

ELASTICSEARCH

INTRODUCTION TO



What is this about?

what is this about **blank**

what is this about **the world ending**

what is this about **a girl eating a tampon**

[I'm Feeling Lucky »](#)

what is this about **the world ending tomorrow**



Jurriaan Persyn
@oemebamo – CTO Engagor



Aa Text | Photo | Quote | Link | Chat | Audio | Video

startupshizzle



When I have to write a commit message.

startupshizzle



When I have to order some extra sandwiches on top of our phone order at the sandwich place.

startupshizzle



TAGS

- #funny gifs
- #funny gifs tumblr

BLOGS

- Funny Gif and Video... funny-gifs-videos.tumblr.com
- Funny Gifs 4 U funny-gifs-4-u.tumblr.com
- Funny Gifs funniestgifsinlife.tumblr.com
- Funny Gifs extremefunnygifs.tumblr.com
- Funny Gifs From Ol... funnygifsfromoldsadmovies...

More blogs with "funny gifs"

curiae
Tumblr Radar

and ▼

or ▼

and ▼

Content ▼ contains any ▼ phpbenelux - + ≡

Content ▼ contains all ▼ php belgium - + ≡

Beer ▼ is ▼ Plenty - + ≡

Language ▼ is ▼ Dutch English - + ≡

Saved Searches ▲ ▼ Save

Reset Search

Introducing Graph Search

🔍 Photos before 1990

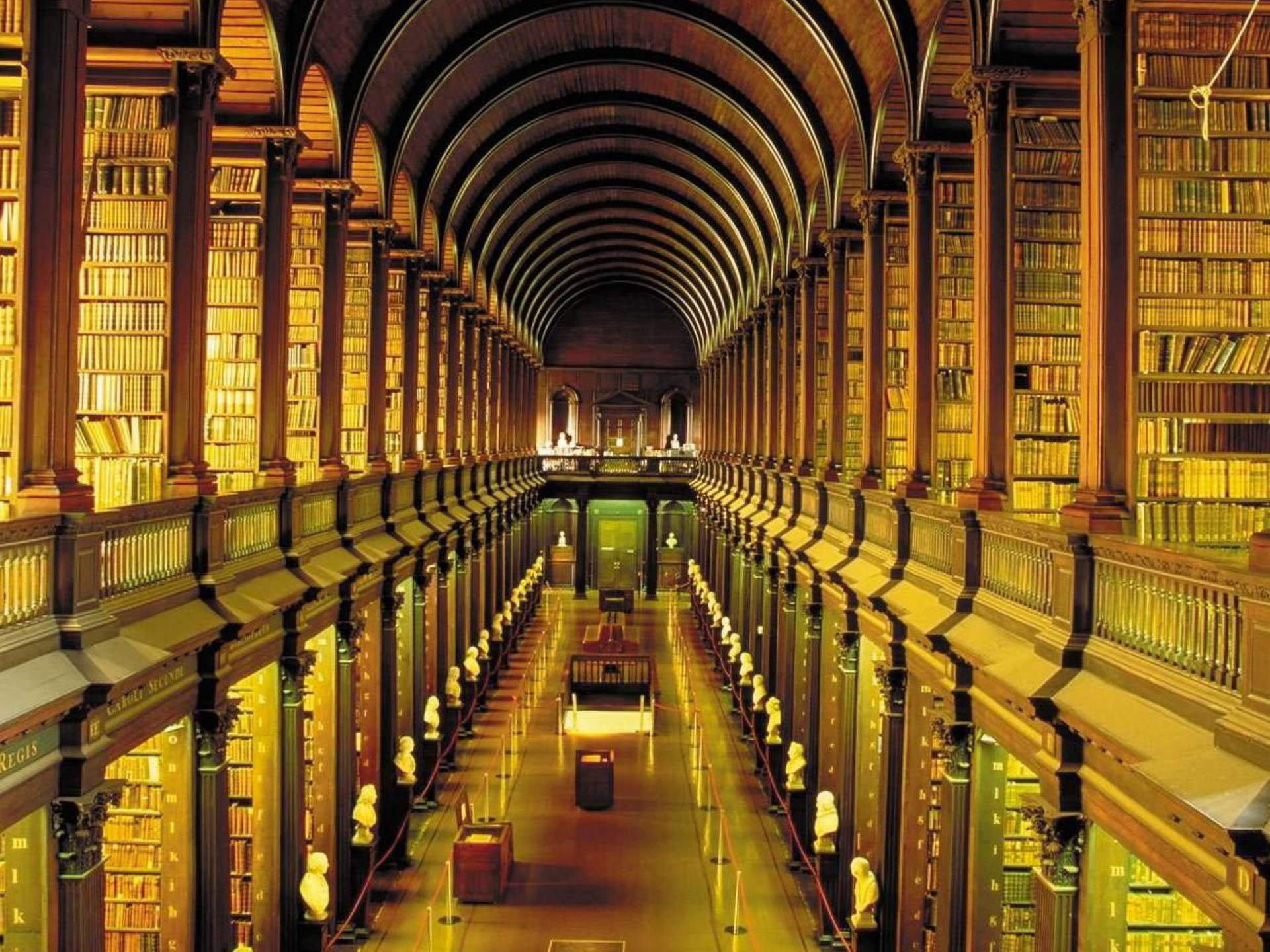


Explore your world through photos

Now you can use simple, specific phrases like “Photos my friends took in New York City” to find anything you want.



```
SELECT *  
FROM myawesomewebsite  
WHERE `text` LIKE  
„%shizzle%“
```



F G C D E
E T U D VV7
CERTALD
T II V W VV7
C P C M Mc N
K I H I D
F B C D E



TEXT SEARCH WITH PHP/MYSQL

- FULLTEXT index
 - Only for CHAR, VARCHAR & TEXT columns
 - For MyISAM & InnoDB tables
 - Configurable Stop Words
 - Types:
 - Natural Language
 - Natural Language with Query Expansion
 - Boolean Full Text

MYSQL FULLTEXT BOOLEAN MODE

Operators:

- + AND
- - NOT
- OR implied
- () Nesting
- * Wildcard
- " Literal matches

TEXT SEARCH WITH PHP/MYSQL (CONT'D)

- Typical columns for search table:
 - Type
 - Id
 - Text
 - Priority
- Process:
 - Blog posts, comments, ...
 - Save (filtered) duplicate of text in search table.
 - When searching ...
 - Search table and translate to original data via type/id

This is how most php/mysql sites implement their search, right?

```
SELECT *  
FROM mysearchtable  
WHERE MATCH(text)  
AGAINST („shizzle“)
```

```
SELECT *  
FROM mysearchtable  
WHERE MATCH(text)  
AGAINST („+shizzle –”ma  
nizzle”” IN BOOLEAN MODE)
```

```
SELECT * FROM jobs
WHERE role = „DEVELOPER“
AND MATCH(job_description)
AGAINST („node.js“)
```

```
SELECT * FROM jobs j
JOIN jobs_benefits jb ON j.id =
    jb.job_id
WHERE j.role = „DEVELOPER“
AND (MATCH(job_description)
    AGAINST („node.js -asp“ IN
        BOOLEAN MODE))
AND jb.free_espresso = TRUE
```



OH BOY!

WHAT IS A SEARCH ENGINE?

- Efficient indexing of data
 - On all fields / combination of fields
- Analyzing data
 - Text Search
 - Tokenizing
 - Stemming
 - Filtering
 - Understanding locations
 - Date parsing
- Relevance scoring

TOKENIZING

- Finding word boundaries
 - Not just explode(„“, „“, \$text);
 - Chinese has no spaces. (*Not every single character is a word.*)
- Understand patterns:
 - URLs
 - Emails
 - #hashtags
 - Twitter @mentions
 - Currencies (EUR, €, ...)

STEMMING

- *“Stemming is the process for reducing inflected (or sometimes derived) words to their stem, base or root form.”*
 - Conjugations
 - Plurals
- Example:
 - Fishing, Fished, Fish, Fisher > Fish
 - Better > Good
- Several ways to find the stem:
 - Lookup tables
 - Suffix-stripping
 - Lemmatization
 - ...
- Different stemmers for every language.

FILTERING

- Remove stop words
 - Different for every language
- HTML
 - If you're indexing web content, not every character is meaningful.

UNDERSTANDING LOCATIONS

- Reverse geocoding of locations to longitude & latitude
- Search on location:
 - Bounding box searches
 - Distance searches
 - Searching nearby
 - Geo Polygons
 - Searching a country

(Note: MySQL also has geospatial indexes.)

RELEVANCE SCORING

- From the matched documents, which ones do you show first?
- Several strategies:
 - How many matches in document?
 - How many matches in document as percentage of length?
 - Custom scoring algorithms
 - At index time
 - At search time
 - ... A combination

Think of Google PageRank.

“There’s an ~~app~~ **software** for
that.”

Success

APACHE LUCENE

- *“Information retrieval software library”*
- Free/open source
- Supported by Apache Foundation
- Created by Doug Cutting
- Written in 1999

~~“There’s software a Java library~~
for that.”





ELASTICSEARCH

- “*You know, for Search*”
- Also Free & Open Source
- Built on top of Lucene
- Created by Shay Banon @kimchy
- Versions
 - First public release, v0.4 in February 2010
 - A rewrite of earlier “Compass” project, now with scalability built-in from the very core
 - Now stable version at 0.20.6
 - Beta branch at 0.90 (*working towards 1.0 release*)
- In Java, so inherently cross-platform

WHAT DOES IT ADD TO LUCENE?

- **RESTfull Service**
 - JSON API over HTTP
 - Want to use it from PHP?
 - CURL Requests, as if you'd do requests to the Facebook Graph API.
- **High Availability & Performance**
 - Clustering
- Long Term **Persistency**
 - Write through to persistent storage system.

```
$ cd ~/Downloads
```

```
$ wget https://github.com/.../elasticsearch-0.20.5.tar.gz
```

```
$ tar -xzf elasticsearch-0.20.5.tar.gz
```

```
$ cd elasticsearch-0.20.5/
```

```
$ ./bin/elasticsearch
```

```
$ cd ~/Downloads
$ wget https://github.com/.../elasticsearch-0.20.5.tar.gz
$ tar -xzf elasticsearch-0.20.5.tar.gz
$ git clone https://github.com/elasticsearch/elasticsearch-
servicewrapper.git elasticsearch-servicewrapper
$ sudo mv elasticsearch-0.20.5 /usr/local/share
$ cd elasticsearch-servicewrapper
$ sudo mv service /usr/local/share/elasticsearch-0.20.5/bin
$ cd /usr/local/share
$ sudo ln -s elasticsearch-0.20.5 elasticsearch
$ sudo chown -R root:wheel elasticsearch
$ cd /usr/local/share/elasticsearch
$ sudo bin/service/elasticsearch start
```

```
$ sudo bin/service/elasticsearch start
```

```
Starting ElasticSearch...
```

```
Waiting for ElasticSearch...
```

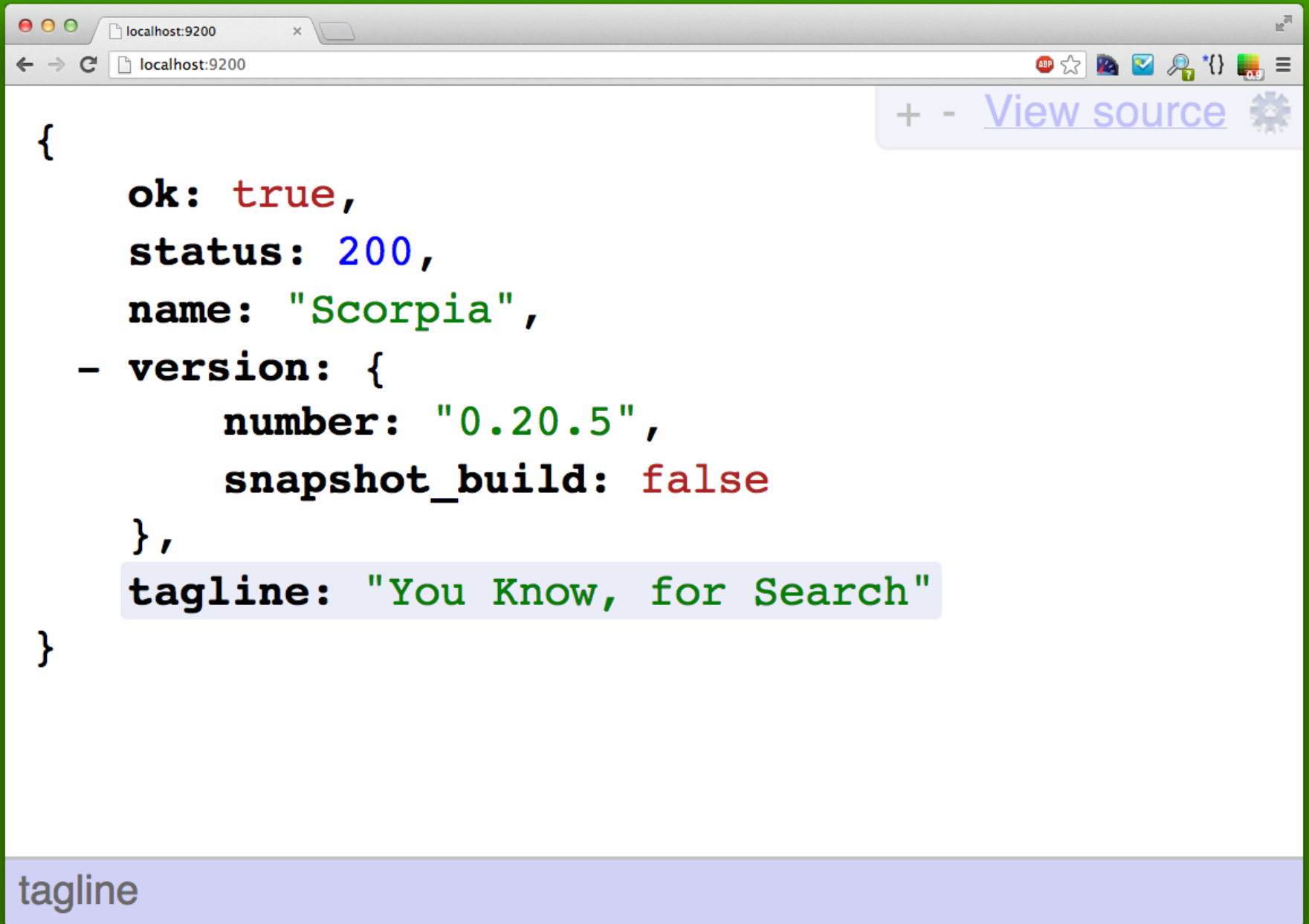
```
.
```

```
.
```

```
.
```

```
running: PID:83071
```

```
$
```



```
{
  ok: true,
  status: 200,
  name: "Scorpia",
  - version: {
    number: "0.20.5",
    snapshot_build: false
  },
  tagline: "You Know, for Search"
}
```

tagline

```
$ curl -XPUT http://localhost:9200/test/stupid-hypes/planking  
-d '{"name":"Planking", "stupidity_level":"5"}'
```

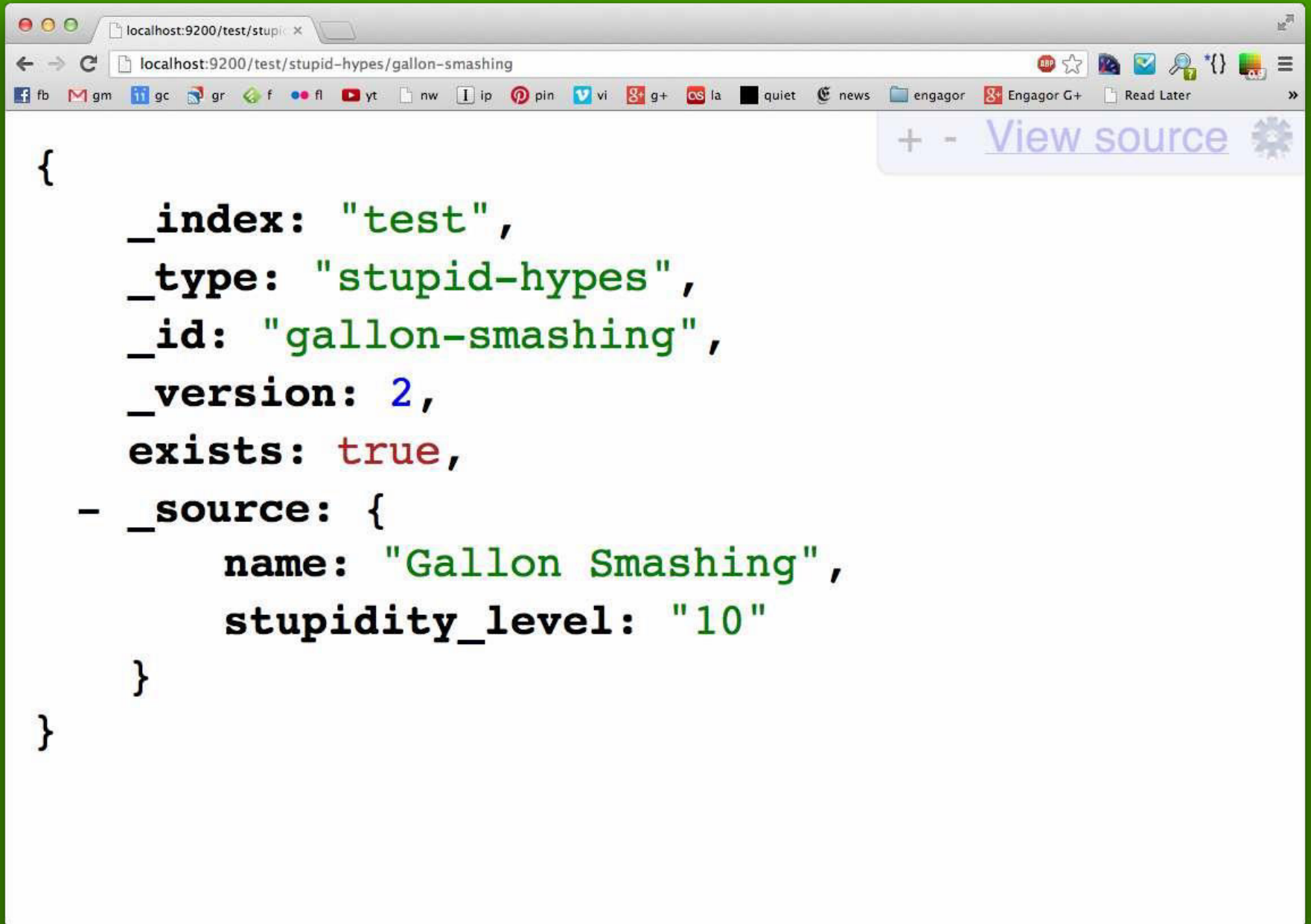
```
{"ok":true,"_index":"test","_type":"stupid-  
hypes","_id":"planking","_version":1}
```

```
$ curl -XPUT http://localhost:9200/test/stupid-hypes/gallon-smashing -d '{"name":"Gallon Smashing", "stupidity_level":"5"}'
```

```
{"ok":true,"_index":"test","_type":"stupid-hypes","_id":"gallon-smashing","_version":1}
```

```
$ curl -XPUT http://localhost:9200/test/stupid-hypes/gallon-smashing -d '{"name":"Gallon Smashing", "stupidity_level":"10"}
```

```
{"ok":true,"_index":"test","_type":"stupid-hypes","_id":"gallon-smashing","_version":2}
```



The image shows a web browser window with the address bar displaying `localhost:9200/test/stupid-hypes/gallon-smashing`. The page content is a JSON object representing a hype entry. The JSON is displayed with syntax highlighting: strings are green, numbers are blue, and booleans are red. A 'View source' button is visible in the top right corner of the page content area.

```
{
  _index: "test",
  _type: "stupid-hypes",
  _id: "gallon-smashing",
  _version: 2,
  exists: true,
  - _source: {
    name: "Gallon Smashing",
    stupidity_level: "10"
  }
}
```

```
$ curl -XPUT http://localhost:9200/test/stupid-hypes/gallon-smashing -d '{"name":"Gallon Smashing", "stupidity_level":"10", "lifetime":30}'
```

```
{"ok":true, "_index":"test", "_type":"stupid-hypes", "_id":"gallon-smashing", "_version":3}
```

```
localhost:9200/test/stupid-hypes/gallon-smashing
+ - View source
{
  _index: "test",
  _type: "stupid-hypes",
  _id: "gallon-smashing",
  _version: 3,
  exists: true,
  - _source: {
    name: "Gallon Smashing",
    stupidity_level: "10",
    lifetime: 30
  }
}
```

`_source.lifetime`

SCHEMALESS, DOCUMENT ORIENTED

- No need to configure schema upfront
- No need for slow ALTER TABLE –like operations
- You can define a mapping (schema) to customize the indexing process
 - Require fields to be of certain type
 - If you want text fields that should not be analyzed (stemming, ...)

“Ok, so it's a **NoSQL** store?”

```
{
  took: 2,
  timed_out: false,
  - _shards: {
    total: 5,
    successful: 5,
    failed: 0
  },
  - hits: {
    total: 1,
    max_score: 0.15342641,
    - hits: [
      - {
        _index: "test",
        _type: "stupid-hypes",
        _id: "gallon-smashing",
        _score: 0.15342641,
        - _source: {
          name: "Gallon Smashing",
          stupidity_level: "10",
          lifetime: 30
        }
      }
    ]
  }
}
```

hits.hits[0]._source.name

```
<?php
```

```
function searchElastic($index, $type, $query) {  
    // create curl resource  
    $ch = curl_init();  
  
    // set url  
    curl_setopt($ch, CURLOPT_URL, "http://localhost:9200/" . $index . "/" . $  
        type . "_search/?q=" . urlencode($query));  
  
    //return the transfer as a string  
    curl_setopt($ch, CURLOPT_RETURNTRANSFER, 1);  
  
    // $output contains the output string  
    $output = curl_exec($ch);  
  
    $error = curl_error($ch);  
    if ($error) {  
        throw new Exception($error);  
    }  
  
    // close curl resource to free up system resources  
    curl_close($ch);  
  
    $output = json_decode($output, true);  
  
    return $output;  
}  
  
var_dump(search('test', 'stupid-hypes', 'smashing'));
```

TERMINOLOGY

MySQL	Elastic Search
Database	Index
Table	Type
Row	Document
Column	Field
Schema	Mapping
Index	Everything is indexed
SQL	Query DSL
SELECT * FROM table ...	GET http://...
UPDATE table SET ...	PUT http://...

DISTRIBUTED & HIGHLY AVAILABLE

- Multiple servers (nodes) running in a **cluster**
 - Acting as single service
 - Nodes in cluster that store data or nodes that just help in speeding up search queries.
- **Sharding**
 - Indexes are sharded (# shards is configurable)
 - Each shard can have zero or more replicas
 - Replicas on different servers (server pools) for failover
 - One in the cluster goes down? No problem.
- **Master**
 - Automatic Master detection + failover
 - Responsible for distribution/balancing of shards

engagor

- Errors
- Fetchers
- ElasticSearch
- TOPIC MATCHER
 - Overview
 - Twitter
 - Topics
- INFO
 - Tasks
 - Shards
 - Topic Details
- STATS
 - Topics (Broken)
 - Connected Services
 - Help Wizards
 - Registration

Dashboard Performance Manage Metrics Tools Test

Shards primary: 11157
total: 22324

Health

server02	Khonshu	server02	server03	server04	server05	server06	server08	server09	server10	server11	server12	server13	server14
server03	Vanguard	server02	server03	server04	server05	server06	server08	server09	server10	server11	server12	server13	server14
server04	Victor Mancha	server02	server03	server04	server05	server06	server08	server09	server10	server11	server12	server13	server14
server05	Living Laser	server02	server03	server04	server05	server06	server08	server09	server10	server11	server12	server13	server14
server06	Gaea	server02	server03	server04	server05	server06	server08	server09	server10	server11	server12	server13	server14
server08	Diamondhead	server02	server03	server04	server05	server06	server08	server09	server10	server11	server12	server13	server14
server09	Ariann	server02	server03	server04	server05	server06	server08	server09	server10	server11	server12	server13	server14
server10	Human Torch	server02	server03	server04	server05	server06	server08	server09	server10	server11	server12	server13	server14
server11	Helleyes	server02	server03	server04	server05	server06	server08	server09	server10	server11	server12	server13	server14
server12	Agent Zero	server02	server03	server04	server05	server06	server08	server09	server10	server11	server12	server13	server14
server13	Radioactive Man	server02	server03	server04	server05	server06	server08	server09	server10	server11	server12	server13	server14
server14	Grim Reaper	server02	server03	server04	server05	server06	server08	server09	server10	server11	server12	server13	server14

SCALING ISSUES?

- No need for an external load balancer
 - Since cluster does its own routing.
 - Ask any server in the cluster, it will delegate to correct node.
- *What if ...*
 - More data > More shards.
 - More availability > More replicas per shard.

PERFORMANCE TWEAKING

- Bulk Indexing
- Multi-Get
 - Avoids network latency (HTTP Api)
- Api with administrative & monitoring interface
 - Cluster's availability state
 - Health
 - Nodes' memory footprint
- Alternatives voor HTTP Api?
 - Java library
 - PHP wrappers (Sherlock, Elastica, ...)
 - But simplicity of HTTP Api is brilliant to work with, latency is hardly an issue.

Still with me?



Some Examples



Engagor | Jurriaan Persyn

app.engagor.com/inbox/2870

engagor

Inbox 143 Insights

Jurriaan Persyn

MENTIONS

- Inbox 143
- Outbox 70
- Followup
- Activity

PROJECTS & TOPICS

- Jurriaan Persyn 143
- No Worries

FILTERS

- Profiles
- Sentiment ?
- Source
- Type

Jurriaan Persyn Tools

Johan Ronsse @wolfr_ replied to a post 2h
@oemebamo OK, point taken.

Jurriaan Persyn

Jurriaan Persyn @oemebamo replied to a post 2h
@wolfr_ @showpad blijft in Gent; ze openen enkel sales office in NY, net als @engagor in SFO. #nodrama

Jurriaan Persyn Assign

Server Density @serverdensity replied to a post 3h
@oemebamo Yeh for long metric names. We'll have a fix for it in our UI revamp which is happening next month.

Jurriaan Persyn

PHPBenelux @phpbenelux 3h
one more day to register for our PHPBenelux meeting tomorrow with Elastic Search and Vagrant: ow.ly/jnwMo
Retweeted by Jurriaan Persyn @oemebamo

Jurriaan Persyn

Pieterjan Bouten @pieterjan replied to a post 5h
@Hanssim @oemebamo @Folke @tjjs Thx. Now the hard work really starts!

Jurriaan Persyn

app.engagor.com/profiles/2870/mention/10555/2810246167479189688

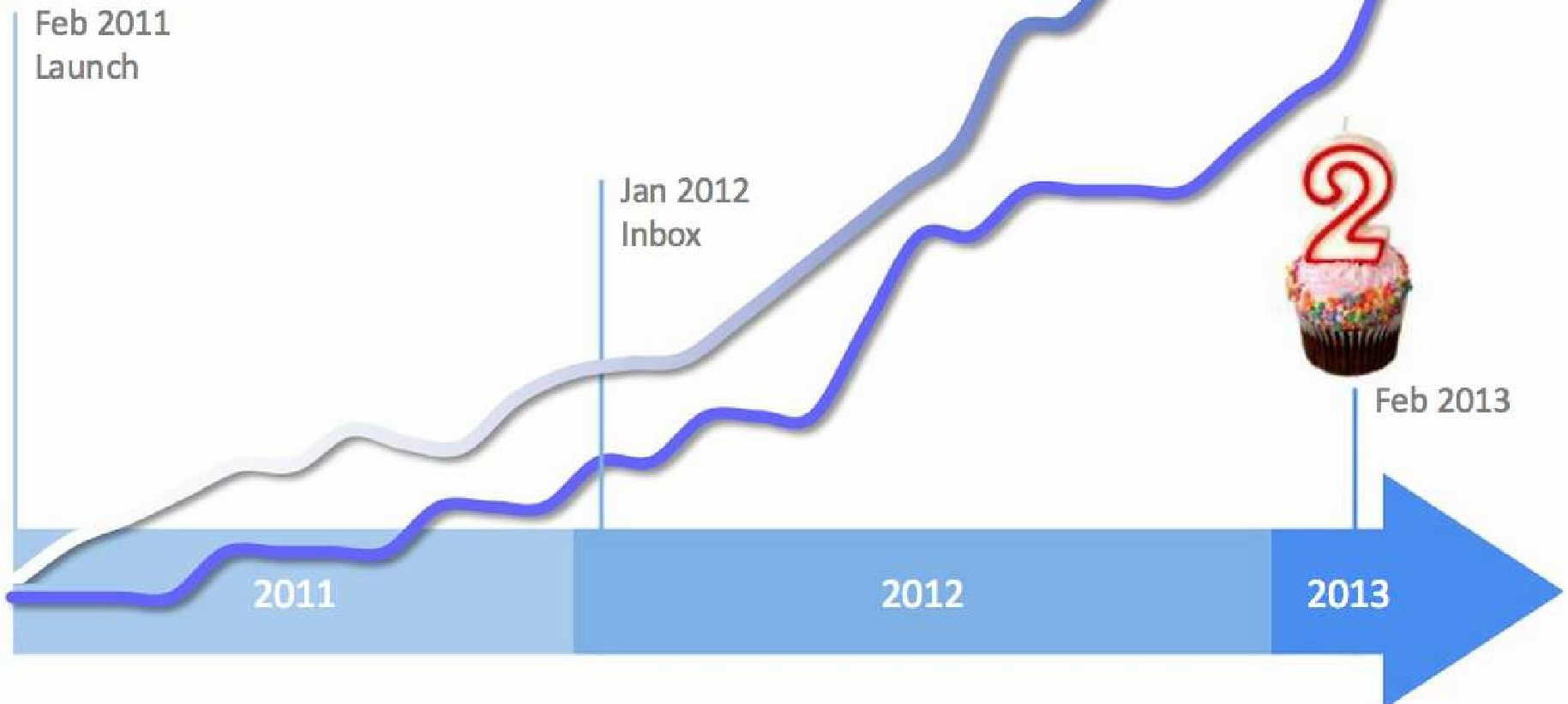
Live Help: Offline +

How we got here

>1.000.000.000 conversations
delivered in Engagor

200+ customers

16 people



```
elastic.engagor.com/shard2005-1/mention/_search?q=calldimi
View source

{
  took: 82,
  timed_out: false,
  - _shards: {
    total: 1,
    successful: 1,
    failed: 0
  },
  - hits: {
    total: 260,
    max_score: 1.2489464,
    - hits: [
      - {
        _index: "shard2005-1",
        _type: "mention",
        _id: "2012-6-138",
        _score: 1.2489464,
        - _source: {
          query: "204",
          ymid: 2161727821137838300,
          dateadd: 1339058215000,
          provider: "SearchProviderTwitter",
          - entities: {
            - urls: [
              "http://twitter.com/calldimi/statuses/210651538039914496"
            ],
            - persons: [
              "engagor"
            ]
          },
          has_photo: false,
          has_video: false,
          - tags: [
            "sentiment.positive"
          ],
          domain: "twitter.com",
          url: "http://twitter.com/calldimi/statuses/210651538039914496",
          title: "Engagor supports 'The Social Traveler' - via @engagor",
          content: "Engagor supports 'The Social Traveler' - via @engagor",
          category: "social",
          language: "en",
          - author: {
            name: "calldimi",
            file_pic_looking_forward_normal.jpg",

```

Engagor | Engagor Demo | x
app.engagor.com/inbox/857

engagor | **Inbox** 22138 | **Insights** | Engagor Demo Data

Search: `sent:negative AND author.followers:[1000 TO *]`

or

- Language **is** **Dutch**
- Country **is** **Belgium** **Unknown**
- Sentiment **is** **Negative**
- Followers **is** 1000 to *

Belgacom | Mobile Vikings

Mobile Vikings BE @mobilevikingsbe replied to a post
@Dxmn_Rhamad What's wrong? Can I help you with something? *BS

Sander V.d.M. @sandervdm replied to a post
@AxelWeydts Zou spijtig zijn maar met zo'n aanhoudende problemen begint een mens te denken... @mobilevikings

Sander V.d.M. @sandervdm
Zeg @mobilevikings, al dagen problemen met 3G in verschillende steden. Herstart of tijdelijk switchen op 2G geen avance. .
#oplossinggraag

3G

Live Help: Offline +

MENTIONS

- Inbox 134
- Outbox 79
- Followup 4
- Activity

PROJECTS & TOPICS

- Belgacom 53
- Mobile Vikings 81
- More

FILTERS

- Profiles
- Sentiment
- Negative 134**
- Source
- Type
- Tags
- Saved Searches

Query DSL Example:

```
(language:nl OR location.country:be OR location.country:aa)
(tag:sentiment.negative) author.followers:[1000 TO *] (-
sub_category:like) ((-status:857.assigned) (-status:857.done))
```

engagor

← overview

CHART BUILDER

- Overview
- Create Chart**

YOUR CHART TEMPLATES

- Overview
- Topic Volume Over Time

EXAMPLE CHART TEMPLATES

- Compare Topic Volume O...
- Stacked Comparison of T...
- Pie Chart of Data Sources
- Sentiment Per Week
- Compare Total Topic Volume
- Average Response Time
- Average Handle Time

Chart Table KPI's Share API Parameters

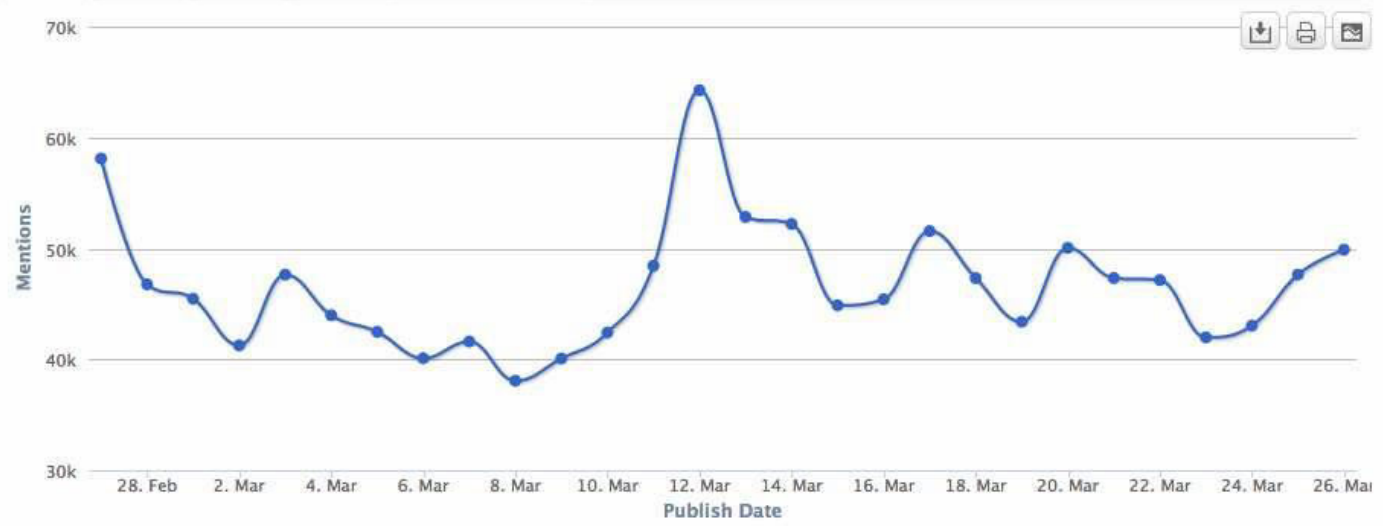


Chart Title

Source Data

- Type Mentions (Statistics about the tracked messages)
 Monitored Profile KPI's (Insights data specific for monitored profiles)

engagor Inbox 1298708 Insights Engagor Demo Data

MENTIONS

- Inbox** 1298708
- Outbox 2091
- Followup 4
- Activity

Tools

DJ KAKÁ @djkakaprod_ 9s

Nunca mas tomare coca Cola, Decepcionante :/ fb.me/2mUsScszS

Coca Cola

J Peace and love @steven_leivaa 11s

#50CosasSobreMi 49. Soy vicioso a la Coca Cola!

Coca Cola Assign

your reason to be. @harryspitty_ 11s

20:Sei imbarazzata da queste domande?OuO lol Scegli una delle due risposte. 21:The[x] CocaCola] 22:Biscotti] Torta[x] 23:UK] USA[x]

Logs (server01) 915ms - 17.81 MB 63 Memcache 13 Database 1 Uri 1 Elastic 0 Browser

INFO	URL	824ms	Util::Fetch():113
		<p>915ms - 17.81 MB</p> <p>get:61 (41ms) sget:56 (37ms) mget:5 (5ms) mget#:17 (5ms) set:2 (6ms) doubles:2 (0ms)</p> <p>connect:2 (1ms) query:11 (6ms)</p> <p>915ms - 17.81 MB</p>	<p>Util::Fetch():113</p> <p>2s</p> <p>6s</p>

```

    Fetched url in 0.733s (server02:9200/shard11535-23,shard11535-22,shard11535-1/mention/_search)
    array(6)
      "error" => false
      "status" => (int) 200
      "content" => array(5)
        "took" => (int) 731
        "timed_out" => false
        "_shards" => array(3)
          "total" => (int) 3
          "successful" => (int) 3
          "failed" => (int) 0
      "hits" => array(3)
  
```

Waiting for a0.twimg.com...

Engagor | Engagor Demo | x Engagor | Engagor Demo | x Engagor | Settings » Jurria | x

debug.app.engagor.com/inbox/857/unread/?debuginbox=1&q=thirsty

engagor **Inbox** 1298682 **Insights** Engagor Demo Data

Tools

MENTIONS

- Inbox** 1298682
- Outbox** 2091
- Followup** 4
- Activity**

Clarissa Ruiz @clar112398 13m

#coca #cola #thirsty #drinks instagram.com/p/XVTW4iS0KE/

Coca Cola

Tim Clark @freestyling80 50m

When you finally decide to bring me some lunch, could you also bring a coca-cola cause I'm just as thirsty as I am hungry. Thx.

Coca Cola

Aaron Golding @gold_k1ng 3h

Cherry Coca Cola.....ahhhh its been to too long my fizzy friend #CocaCola #Thirsty

Logs (server01) 548ms - 18.09 MB

68 Memcache get:66 (51ms) sget:60 (46ms) mget:6 (6ms) mget#:18 (6ms) set:2 (7ms) doubles:2 (0ms)

13 Database connect:2 (2ms) query:11 (6ms)

1 Elastic search:1 (362ms)

0 Browser

INFO	URL	450ms	Util::Fetch():113
		450ms	Util::Fetch():113

```

  548ms - 18.09 MB
  get:66 (51ms) sget:60 (46ms) mget:6 (6ms) mget#:18 (6ms) set:2 (7ms) doubles:2 (0ms)
  connect:2 (2ms) query:11 (6ms)
  search:1 (362ms)
  Util::Fetch():113
  Fetched url in 0.357s (server14:9200/shard11535-23,shard11535-22,shard11535-1/mention/_search)
  array(6)
    "error" => false
    "status" => (int) 200
    "content" => array(5)
      "took" => (int) 353
      "timed_out" => false
      "_shards" => array(3)
        "total" => (int) 3
        "successful" => (int) 3
        "failed" => (int) 0
    "hits" => array(3)
  
```

FACETS

- Instead of returning the matching documents ...
- ... return data about the distribution of values in the set of matching documents
 - Or a subset of the matching documents
- Possibilities:
 - Totals per unique value
 - Averages of values
 - Distributions of values
 - ...

TERMINOLOGY (CONT'D)

MySQL	Elastic Search
SELECT field, COUNT(*) FROM table GROUP BY field	Facet

engagor

← project

TOPIC » PULP FICTION

- Overview
- Trends
- People
- Content
- Sources
- Events
- Sentiment**
- Timeline
- Explore

MONITORED PROFILES

- Overview
- Twitter Accounts
- Facebook Pages
- Foursquare Venues
- Google+ Profiles
- Instagram Profiles
- LinkedIn Groups
- Pinterest Profiles
- Youtube Profiles

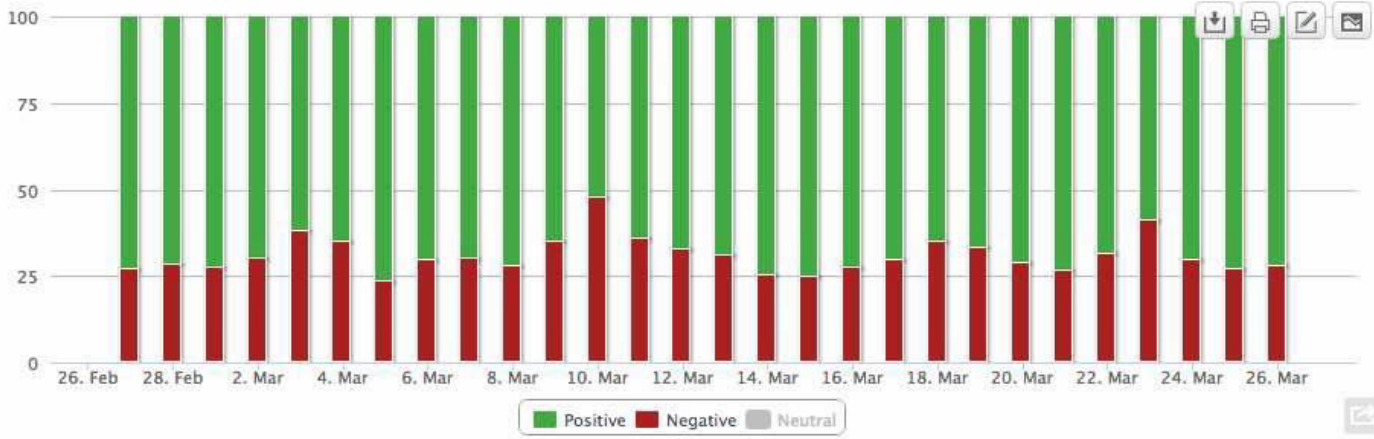
Sentiment ?

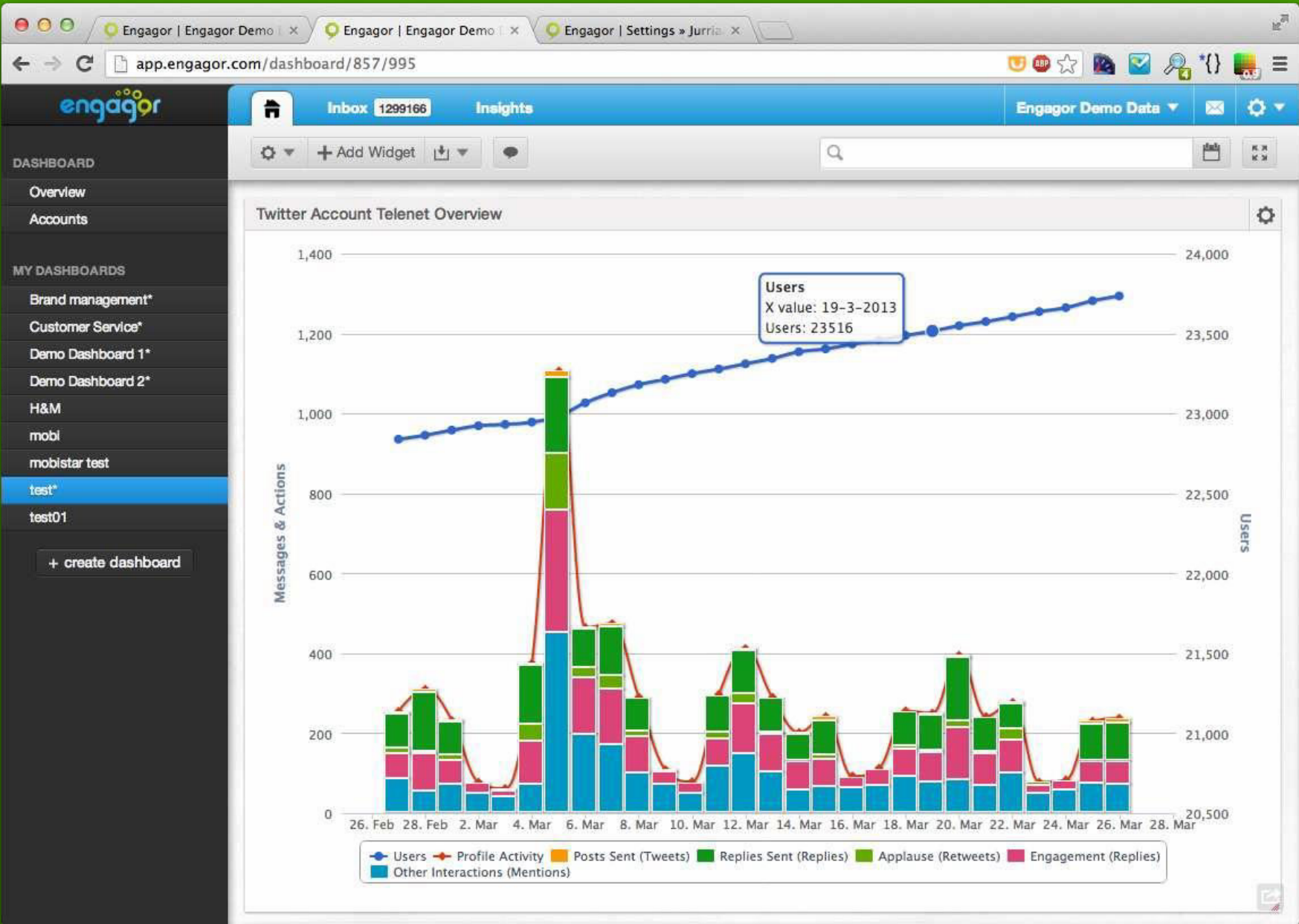


- 😊 Positive Mentions **21,226**
- 😐 Neutral Mentions **41,131**
- 😞 Negative Mentions **9,998**

Sentiment unknown for 21,348 mentions

Sentiment Per Day ?





ADVANCED FEATURES

- **Nested documents** (Child-Parent)
 - *Like MySQL joins?*
- **Percolation Index**
 - Store queries in Elastic
 - Send it documents
 - Get returned which queries match
- **Index Warming**
 - Register search queries that cause heavy load
 - New data added to index will be *warmed*
 - So next time query is executed: *pre cached*

WHAT ARE MY OTHER OPTIONS?

- **RDBMS**
 - MySQL, ...
- **NoSQL**
 - MongoDB, ...
- **Search Engines**
 - Solr
 - Sphinx
 - Xapian
 - *Lucene itself*
- **SaaS**
 - Amazon CloudSearch

... VS. SOLR

- +
 - Also built on Lucene
 - So similar feature set
 - Also exposes Lucene functionality, like Elastic Search, so easy to extend.
 - A part of Apache Lucene project
 - Perfect for Single Server search
- -
 - Clustering is there. But it's definitely not as simple as ElasticSearch"
 - Fragmented code base. (Lots of branches.)

Engagor used to run on Solr.

... VS. SPHINX

- +
 - Great for single server full text searches;
 - Has graceful integration with SQL database;
 - (*Eg. for indexing data*)
 - Faster than the others for simple searches;
- -
 - No out of the box clustering;
 - Not built on Lucene; lacks some advanced features;

Netlog & Twoo use Sphinx.

WANT TO USE IT?

- In an existing project:
 - As an extra layer next to your data ...
 - Send to both your database & elasticsearch;
 - Consistency problems?;
 - Or as replacement for database
 - Elastic is as persistent as MySQL;
 - If you don't need RDBMS features;
 - *@Engagor: Our social messages are only in Elastic*

“Users are incredibly bad at finding and researching things on the web.”

Nielsen *(March 2013)*

<http://www.nngroup.com/articles/search-navigation/>

“Pathetic and useless are words that come to mind after this year’s user testing.”

Nielsen (*March 2013*)

<http://www.nngroup.com/articles/search-navigation/>

“I’m searching for apples and pears.”

“apples **AND** pears”



“apples **OR** pears”

“It’s too young. Is it even **stable** enough?”

Your boss (*Tomorrow Morning*)







Aug 09th

foursquare now uses Elastic Search (and on a related note: Slashem also works with Elastic Search)!

Historically venue search used to be done in Mongo at foursquare; we later migrated it to [Solr](#) and now we are using [Elastic Search](#). While Solr has been great (and we continue to use it for users and to-dos) we want to make sure that we can keep up with the growth of our venues database as our wonderful users keep adding more places.

Elastic Search provides a really nice way to shard search data and manage the related problems around this. By sharding search data we are able to split it up the documents into groups which can easily fit on a single machine and then get the results back from a collection of machines. Elastic Search and Solr have quite a few similarities that come from both being built on top of [lucene](#). The Elastic Search team has been super helpful during the migration process, providing ideas on how to tweak our queries to get better performance out of Elastic Search. We got a non-trivial performance improvement by switching the field types for our geohash to allow us to skip the analyzer. One of the biggest improvements for performance we found was parameterizing our ranking scripts to avoid the per-search compilation over-head. We also saw an improvement when [using a custom scoring plugin](#)

Recent Posts

- Native app integration like never before: The Foursquare for BlackBerry 10 SDK
- The Opportunity Gap: learn about the public schools in your neighborhood with @ProPublica and Foursquare
- A Foursquare Hackathon for the new year – sign up and build the next amazing hack!
- MongoDB at Foursquare: From the the cloud to bare metal
- Even more ways to personalize Foursquare – check out the latest additions to our app gallery

Recent Comments

- Hladvd on Native app integration like never before: The Foursquare for BlackBerry 10 SDK
- Domain Registration India on A



LEARN FROM THE EXPERTS

Register now for early bird discounts.



[View Training Dates](#)

Sign up today for updates on training events in your area, discounts on support, and consulting services.

 [>](#)

Welcome to **Elasticsearch**, your search stops here!

Elasticsearch is the most advanced search engine to date. But don't take our word for it, try it yourself:

[Download](#)



elasticsearch.org

irc.freenode.net

[#elasticsearch](http://irc.freenode.net/#elasticsearch)

elasticsearch

node.js

socket.io

real time notifications

rabbitmq

backbone.js

gearman

jobs@engagor.com

START-UP SHIZZLE

JUST ANOTHER DAY AT A SOCIAL MEDIA MANAGEMENT TOOL
START-UP.

ARCHIVE RANDOM RSS SEARCH

22ND MAR 2013



```
$ cd /usr/local/share/elasticsearch
```

```
$ sudo bin/service/elasticsearch stop
```

**THAT'S IT,
THANKS**

HOPE YOU'RE INSPIRED NOW



Sources include:

- <http://www.elasticsearch.org/videos/2010/02/07/es-introduction.html>
- <http://www.elasticsearchtutorial.com/>
- <http://www.slideshare.net/clintongormley/cool-bonsai-cool-an-introduction-to-elasticsearch>
- <http://www.slideshare.net/medcl/elastic-search-quick-intro>
- <http://www.slideshare.net/macrochen/elastic-search-apache-solr-10881377>
- http://www.slideshare.net/cyber_jso/elastic-search-introduction
- <http://www.slideshare.net/infochimps/elasticsearch-v4>
- <http://engineering.foursquare.com/2012/08/09/foursquare-now-uses-elastic-search-and-on-a-related-note-slashem-also-works-with-elastic-search/>
- <http://stackoverflow.com/questions/10213009/solr-vs-elasticsearch>
- <http://stackoverflow.com/questions/11115523/how-does-amazon-cloudsearch-compares-to-elasticsearch-solr-or-sphinx-in-terms-o>
- <http://blog.socialcast.com/realtime-search-solr-vs-elasticsearch/>