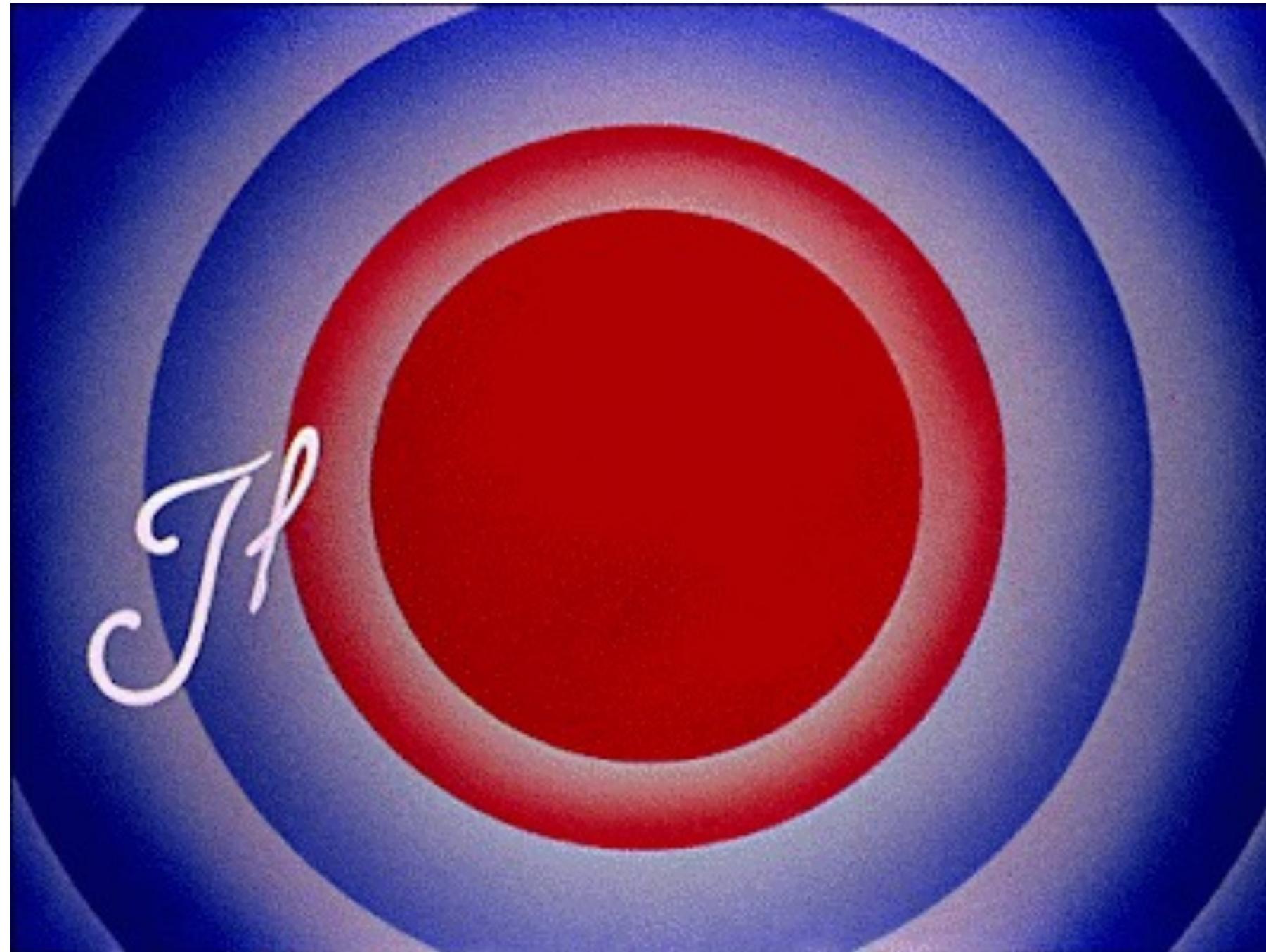
What about Drupal?

Set it up!

```
class ChangeIndexEventSubscriber implements EventSubscriberInterface {  
    public function prepareIndex(PrepareIndexEvent $event) {  
        $indexConfig = $event->getIndexConfig();  
        $indexConfig['body']['settings']['analysis']['analyzer'] = [  
            'whitespace_lowercase' => [  
                'tokenizer' => 'whitespace',  
                'filter' => [ 'lowercase' ],  
            ],  
        ];  
        ...  
        $event->setIndexConfig($indexConfig);  
    }  
    ...  
}
```

What about Drupal?

And...



Conclusion

- Anything could change the score
- Refine as you go

Sources 1/2

https://en.wikipedia.org/wiki/Vector_space_model

https://lucene.apache.org/core/3_5_0/api/core/org/apache/lucene/search/Similarity.html

<https://www.youtube.com/watch?v=UhONe6GSfGQ&list=PL9zDdgiGjkIcNOfBpm7NX3ZC5Fh7e0OMj>

<https://www.elastic.co/guide/en/elasticsearch/reference/current/analyser-anatomy.html>

<https://www.youtube.com/watch?v=ajNfOPeWiAY>

<https://www.elastic.co/fr/blog/how-to-improve-elasticsearch-search-relevance-with-boolean-queries>

https://en.wikipedia.org/wiki/Inverted_index

Sources 2/2

<https://www.elastic.co/fr/blog/how-to-improve-elasticsearch-search-relevance-with-boolean-queries>

<https://www.elastic.co/fr/blog/practical-bm25-part-2-the-bm25-algorithm-and-its-variables>

<https://medium.com/elasticsearch/introduction-to-analysis-and-analyzers-in-elasticsearch-4cf24d49ddab>



Nerea Enrique

@ nerea.enrique@gmail.com

Thank you. Merci. 謝謝. dhany



<https://github.com/Curryed/search>