Google

Realtime Events

May 16 2017 - felipeg@google.com

Design Doc: <u>go/realtime-events-design</u> Demo: <u>go/realtime-events</u>

Spike Cluster

New Publication

Document

Event detection

Freshdocs

Time Structure

Group of Documents

RealtimeBoost / Hivemind **Group of Spikes**

Clusterizer-Service

Structured Metadata Checkpoints Timelines

Event

Serving to UI/Clients

Model-T

Every step in Realtime

Event detection



lst
d
_
liu
liu
liu
can
diu car
can
car



200.54

110.09

96.07

85.86

76.78

48.63

48.61

45.45

45.05

42.93

180.70

138.02

114.37

105.08

96.92

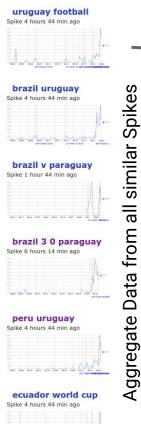
94.18

58.09

55.81

55.81

51.76



new loca see see and see on and

Cluster

Event

Event Label

Brazil 1st to gualify for world cup

Event Summary

A resurgent Brazil squad under new management became the first team to qualify for the World Cup on Tuesday.

Oueries

brazil vs paraguay live streaming

Ouestions

who scored in brazil vs paraguay



Locations (from cv and nb)



Popularity (CV and NB lift)

Google

Event detection

Cluster



Spike

Stored in Model-T

Metadata to be served to the UI



Spike

brazil v paraouav	
	٢.

Event

Event Label

Brazil 1st to qualify for world cup

Event Summary

A resurgent Brazil squad under new management became the first team to qualify for the World Cup on Tuesday.

Queries

brazil vs paraguay live streaming

Questions

who scored brazil vs paraguay



Locations (from cv and nb)



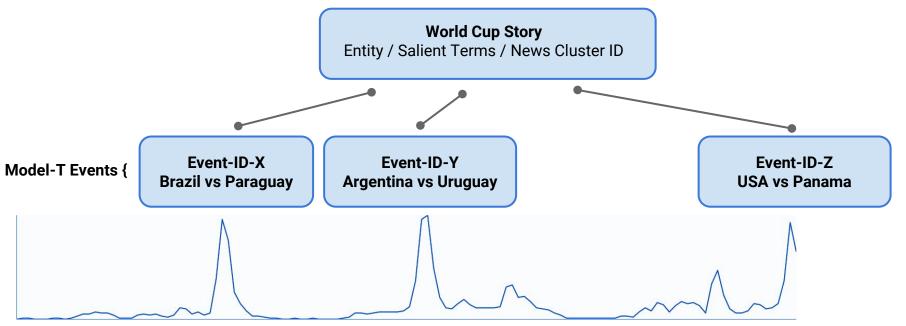
Popularity (CV and NB lift)

Google

Timelines - Hierarchy

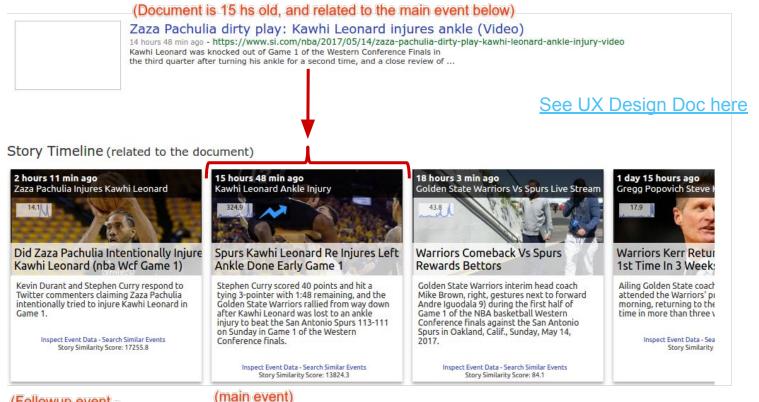
Timelines

Connecting multiple events that are separate in time



Checkpoints, Updates and Timelines



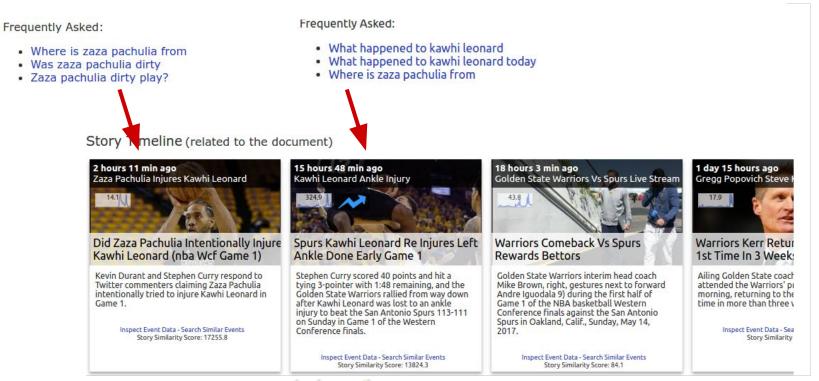


(Followup event - (main event) people are wondering if the other player intentionally tried to injure Kawhi)

Google

Events - Questions





(Followup event - (main event) people are wondering if the other player intentionally tried to injure Kawhi)

Events - Questions - Temporal Topicality

Frequently Asked:

- Where is zaza pachulia from
- Was zaza pachulia dirty
- Zaza pachulia dirty play?

Frequently Asked:

- What happened to kawhi leonard
- What happened to kawhi leonard today
- Where is zaza pachulia from

Temporal Topicality

- How topical is this question for the current time.
- Rank the questions based on the salient terms of the Spikes in the Event.
- The Spike salient terms are ranked by hivemind lift-score.
 - "How important the salient term compared to the background distribution".
 - Rank higher if they are specific the exact topic happening during the time of the query.

2 days later - More updates

- Pachulia is out of Game 2 for dangerous play ____
- Spurs coach is irate with Patchulia –
- Warriors vs Spurs Game 2 is on ¬

Story Timeline



the Spurs missing Kawhi Leonard, running away from San Antonio for a 136-100 rout Tuesday night and 2-0 lead in the Western Conference finals.

> Inspect Event Data - Search Similar Events Story Similarity Score: 148.5



(The Sports Xchange) – San Antonio star forward Kawhi Leonard will miss Tuesday night's Game 2 of the Western Conference finals at the Golden State Warriors, leaving Spurs coach Gregg Popovich irate over what he felt was an avoidable injury.

> Inspect Event Data - Search Similar Events Story Similarity Score: 9233.2

1 day 19 hours ago Gregg Popovich Zaza Pachulia

Gregg Popovich Says Kawhi Leonard Likely Out For Game 2 Rips Zaza Pachulia For 'dangerous' Play

NBA analyst Sekou Smith joins Tiki and Tierney to discuss if Golden State Warriors center Zaza Pachulia's play on San Antonio Spurs forward Kawhi Leonard should be considered dirty.

> Inspect Event Data - Search Similar Events Story Similarity Score: 7555.0





Did Zaza Pachulia Intentionally Injure Kawhi Leonard (nba Wcf Game 1)

Kevin Durant and Stephen Curry respond to Twitter commenters claiming Zaza Pachulia intentionally tried to injure Kawhi Leonard in Game 1.

> Inspect Event Data - Search Similar Events Story Similarity Score: 37708.0

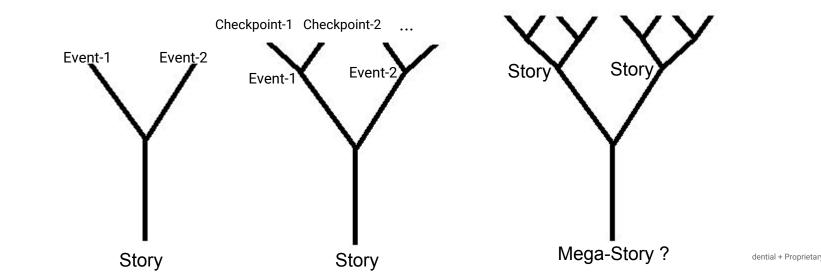


Story vs Event vs Checkpoint: Granularity

- Story can develop over many days, even months
- Event is a new development inside a Story
 - Takes place over a few hours, not more
- Checkpoints are intra-Event "snapshots"

Google

• Can be used to track the user's state in the Event/Story on a minute level



RTB Story Granularity

Main Story: Legarrette goes from Patriots to Eagles



Very Specific

Only about Eagles and Patriots dealings players' contracts (high similarity threshold - <u>query</u>)

Story Timeline



Best Fits For The Top 7 Nfl Free Agents Still Available

The New England Patriots issued \$1.1 million tender to free agent running back LeGarrette Blount on Tuesday, multiple media outlets reported Wednesday.

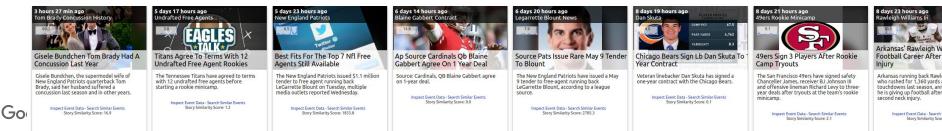
> Inspect Event Data - Search Similar Events Story Similarity Score: 1833.9



Inspect Event Data - Search Similar Events Story Similarity Score: 2785.3

Very Broad - Everything about Football - (low similarity threshold - <u>query</u>)

Story Timeline



Trump Fires Comey - News Cluster

(news cluster link)

The Original Cluster has been transformed into everything Comey/Russia/Trump

Mishmash of new developments:

- Comey Memo / Russia probe
- FBI next pick

Gooale

• Russia Leaks / Classified info

News Cluster is active for 8 days

• 18,549 articles so far

Realtime Coverage

Chaffetz: I'm ready to subpoena for Comey memo

fox8.com - 35 minutes ago

House oversight committee chairman Jason Chaffelz said Tuesday that if a memo FBI Director James Comey wrote detailing President Donald Trump's request to end an investigation exists, then his committee is willing to issue subpoenas to get it. *.



If the president isn't hiding anything, what is the need for his secrecy? Las Vegas Son - 4 hours ago Trump: FBI job hunt on fast track Northwest Arkanas Democrate Gazette - May 16, 2017 Americans Deserve To Know About Russian US Election Meddling | Senator Casey Newweek - May 15, 2017 Most important editorial of the weekend: It's time to consider impeachment The Progressive Pulse - May 15, 2017 Democrate continue to pressure Trump over FBI boss' axing SBS - May 14, 2017 Show recently hidden articles In Depth

Key Moments of Trump-Comey Clash and What's Come After

New York Times - 5 hours ago WASHINGTON — The standoff between President Donald Trump and his former FBI director intensified with the news of James Corney's memo saying that Trump asked him to shut down the agency's investigation into ousted national security adviser Michael ...

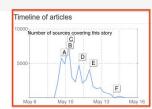
Will Firing Comey Cost Trump?

Daily Beast - May 14, 2017 While supporters of the president are arguing that he was within his legal right to fire Corney, there could be other repercussions. Michael Waldman. 05.13.17 11:29 PM ET. In the frenzied days since Donald Trump fired James Corney, all the original ...

Trump, GOP control FBI chief pick despite Democrats' calls

Washington Post - May 15, 2017 WASHINGTON — While Democrats may trot out any number of demands or maneuvers to influence the selection of the next director of the FBI, here's a reality check: Republican President Donald Trump fired James Comey, and he and his party will decide ...

Yes, Trump Thinks He Can Defeat The Russia Probe



- F Clapper: US govt 'under assault' by Trump after Comey firing Boston.com - May 14, 2017
- E The Daily 202: Trump's warning to Comey deepens doubts about his respect for the rule of law

Washington Post - May 12, 2017

- D Comey firing could spur new review of Clinton case, immunity deals, ex-agent says Fox News - May 11, 2017
- C Late night comes for President Trump over James Comey firing USA TODAY - May 10, 2017
- B The Latest: Pence says firing unrelated to Russia probe New York Daily News - May 10, 2017

A Trump ax falls on FBI's Comey in midst of Russia probe Atlanta Journal Constitution - May 9, 2017

Trump Fires Comey - RTB Event

- Original Event is only about Trump firing Comey
 - Precise in time and in topic
 - Trump leaking confidential information to Russia is a different story
- Timeline brings new developments
 - Still very related to Trump firing Comey



Story Timeline



Trump pulling out of Paris agreement



White House Official Says Trump Expected To Withdraw Us From Paris Climate Accord

President Donald Trump is expected to withdraw the United States from the Paris climate accord, a White House official said Wednesday, confirming a move certain to anger allies that spent years negotiating the landmark agreement to reduce carbon emissions.

Inspect Event Data - Inspect Cluster Data - Search Similar Events

1fd75f9990c3cb68

ac183440/381de63

6053dF2F7d010cc7

Frequently Asked:

- Why withdraw paris climate agreement?
- What is the paris climate accord?
- What is the paris climate deal?

News Cluster IDs: 00003000afebe4d2 - See in Google News 00003000b02e4f96 - See in Google News 00003000b038c45c - See in Google News CV Documents: Trump paris climate accord 1fd751 Paris climate accord ac183 Paris climate accord 6053d

NB Documents: 4ce7223de4932218 605cd4c6c25ad68c ff9b4bfe1e82bfae

Story Timeline



G 7 Source Us Gets More Time On In Out Climate Pact Answer

Seven wealthy democracies ended their summit Saturday in Italy without a unanimous agreement on climate change, as the Trump administration plans to take more time to say whether the U.S. Is going to remain in the landmark Paris accord that fights climate change.

Inspect Event Data - Search Similar Events

Story Similarity Score: 2554.2

Google



Trump Says He Ll Decide On Paris Climate Deal Next Week

President Donald Trump faced new controversies over his team's possible ties to Russia as he closed out his first trip abroad Saturday, a grueling five-stop sprint that ended with the promise of an imminent decision on the much-discussed Paris climate accord.

> Inspect Event Data - Search Similar Events Story Similarity Score: 1111.4



Europeans Making Sales Pitch To Trump On Climate Accord

White House Correspondent Julie Pace reports that President Donald Trump's views on climate change are "evolving" following discussions with European leaders who are pushing for him to stay in the Paris climate accord.

> Inspect Event Data - Search Similar Events Story Similarity Score: 401.2



Migration Crisis Loo Meeting | Dw News

> Talks between U.S. Presic and other leaders of the at the G7 summit on Fridbe "robust" and "challeng lambasted NATO allies ar German trade policies a c

> > Inspect Event Data - Sea Story Similarity Sc

(query)

And 21 days earlier, he said he would decide on the agreement after the G7 summit, which he just did. (also from the timeline)



Realtime Boost Events

• **Realtime** - Realtime detection of news-events - Under 5 minutes after event started.

• Event Understanding

- Build Correlations in multidimensional space
- Temporal locality
- Temporal topicality
 - Entities / Salient-Terms / Summary / Label / Queries / Questions / Videos, etc...
 - Event Popularity (CV / NB) and importance
- **Query Understanding** Search currently trending Events for the Query
 - Also for Query-less, using list of Entities / Salient Terms / Google-Now profile
- **Statefulness** Understand Story updates and checkpoints
- Fine Grained and precise in time

News Clusters catch-up plan?

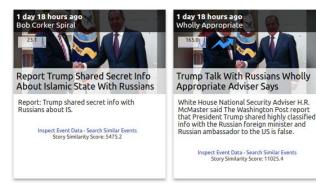
If we were to improve News Clusters to get to where RTBoost Events is right now. (Hypothetically without using Hivemind or RTBoost Spikes for the sake of argument.)

- Use modern-day document tokens: Entities and Salient Terms
- Improve Granularity and Time precision
 - Understand them in much more detail than we currently do
 - Spikes inside the current NewsCluster granularity ?
- Make cluster id stable (no splitting or merging)
- Build links between clusters (Stories)
- Integrate NB, CV, MiniSessions data pipelines into NewsClusters
- Build correlations with Temporal topicality
 - Use TF-DF over background distribution of terms in the corpus (in realtime?)

We would need to work on all of these integrations, to get to the same point we are with RTBoost Events.

RealtimeBoost Events - Known Issues

- **Underclustering** Sometimes too fine grained
 - likely a bug in the clustering



• Entity Intrusion

- Some articles contains Anchors to other non-related headlines in the middle of the centerpiece text.
- Fix is not hard.



Boy died of caffeine overdose.

Some articles contained a link to the SpaceX Cape Canaveral launch that just happened. The entity leaked to the caffeine Event.

• Local News coverage

Model-T Cost

Currently we use the Freebie Quota: 10 QPS We already requested 40KQPS for production deployment.

- Average Document Size:
 - Realtime Event Corpus: 400KB / doc
 - Can be much smaller if we want
 - Newsstand corpus : **45KB / doc**
- Indexing Memory Usage
 - Realtime Event: **4GB**
 - Newsstand: 64GB
- Muppet CPUs Per Query
 - Realtime Events: **0.4 CPUs** / Query [link]
 - Also can be improved
 - Newsstand: 1.1 CPUs / Query [link]

Ready for Launch

Need greenlight from leads to go ahead with 40KQPS for production deployment.

ModelT Query API is ready and available <u>here</u>.

Integration inside FCS Stories is to be decided with AlexK.

ATTIC

Model-T Events Retrieval

Online Retrieval from Model-T

Can be queried by:

- SQuery / Entity / Terms
- **Spike** (bring Events for given spike)
- Weighted set of **Entities** (Google Now profile)
- Weighted set of **Locations** (What is trending here)
- News Cluster ID (bring Spike Events containing this news-cluster)

Document annotation

(Indexing in regular muppet)

 Annotate each news document with Event ID Similar to what google news clusters do

Event

Event Label

Brazil 1st to qualify for world cup

Event Summary

A resurgent Brazil squad under new management became the first team to qualify for the World Cup on Tuesday.

Queries

brazil vs paraguay live streaming

Questions

who scored brazil vs paraguay



Locations (from cv and nb)



Popularity (CV and NB lift)

Model-T Events Retrieval

Online Retrieval

Flexible (can do complex queries)

• Trending events per Location etc

Breaking news events is guaranteed to be quickly indexed, retrieved, ranked

Can't index singleton-documents or non-spiking topics

Document annotation

Plays well with current ranking systems (FCS etc)

Rank singleton-docs and topics vs Events

Already contains all per-doc ranking signals we want

Might miss very fresh breaking news due to lack of signals on fresh docs

• Can be fixed with Spark / Crowding per Event-ID etc...

Model-T Events Checkpointing

Event-ID-X (timestamp 1)

Event Label Brazil is playing for world cup

Event Summary

Brazil must win to qualify for the World Cup.

Queries

brazil vs paraguay live streaming

Questions

When does the match start



Locations (from cv and nb)



Event-ID-X (timestamp 2)

Event Label Brazil 1st to qualify for world cup

Event Summary

A resurgent Brazil squad under new management became the first team to qualify for the World Cup on Tuesday.

Queries

brazil vs paraguay live streaming

Questions

who scored brazil vs paraguay



Locations (from cv and nb)

Popularity (CV and NB lift)

Keep checkpoints of the same Event-ID for multiple timestamps

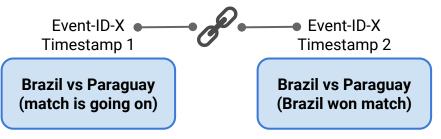
Retrieve the checkpoint based on user-state

Works for bigger / long-running events Might be harder for short or smaller events (Less documents distributed across time)

Checkpointing VS Timelines - Hierarchy

Checkpoints

Multiple timestamps for same Event under development



Timelines

Google

Multiple Events connected by common topic



Timelines - Hierarchy

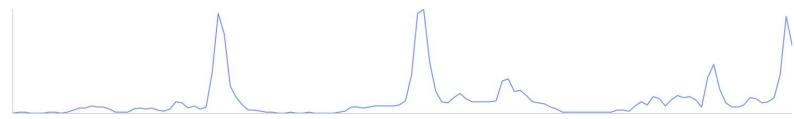
Timelines

- Model-T allows retrieval by cosine similarity of
 - Entities / Salient Terms / News Cluster ID / etc...
 - Start with Event-ID-X and retrieve Events by similar Entities etc ...

RealtimeBoost Spike Events vs News Clusters

- **Coarse grained** News Clusters has longer time-span and contains multiple spikes
- **Fine grained** Spikes are punctual in time, lasting 3 to 4 hours

We can use these ideas to build the Hierarchy



Event Detection - Current State

Clusterization Binary

- Index Events in Model-T (in realtime)
- Under development ETA beginning of April

Model-T

• Already in place with Flex quota - <u>Getting Prod resources now</u>

Model-T Query Service

- Serve Events to Clients
- Simpler Query API than querying directly Model-T
- Under development ETA beginning of April

Event Detection - Current State

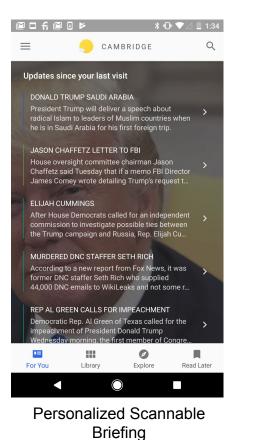
Current Features

- Entities / Salient Terms / Images / DocIDs / URLs / Titles / Summaries / Labels
- Locations by CV and NB
- HasVideo Tag
- Queries by NB and Mini-Sessions
- Popularity by CV and NB
- News Cluster ID
- Questions

Planned Features

- Fact Checking / Sentence Understanding ETA TBD
- Youtube Views ETA TBD
- Position Ranking signal ETA April

RT Events in 360



Google

◕▥◯▤ऩॕ▥▥◙▻ ← Comey Memos Z \sim Updates since your last visit The Senate intelligence committee first requested from the FBI copies of his memos, said the panel's top Democrat, Sen, Mark Warner, as well as asking ... Why was flynn fired? SEARCH RESULTS Can the president be forced to testify before congress? SEARCH RESULTS \bigcirc Story Summary and Questions

ਿ∎ਿ≣ਜਿ⊠	🖻 🗟 🔈 🔺 🖨 💎 ⊿ 🛢 10:16
Does bertha burn	n down thornfield hall?
	SEARCH RESULTS
What happened on fox news web	
	SEARCH RESULTS
Story Timeline	
d its possible omey context ^{rday}	Republican leader recorded sa he thinks Putin pays Trump New York Daily News Yester
•	

Timeline