

Elasticsearch as we know it

The Damerau-Levenshtein distance

[Elasticsearch] uses the Damerau-Levenshtein distance to find all terms with a maximum of two changes, where a change is the insertion, deletion or substitution of a single character, or transposition of two adjacent characters.

The default edit distance is 2, but an edit distance of 1 should be sufficient to catch 80% of all human misspellings

From:

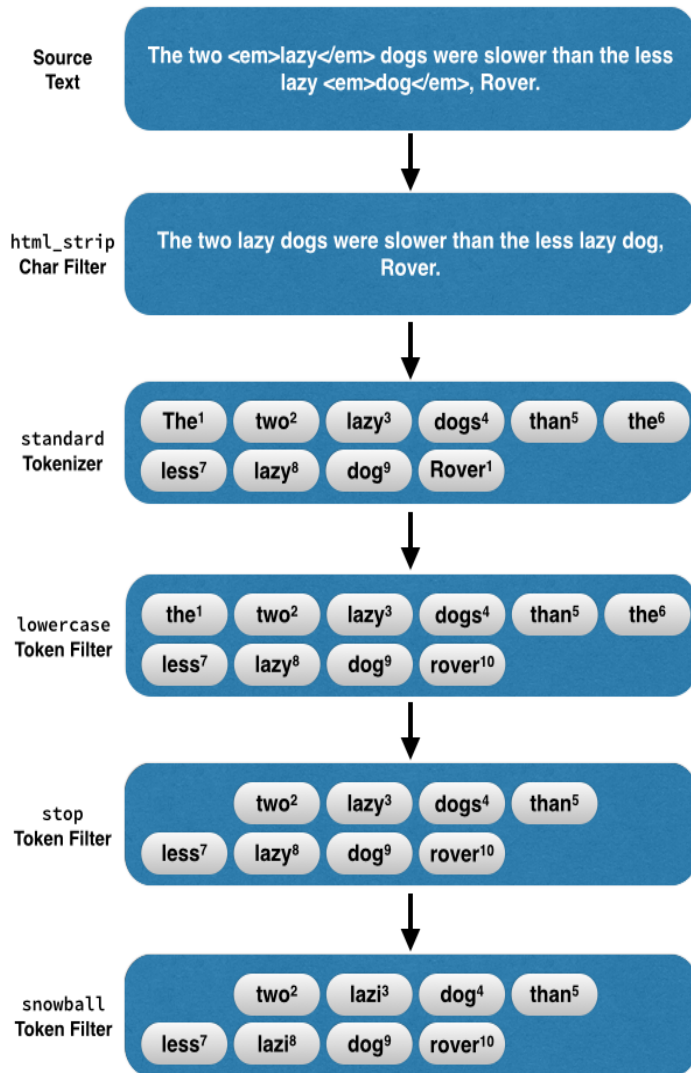
https://www.elastic.co/guide/en/elasticsearch/reference/current/query-dsl-query-string-query.html#_fuzziness

The Components of an Analyzer



From: <https://www.elastic.co/blog/found-text-analysis-part-1>

Custom Analysis Flow



From: <https://www.elastic.co/blog/found-text-analysis-part-1>

Standard analyzer

The standard analyzer is the default analyzer that Elasticsearch uses. It is the best general choice for analyzing text that may be in any language. It splits the text on word boundaries, as defined by the Unicode Consortium, and removes most punctuation. Finally, it lowercases all terms.

Set the shape to semi-transparent by calling `set_trans(5)`
becomes

```
set, the, shape, to, semi, transparent, by, calling,  
set_trans, 5
```

From:

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

```
{
  "followers": [
    { "age": 35, "name": "Mary White"},
    { "age": 26, "name": "Alex Jones"},
    { "age": 19, "name": "Lisa Smith"}
  ]
}
```

The result will look like this:

```
{
  "followers.age": [19, 26, 35],
  "followers.name": [alex, jones, lisa, smith, mary, v
}
```

The correlation between **{age: 35}** and **{name: Mary White}** is lost.

Correlated inner objects, which are able to answer queries like these, are called *nested objects*.

From: