



# Realtime Boost

felipeg@google.com

<http://go/realtime-boost>

<http://go/realtime-boost-demo>

<http://go/realtime-boost-howto>



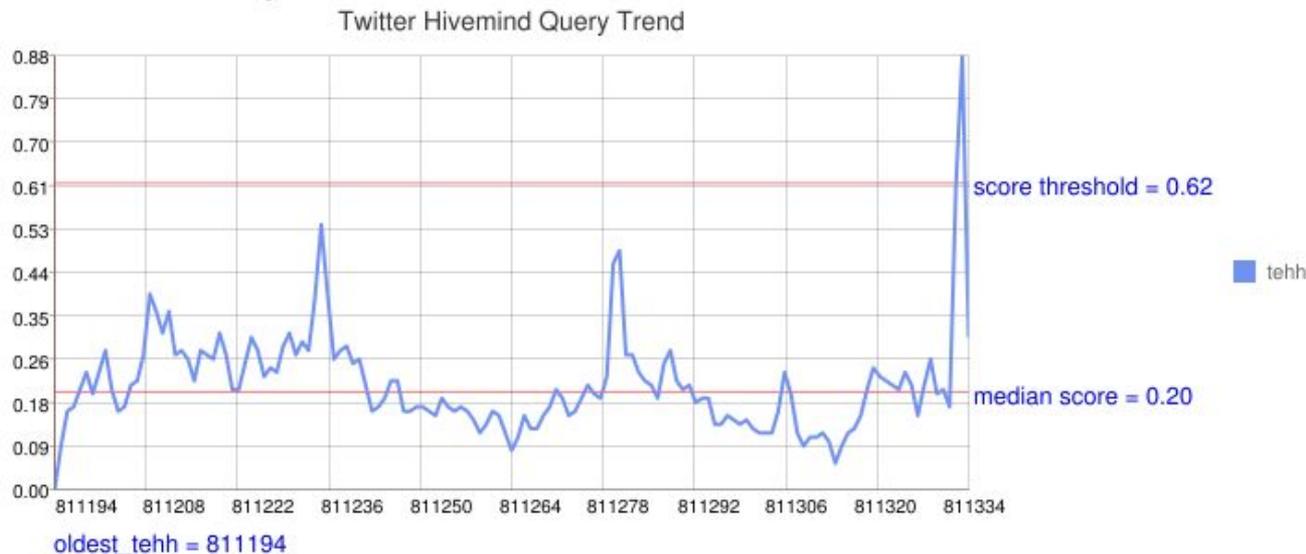
# Goals

- Detect real world events
- Is this Query Trending ?
- Fast triggering: < 5 mins after the event
- Fast serving: 5 ms average / 40ms 99.9 percentile
  - Run in QRewrite - Critical path on search

# Realtime Carousel with Twitter Data

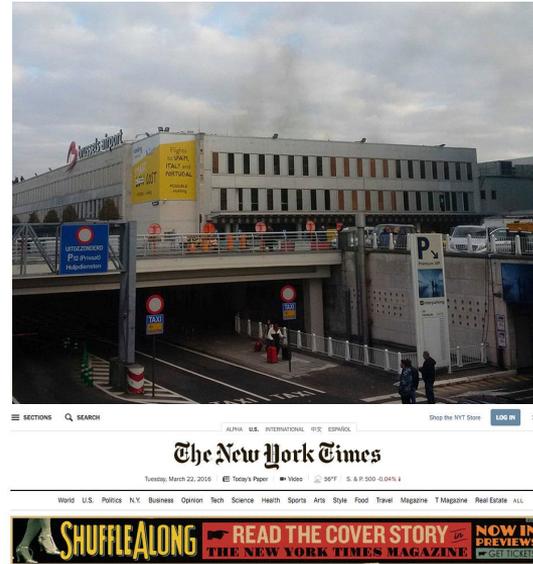
- Show tweets only when relevant
- Twitter is a great source of news
- Trigger on spikes on number of tweets [Breaking News]

Twitter Hivemind Query Trend:



# How NYT covers breaking news [\[link\]](#)

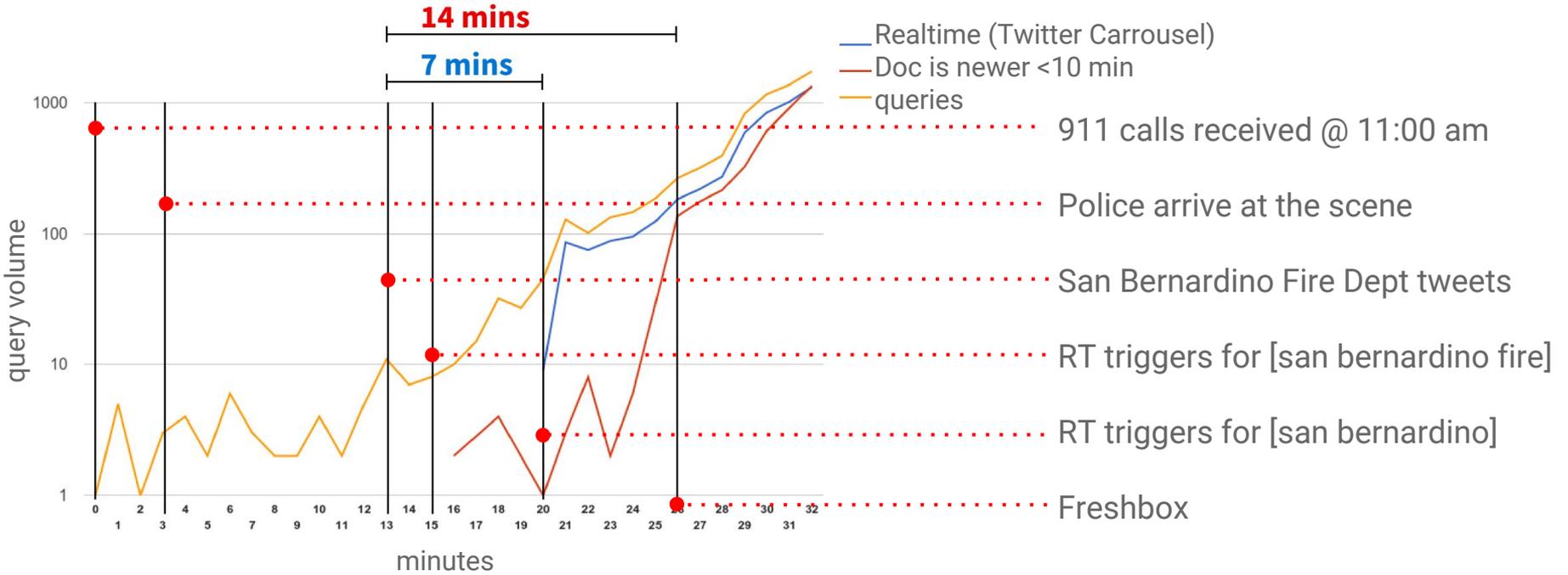
- 3:18 a.m. E.D.T.  
Editor in London Sees Tweet  
(smoke at Brussels Airport)



- 3:33 a.m. E.D.T.  
Article Posted on NYTimes.com



# [san bernardino]



# REALTIME ALL THE THINGS ...



# REALTIME ALL THE THINGS ...



# Realtime Boost Signal ([go/realtime-boost](https://go/realtime-boost))

Proprietary + Confidential

## Spikes and Correlations on **Content Creation**

- Location (S2), Entities, Salient Terms, NGrams ...

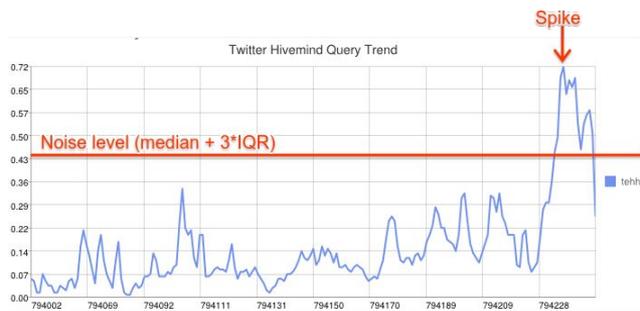
## Sources

- Freshdocs-instant
- Chrome Visits (soon)
- Instant Navboost (soon)

Not restricted by Twitter contract

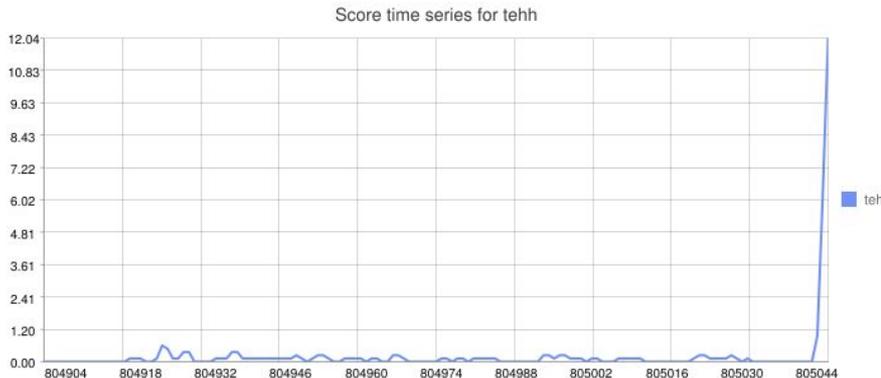
## Run in Query Rewriter

- Can be used anywhere: Freshbox, Stream ...



## Query [Dilma]

Spike 5 mins after impeachment process announced



# Realtime Boost - Indexing

## Freshdocs-instant Crawl

Most docs are crawled in under 60s

[/quality/realtime/hivemind/indexing/realtime\\_freshdocs/firstseen\\_age](#) (cumulative ☺): [Distribution =](#)

Range	Bucket Size	Percent	Cumulative Percent	
[0...1)	1537	0.003%	0.003%	
[1...2)	25477	0.047%	0.050%	
[2...4)	647500	1.204%	1.254%	
[4...8)	1984768	3.689%	4.943%	
[8...16)	3247774	6.037%	10.980%	
[16...32)	9967434	18.528%	29.508%	
[32...64)	7631145	14.185%	43.694%	
[64...128)	2071684	3.851%	47.545%	
[128...256)	1283314	2.386%	49.930%	
[256...512)	1124126	2.090%	52.020%	
[512...1024)	1646429	3.060%	55.080%	
[1024...2048)	1901266	3.534%	58.615%	
[2048...4096)	2295745	4.267%	62.882%	
[4096...8192)	2103170	3.910%	66.791%	
[8192...16384)	2169897	4.034%	70.825%	
[16384...32768)	3317442	6.167%	76.992%	
[32768...65536)	3767721	7.004%	83.995%	
[65536...131072)	4714307	8.763%	92.759%	
[131072...262144)	3895550	7.241%	100.000%	

First crawl

Recrawl

# Realtime Boost - Indexing

## Freshdocs-instant Crawl

**Dominicans of Haitian descent tired of bias at the ballot box**

By **Mariano Castillo**, CNN  
Updated 6:01 AM ET, Sat April 9, 2016



Haitian workers spread fertilizer on a rice field in Esperanza, Dominican Republic, in July 2015. Their children were born in the Dominican Republic and are citizens, though civil rights groups say the government discriminates against both parents and their Dominican-born children.

**Story highlights** (CNN) — A month away from presidential elections, one segment of Dominican Republic society fears it will be denied the right to vote.

OAS commission hears grievances about Dominican vote process

Country denies it is rejecting voting by those with Haitian descent

thousands of Dominicans.

A crisis over the sudden statelessness of so many led the Dominican government to provide a path for the affected to have their Dominican citizenship restored.

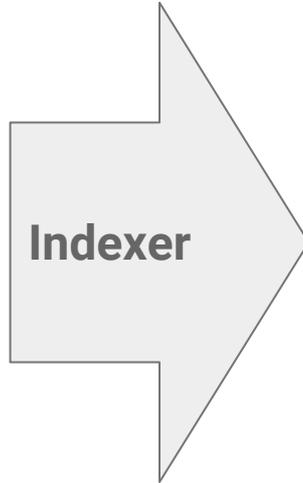
Civil rights groups representing Dominicans of Haitian descent testified before the Inter-American Commission on Human Rights in Washington on Friday, arguing that this segment of voters will be disenfranchised.

Despite government claims that everyone who qualified to have their citizenship restored did so, and that no Dominicans will be denied the right to vote, the witnesses told the commission that the reality on the ground is very different.

### Aftermath of a crisis

Dominicans of Haitian descent have long complained that others treat them as second-class citizens.

## Pubsub CDOC



Time  
Unigram  
Location  
Entity  
Salient Terms

## Hivemind - Tokenized - Muppet Index

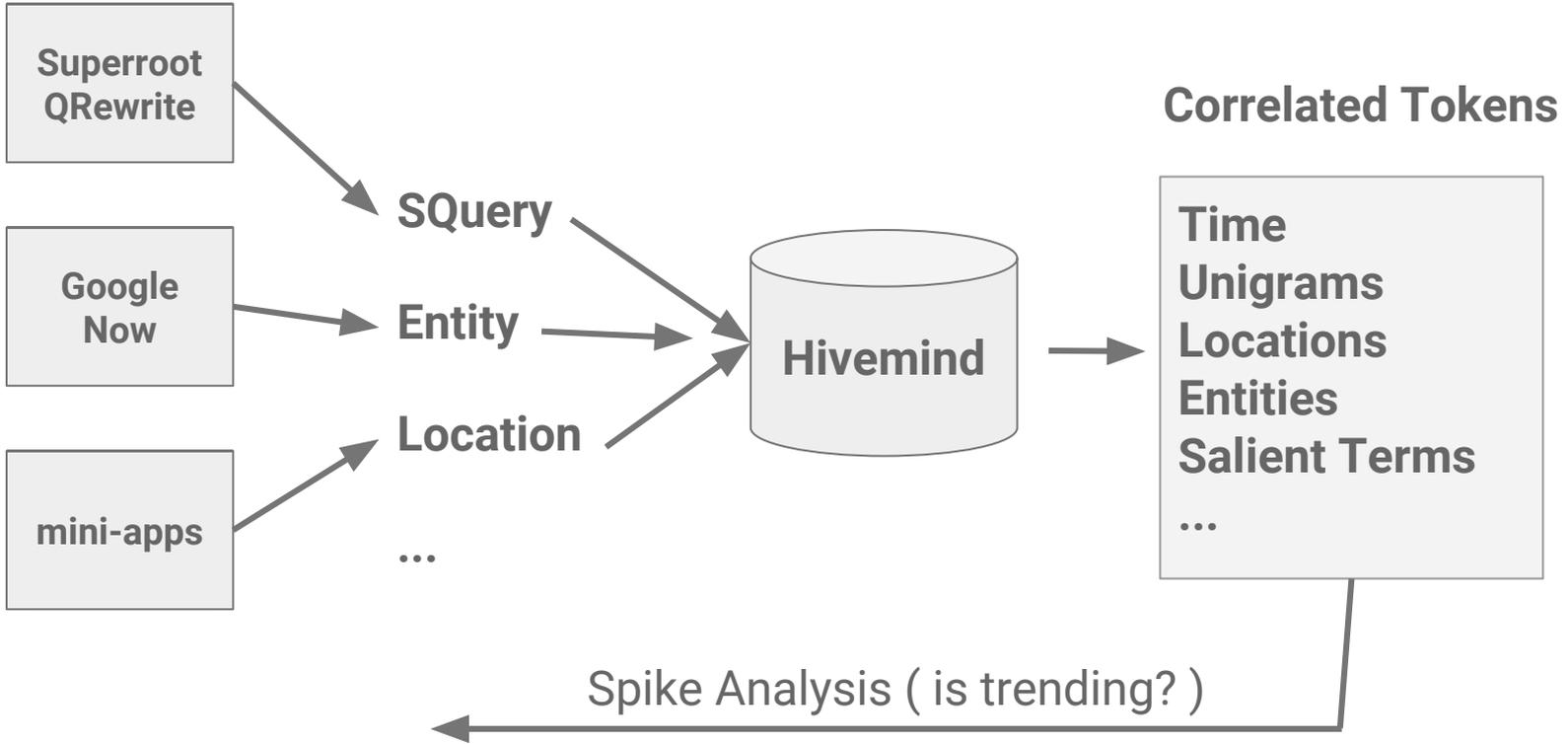
Global ExtendedDocId: 10529037701726682924::0::0

LocalDocId: 1029292, Global ExtendedDocId: 10529037701726682924

Section body (Pos 131749376 -> 131749492)

TokenPos	TokenId	token
131749376	2983	[ <global_delimiter>]
131749378	6362733	[ rthh=811207]
131749379	4023159	[ rthh=811208]
131749380	3138	[ rtl=eng]
131749381	77581	[ rtw=voting]
131749382	4097	[ rtw=cnn]
131749383	4025558	[ rtw=dominicans]
131749384	140670	[ rtw=haitian]
131749385	100855	[ rtw=descent]
131749386	108206	[ rtw=tired]
131749387	240848	[ rtw=bias]
131749388	86357	[ rtw=mariano]
131749389	86358	[ rtw=castillo]
131749390	69870	[ rtw=img]
131749391	4424	[ rtw=updated]
131749395	6157	[ rtw=edt]
131749398	9657	[ rtw=month]
131749399	4140	[ rtw=away]
131749400	8144	[ rtw=presidential]

# Realtime Boost - Serving





# RealtimeBoost overview

## Clients

Google Now

Superroot QRewrite

Mini-Apps

Entity

SQuery

Location



Spike Analysis & Correlations

RealtimeBoost Library/Node

Serving

Indexing

Web

Crawl

Fresdocs-instant  
(crawl new page in 60s avg)

Pubsub CDoc

RTBoost indexer

MDUs

Union

Hivemind  
(muppet leaves)

Hivemind  
(muppet root)

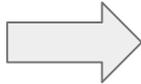
Tokenize

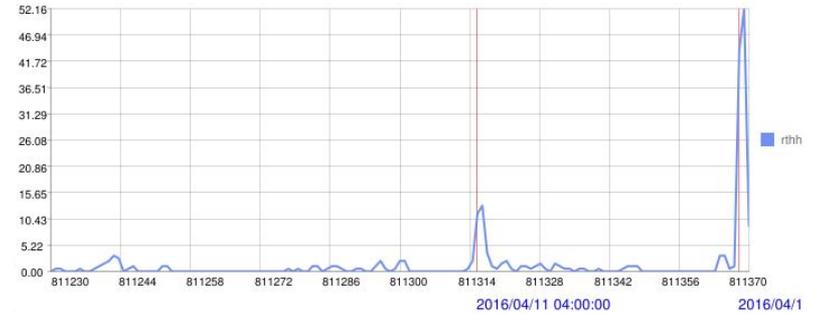
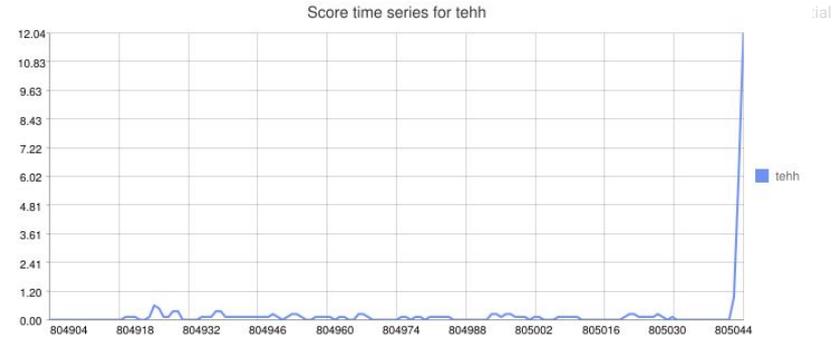
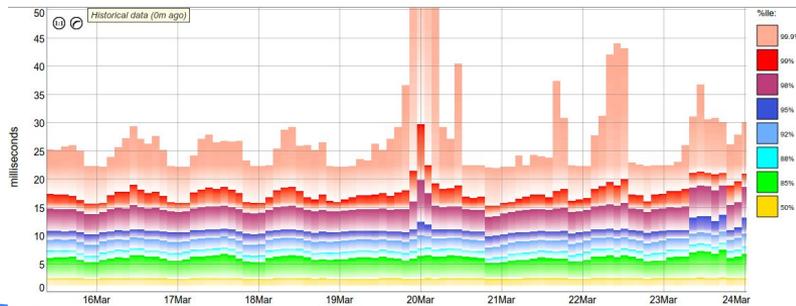
Time  
Unigram  
Location  
Entity  
Salient Terms  
...

Tokenized  
muppet  
doc

```
Global ExtendedDocId: 10529037701726682924::0:0
LocalDocId: 1029292, Global ExtendedDocId: 10529037701726682924::0:0
Section body (Pos 131749376 -> 131749492)
TokenPos  TokenId  [
131749376  2983  [ <global_delimi
131749378  6362733 [ rthh=81
131749379  4023159 [ rthh=81
131749380  3138  [ rtl
131749381  77581  [ rtw=vo
131749382  4097  [ rtw
131749383  4025558 [ rtw=domini
131749384  140670 [ rtw=hai
131749385  100855 [ rtw=des
131749386  108206 [ rtw=t
131749387  240848 [ rtw=
131749388  86357  [ rtw=ma
```

# Realtime Boost - Latencies

- Freshdocs crawl
    - < 60 seconds for most docs
  - Indexing < 1s
  - Triggering latency: It depends 
  - Serving: 5ms
- Hivemind latency



# Realtime Hivemind Scoring ([go/hivemind-scoring](https://go/hivemind-scoring))

$$\text{LiftScore} = \frac{P(\text{event} | \text{query})}{P(\text{event})}$$

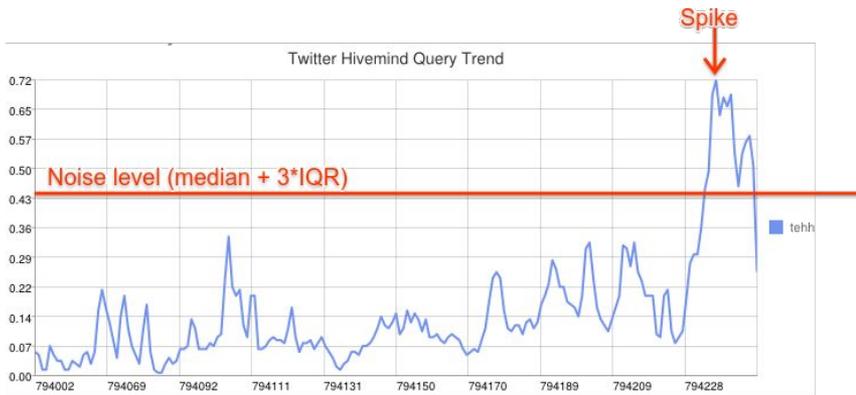
Event is the time of the document publication, quantized to 30 minutes e.g. time=794237.

LiftScore can be interpreted as how much more likely to see document matching the query at the given time, comparing to seeing random document at the given time.

$$\text{LiftScore} = \frac{P(\text{time}=794237 | [\text{donald trump}])}{P(\text{time}=794237)}$$

$$P(\text{time} = 794237) = \frac{N(\text{docs at time}=794237)}{N(\text{docs total})}$$

$$P(\text{time} = 794237 | [\text{donald trump}]) = \frac{N(\text{docs at time}=794237 \text{ matching } [\text{donald trump}])}{N(\text{docs matching } [\text{donald trump}])}$$

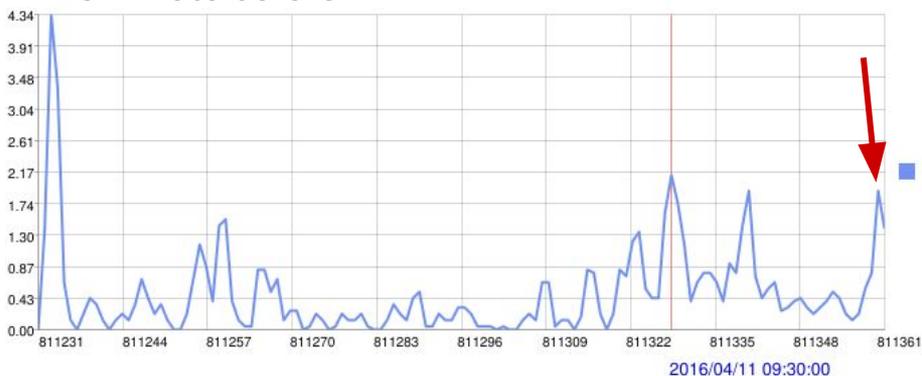


# Realtime Boost - Hivemind time token

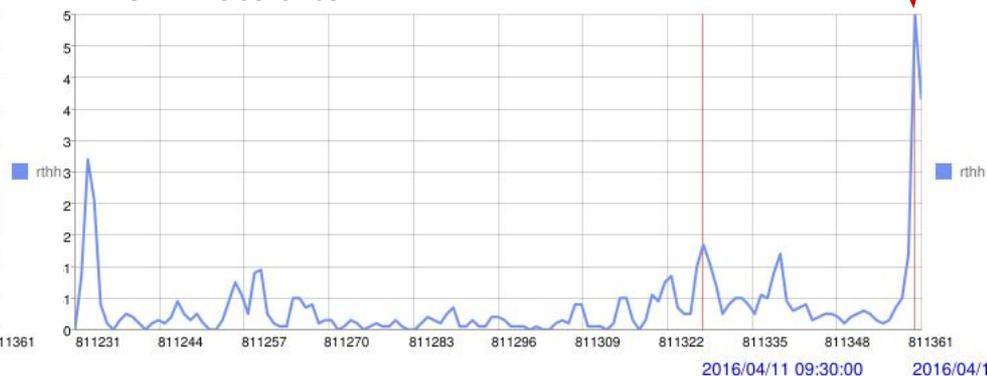
Proprietary + Confidential

- Incomplete bucket (current half-hour)
  - Normalized
  - $\text{Score} = P(Q,T) / P(T)$
  - Numerator = number of query-matched documents in the half-hour bucket
  - Denominator = total number of documents in the half-hour bucket

The minute before



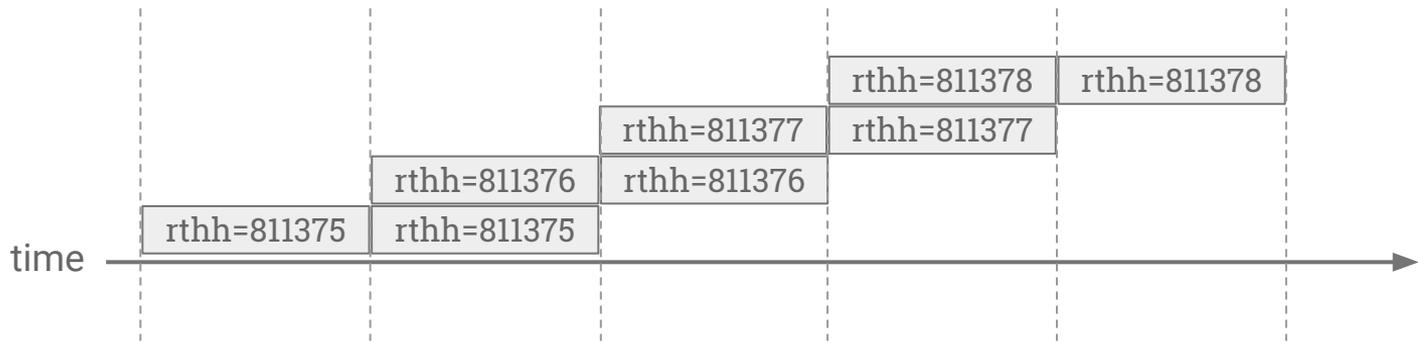
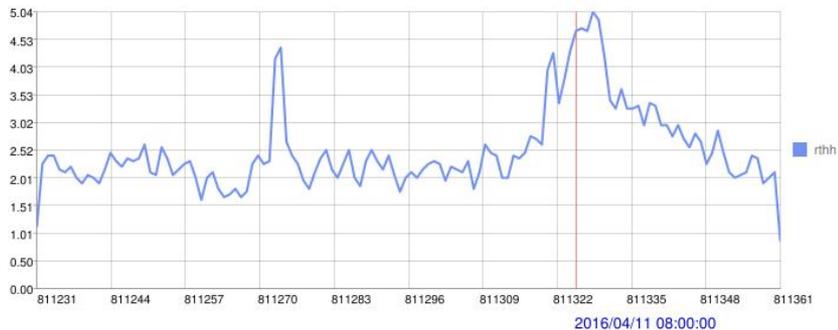
The minute after



# Realtime Boost - Hivemind time token

- Half Hour Buckets

- Overlapping 1 hour
- Smoothing (sliding window)
- Add new tokens to previous half-hour bucket
- Helps the incomplete bucket (current half-hour)



# Realtime Boost ([go/realtime-boost](https://www.google.com/realtime-boost))

## Query [Bruce Springsteen]

- Announced the tour for the album **The River**
- Map shows next cities where the tour will be

### Realtime KG entities (rtkg)

Matched sessions: 4260 of 147530340

Search:

Show **25** entries

[First] [Previous] 1 2 3 4 5 [Next] [Last]

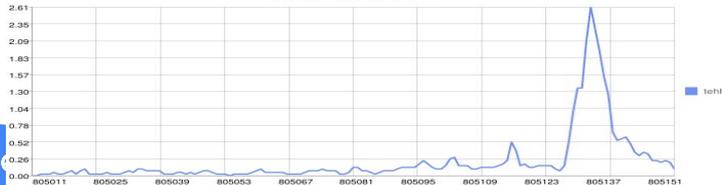
#	score	matched	count	bits	Go	HM	Label
1	1277.1569	3191	3489	32	<a href="#">Q</a>		/m/0gcs9
2	733.3225	790	992	32	<a href="#">Q</a>		/m/09j_c4
3	346.4903	218	423	31	<a href="#">Q</a>		/m/01hmkd9
4	315.8387	299	778	32	<a href="#">Q</a>		/m/0c83_
5	275.6321	130	266	27	<a href="#">Q</a>		/m/07y8z
6	202.1766	133	429	31	<a href="#">Q</a>		/m/0gpvt
7	189.8295	273	1404	32	<a href="#">Q</a>		/m/0f6c3
8	169.7850	1053	12251	32	<a href="#">Q</a>		/m/02py3
9	106.4968	35	127	19	<a href="#">Q</a>		/m/0lhr6r
10	100.9163	32	118	18	<a href="#">Q</a>		/m/04051
11	73.0128	30	168	19	<a href="#">Q</a>		/m/0lhr7v
12	67.5626	26	53	8	<a href="#">Q</a>		/m/0bzj8
13	54.9956	21	64	12	<a href="#">Q</a>		/m/0w9y5
14	49.9862	20	57	14	<a href="#">Q</a>		/m/04zz
15	49.0301	20	53	2	<a href="#">Q</a>		/m/05f7wv

NO IMAGE

**The River**  
/en/the\_river

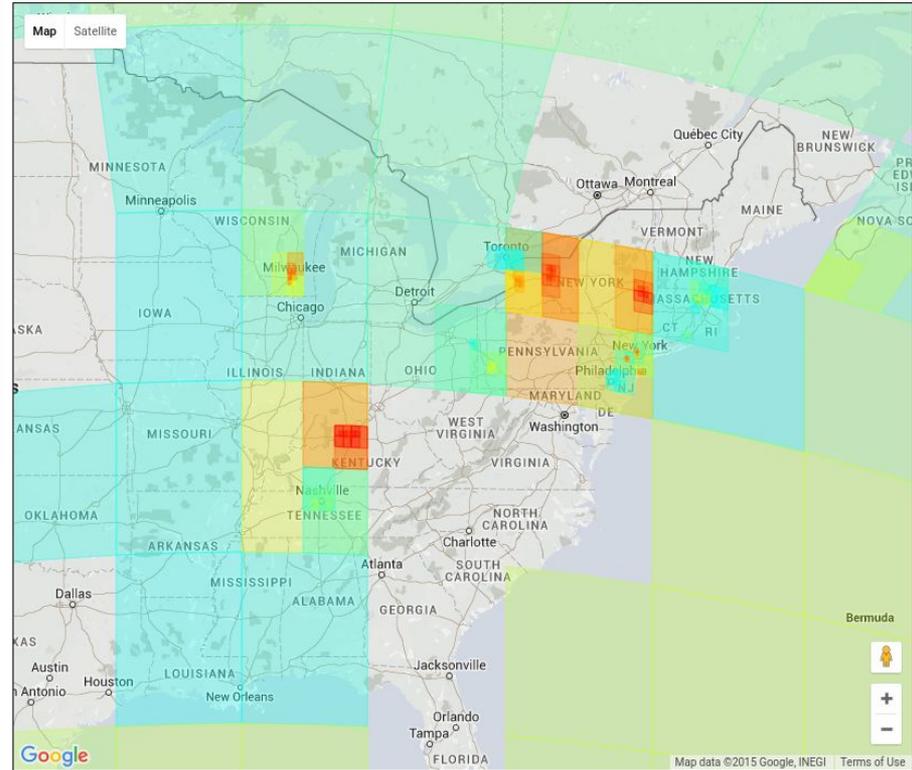
The River is the fifth studio album, and the first double album by Bruce Springsteen. The album was released on October 17, 1980. Rolling Stone ranked it at number 253 on their list of the greatest albums of all time.

Score time series for teh



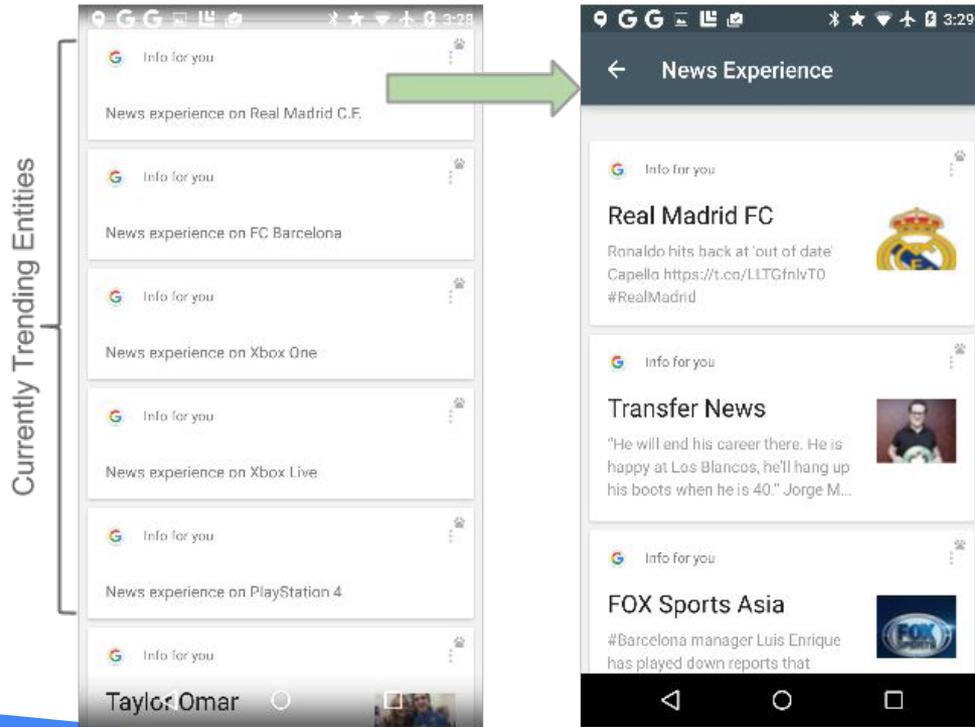
## Map

Location (s2) Heatmap



# Google Now News Experiment

## Notification from Spike on User-Interests (entities)



CTR from 10% to 22%  
(on fishfood)

Rasta

Studies Metrics Configs Sizing

Table & Graphs  
How to show graphs?

Executive Summary

42

Time range: 2016-03-07 00:00 ~ 2016-04-30

Queried time ranges = (2016/03/07 00:00 ~ 2016/04/03 23:59) Confidence Interval Level: 95%

Experiment	Users	BooleanViewR (%)	ViewR(%)	BooleanCTR(%)	CTR(%)	DismissR
All Traffic 42	6,691	17.6523	23.1093	8.0465	10.6481	5.1525
TOTAL * SCARY						
All Traffic 42	6,691	23.7720	34.3837	19.8830	22.4172	8.1871
NEWS_EXPERIENCE_SECONDARY_PAGE						

# Query Suggest

[p] -> [Prince]

A screenshot of a Google search interface. The search bar contains the letter 'p'. Below the search bar, a list of suggestions is shown: 'python2 fetch url content', 'python urllib importerror: no module named request', 'prince', and 'prince dead'. The 'prince' and 'prince dead' suggestions are highlighted with a red rectangular box. To the right of each suggestion is a 'Remove' link. Below the suggestions, it says 'Press Enter to search.'

Google suggest

Raw ASCII [Raw XML](#)

ds:  lang: en country: us client: psy-ab partner\_id:  expid:   
 Show maps options  Show debug options  Set client options params  Set reference client options params  Set Suggest-Generic request options params

[p]

Suggestion	Type	info	frequency	subtype_string												
<input type="text" value="prince"/>	0	<div><p><i>message_set_extension</i></p><table><tr><td>fresh_clicks_ratio</td><td>0.84232813</td></tr><tr><td><b>is_spiking</b></td><td>true</td></tr><tr><td>debug_string</td><td>prince</td></tr></table><p><i>spike_result</i></p><table><tr><td>score</td><td>10</td></tr><tr><td>confidence</td><td>0.93055558</td></tr></table><p><i>time_range</i></p><table><tr><td>start_secs</td><td>1461256200</td></tr></table></div>	fresh_clicks_ratio	0.84232813	<b>is_spiking</b>	true	debug_string	prince	score	10	confidence	0.93055558	start_secs	1461256200	363,394,080	FRESH_NEWS, FRESH
fresh_clicks_ratio	0.84232813															
<b>is_spiking</b>	true															
debug_string	prince															
score	10															
confidence	0.93055558															
start_secs	1461256200															
prince dead	0		57,862,620	FRESH												
prince died	0		29,106,900	FRESH												

# Query Suggest

[q] -> [Queen Elizabeth]

(it was her anniversary)

Google suggest

Raw ASCII [Raw XML](#) [Update URL](#)

ds:  lang: en country: us client: psy-ab partner\_id:  expid:  location\_id [met/reg]:  0 debug:  show info:

Show maps options  Show debug options  Set client options params  Set reference client options params  Set Suggest-Generic request options params

[q]

Suggestion	Type	frequency	subtype_string	info
quizlet	0	5,886,116	FRESH_RERANKED	

queen elizabeth

0 405,988 FRESH\_NEWS, FRESH\_RERANKED, DISAMBIGUATED\_BY\_ENTITIES

**message\_set\_extension**

fresh_clicks_ratio	0.54415619
is_spiking	true
debug_string	uk s cameron hails steadfast queen on her 90th birthday

**spike\_result**

score	4.51831
confidence	0.6875
<b>time_range</b>	
start_secs	1461193200
end_secs	1461195000

# Query Suggest

[Alp] -> [AlphaGo]

Google suggest

Raw ASCII [Raw XML](#)

ds:  lang: en country: us client: psy-ab partner\_id:  expid:  location\_id [metreq]:  debug:  show info:

Show maps options  Show debug options  Set client options params  Set reference client options params  Set Suggest-Generic request options params

[alp]

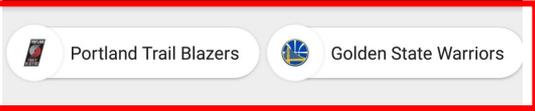
Suggestion	Type	info	frequency	subtype_string	
alphago	0	<i>message_set_extension</i>	2,048,580	FRESH_NEWS, FRESH_NEWS, FRESH	
		fresh_clicks_ratio			0.82746923
		snippet			Spiking in the news.
		source_name			google s software beats human go champion in first match
alphabet	0		361,108	FRESH_RERANKED	
alphago vs lee	0	<i>message_set_extension</i>	492,660	FRESH_NEWS, FRESH_NEWS, FRESH	
		fresh_clicks_ratio			0.96949667
alprazolam	0		168,546		

BoldSuggestions: 1  
complete\_usec: 76086  
experiment\_triggered: f  
word\_boundary\_probability: 0.010128212872179961  
num\_get\_completion\_requests: 1  
query\_boundary\_probability: 0.0032958192125247558

Debug o  
 get  
 sxs  
 trac  
 disa  
 byp  
num\_sug  
fresh\_m  
starlight  
starlight

# Newsstand - Pivots topics

Proprietary + Confidential



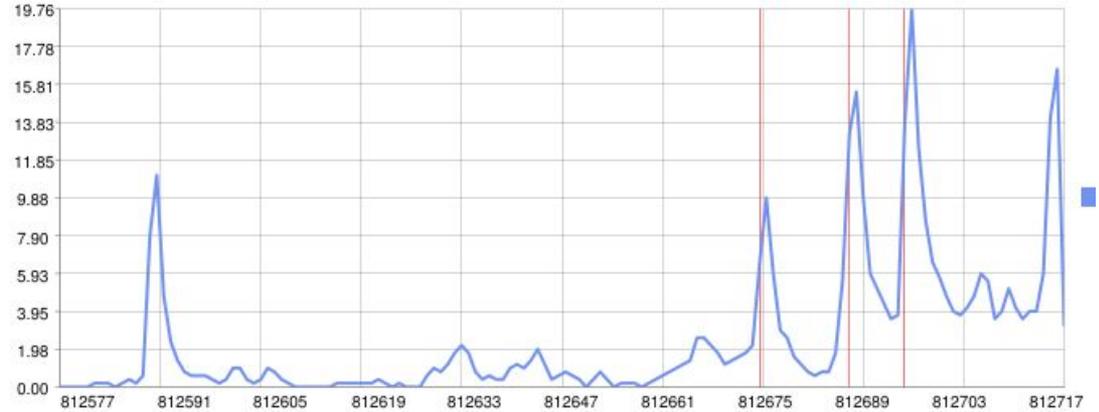
The Huffington Post NBA



ASSOCIATED PRESS

Stephen Curry Is The First Unanimous MVP In NBA History

51 minutes ago

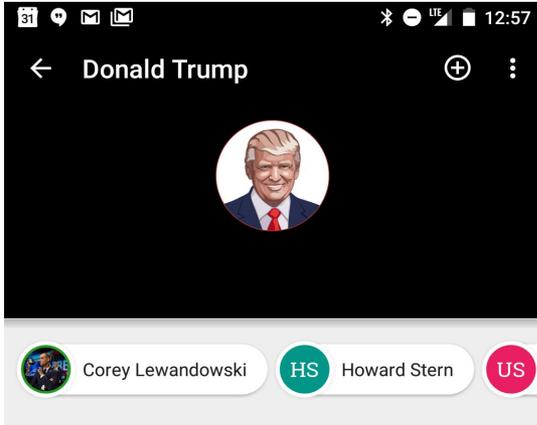


## Related Entities

Img	Name	Score	Numerator	Denominator
	<b>Stephen Curry</b> /m/0dw15r Basketball Shooting guard	482.49	1134	8578
	<b>Point</b> /m/0bd4bf	421.54	56	95
	<b>Portland Trail Blazers</b> /m/0jmm4 Professional Sports Team	346.11	743	7329
	<b>Golden State Warriors</b> /m/0jmj7 Professional Sports Team	220.25	507	7900
	<b>NBA</b> /m/05jvx Sports Association	51.07	150	10088
	<b>NBA Playoffs</b> /m/05_10r Recurring competition	23.79	28	2021

# Newsstand - Pivots topics

Proprietary + Confidential



Watch Will Ferrell & Ryan Gosling Take Cuts At Donald Trump



Panama Papers Show 200,000 Offshore Accounts  
19 hours ago

DEMO

<http://go/realtime-boost-demo>

attic

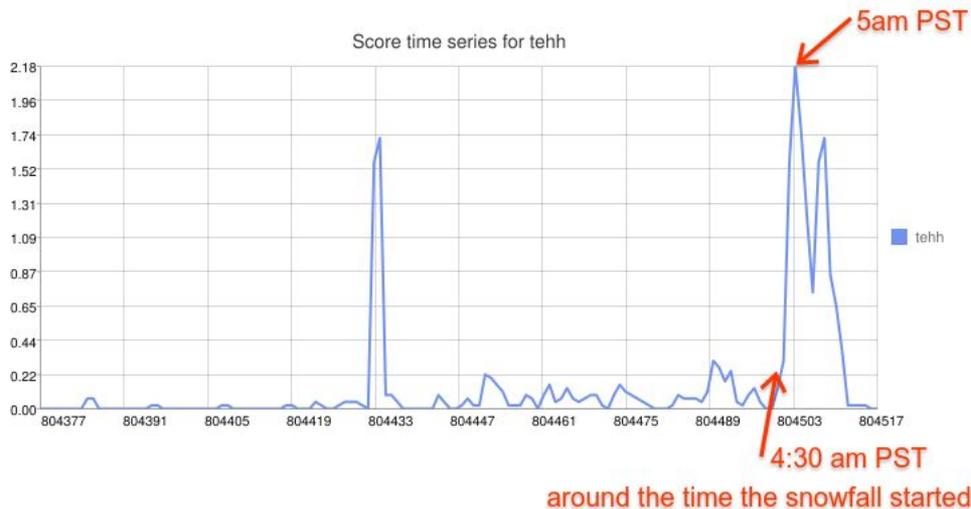
# Realtime Boost ([go/realtime-boost](https://go/realtime-boost))

Realtime Boost can be faster than Twitter

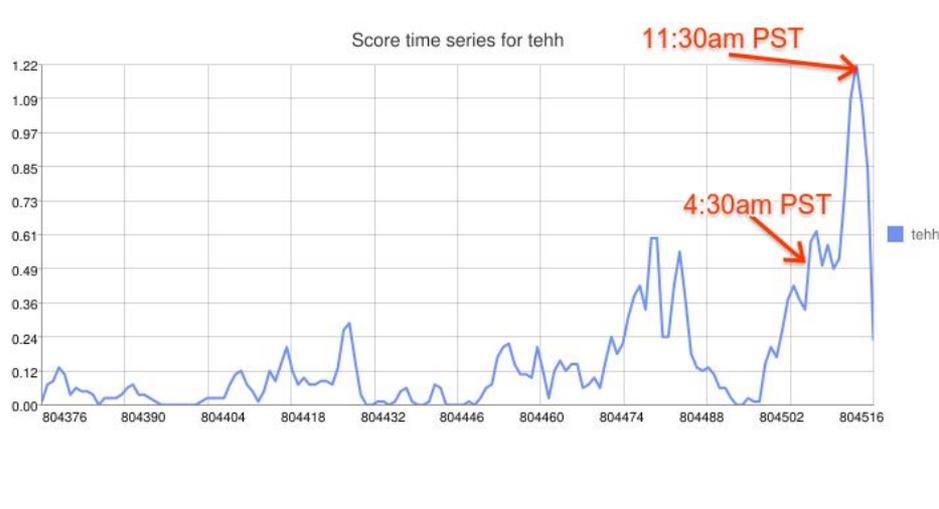
Query [Weather Chicago]

- Huge snowstorm arrived in the middle of the night

### Realtime Boost



### Twitter Hivemind



# Realtime Boost ([go/realtime-boost](https://www.google.com/realtime-boost))

We can replace Twitter Hivemind with Realtime Boost

## Query [John Lennon]

- His death anniversary
- Political figures took the opportunity to talk about gun law

Realtime Boost



Twitter Hivemind

