



elasticsearch



me

Developer at Twinslash

email: dima.zhlobo@gmail.com

skype: [dima.zhlobo](https://www.skype.com/en/contacts/dima.zhlobo)

github: [proghat](https://github.com/proghat)

twitter: [@proghat](https://twitter.com/proghat)



Dmitry Zhlobo

search is hard

- speed vs. relevancy

search is hard

- speed vs. relevancy
- real time

search is hard

- speed vs. relevancy
- real time
- different kinds of data

usual approach

- `SELECT * FROM posts WHERE `body`
LIKE '%query%'`

usual approach

- `SELECT * FROM posts WHERE `body`
LIKE '%query%'`

- `gem 'thinking-sphinx'`

...

`Article.search(params[:q])`

usual approach

- `SELECT * FROM posts WHERE `body`
LIKE '%query%'`
- `gem 'thinking-sphinx'`
- ...

`Article.search(params[:q])`



how search works?

- document 1:
flexible and powerful open source, distributed real-time search and analytics engine for the cloud...
- document 2:
Apache Mahout has implementations of a wide range of machine learning and data mining...
- document 3:
Our core algorithms for clustering, classification and batch based collaborative filtering are implemented on top of Apache Hadoop using the MapReduce...

how search works?

data

analysis

learning

classification

recommenders

mapreduce

how search works?

data

analysis

learning

classification

recommenders

mapreduce

1

2

3

1

2

2

3

2

3

2

3

how search works?

data

analysis

learning

classification

recommenders

mapreduce

1

2

3

1

2

2

3

2

3

2

3

how search works?

data

analysis

learning

classification

recommenders

mapreduce

1

2

3

1

2

2

3

2

3

2

3

how search works?

data	1	2	3
analysis	1	2	
learning	2	3	
classification	2	3	
recommenders	2		
mapreduce	3		

elasticsearch

elasticsearch

- full text search
- real time data
- opensource
- restful api
- distributed
- schema free & document oriented

analysis

Flexible and powerful **search** engine

analysis

**char
filters**

Flexible and powerful `<strong>search</strong>` engine

Mapping

Pattern Replace

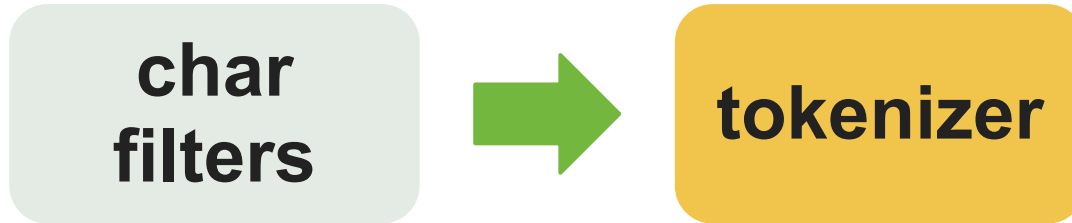
HTML Strip

analysis

**char
filters**

Flexible and powerful search engine

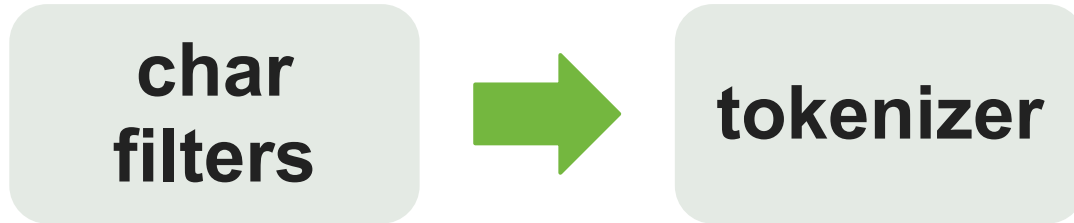
analysis



Flexible and powerful search engine

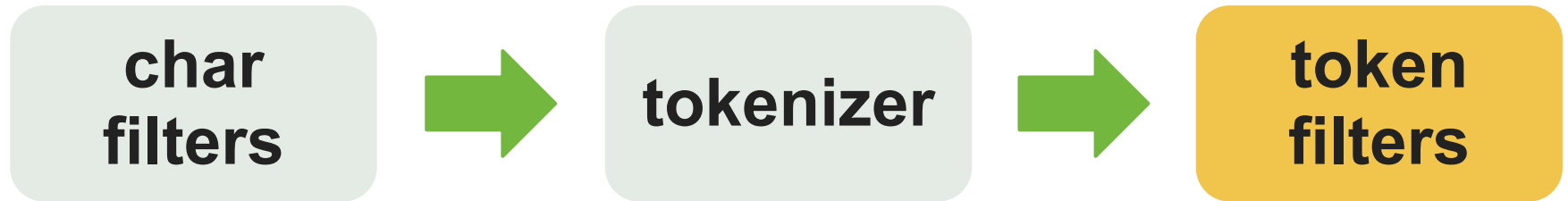
Edge NGram Letter Standard Path Hierarchy
Keyword Whitespace
NGram Lowercase Pattern

analysis



Flexible and powerful search engine

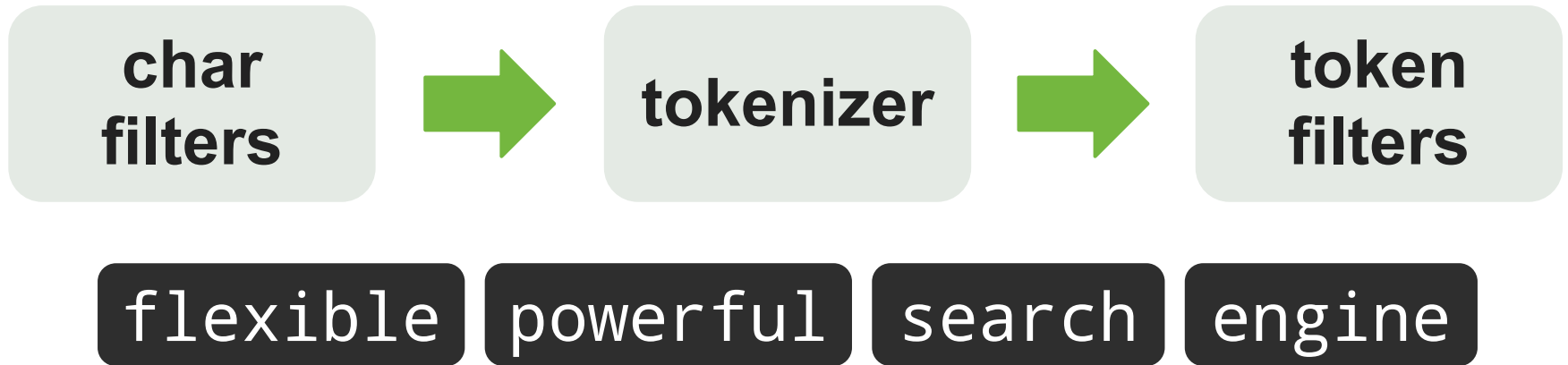
analysis



Flexible and powerful search engine

Stop Lowercase Stemmer Truncate
Synonym Shingle
Snowball
Unique Trim Normalization Reverse

analysis



russian morphology

“Он отлично рассказал о лучшем поисковом движке”

russian morphology

“Он **отлично** рассказал о лучшем поисковом движке”

отлично

отличный

russian morphology

“Он отлично **рассказал** о лучшем поисковом движке”

отлично

отличный

рассказать

russian morphology

“Он отлично рассказал о **лучшем** поисковом движке”

отлично

отличный

рассказать

хороший

russian morphology

“Он отлично рассказал о лучшем **ПОИСКОВОМ** движке”

отлично

отличный

рассказать

хороший

ПОИСКОВЫЙ

russian morphology

“Он отлично рассказал о лучшем поисковом **движке**”

отлично

отличный

рассказать

хороший

поисковый

движок

russian morphology

“Он отлично рассказал о лучшем поисковом движке”

отлично

отличный

рассказать

хороший

поисковый

движок

russian morphology

“Он отлично рассказал о лучшем поисковом движке”

отлично

отличный

рассказать

хороший

поисковый

движок

“хороший поисковой движок”

хороший

поисковый

движок

russian morphology

“Он отлично рассказал о лучшем поисковом движке”

отлично

отличный

рассказать

хороший

поисковый

движок

“хороший поисковой движок”

хороший

поисковый

движок

phonetic analysis



phonetic analysis

Eyjafjallajökull

phonetic analysis

Eyjafjallajökull

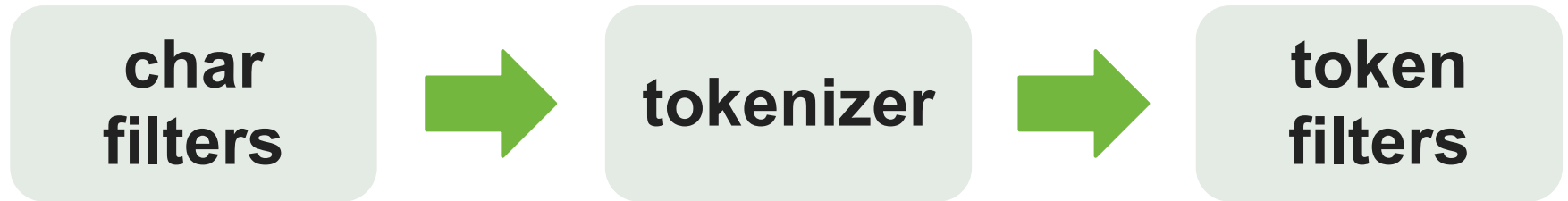


iofiolDkul

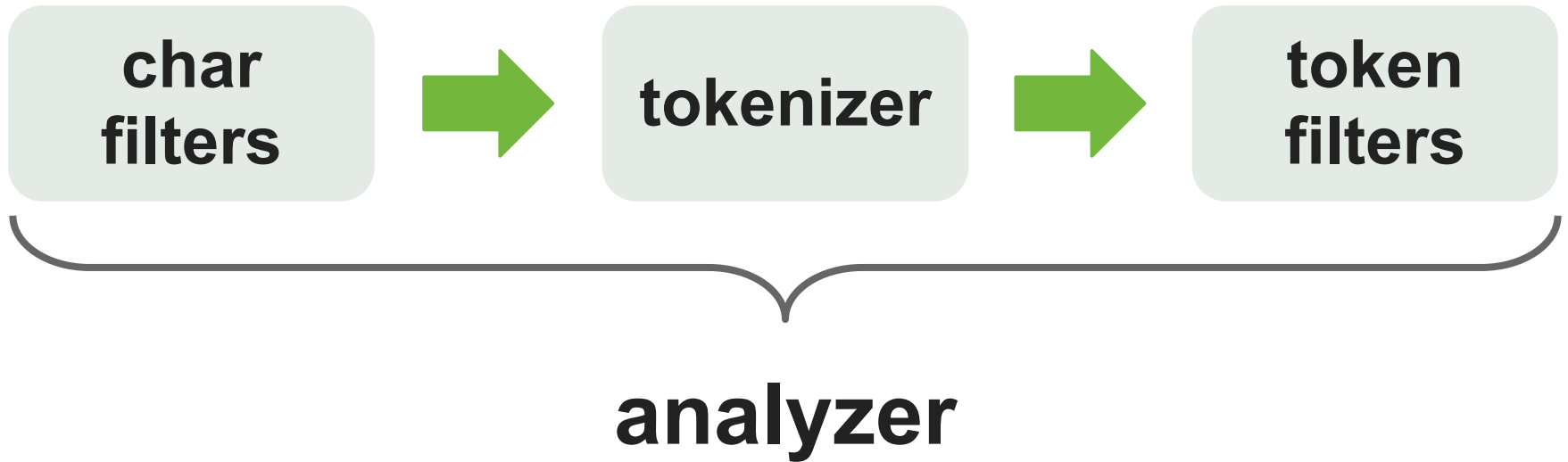


Eiyafyalayokul

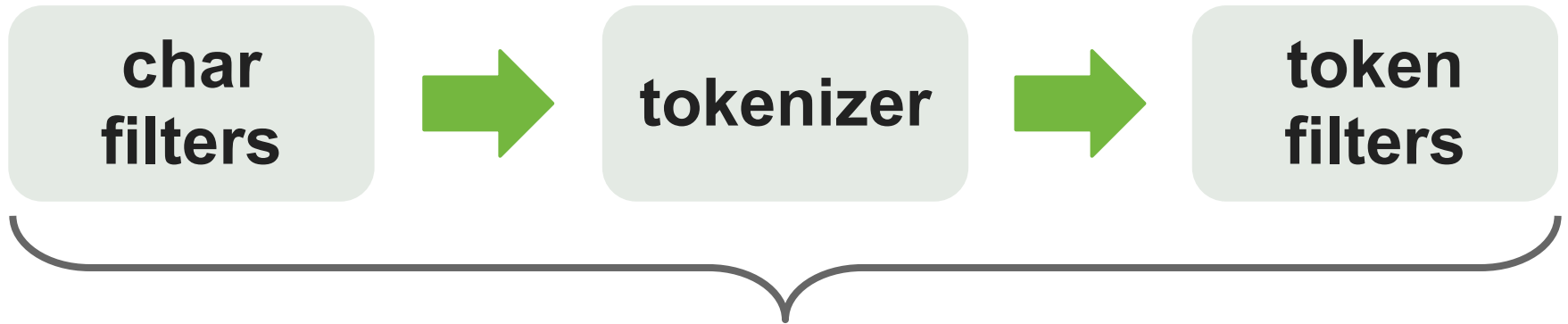
analysis



analysis



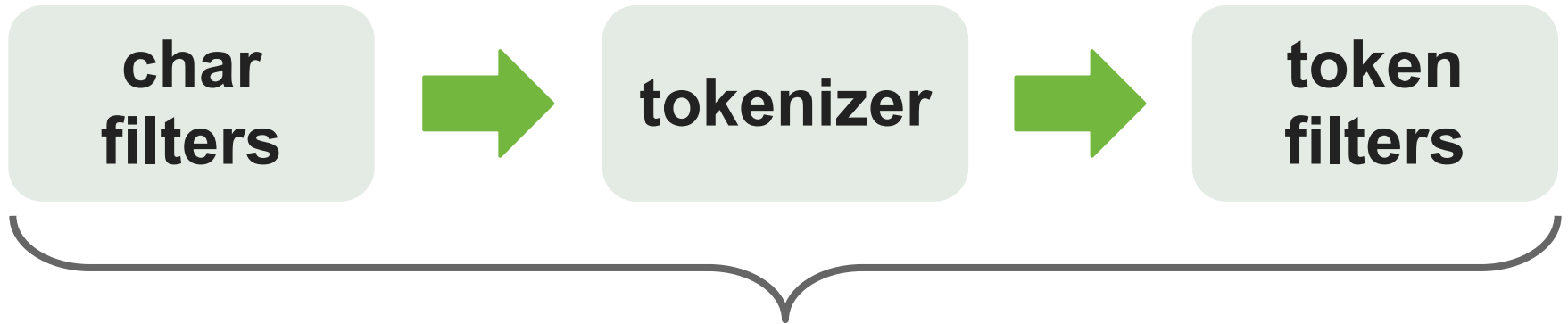
analysis



analyzer

- has name

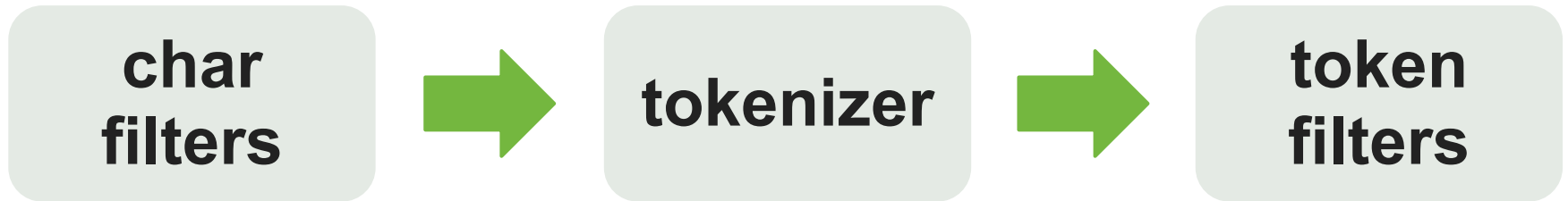
analysis



analyzer

- has name
- reusable

analysis



analysis:

analyzer:

rus_morphology:

type: "custom"

char_filter: ["html_strip"]

tokenizer: "standard"

filter: ["lowercase", "russian_morphology", "stopwords"]

getting started

Add document to index

```
curl -XPOST "localhost:9200/posts/post/1" -d '{ title: "Search" }'
```

getting started

Add document to index

```
curl -XPOST "localhost:9200/posts/post/1" -d '{ title: "Search" }'
```

getting started

Add document to index

```
curl -XPOST "localhost:9200/posts/post/1" -d '{ title: "Search" }'
```

getting started

Add document to index

```
curl -XPOST "localhost:9200/posts/post/1" -d '{ title: "Search" }'
```

getting started

Add document to index

```
curl -XPOST "localhost:9200/posts/post/1" -d '{ title: "Search" }'
```

```
curl -XPOST "localhost:9200/posts/post/whatever" -d '{ title: "ES" }'
```

getting started

Add document to index

```
curl -XPOST "localhost:9200/posts/post/1" -d '{ title: "Search" }'
```

```
curl -XPOST "localhost:9200/posts/post/whatever" -d '{ title: "ES" }'
```

Search documents

```
curl -XGET "localhost:9200/posts/_search?q=data"
```

```
curl -XGET "localhost:9200/posts/_search?q=title:elasticsearch"
```

getting started

Add document to index

```
curl -XPOST "localhost:9200/posts/post/1" -d '{ title: "Search" }'
```

```
curl -XPOST "localhost:9200/posts/post/whatever" -d '{ title: "ES" }'
```

Search documents

```
curl -XGET "localhost:9200/posts/_search?q=data"
```

```
curl -XGET "localhost:9200/posts/_search?q=title:elasticsearch"
```

Update and delete documents

```
curl -XPUT "localhost:9200/posts/post/1" -d '{ title: "Data" }'
```

```
curl -XDELETE "localhost:9200/posts/post/whatever"
```

mapping

```
{  
  repository: "elasticsearch/elasticsearch",  
  description: "Open Source, Distributed, RESTful Search Engine",  
  maintainer: {  
    name: "Shay Banon",  
    email: "kimchy@gmail.com"  
  },  
  languages: ["Java", "Shell"],  
  created_at: "2010-02-08"  
}
```

mapping

```
{  
  repository: "elasticsearch/elasticsearch",  
  description: "Open Source, Distributed, RESTful Search Engine",  
  maintainer: {  
    name: "Shay Banon",  
    email: "kimchy@gmail.com"  
  },  
  languages: ["Java", "Shell"],  
  created_at: "2010-02-08"  
}  
  
repository: {  
  type: "string"  
}
```

mapping

```
{
  repository: "elasticsearch/elasticsearch",
  description: "Open Source, Distributed, RESTful Search Engine",
  maintainer: {
    name: "Shay Banon",
    email: "kimchy@gmail.com"
  },
  languages: ["Java", "Shell"],
  created_at: "2010-02-08"
}

repository: {
  type: "string",
  boost: 5
}
```

mapping

```
{
  repository: "elasticsearch/elasticsearch",
  description: "Open Source, Distributed, RESTful Search Engine",
  maintainer: {
    name: "Shay Banon",
    email: "kimchy@gmail.com"
  },
  languages: ["Java", "Shell"],
  created_at: "2010-02-08"
}

repository: {
  type: "string",
  boost: 5,
  analyzer: "repo_name"
}

repo_name: {
  tokenizer: "letter",
  filter: ["lowercase", "phonetic"]
}
```

mapping

```
{
  repository: "elasticsearch/elasticsearch",
  description: "Open Source, Distributed, RESTful Search Engine",
  maintainer: {
    name: "Shay Banon",
    email: "kimchy@gmail.com"
  },
  languages: ["Java", "Shell"],
  created_at: "2010-02-08"
}

description: {
  type: "string"
}
```

mapping

```
{
  repository: "elasticsearch/elasticsearch",
  description: "Open Source, Distributed, RESTful Search Engine",
  maintainer: {
    name: "Shay Banon",
    email: "kimchy@gmail.com"
  },
  languages: ["Java", "Shell"],
  created_at: "2010-02-08"
}

description: {
  type: "string",
  analyzer: "english_text"
}

english_text: {
  tokenizer: "standard",
  filter:
    ["lowercase", "stopwords", "snowball"]
}
```

mapping

```
{  
  repository: "elasticsearch/elasticsearch",  
  description: "Open Source, Distributed, RESTful Search Engine",  
  maintainer: {  
    name: "Shay Banon",  
    email: "kimchy@gmail.com"  
  },  
  languages: ["Java", "Shell"],  
  created_at: "2010-02-08"  
}  
  
maintainer: {  
  properties: {  
  
  }  
}
```

mapping

```
{
  repository: "elasticsearch/elasticsearch",
  description: "Open Source, Distributed, RESTful Search Engine",
  maintainer: {
    name: "Shay Banon",
    email: "kimchy@gmail.com"
  },
  languages: ["Java", "Shell"],
  created_at: "2010-02-08"
}

maintainer: {
  properties: {
    name: {
      type: "string"
    }
  }
}
```

mapping

```
{
  repository: "elasticsearch/elasticsearch",
  description: "Open Source, Distributed, RESTful Search Engine",
  maintainer: {
    name: "Shay Banon",
    email: "kimchy@gmail.com"
  },
  languages: ["Java", "Shell"],
  created_at: "2010-02-08"
}

maintainer: {
  properties: {
    name: {
      type: "string",
      analyzer: "phonetic"
    }
  }
}

phonetic: {
  tokenizer: "standard",
  filter: ["lowercase", "stopwords",
    "beidermorse"]
}
```

mapping

```
{
  repository: "elasticsearch/elasticsearch",
  description: "Open Source, Distributed, RESTful Search Engine",
  maintainer: {
    name: "Shay Banon",
    email: "kimchy@gmail.com"
  },
  languages: ["Java", "Shell"],
  created_at: "2010-02-08"
}

maintainer: {
  properties: {
    email: {
      type: "string"
    }
  }
}
```

mapping

```
{
  repository: "elasticsearch/elasticsearch",
  description: "Open Source, Distributed, RESTful Search Engine",
  maintainer: {
    name: "Shay Banon",
    email: "kimchy@gmail.com"
  },
  languages: ["Java", "Shell"],
  created_at: "2010-02-08"
}

maintainer: {
  properties: {
    email: {
      type: "string",
      index: "not_analyzed"
    }
  }
}
```

mapping

```
{  
  repository: "elasticsearch/elasticsearch",  
  description: "Open Source, Distributed, RESTful Search Engine",  
  maintainer: {  
    name: "Shay Banon",  
    email: "kimchy@gmail.com"  
  },  
  languages: ["Java", "Shell"],  
  created_at: "2010-02-08"  
}  
language: {  
  type: "string"  
}
```

mapping

```
{
  repository: "elasticsearch/elasticsearch",
  description: "Open Source, Distributed, RESTful Search Engine",
  maintainer: {
    name: "Shay Banon",
    email: "kimchy@gmail.com"
  },
  languages: ["Java", "Shell"],
  created_at: "2010-02-08"
}

language: {
  type: "string",
  analyzer: "programming_lang"
}

programming_lang: {
  tokenizer: "keyword",
  filter: ["lowercase"]
}
```

mapping

```
{
  repository: "elasticsearch/elasticsearch",
  description: "Open Source, Distributed, RESTful Search Engine",
  maintainer: {
    name: "Shay Banon",
    email: "kimchy@gmail.com"
  },
  languages: ["Java", "Shell"],
  created_at: "2010-02-08"
}

created_at: {
  type: "date",
  format: "YYYY-MM-DD"
}
```

mapping

```
curl -XPOST "localhost:9200/repositories" -d '{
  settings: {
    analysis: {
      analyzer: { ... },
      filter: { ... }
    }
  },
  mappings: {
    repository: {
      properties: { ... }
    }
  }
}'
```

mapping

```
curl -XPOST "localhost:9200/repositories" -d '...
```

```
curl -XPOST "localhost:9200/repositories/repository" -d '{
  repository: "elasticsearch/elasticsearch",
  description: "Open Source, Distributed, RESTful Search Engine",
  maintainer: {
    name: "Shay Banon",
    email: "kimchy@gmail.com"
  },
  languages: ["Java", "Shell"],
  created_at: "2010-02-08"
}'
```

search

```
curl -XGET "localhost:9200/repositories/repository/_search?q=engine"
```

search

```
curl -XPOST "localhost:9200/repositories/repository/_search" -d '{
  query: {
    match: {
      description: "search"
    }
  }
}'
```

search

```
"hits" : {  
  "total" : 3,  
  "hits" : [  
    {  
      "_score" : 0.22295055,  
      "_source" : { repository: "elasticsearch/elasticsearch" }  
    }, {  
      "_score" : 0.22295055,  
      "_source" : { repository: "ankane/searchkick" }  
    }, {  
      "_score" : 0.095891505,  
      "_source" : { repository: "karmi/tire" }  
    }  
  ]  
}
```

search

```
curl -XPOST "localhost:9200/repositories/repository/_search" -d '{
  query: {
    match: {
      _all: "elasticsearch"
    }
  }
}'
```

search

```
"hits" : {  
  "total" : 2,  
  "hits" : [  
    {  
      "_score" : 5.46875,  
      "_source" : { repository: "elasticsearch/elasticsearch" }  
    }, {  
      "_score" : 0.04746387,  
      "_source" : { repository: "karmi/tire" }  
    }  
  ]  
}
```

facets

```
curl -XPOST "localhost:9200/repositories/repository/_search" -d '{
  query: {
    match: {
      _all: "search"
    }
  },
  facets: {
    language: {
      terms: {
        field: "languages"
      }
    }
  }
}'
```

facets

```
"hits" : {  
  "total" : 2,  
  "hits" : [ ... ]  
},  
"facets" : {  
  "language" : {  
    "terms" : [  
      { "term" : "ruby", "count" : 2 },  
      { "term" : "shell", "count" : 1 },  
      { "term" : "java", "count" : 1 }  
    ]  
  }  
}
```

filters

```
curl -XPOST "localhost:9200/repositories/repository/_search" -d '{
  query: {
    filtered: {
      query: {
        match: {
          _all: "search"
        }
      },
      filter: {
        term: { "languages": "java" }
      }
    }
  }
}'
```

filters

```
"hits" : {  
  "total" : 1,  
  "hits" : [  
    {  
      "_score" : 5.46875,  
      "_source" : { repository: "elasticsearch/elasticsearch" }  
    }  
  ]  
}
```

performance and scaling

performance and scaling

elasticsearch is web scale



random facts

- bulk operations
- real time
- highlights
- geo types and geo distance facets
- attachments
- “did you mean?” and completions
- common terms
- filters and caching
- river

You know. For search.