# Elasticsearch 0.90.9 with RankingAlgorithm 1.5.x

By

Nagendra Nagarajayya
tgels, inc
http://elasticsearch-ra.tgels.org

# Elasticsearch 0.90.9 with RankingAlgorithm 1.5.2

By

Nagendra Nagarajayya
http://elasticsearch-ra.tgels.org
updated 01/05/2014

## 1. Introduction

elasticsearch a very fast, elastic, open source search platform can now work with a new search library RankingAlgorithm along with Lucene. elasticsearch with RankingAlgorithm search seems to be comparable to Google site search (see Perl index comparison results) for certain queries and much better than Lucene. RankingAlgorithm 1.5.2 enables elasticsearch to rank product searches very accurately.

Multiple Algorithms are available SIMPLE, SIMPLE1, COMPLEX, COMPLEX1 and COMPLEX-LSA. SIMPLE is a very fast algorithm and can return queries in <100ms on a 10m wikipedia index (complete index). It can also scale to 100m docs or maybe more. COMPLEX is a more complex algorithm so is a little slower compared to the SIMPLE, but can also still return queries in < 100ms on a 10m wikipedia index (complete index). COMPLEX is more accurate and should be able to give you the best rankings as compared to SIMPLE.

RankingAlgorithm has been integrated into elastisearch in such a way that either Lucene or the RankingAlgorithm can be used to do the search. RankingAlgorithm scoring does not break any of the existing functionality. So elasticity, shrad, faceting, highlighting, etc. still work as before.

## 2. Using elasticsearch with RankingAlgorithm

There is no change in the way you access elasticsearch. All searches work the same as before.

So a elasticsearch such as:

```
curl -XGET 'http://localhost:9200/twitter/tweet/_search?q=california+gold&pretty=true'
```

should still work as before. The only difference is that elasticsearch instead of using Lucene library for search, uses the RankingAlogrithm library to search and ranks the documents. The returned scores are different from Lucene and reflects the relevancy of a document.

As said above, RankingAlgorithm scores in two different modes, Document mode and Product mode. In Document mode, the scoring is for relevance and in Product mode, scoring is for occurrence. Document mode is suitable for general purpose searches such as Wikipedia docs, HTML, Word/PDF or similar docs. The Document mode is the default. Product mode is for searches found on retail stores, online store/shopping/comparison/auction websites, etc, including short text sites like tweeter.

Product mode takes a term occurrence into account and scores accordingly. For eg. a search for “wii console” will show titles starting with “wii console” are first, and the others rank lower as the occurrence of “wii console” shifts in the title or gets reversed, see below:

**Wii Console** and Wii Fit Plus with Balance Board Bundle (Nintendo Wii)
**Wii Console** System with Wii Sports Resort Game with TWO MotionPlus Attachments
Nintendo **Wii Console** w/ Bonus Wii Sports Resort Bundle, Black
Pelican Accessories **Wii Console** Stand - Nintendo Wii
Grafitti Skin for Nintendo **Wii Console**
NEW AC Adapter Cable Cord Power Supply For NINTENDO **WII** Gaming **Console**
**Wii** Remote Charging **Console** Stand
Nintendo **Wii** Skin - System **Console** Skin and two Wii Remote Skins - Blue Vortex
CET Domain 10301901 **Console** Stand Station for Nintendo **Wii**

There is also a scan attribute, where the scan can be a fast scan, medium scan or a full scan. Scan is the depth of the search so can be fast, slower or slow. The default is fast scan.

## *2.1 Enabling different algorithms*

To use the SIMPLE algorithm (default), use:

export ES_JAVA_OPTS=”-DaIgorithm=simple”
restart elasticsearch

To use the COMPLEX algorithm in document mode, use:

export ES_JAVA_OPTS=”-Dalgorithm=complex -Dmode=document”
restart elasticsearch

To use the COMPLEX algorithm in product mode, use:
export ES_JAVA_OPTS=”-Dalgorithm=complex -Dmode=product”
restart elasticsearch

Default is document mode.

SIMPLE algorithm functions only in document mode. COMPLEX algorithm is more accurate but a little slow compared to SIMPLE algorithm. SIMPLE is very fast and can return queries on the wikipedia index in < 50 ms. Both SIMPLE/COMPLEX can scale to 2 billion documents.

### *2.2 Enabling lucene library*

To use lucene library:

export ES_JAVA_OPTS=-Dlibrary=lucene
restart elasticsearch

## 3. Installing elastisearch with the RankingAlgorithm

You can install elasticsearch with RA in two different ways. You can download elasticsearch with RA.zip a bundle of elasticsearch 0.90.9 and Ranking Algorithm (a big download) or just download the elastisearch.zip from here, install as with instructions from here, and then download Download RankingAlgorithm exposed as lucene-core-4.6 below, and then download RankingAlgorithm 1.5.2 itself. See below for more details.

### *3.1 Download elasticsearch 0.90.9 with RA.zip (bundle)*

Installation is the same as elasticsearch. See here

Step1:

Download elasticsearch with RA.zip from here

Step2:

Unzip it to a directory

Step3:

bin/elasticsearch

### *3.2 Download elastisearch and RankingAlgorithm separately*

Download elasticsearch from here as before. Install elasticsearch as in the instructions here.

- Follow below steps:
    - Download RankingAlgorithm exposed as lucene-core-4.6 from here .
    - Download RankingAlgorithm 1.5.x from here
    - mv elasticsearch/lib/lucene-core-4.6.jar to /tmp
    - mv downloads/lucene-ra-core-4.6.jar elasticsearch/lib
    - mv downloads/rankingalgorithm40_1.5.2 elasticsearch/lib

## 4. Conclusion

Elasticsearch with RankingAlgorithm offers a new search library along with lucene. RankingAlgorithm ranking seems to be comparable to Google site search (see Perl index comparison results) and much better than Lucene.

SIMPLE algorithm returns queries on a 10m wikipedia index in <100 ms. COMPLEX algorithm is more accurate but a little slower and can also return queries in <100ms. In document mode RankingAlgorithm ranks documents relevantly while ranking very accurately and precisely in the product mode. Document mode is very well suited for searching html, wikipedia, pdf/word type documents, while product works very well with short text as in retail websites, product comparison websites, short text messaging like twitter, etc. RankingAlgorithm with document and product mode is very well suited for the enterprise as well as the retail, ecommerce and websites.

## 5. References

1. elastisearch with RA, http://elasticsearch-ra.tgels.org
2. RankingAlgorithm, http://rankginalgorithm.tgels.org
3. elasticsearch, http://elasticsearch.org