

Building Trust and Fighting Misinformation @google.com

Corinna Cortes, Head Google Research, NYC

The Way We Experience News Today is Complex

Overwhelming

48K new documents
every minute



Trump says "military option" is possible in Venezuela

Miami Herald · 5h ago



Trump Warns North Korea 'Military Solutions' Are 'In Play'

Wall Street Journal · 5h ago

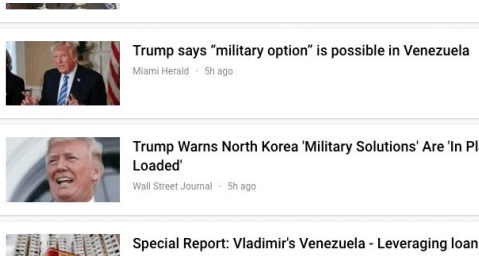


Special Report: Vladimir's Venezuela - Leveraging loan:

The Way We Experience News Today is Complex

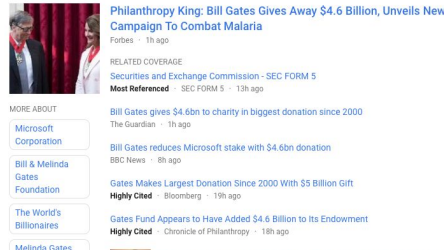
Overwhelming

48K new documents
every minute



Repetitive

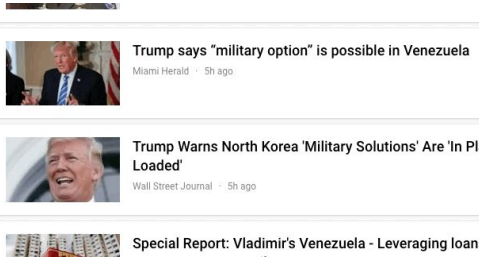
News has been
commoditized



The Way We Experience News Today is Complex

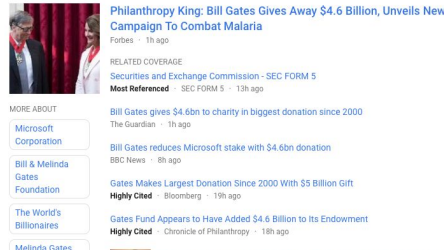
Overwhelming

48K new documents
every minute



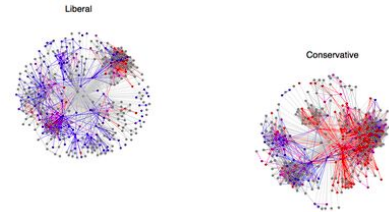
Repetitive

News has been
commoditized



Polarizing

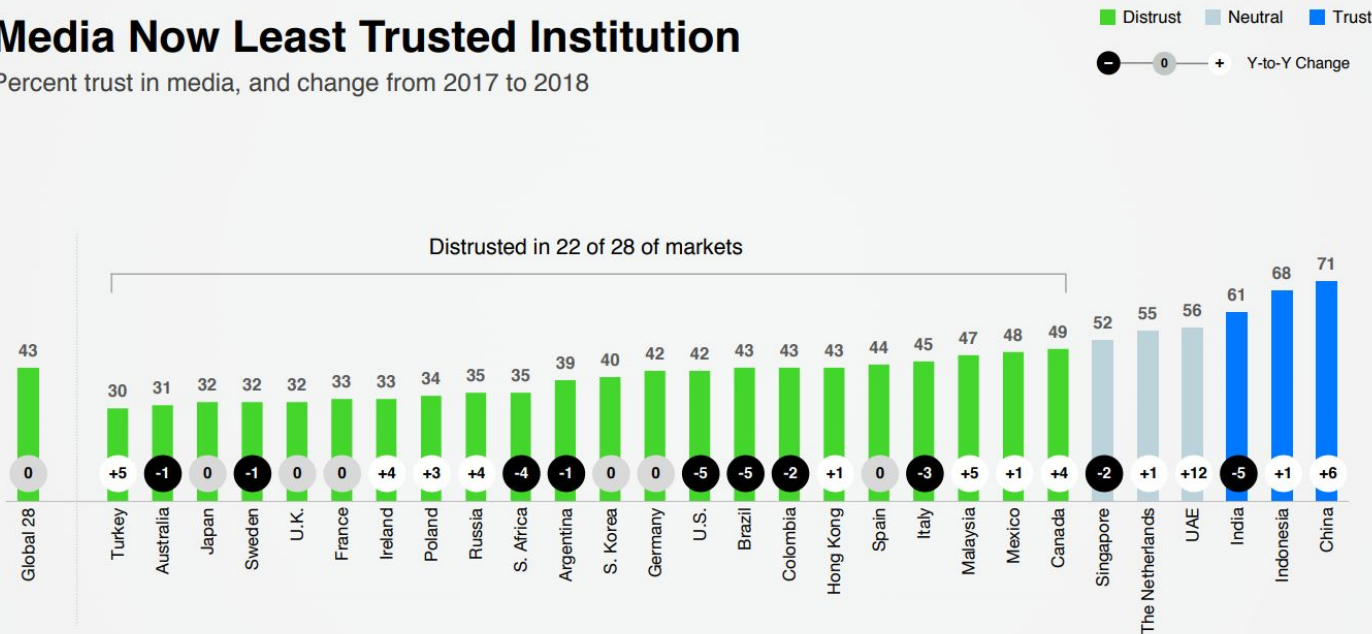
Filter bubbles
abound



The Trust in Media has Dropped to All-time Low in 2018

Media Now Least Trusted Institution

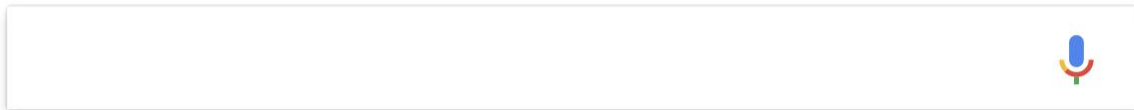
Percent trust in media, and change from 2017 to 2018



Source: 2018 Edelman Trust Barometer. TRU_INS. [MEDIA IN GENERAL] Below is a list of institutions. For each one, please indicate how much you trust that institution to do what is right using a nine-point scale where one means that you "do not trust them at all" and nine means that you "trust them a great deal." (Top 4 Box, Trust) General population, 28-market global total.

Lets start with this:

Google Search is a **search tool,
not a ledger of absolute truth**



Search quality

relevance

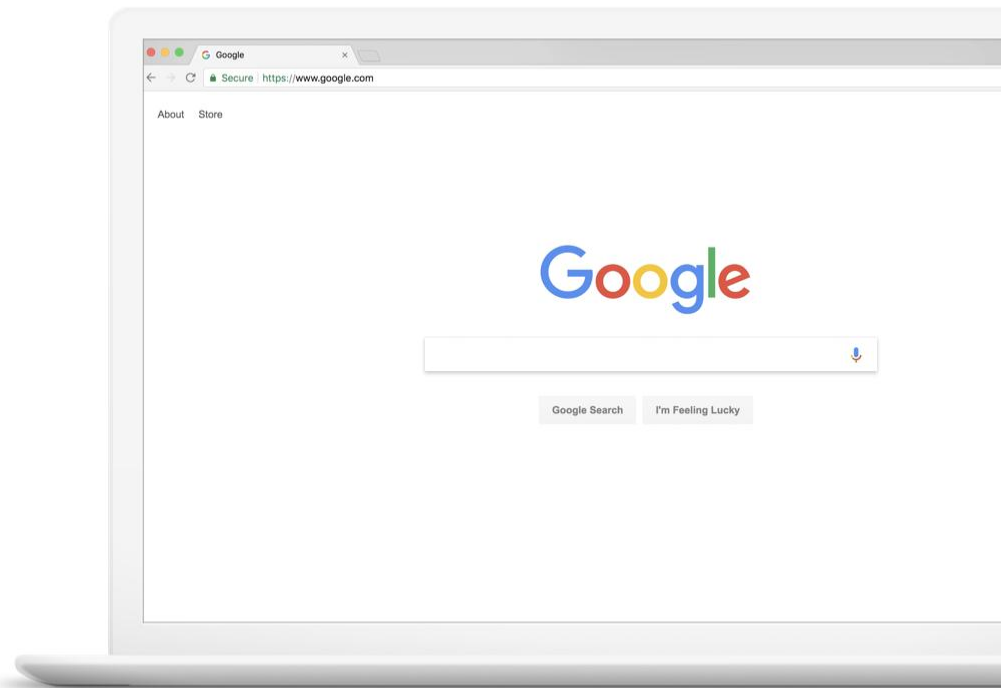
rel•e•vance /ˈreləv(ə)ns/ *noun*

How closely the words in a website match
the words in your search query

authoritativeness

au•thor•i•ta•tive•ness /əˈTHôrə,tādiv/ *noun*

How much expertise the result is based on
and how much people trust the source



Search quality

relevance

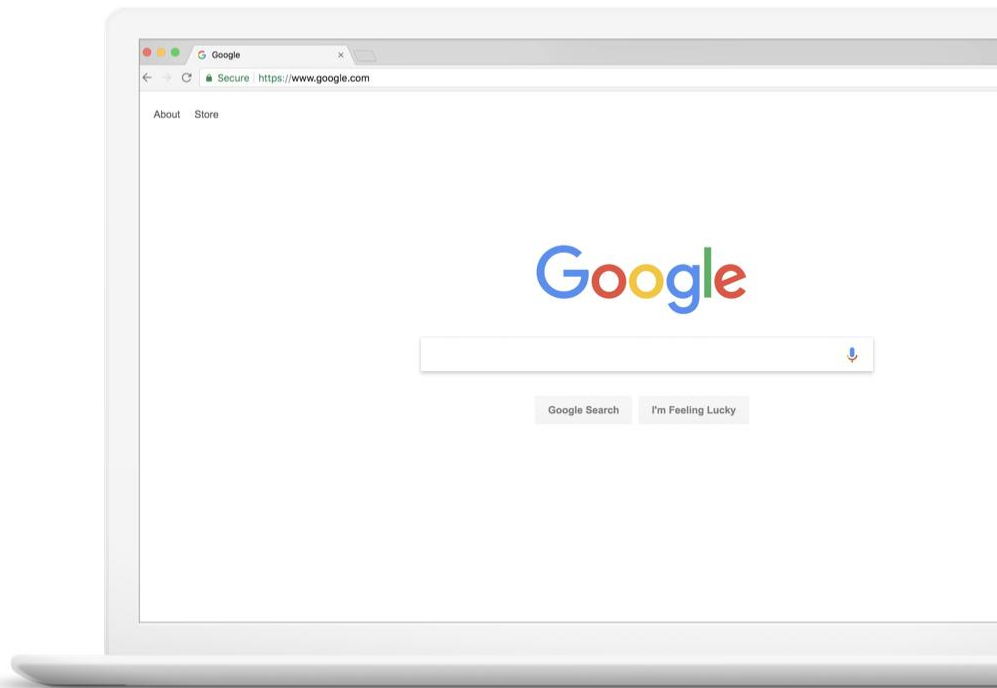
rel•e•vance /ˈreləv(ə)ns/ *noun*

How closely the words in a website match
the words in your search query

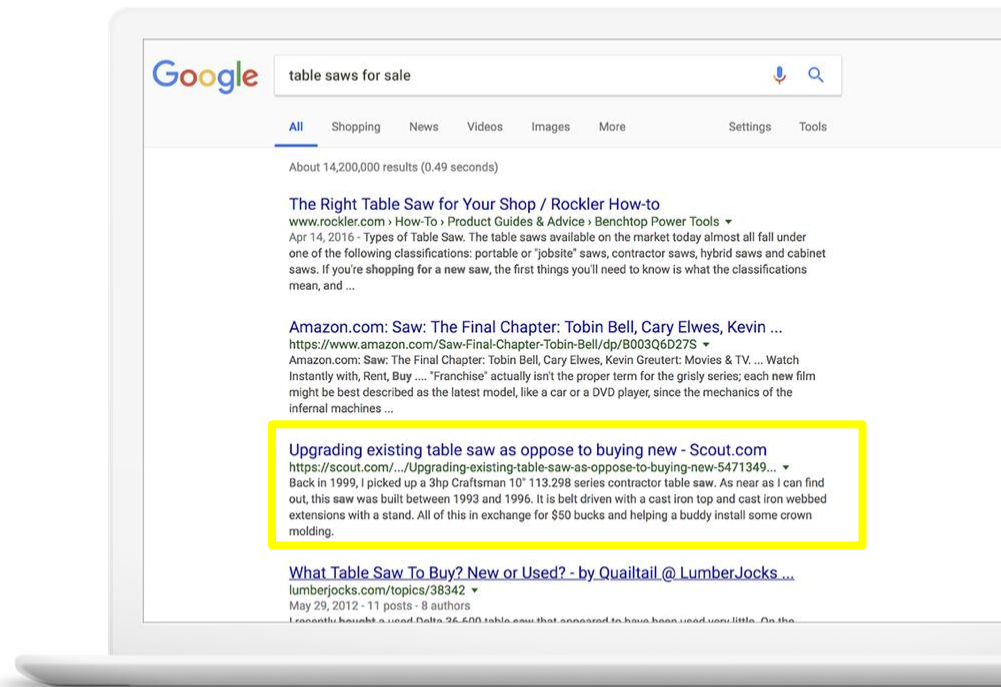
authoritativeness

au•thor•i•ta•tive•ness /əˈTHôrə,tādiv/ *noun*

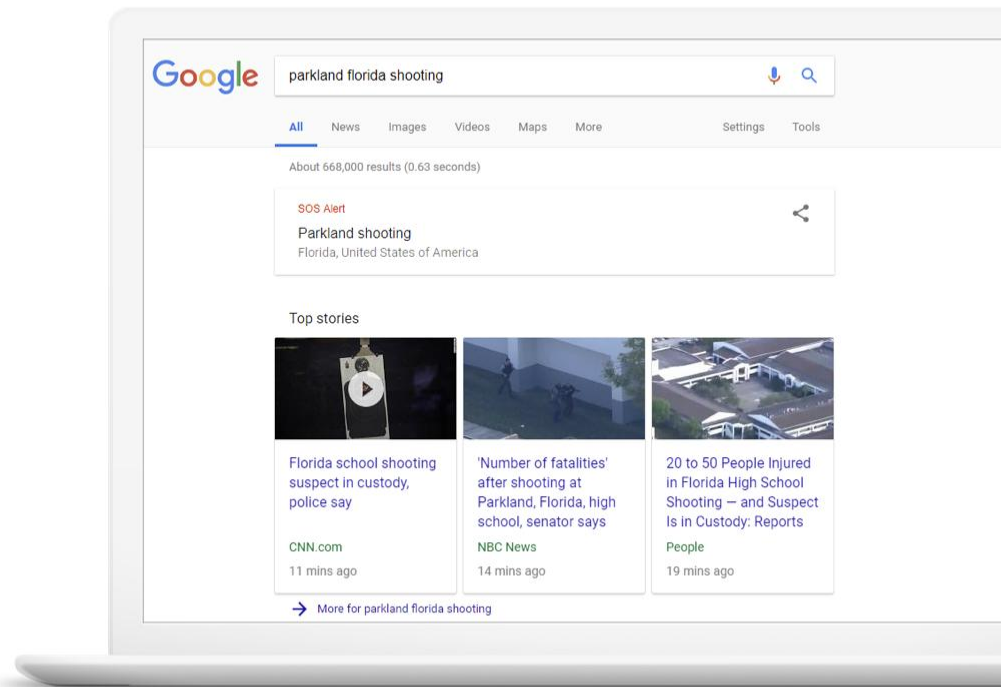
How much expertise the result is based on
and how much people trust the source



Balance between
relevance and
authoritativeness
varies depending on
the nature of the query



Surfacing more
authoritative content
during breaking
news situations



Bad Actors Targeting High-Interest Events

Sections

The Washington Post
*Democracy Dies in
Darkness*

The Fix

Google's top news link for 'final election results' goes to a fake news site with false numbers

By JENNIFER EARL / CBS NEWS / November 14, 2016, 5:57 PM

Google's top search result for "final election numbers" leads to fake news site

In the news



FINAL ELECTION 2016 NUMBERS: TRUMP WON BOTH POPULAR (62.9 M -62.2 M) AND ELECTORAL COLLEGE ...

70news - WordPress.com - 2 days ago

CNN NUMBERS AS OF 11/13/16: <http://edition.cnn.com/election/results>.
trump headshot ...

Clinton vs. Trump Popular Vote: Are There Still Uncounted Ballots?

Heavy.com - 23 hours ago

Laramie County 2016 Election Results Final

Kgab - 23 hours ago

3/20/2018 Announcement

Launching today: online courses

firstdraftnews.org/learn

FIRST DRAFT

HARVARD Kennedy School
SHORENSTEIN CENTER
on Media, Politics and Public Policy

ABOUT

EDUCATION

NEWS

PROJECTS

MANAGE CURRICULUM USERS

DASHBOARD



Course:

Verification Training For Journalists

In this course, we'll teach you the steps involved in verifying the eyewitness media, fabricated websites, visual memes and manipulated videos that emerge on social media. The course is designed so that anyone can take the course from start to finish online or educators can take

Unit List

GETTING STARTED WITH
VERIFICATION [+]

VERIFICATION AS PROCESS
[+]

INVESTIGATING DIGITAL
FOOTPRINTS [+]

LEARNING DATE AND TIME
STAMPS [+]

UNDERSTANDING GEO-
LOCATION [+]

Download all PDFs

Start

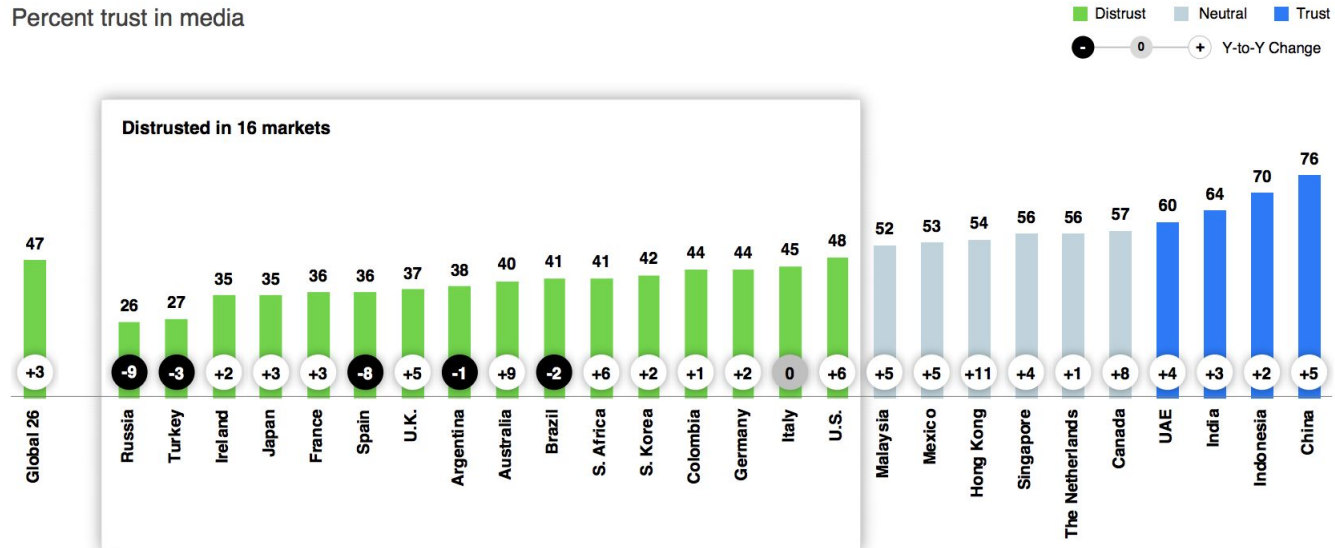
FIRST DRAFT
FirstDraftNews.org

HARVARD Kennedy School
SHORENSTEIN CENTER
on Media, Politics and Public Policy

Trust in Media increase slightly in 2019

TRUST IN MEDIA RISES, REMAINS LEAST-TRUSTED INSTITUTION

Percent trust in media



2019 Edelman Trust Barometer. TRU_INS. [MEDIA IN GENERAL] Below is a list of institutions. For each one, please indicate how much you trust that institution to do what is right using a nine-point scale where one means that you "do not trust them at all" and nine means that you "trust them a great deal." 9-point scale; top 4 box, trust. General population, 26-market average.

What else can we do?

Outline

ClaimReview Schema

Knowledge Card for publishers

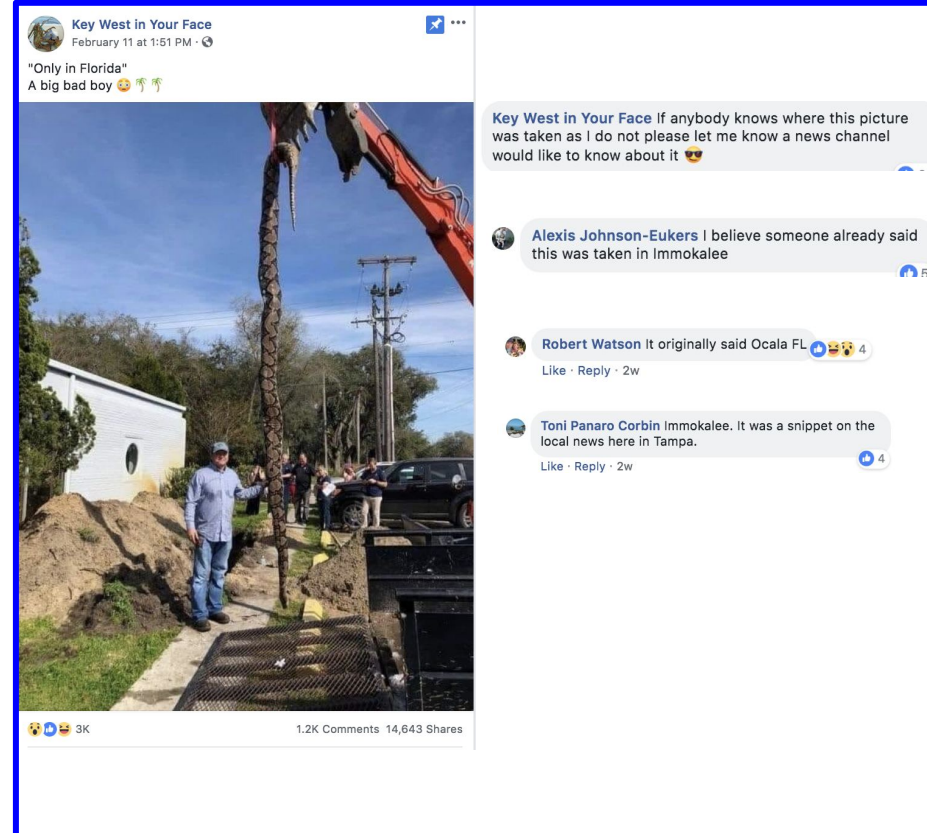
Finding related and relevant stories

Automatic fact checking

64 Fact Checking Organizations Around the World, IFCN



Fact Check Pair (FC-ing Article, FC-ed Claim)



Fact Check Pair (FC-ing Article, FC-ed Claim)

News / Fact Check / Fact Check: This is not a video of heavy firing in Pakistan but a gun event in US

Fact Check: This is not a video of heavy firing in Pakistan but a gun event in US

A Facebook page called 'Voice of Pashtun' has posted a video of heavy firing, claiming it to be of "the current war going on in Pakistan's Sialkot region".



Chayan Kundu
New Delhi

February 28, 2019 UPDATED: February 28, 2019 15:29 IST

In the video, we can see people moving casually behind the guns and taking shots on mobile devices, which is very unlikely of a war-like situation.

@meskal - (4K) EPIC Machine Gun Night Shoot at the Knob Creek Gun Range

[Short Version](#)

Facebook page [Voice of Pashtun](#) posted a 48-minute video of heavy firing at night with the caption: "Breaking news!... The current war is now going on in Pakistan's Sialkot region#PashtunsStand4Peace".

This [post](#) has been viewed by 93,000 people and shared nearly 2,000 times at the last count.



The Fact Checking Schema

ClaimReview from schema.org

- Fields: **claim**, claimant, **verdict**
- Collaboration with the Fact Checking Community, e.g., Bill Adair

ClaimReview

Canonical URL: <http://schema.org/ClaimReview>

[Thing](#) > [CreativeWork](#) > [Review](#) > [ClaimReview](#)

Research, Jigsaw, and Google News
Surfaced by Google and Bing

Properties

@type

ClaimReview

author

The publisher of the fact check article ...

name

url

claimReviewed

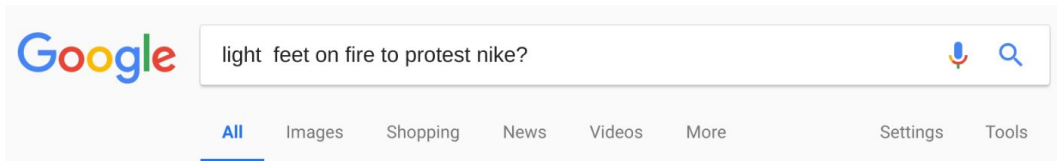
A short summary of the claim being evaluated ...

ClaimReview in Google Products

Fact Checking Corpus

- Tens of thousands of articles
- 100+ publishers

Surfaced in Search and News



Fact Check: Did Someone Light Their Feet on Fire to Protest Nike?

<https://www.weeklystandard.com/.../fact-check-did-someone-light-their-feet-on-fire-to...>

Claim: Did Someone Light Their Feet on Fire to Protest Nike?

Claimed by: 12up.com

Fact check by The Weekly Standard: FALSE

Feedback



Fact check

Kavanaugh File: Executive Privilege

FactCheck.org

Donald Trump wrongly says Social Security and Medicare are stronger

PolitiFact

Did the Widow of 'American Sniper' Chris Kyle Criticize Nike's Colin Kaepernick Campaign?

Snopes.com

Fact Check: Did Someone Light Their Feet on Fire to Protest Nike?

The Weekly Standard

Online Joke Circulates False Information About US Visa Application for Russian Citizens

Polygraph.info

Google News

Search for topics, locations & sources



ories

ou

ites

l searches

I

ess

iology

tainment

s

ce

• Doomed Ethiopian Airlines flight seen 'swerving and dipping' in moments before crash

New York Post • Yesterday

[View full coverage](#)

Trump Lauded Farmers, Medicare and AIDS Programs. Then He Unsheathed the Budget Knife.

The New York Times • 5 hours ago

• 'We're tired of being right' about the border, budget director says on Trump's new wall money request

CNBC • Yesterday

[View full coverage](#)



Vitriol and deflection: The return of the White House press briefing

CNN • 1 hour ago

• The Hill's Morning Report - Dems look to rebuild 'blue wall' | TheHill

The Hill • 4 hours ago

[View full coverage](#)



Fact check

Comparing Manafort's sentence to life sentence for man who sold pot

PolitiFact

Did the Two American Mormons Detained in Russia Violate Russian Law?

Polygraph.info

Henry Winkler, Democratic Donor, Still Alive

FactCheck.org

Did the Girls Scouts Give Their 'Highest Award' to a Girl for Her Pro-Choice Activism?

Snopes.com

No, this video does not show the moment the Ethiopian Airlines plane crashed near Addis Ababa in March 2019

AFP Factcheck



fact checking youtube



All

Videos

News

Images

Shopping

More

Settings

Tools

About 157 000 000 results (0,55 seconds)

Top stories



YouTube will provide users with a 'fact-check' on sensitive search topics

TechRadar

2 hours ago



YouTube is testing a fact-check tool that debunks conspiracy theories

The Drum

2 days ago



YouTube testing 'fact check' pop-up

The Hindu

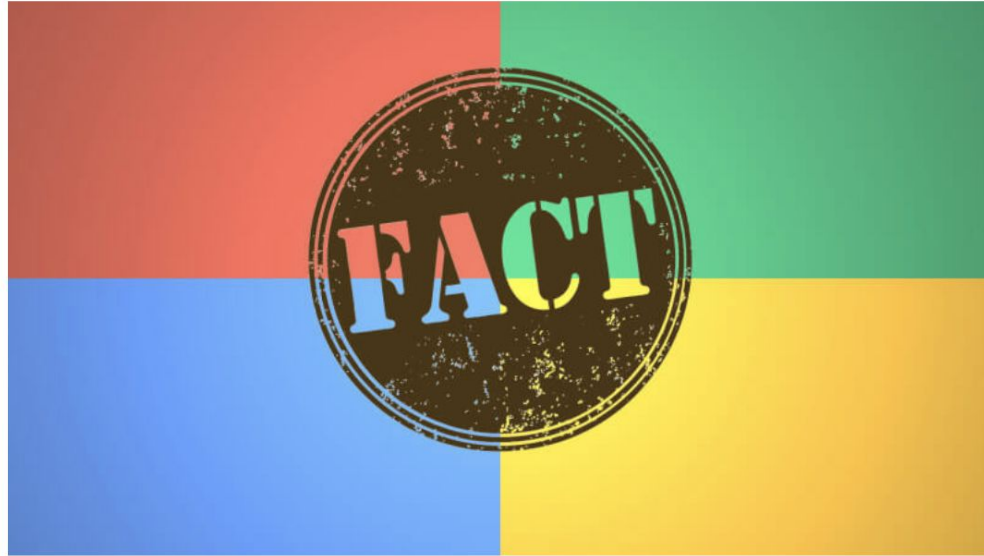
45 mins ago

→ [More for fact checking youtube](#)

Google adds new knowledge panel to provide information about news publishers

In an effort to combat fake news and equip searchers with more data about news sources, Google has released a new publisher knowledge graph.


Barry Schwartz on November 7, 2017 at 8:09 pm



Google has [announced](#) a new knowledge graph card for news publishers in which searchers can learn more about a specific news publication directly in the search results.

Google said this will help searchers learn about “a publication you’re not familiar with or one you wanted to learn more about.” The knowledge panels also give searchers faster access to information about a publisher and can help Google address the [misinformation](#) and [fake news](#) issues it has been fighting for some time now.

Fact Checking Part of KPs for Publishers



[All](#) [News](#) [Shopping](#) [Videos](#) [Books](#) [More](#) [Settings](#) [Tools](#)

About 10,100,000 results (0.67 seconds)

Breitbart News Network

www.breitbart.com/ ▼
Syndicated news and opinion website providing continuously updated headlines to top news and analysis sources.

Big Government

Big government is a term used by conservatives to describe a ...

Breitbart News, Author at ...

Breitbart News, Author at Breitbart. ... crowd at the Citadel ...

Breitbart London


Breitbart London is the home of the hottest video on politics, world ...

Breitbart TV

Breitbart TV is the home of the hottest video on politics, world ...


[More results from breitbart.com](#)

Top stories from breitbart.com




Remembrance Sunday: Britain Remembers the Fallen

Breitbart · 9 hours ago



Donald Trump on Kim Jong-un: 'I Would NEVER Call Him "Short and Fat"'

Breitbart · 16 hours ago



Pat Caddell: Global Elites Want to 'Manage' the Country's Decline

Breitbart · 17 hours ago

Breitbart News

Website



breitbart.com

Breitbart News Network is a far-right American news, opinion and commentary website founded in 2007 by conservative commentator Andrew Breitbart. [Wikipedia](#)

[WRITES ABOUT](#) [REVIEWED CLAIMS](#)

Claim: Planned Parenthood and the Satanic Temple have worked together to oppose laws restricting abortion rights in...

Claimed by: Breitbart

Fact check by Snopes.com: False

FACT CHECK: Did Planned Parenthood 'Team Up' With...
<https://www.snopes.com/plan...>

Claim: 80 percent of Americans oppose sanctuary cities.

Claimed by: Breitbart

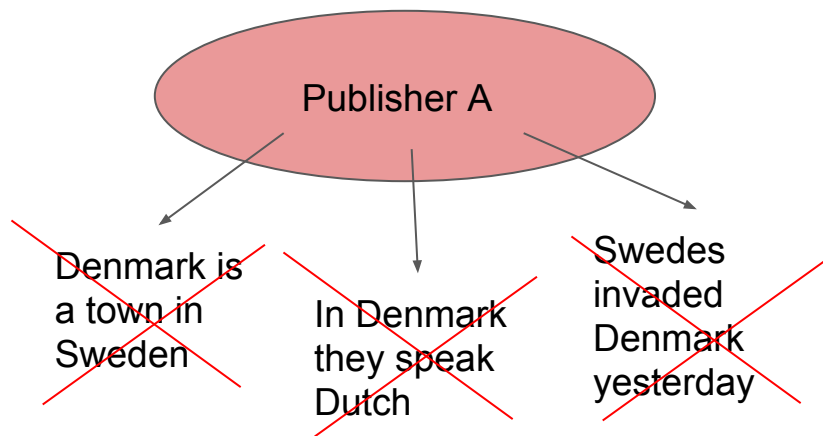
Fact check by PolitiFact: Polls depend on wording

Anatomy of a statistic: Do 80 percent of Americans oppos...
www.politifact.com/truth-o-...

www.politifact.com/truth-o-meter/article/2017/feb/24/anatomy-statistic-do-80-percent-americans-oppose-s/

Knowledge Graph Panel for Publishers

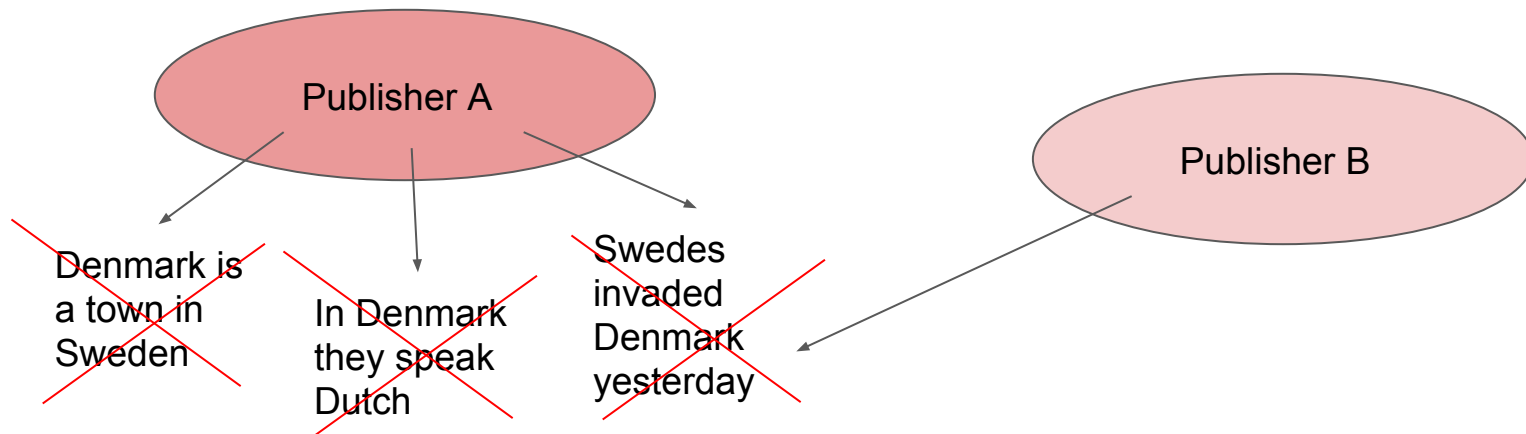
Can we rate how authoritative a publisher is based on the agreements with the fact checking articles?



Knowledge Graph Panel for Publishers

Can we rate how authoritative a publisher is based on the agreements with the fact checking articles?

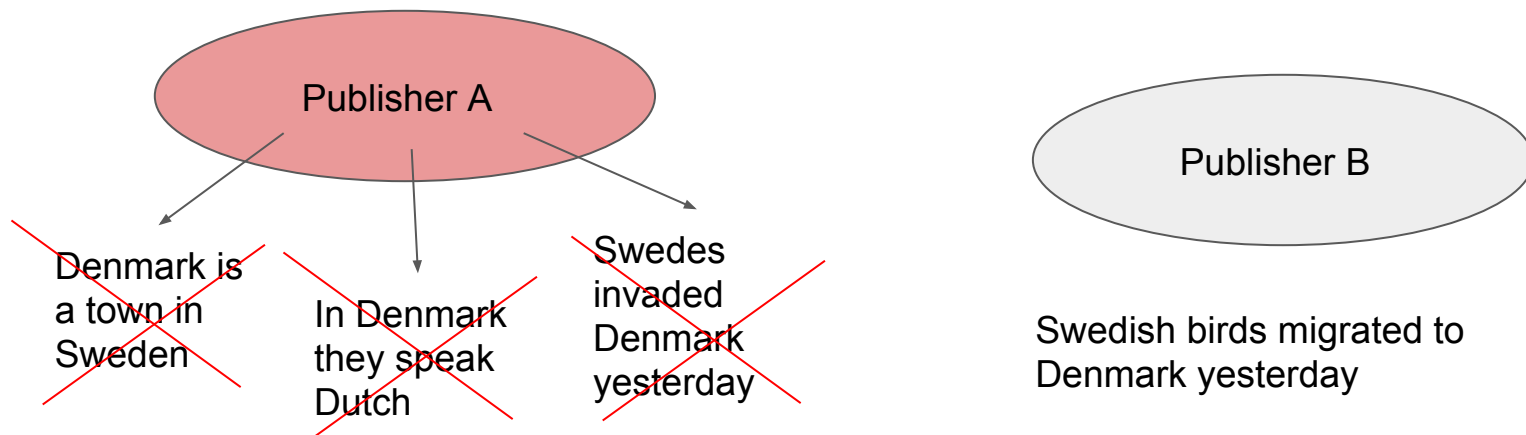
Given a fact checking article can we find articles by other publishers on exactly the same claim?



Knowledge Graph Panel for Publishers

Can we rate how authoritative a publisher is based on the agreements with the fact checking articles?

Given a fact checking article can we find articles by other publishers on exactly the same claim?



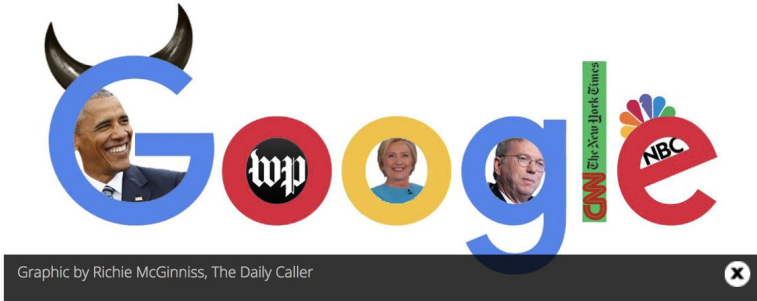
Fact Checking Part of KPs for Publishers



THE **DAILY CALLER**

MEDIA

THE **DAILY CALLER**
NEWS FOUNDATION



Google's New Fact-Check Feature Almost Exclusively Targets Conservative Sites

- Fact checking primarily done on one side of political spectrum
- Claims can have small but crucial differences

Fact Checking Part of KPs for Publishers



THE **DAILY CALLER**

TECH

THE **DAILY CALLER**
NEWS FOUNDATION



Google Suspends Fact Check Project,
Crediting TheDCNF Investigation With
Decision

 **mrc NewsBusters™**
EXPOSING & COMBATING LIBERAL MEDIA BIAS

[MRC Business](#) [MRC Culture](#) [MRC Latino](#) [Donate](#)

[John Bolton](#) [Stormy Daniels](#) [Guns](#) [Trump Coverage Studies](#)

Google Discontinues Flawed Fact-Checking Effort; Poisonous Corporate Culture Remains

f 15 Shares

Tweet

18



By **Tom Blumer** | January 19, 2018 4:48 PM EST

Outline

ClaimReview Schema

Knowledge Card for publishers

Finding related and relevant stories

Automatic fact checking

How Do We Find Claim-Relevant Documents?

Claim-relevant discovery process to help identify online misinformation:

1. Retrieve related candidate documents;
2. Score for relevance;
3. Stance detection: contradict, support, discuss the claim.

Relevant Document Discovery for Fact-Checking Articles,
WWW2018, Xuezhi Wang, Cong Yu, Simon Baumgartner, Flip

Retrieve Candidate Documents

24 August 2017



Fact Check > Science

Were Enormous Pyramids Just Discovered In Antarctica?

Viral images attempting to sell pyramid-shaped features as evidence of an ancient civilization in Antarctica show nothing but mountains.



Google™

CLAIM

Recently discovered satellite photos from Google Earth show that an ancient civilization built pyramids in Antarctica 100 million years ago.

RATING



ORIGIN

The most recent version of this story comes from a [post](#) filed to the “consciousness” section of new-age web site “Earth. We Are One” (EWAO), which suggests that a photograph from a mountain range first [discovered](#) in 1935 is now breaking news:

Recently, Google Earth satellite imagery discovered a constellation of three snow-covered pyramids in Antarctica. Scientists and researchers are, admittedly, rather taken aback. Two of the three pyramids are about 10 miles inland, while the third is directly near the coastline.

The implications? Well, this discovery could end up re-writing our entire history. No research has ever mentioned a civilization in Antarctica, much less one with the technological know-how to construct pyramids. So, how did these pyramids get there?

Retrieve Candidate Documents: Collect Candidates

Automation:

- Search Google with
 - Title and claim text
 - Entities from title and claim text
 - Queries from click-graph, top 50
- Top 100 search results
- Links from fact checking article

Total of 2,000+ candidates per claim.

 **Human research, skilled workers:** Automation gets 80% recall.

[https://www.google.com/search? q=Enormous+Pyramids+Just+Discovered+In+Antarctica](https://www.google.com/search?q=Enormous+Pyramids+Just+Discovered+In+Antarctica)

Enormous Pyramid Just Discovered in Antarctica Could ... - SOULSpot

soulspottv.com/.../enormous-pyramid-just-discovered-in-antarctica-could-change-how... ▼
Aug 19, 2017 - Google Earth: It's good for more than **just** looking at aerial views of where your ex is living now. As it turns out, it's good for a whole lot more, actually: **Recently**, Google Earth satellite imagery **discovered** a constellation of three snow-covered **pyramids** in **Antarctica**. Scientists and researchers are admittedly

Mysterious Void Discovered in Egypt's Great Pyramid - Latest Stories

<https://news.nationalgeographic.com/.../great-pyramid-giza-void-discovered-khufu-ar...> ▼
Nov 2, 2017 - The void is the first **large** inner structure **discovered** within the 4,500-year-old **pyramid** since the 1800s—a find made possible by recent advances in "What's so delightful is that [muons] are like Goldilocks: They lose enough [energy] to detect them but not so much that they **just** get absorbed in the target," ...

US Marines sent to explore Ancient Pyramids discovered in Antarctica ...

www.section51-ufo.com/.../us-marines-sent-to-explore-ancient-pyramids-discovered-i... ▼
Apr 25, 2017 - Three ancient **pyramids** have been **discovered** in the **Antarctic** in 2014 by a team of American and European scientists. The **only** reliable information provided by the scientists was that they were planning an expedition to the **pyramids** to research them more thoroughly and determine for sure whether the ...

The Truth About The Mysterious "Pyramid" Discovered In Antarctica ...

www.iflscience.com/.../the-truth-about-the-mysterious-pyramid-discovered-in-antarcti... ▼
Nov 27, 2016 - It's a new day on the internet so there's a new conspiracy theory: a mysterious new **pyramid** has been **discovered** in **Antarctica** thanks to Google Earth. Except it's **just** a mountain. And even the "new" part of the conspiracy theory is **just** a re-hashed version of a story that has been reported for years on ...

**Same claim,
supporting the
claim**

Different claim

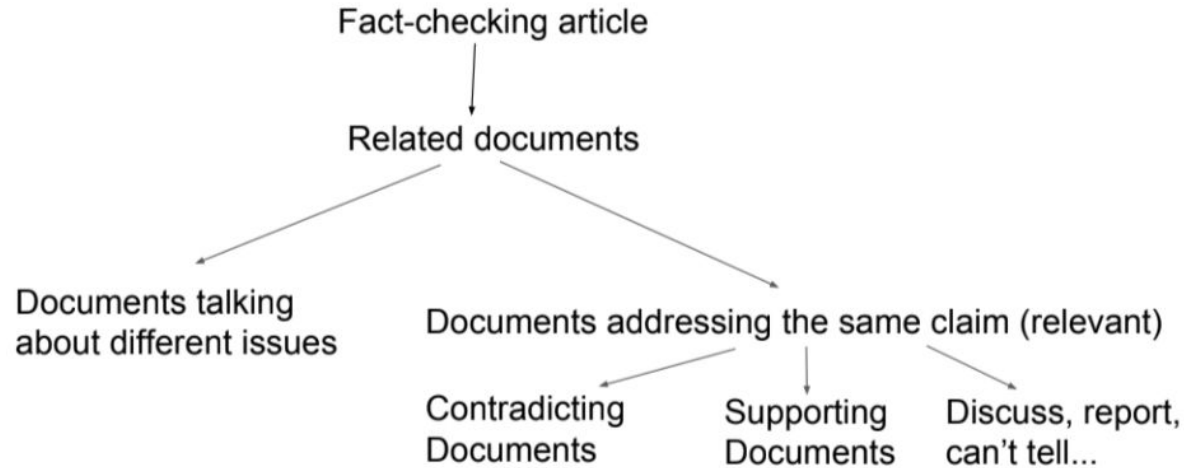
Different claim ???

**Same claim,
contradicting
the claim**

Relevance Classification

- Date difference
- Cosine similarity measure:
 - Entity extraction, (e_1, e_2, \dots, e_K) , with confidence scores, c :
 - Claim document: claim text (c_1, c_2, \dots, c_K)
 - Candidate document: title, heading, paragraphs, sentences $(c'_1, c'_2, \dots, c'_K)$
 - Repeat for all pairs (claim, title), (claim, heading), (claim, paragraphs), (claim, sentences). Keep max score.
- Gradient boosted decision tree.

How Do We Find More Claim-Relevant Documents?



Stance Detection: Contradicting, Supporting, Discussing?

Bag of words do not easily capture contradictions

Look for patterns

POLYGRAPH.info 🔍 ☰

FALSE

Despite Denials, Links Between Cambridge Analytica and Russia Emerge

March 23, 2018

Fact Check > Science

Were Enormous Pyramids Just Discovered In Antarctica?

Viral images attempting to sell pyramid-shaped features as evidence of an ancient civilization in Antarctica show nothing but mountains.



↗ 298

G

f

🐦

</>

+

CLAIM

Recently discovered satellite photos from Google Earth show that an ancient

Stance Detection: Contradicting, Supporting, Discussing?

Extract unigrams and bigrams, 900-dim vocabulary

- Unigrams: "fake", "purportedly", "hoax", "actually", "rumor", "satirical", "false", "fabricated", "misleading", "inaccurate", "manipulated", "fictional", "incorrectly", ...
- Bigrams: "made up", "fake news", "fact check", "not true", "no evidence", ...

900-dimensional vector based on the union of the key components.

Gradient boosted decision tree: supporting/contradicting

YouTube Launch in India

- March 8, 2019
- Information panel at the top



fact checking youtube



All

Videos

News

Images

Shopping

More

Settings

Tools

About 157 000 000 results (0,55 seconds)

Top stories



YouTube will provide users with a 'fact-check' on sensitive search topics

TechRadar

2 hours ago



YouTube is testing a fact-check tool that debunks conspiracy theories

The Drum

2 days ago



YouTube testing 'fact check' pop-up

The Hindu

45 mins ago

→ [More for fact checking youtube](#)

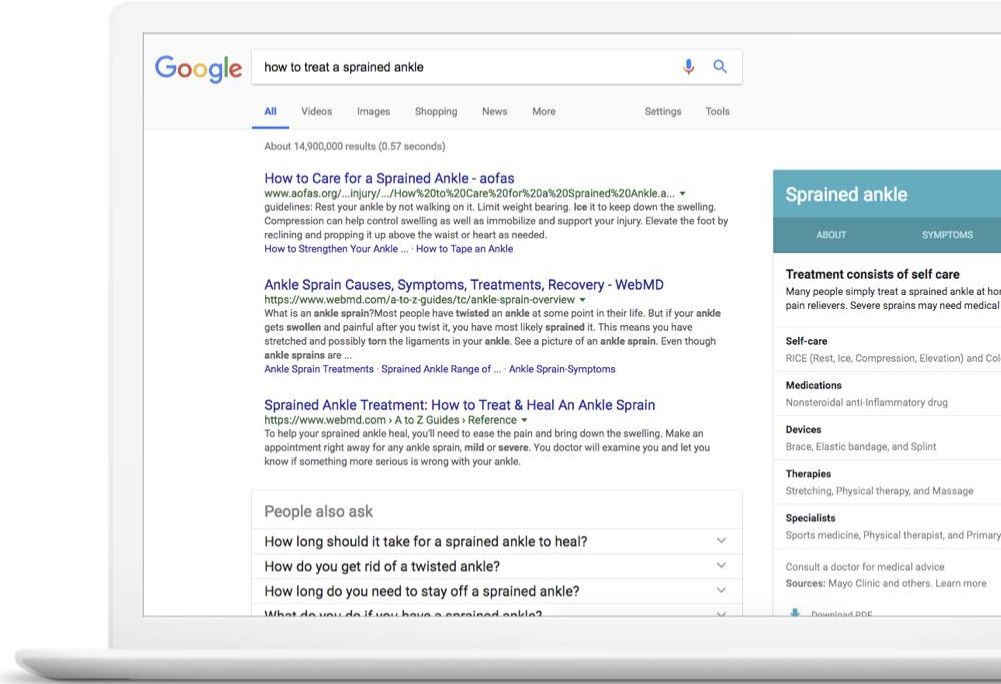
We are expanding fact checks to health information

The National
Academies of

SCIENCES
ENGINEERING
MEDICINE



Memorial Sloan Kettering
Cancer Center.



Collaboration with National Academy of Sciences, Based On
Science

Outline

ClaimReview Schema

Knowledge Card for publishers

Finding related and relevant stories

Automatic fact checking

Can we spot all claims and automatically check them?

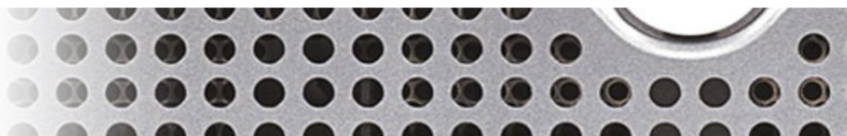
End-to-End FactChecking

ClaimBuster, <http://idir-server2.uta.edu/claimbuster/>

Cornell/Google Automated FactChecking collaboration

<http://www.cs.cornell.edu/database/checking/>

Verifying Text Summaries of Relational Data Sets, SIGMOD
2019, Saehan Jo, Immanuel Trummer, Weicheng Yu, Xuezhi
Wang, Cong Yu, Daniel Liu, Niyati Mehta



Homepage

People

Publications

Education

DB Seminar

Projects

Query Optimization

Data Vocalization

Fact Checking

Text Mining

Past Projects

Youtopia

Data Privacy

Games

Grace

Spatial Indexing

Automated Fact Checking

Data sets are often summarized via natural language text documents. Examples include newspaper articles by data journalists, scientific papers summarizing experimental results, or business reports summarizing quarterly sales. A majority of the population never accesses raw relational data but relies on text summaries alone. In that context, the following question arises: how can we trust such summaries to be consistent with the data?

We are developing approaches for automated and semi-automated fact checking of data summaries to answer that question. A text document, together with an associated data set, form the input for fact checking. Our goal is to identify erroneous claims about the data in the input text. More precisely, we focus on text passages that can be translated into a pair of an SQL query and a claimed query result. A claim is erroneous if evaluating the query yields a result that cannot be rounded to the one claimed in text.

In our first project in this space, we have developed a "fact checker tool" that supports authors in producing accurate data summaries. The tool is similar in spirit to a spell checker: where a spell checker supports users in avoiding erroneous spelling and grammatical mistakes, the fact checker supports users in avoiding erroneous claims. We focus on a restricted class of claims that are at the same time common and error-prone. The fact checker translates text passages into equivalent SQL queries, evaluates them on a database, and marks up potentially erroneous claims. Users obtain a natural language explanation, summarizing the system's interpretation of specific text passages, and can easily take corrective actions if necessary. We have recently used this tool to identify erroneous claims in articles from several major newspapers, some of which had gone by unnoticed for years.

Try our [Fact Checker Online Demo](#)!



DEPARTMENT OF COMPUTER SCIENCE
CORNELL DATABASE GROUP

Fact Checker

VERIFYING NATURAL LANGUAGE SUMMARIES OF RELATIONAL DATABASES

0. VIDEO TUTORIAL

[\[Video Tutorial Link\]](#)

1. TRY ON TESTCASES

Select Testcase:

Please select ▼

Query Processing Strategy:

3. Query Merging + Caching ▼

VERIFY

UPLOAD YOUR OWN TEXT AND DATA



Fact Checker

VERIFYING NATURAL LANGUAGE SUMMARIES OF RELATIONAL DATABASES

0. VIDEO TUTORIAL

[\[Video Tutorial Link\]](#)

2. TRY ON YOUR OWN TEXT AND DATA

Please input text and upload a **csv** file. Then click on Verify to start the fact-checking process.

Input Text:

Choose File

No file chosen

VERIFY

Dataset Search, launched Sep 5, 2018



The Keyword

Latest Stories

Product Updates

Company News

SEARCH

Making it easier to discover datasets

Natasha Noy

Research Scientist, Google AI

Published Sep 5, 2018

In today's world, scientists in many disciplines and a growing number of journalists live and breathe data. There are many thousands of data repositories on the web, providing access to millions of datasets; and local and national governments around the world publish their data as well. To enable easy access to this data, we launched [Dataset Search](#), so that scientists, data journalists, data geeks, or anyone else can [find the data](#) required for their work and their stories, or simply to satisfy their intellectual curiosity.

Similar to how [Google Scholar](#) works, Dataset Search lets you find datasets wherever they're hosted, whether it's a publisher's site, a digital library, or an author's personal web page. To create Dataset search, we developed [guidelines for dataset providers](#) to describe their data in a way that Google (and other search engines) can better understand the

<https://toolbox.google.com/datasetsearch/>



Dataset Search for Tennis

Google Dataset Search

atp tennis



About



Feedback

15 results found



Tennis ATP Results since 2001

public.opendatasoft.com

Updated Oct 17, 2016



ATP Tennis Dataset

www.kaggle.com

Updated Aug 6, 2017



ATP Men's Tour

www.kaggle.com

Updated Sep 27, 2016



A large Tennis dataset for ATP and
ITF betting

www.kaggle.com



ATP Tennis Dataset

All pro tour matches compiled with multiple features from Jan. 2012 to Jul. 2017



Kaggle

275 scholarly articles cite this dataset ([View in Google Scholar](#))

Dataset updated Aug 6, 2017

Authors

V.A. Freeman

License


[CC BY-SA 4.0](#)


Available download formats from providers


zip (418732 bytes)

Description

ATP Tennis


 Dataset






 19

ATP Tennis Dataset

All pro tour matches compiled with multiple features from Jan. 2012 to Jul. 2017

 V.A. Freeman · updated 2 years ago (Version 9)

[Data](#) [Kernels \(1\)](#) [Discussion](#) [Activity](#) Download (409 KB) [New Kernel](#) 

 CC BY-SA 4.0  tennis

Description

Context

Need to maintain a database of ATP Tour data for forecasting and predictions, etc.

Content

Retrieved data from online, but compiled file from 2012 to 2017 through 07/20/2017. Note that there are errors in some 2012 dates, but I revised the initial dataset as most recent version. Also, the W_Odds and L_Odds columns are averages of Odds taken from several online gaming sites including BetOnline365 and others.

Most recent data is usually updated weeks prior to all major grand slam events.

Acknowledgements

Compiled all useful data from the following link: <http://tennis-data.co.uk/alldata.php>

File content

corinna-macbookpro\$ head AAA.csv

```
ATP,Tournament,Tournament_Int,Date,Series,Series_Int,Court,Court_Int,Surface,Surface_Int,Round,Round_Int,Best_of,Winner,Winner_Int,Player1,Player1_Int,Player2,Player2_Int,Player1_Rank,Player2_Rank,Player1_Odds,Player2_Odds,Player1_Implied_Prob,Player2_Implied_Prob

1,BrisbaneInternational,1.2757,42737,ATP250,2.9693,Outdoor,3.6494,Hard,4.4983,1stRound,5.1257,3,ThompsonJ.,6.7926,YmerE.,6.7633,ThompsonJ.,6.7926,160,79,3.5,1.29,0.2857,0.7752

1,BrisbaneInternational,1.2757,42737,ATP250,2.9693,Outdoor,3.6494,Hard,4.4983,1stRound,5.1257,3,MahutN.,6.9297,MahutN.,6.9297,RobertS.,6.9686,39,54,1.54,2.43,0.6494,0.4115

1,BrisbaneInternational,1.2757,42737,ATP250,2.9693,Outdoor,3.6494,Hard,4.4983,1stRound,5.1257,3,FerrerD.,6.3881,TomicB.,6.5792,FerrerD.,6.3881,26,21,1.77,2.01,0.565,0.4975

1,BrisbaneInternational,1.2757,42737,ATP250,2.9693,Outdoor,3.6494,Hard,4.4983,1stRound,5.1257,3,EdmundK.,6.8384,EdmundK.,6.8384,EscobedoE.,6.0929,45,141,1.37,3.01,0.7299,0.3322

1,BrisbaneInternational,1.2757,42737,ATP250,2.9693,Outdoor,3.6494,Hard,4.4983,1stRound,5.1257,3,DimitrovG.,6.5157,JohnsonS.,6.7032,DimitrovG.,6.5157,33,17,2.85,1.41,0.3509,0.7092

1,BrisbaneInternational,1.2757,42738,ATP250,2.9693,Outdoor,3.6494,Hard,4.4983,1stRound,5.1257,3,DonaldsonJ.,6.9054,DonaldsonJ.,6.9054,MullerG.,6.6176,105,34,2.64,1.47,0.3788,0.6803

1,BrisbaneInternational,1.2757,42738,ATP250,2.9693,Outdoor,3.6494,Hard,4.4983,1stRound,5.1257,3,GrothS.,6.9961,HerbertP.H.,6.9637,GrothS.,6.9961,78,180,1.61,2.27,0.6211,0.4405

1,BrisbaneInternational,1.2757,42738,ATP250,2.9693,Outdoor,3.6494,Hard,4.4983,1stRound,5.1257,3,PouilleL.,6.8045,PouilleL.,6.8045,SimonG.,6.6845,15,25,1.83,1.94,0.5464,0.5155

1,BrisbaneInternational,1.2757,42738,ATP250,2.9693,Outdoor,3.6494,Hard,4.4983,1stRound,5.1257,3,TroickiV.,6.8786,NishiokaY.,6.8975,TroickiV.,6.8786,100,29,2.57,1.49,0.3891,0.6711
```



DEPARTMENT OF COMPUTER SCIENCE
CORNELL DATABASE GROUP

Fact Checker

VERIFYING NATURAL LANGUAGE SUMMARIES OF RELATIONAL DATABASES

2. TRY ON YOUR OWN TEXT AND DATA

Please input text and upload a **csv** file. Then click on Verify to start the fact-checking process.

Input Text:

DjokovicN. won 200 matches

Choose File AAA

VERIFY

Dublin Parks

Verified Text:

DjokovicN. won ? 200 matches



FACT
CHECKER

of Claims: 0 / 2
(Verified / Total)

APPLY CHANGES

NEW INPUT

Sentence:DjokovicN. won 200 matches

PICK THE QUERY THAT MATCHES TEXT

Query Description	(Result)
Count all elements where "Player1" = "djokovicn."	(208.00)
Count all elements where "Player1" = "djokovicn.", "Winner" = "djokovicn."	(183.00)
Count all elements where "Player2" = "djokovicn."	(221.00)
Count all elements	(14735.00)
Count all elements where "Winner" = "djokovicn.", "Player2" = "djokovicn."	(197.00)
Count distinct elements in "Date" where "Player1" = "djokovicn."	(208.00)
Count distinct elements in "Player2_Odds" where "Player1" = "djokovicn."	(192.00)
Sum over "Player1_Implied_Prob" where "Player1" = "djokovicn."	(185.05)
Count all elements where "Winner" = "djokovicn."	(380.00)
Sum over "Player1_Odds" where "Player1" = "djokovicn."	(241.13)

Thank you, and please join the effort!