



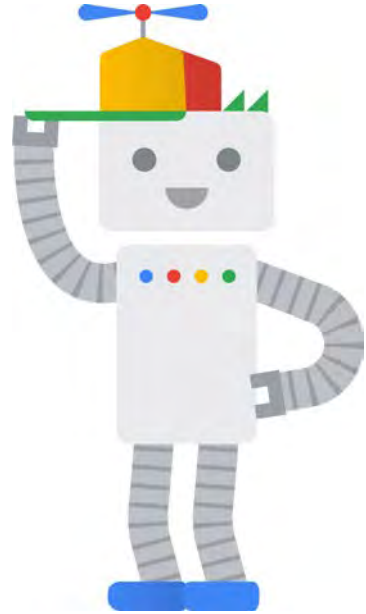
Webmaster Conference

Product Summit

November 4, 2019 | Mountain View, CA



Welcome



Lightning talks

Structured Data

Search Console

Web Deduplication

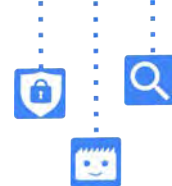
Google Images

Rendering

Titles & Snippets/Result Previews

Googlebot & Web Hosting

Claiming Your Knowledge Panel



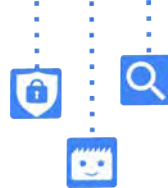
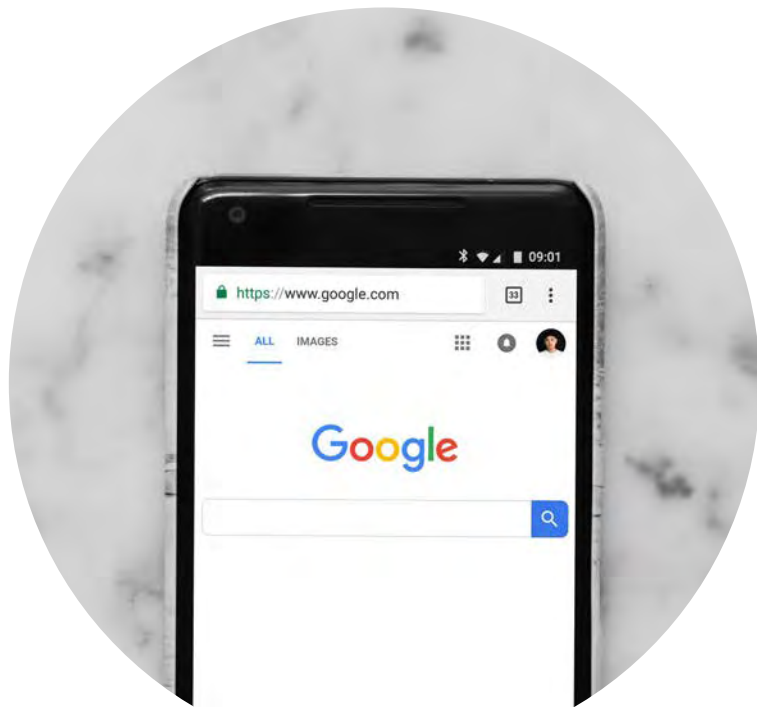
Session

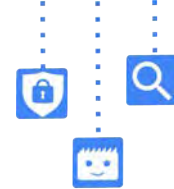
Improving Search Over the Years



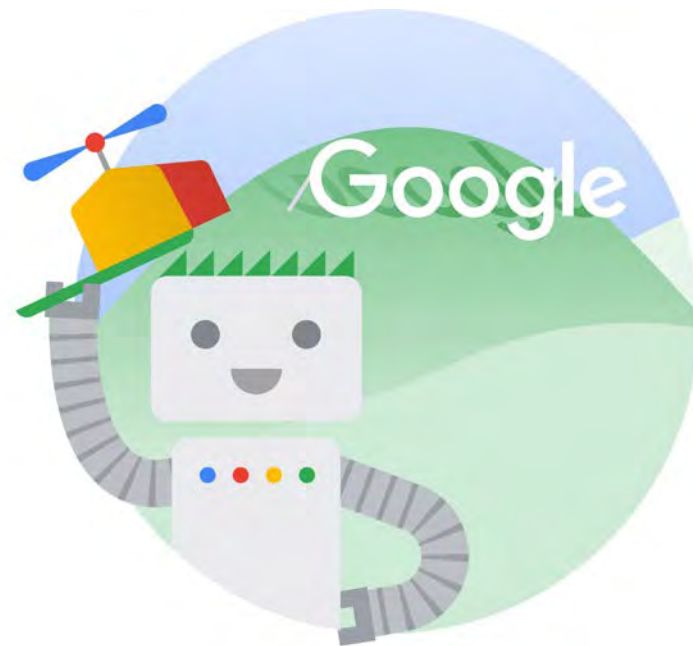
Paul Haahr

Distinguished Engineer, Search

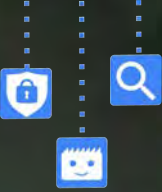




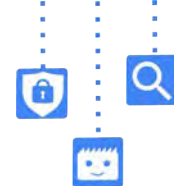
Why this event?



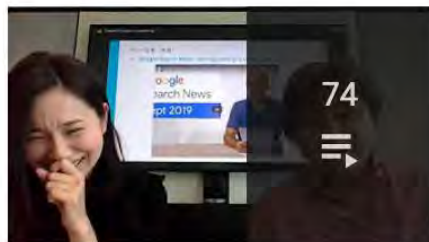
We do get lots of feedback...



We do get lots of feedback...



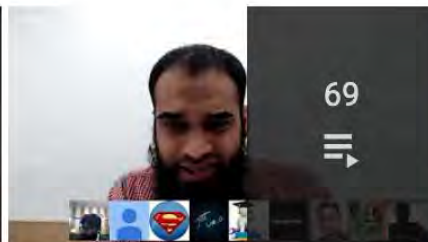
Webmaster Office Hours



Japanese Hangout on Air

Google Webmasters ✓

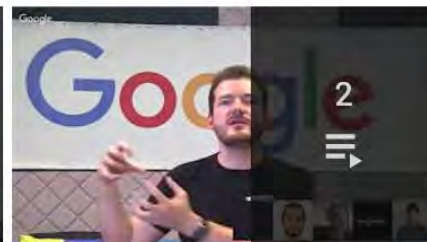
[VIEW FULL PLAYLIST](#)



English/Hindi Hangout on Air

Google Webmasters ✓

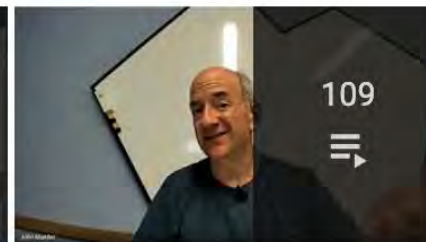
[VIEW FULL PLAYLIST](#)



L'Heure des Webmasters - en Français !

Google Webmasters ✓

[VIEW FULL PLAYLIST](#)



German Hangout on Air

Google Webmasters ✓

Updated yesterday

[VIEW FULL PLAYLIST](#)

And we
really do
think about
webmasters
& content
owners
constantly...



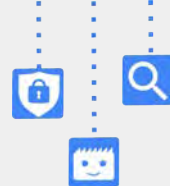
Danny Sullivan ✓

@dannysullivan

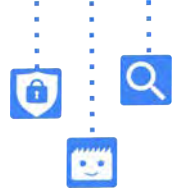
Replying to @dannysullivan and @CyrusShepard

Support of the ecosystem is constantly raised in meetings I'm in. It always comes up. It is a front-line concern with everyone involved with search. Any feature you see, impact on ecosystem has been considered. The hope is that overall, as Google grows, so does the ecosystem....

7:07 AM · Feb 24, 2019 · [Twitter Web Client](#)



An example of that, the **origin** of this event:



“We're interested in organizing a ‘meet the ecosystem’ initiative for the core search PM team, and Emily thought you all would be good thought partners to kick this off.”

We know
there can be
concerns,
but our future
is **with you**.



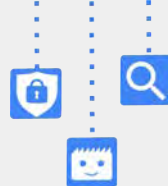
Danny Sullivan ✓

@dannysullivan

Replying to [@CyrusShepard](#)

Since I got asked about this, a couple of things.

Most important, the future of Google Search is to continue supporting the ecosystem. We don't thrive & users don't thrive unless the ecosystem thrives.



But both
our futures
depend on
search
**continuing
to evolve...**

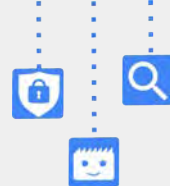


Danny Sullivan 

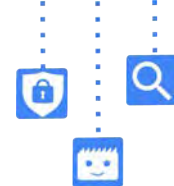
@dannysullivan

Replying to [@dannysullivan](#) and [@CyrusShepard](#)

For everyone to grow, search has to keep evolving.



The world of search has evolved from this...





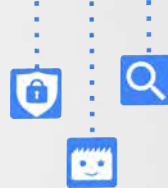
To this.

Traffic could
be one measure
of ecosystem
health as
search evolves.

Google sends
over 24 billion
visits per month
to news publisher
websites.

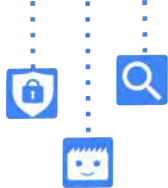
24
billion
visits/month

Google

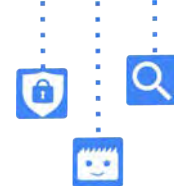


Google Search

I'm Feeling Lucky

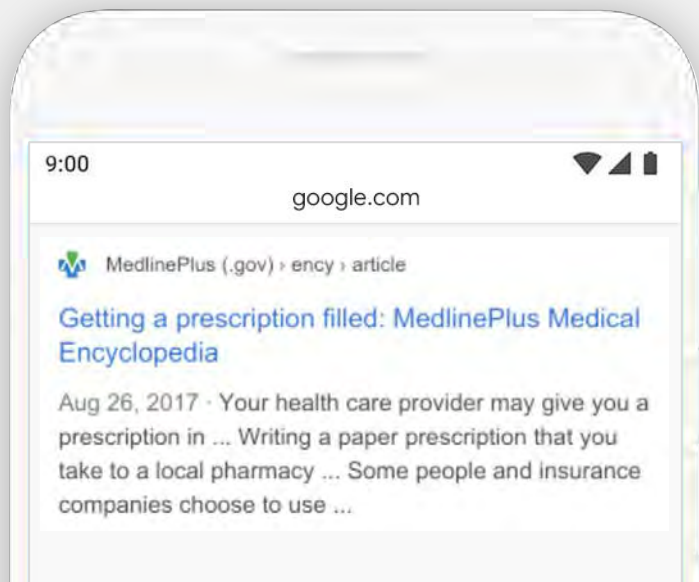


But measuring in
pure traffic can be imperfect.

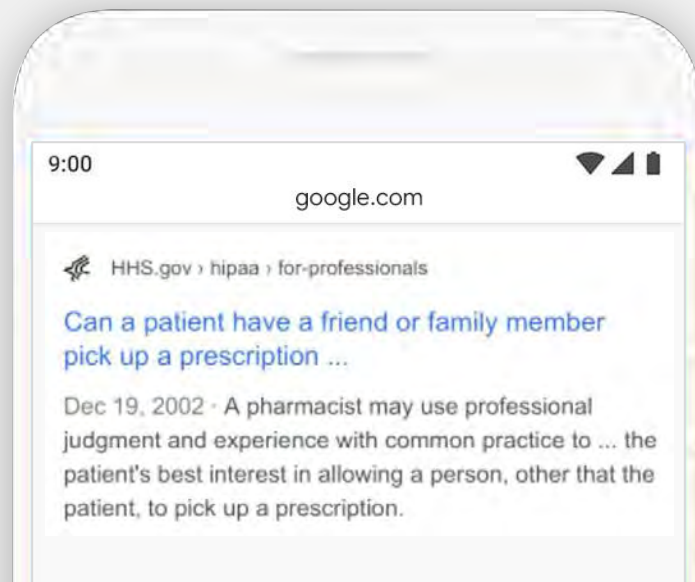


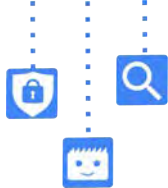
Can you get medicine for someone pharmacy

Before Bert



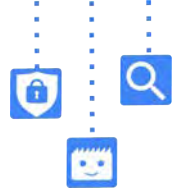
After Bert, better page listed





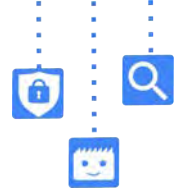
Traffic lost —
but was it **useful traffic**
in the first place?

Feedback on...



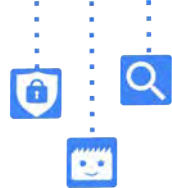
What are ways Google
should consider assessing
how it supports the
open web & ecosystem?

Feedback on...



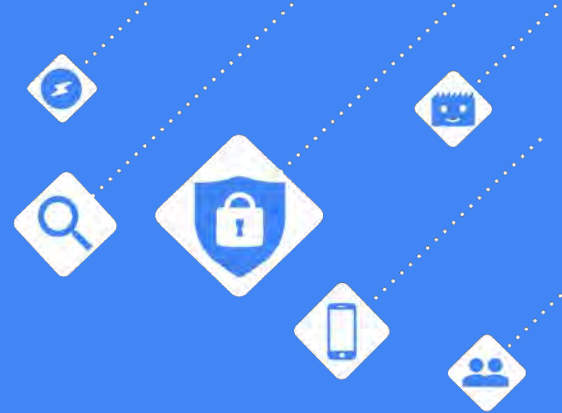
What are ways Google
can improve support **overall**?

Feedback on...



Anything that's on your minds!

Lightning talks



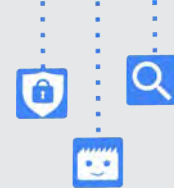
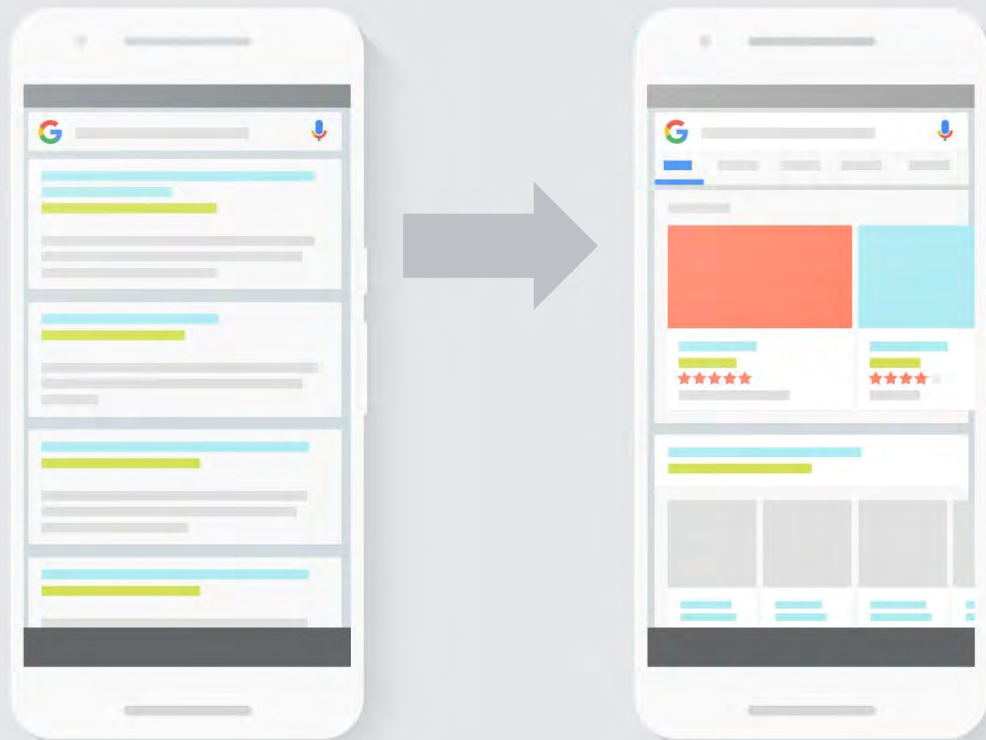
Structured Data



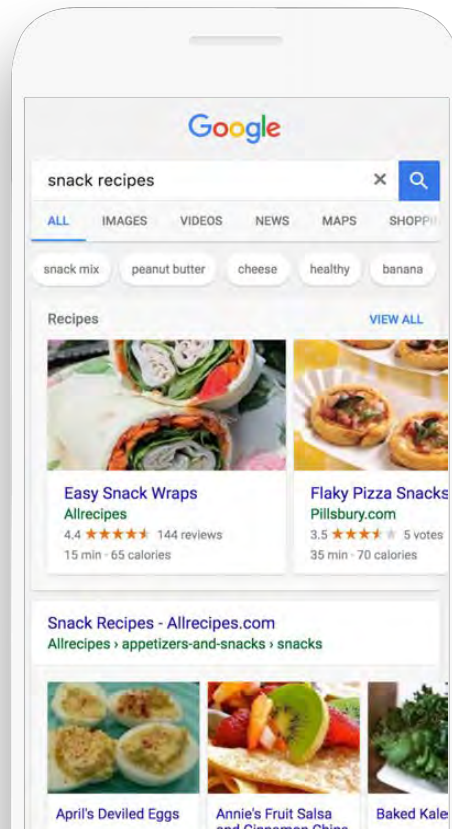
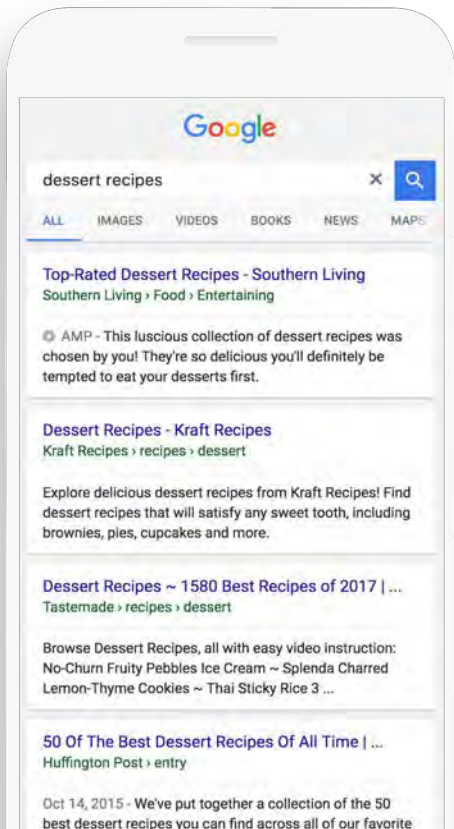
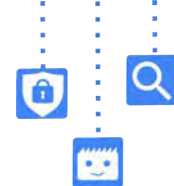
Allen Harvey
Product Manager, Search



Evolution of Search



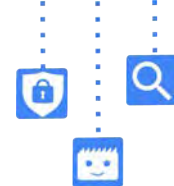
For example, recipe queries



Powered by
structured data!



Example HowTo structured data



```
1 <html>
2 <head>
3 <title>How to tie a tie</title>
4 <script type="application/ld+json">
5 {
6   "@context": "http://schema.org",
7   "@type": "HowTo",
8   "image": {
9     "@type": "ImageObject",
10    "url":
11      "https://example.com/1x1/photo.jpg"
12    },
13   "name": "How to tie a tie",
14   "description": "The four-in-hand knot is a great look for any occasion. From formal suits to casual jacket and tie affairs, it's a simple way to quickly look great. Once you have mastered the knot you can tie it in just a minute, so it's also great when you're in a pinch too. ",
15   "step": [
16     {
17       "@type": "HowToStep",
18       "name": "Preparations",
19       "text": "Button your shirt how you'd
```

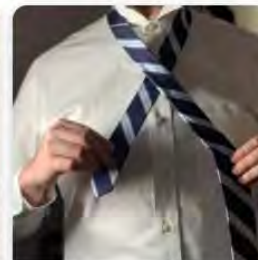
How to Tie a Tie

[Example.com](#) › [fashion](#) › [neckwear](#)

The four-in-hand knot is a great look for any occasion. From formal suits to casual jacket and tie affairs, it's a simple way to quickly look great. Once you have mastered the knot you can tie it in just a minute, so it's also great when you're in a pinch too.



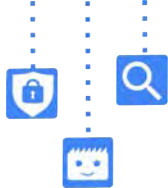
1. Make one end about your neck...



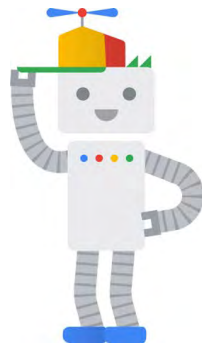
2. Cross the short end over the...



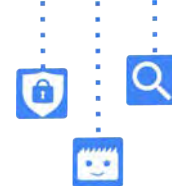
3. C you



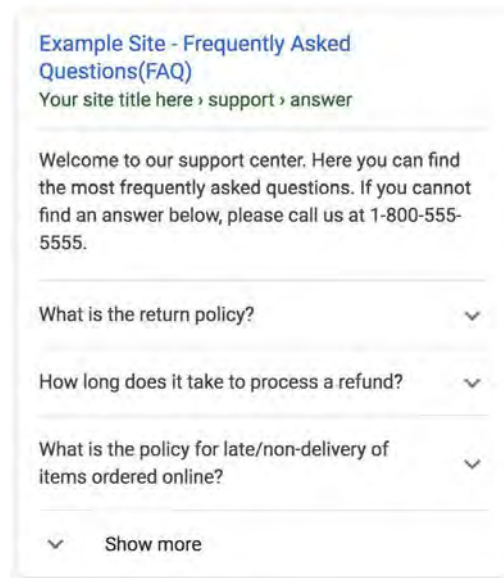
Structured data helps
Google understand
your page and enable
special features.



Makes your pages eligible for Rich Results

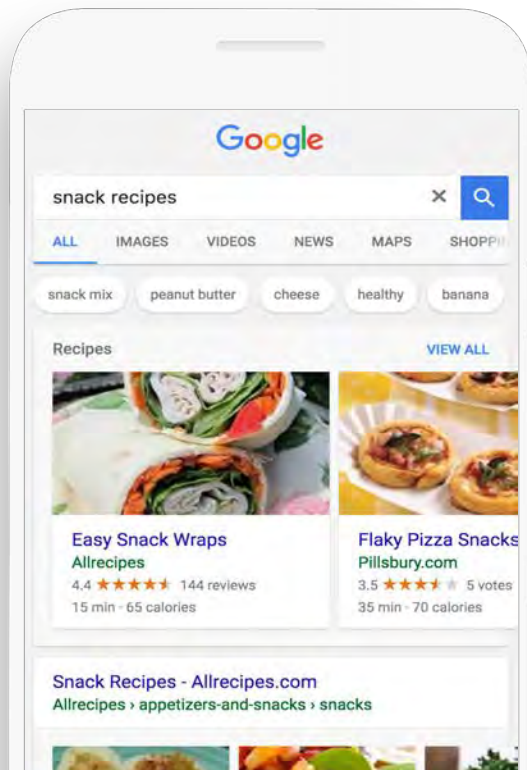
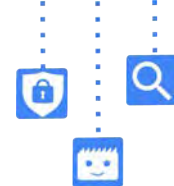


Q&A



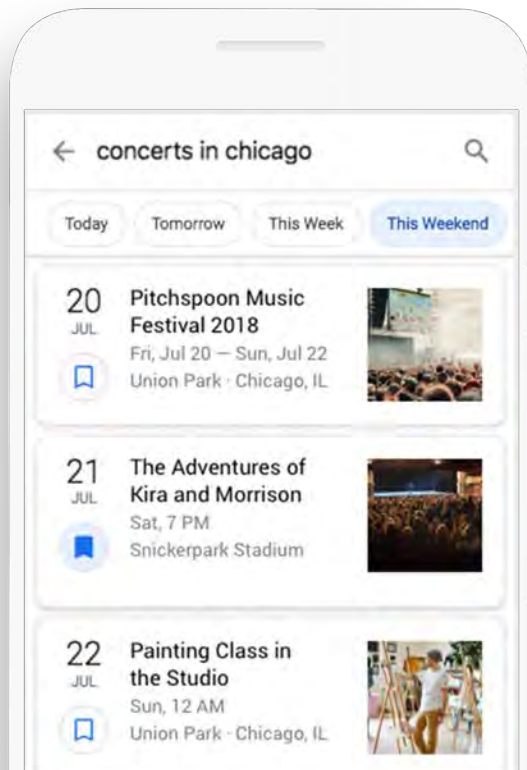
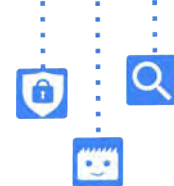
FAQ

Eligible to appear in **carousels** on pages



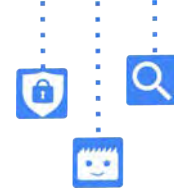
Recipes

Eligible to appear in search experiences



Events

In summary, makes your pages **eligible** to...



1

Appear as
Rich Results

2

Show in
page-level
features

3

Be visible
in Search
experiences

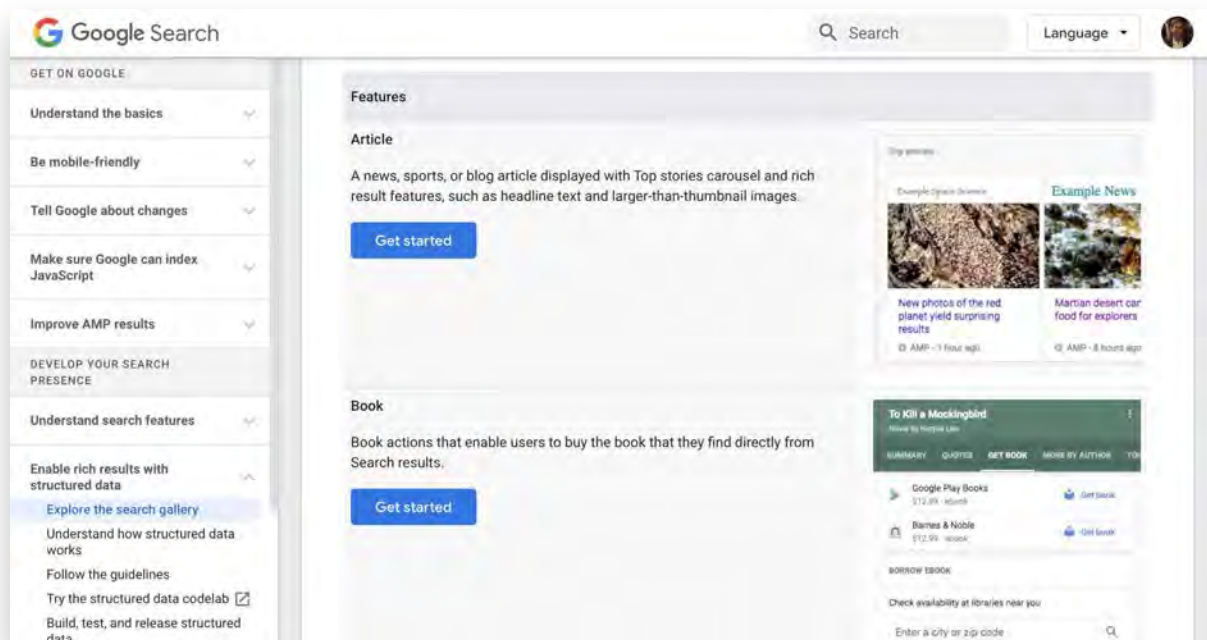
4

Work on the
Assistant

How do I get started?



Find the
structured data
feature most
suitable for
your site in our
Search gallery



goo.gle/searchgallery

View the
example markup
and add your
own content

JSON-LD

Microdata

Here's an example of a how-to page with images for each step using JSON-LD. When there's an image for each step, users may see a preview with a carousel of images.

[See Markup](#)

Does your page support rich results?

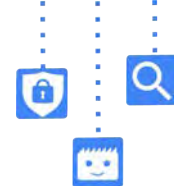
URL

CODE

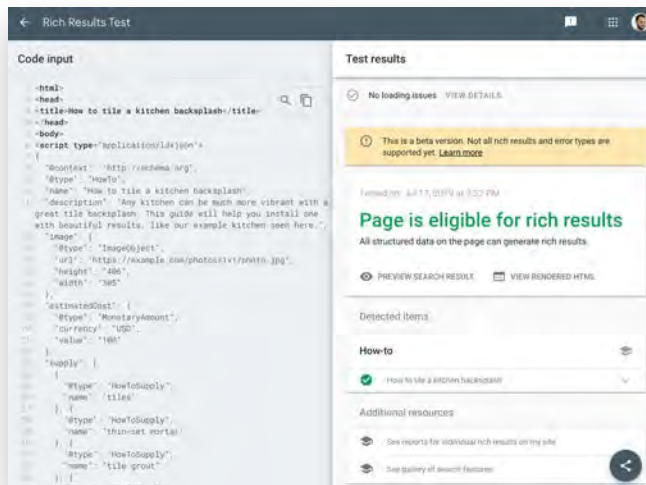
```
1 <html>
2 <head>
3 <title>How to tie a tie</title>
4 <script type="application/ld+json">
5 {
6   "@context": "http://schema.org",
7   "@type": "HowTo",
8   "image": {
9     "@type": "ImageObject",
10    "url": "https://example.com/1x1/photo.jpg"
11  },
12   "name": "How to tie a tie",
13   "description": "The four-in-hand knot is a great look for any occasion. From formal suits to casual jacket and tie affairs, it's a simple way to quickly look great. Once you have mastered the knot you can tie it in just a minute, so
```

Googlebot smartphone

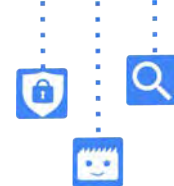
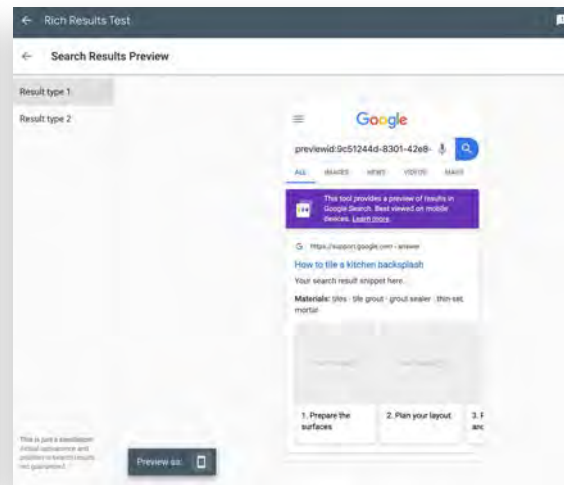
TEST CODE



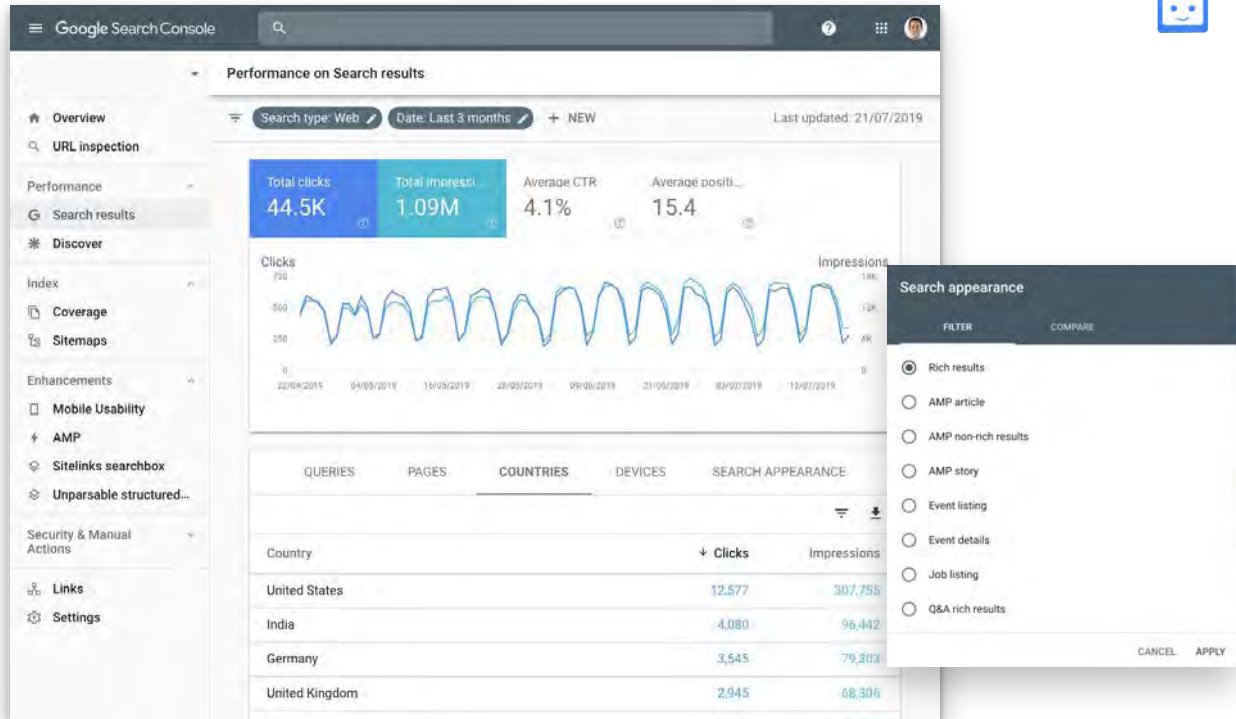
Test and Preview your Structured Data in Rich Results Test



goo.gle/richresults

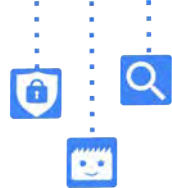


Measure performance in Search Console



g.co/searchconsole

Recap



Structured data helps Google understand your page and makes it eligible to appear in special features.

All structured data features are available at goo.gle/searchgallery, please take a look and find the one right for your site.

Rich Results Test and Search Console can help you get your markup right and monitor performance for your pages.

Thank you

Search Console



Hillel Maoz
Software Engineer



Ofir Roval
Product Manager



Search Console's mission:

Provide **data & tools** to help site owners improve their websites and optimize their appearance on **Google**

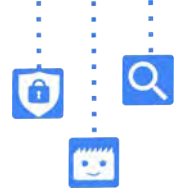


How do we ‘improve’ websites

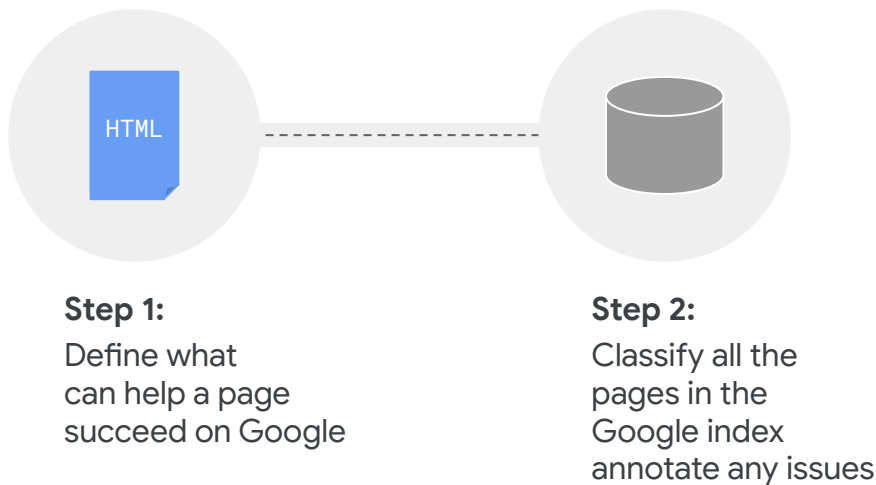
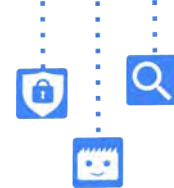


Step 1:

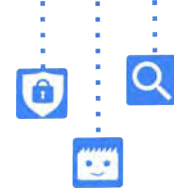
Define what
can help a page
succeed on Google



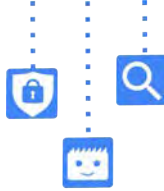
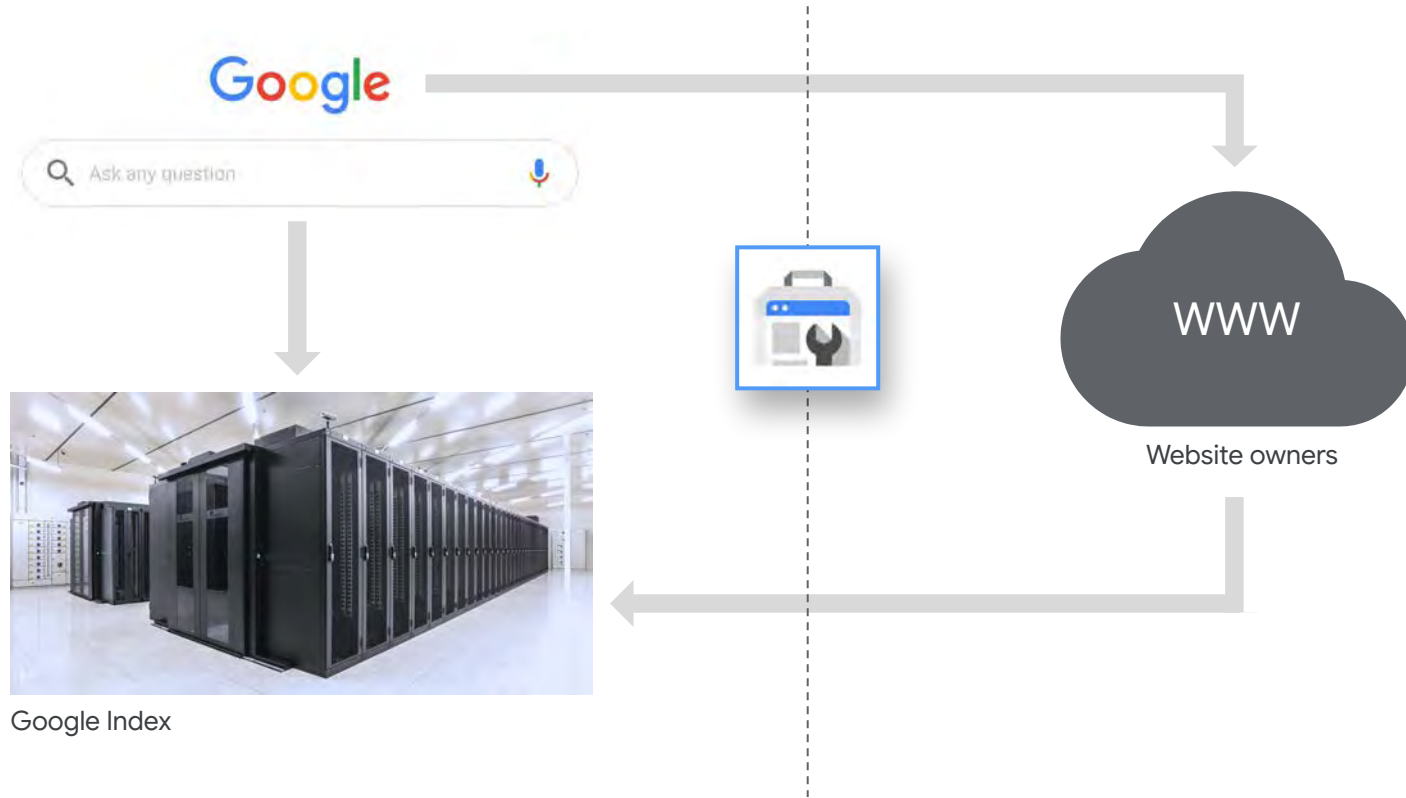
How do we 'improve' websites

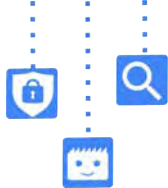


How do we 'improve' websites



How Search Console Works?

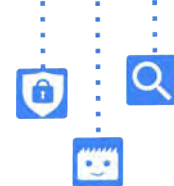




How Search Console Works?

Extract all relevant **information** from Google Search, report it to webmasters and **help them use it** to improve their site.

Search Console Components



Search Analytics

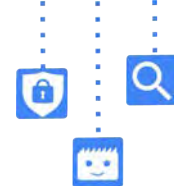
Reporting

Testing Tools

Accounts

Alerts

How does it work?



Search Analytics

- Track all impressions on Search
- Many billions of Search results daily
- 16 months of data
- Support slice and dice to help you find insights and trends

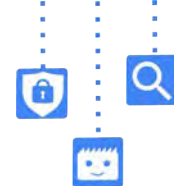
Google



Ask any question



How does it work?

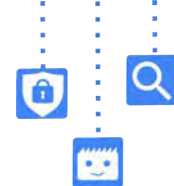


Reporting

- Track all pages crawled by Google
- Many millions of domains
- 100s of signals supporting dozens of Search features
- Actionable reports to help you fix any issues



How does it work?

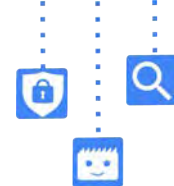


Testing Tools

- Trigger the entire Google index stack for a given URL
- On demand for any URL
- High fidelity to Search
- Support full debugging flow



How does it work?

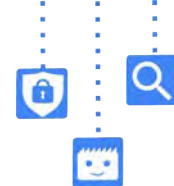


Accounts

- Manage all Search Console sites and owners
- Millions of sites / domains
- Protect website data via verification

WWW

How does it work?



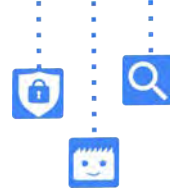
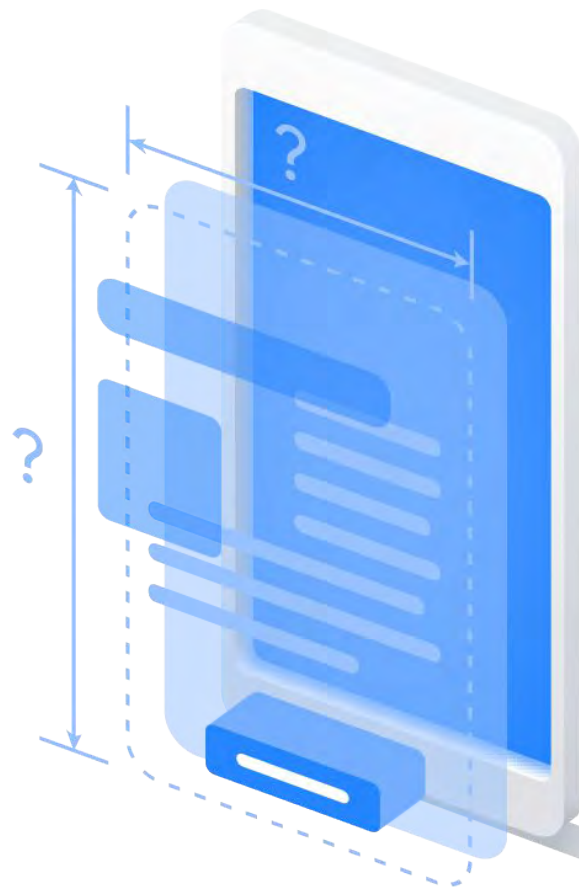
Alerts

- Understand over time site changes
- Send emails to all Search Console websites
- 10s of alert types
- Make sure you never miss any important change with your site

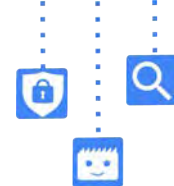
WWW

Launching today:

Speed Report on Search Console



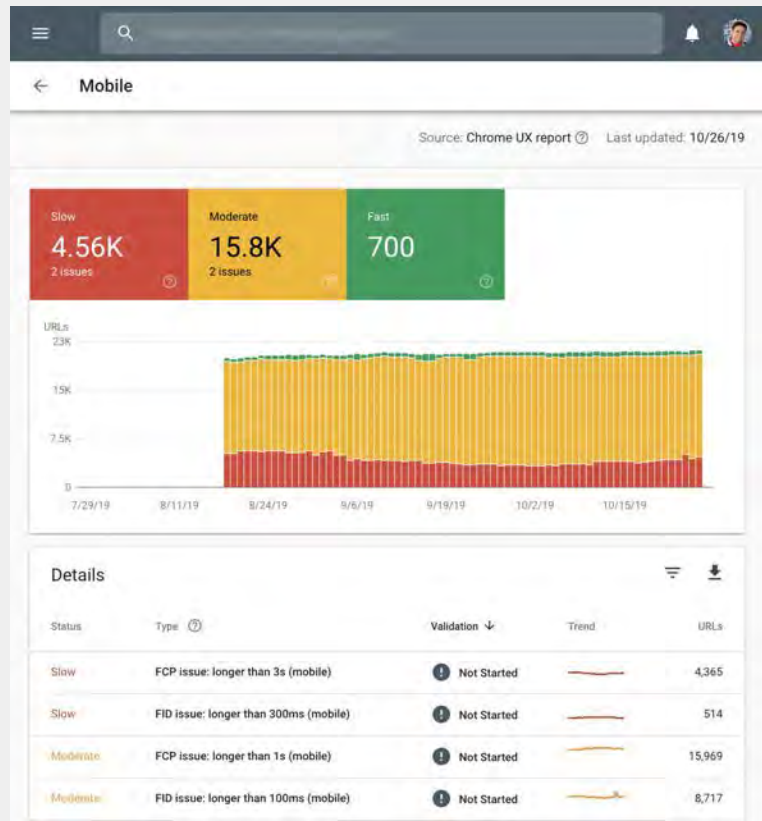
How we created the Speed report



Search Console Speed report

Get an overview on
your site performance:
slow, moderate, fast

Data is based on real user
experience as represented
by Chrome User Experience
Report (CrUX)

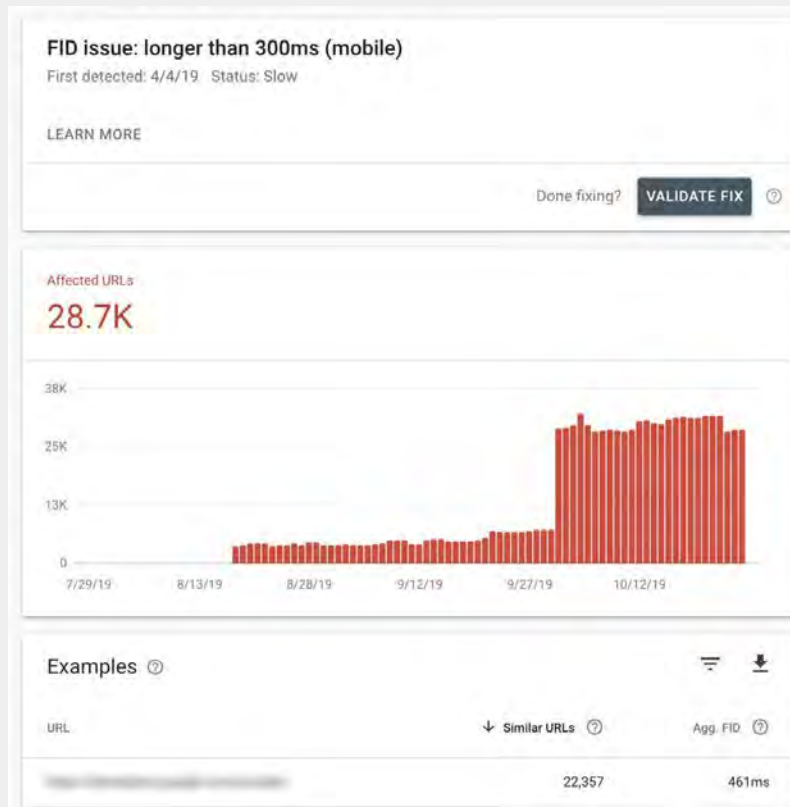
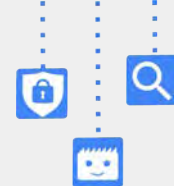


Identify Speed Issues

See examples of pages
with speed related issues

Use **Page Speed Insights** to
get optimization suggestions

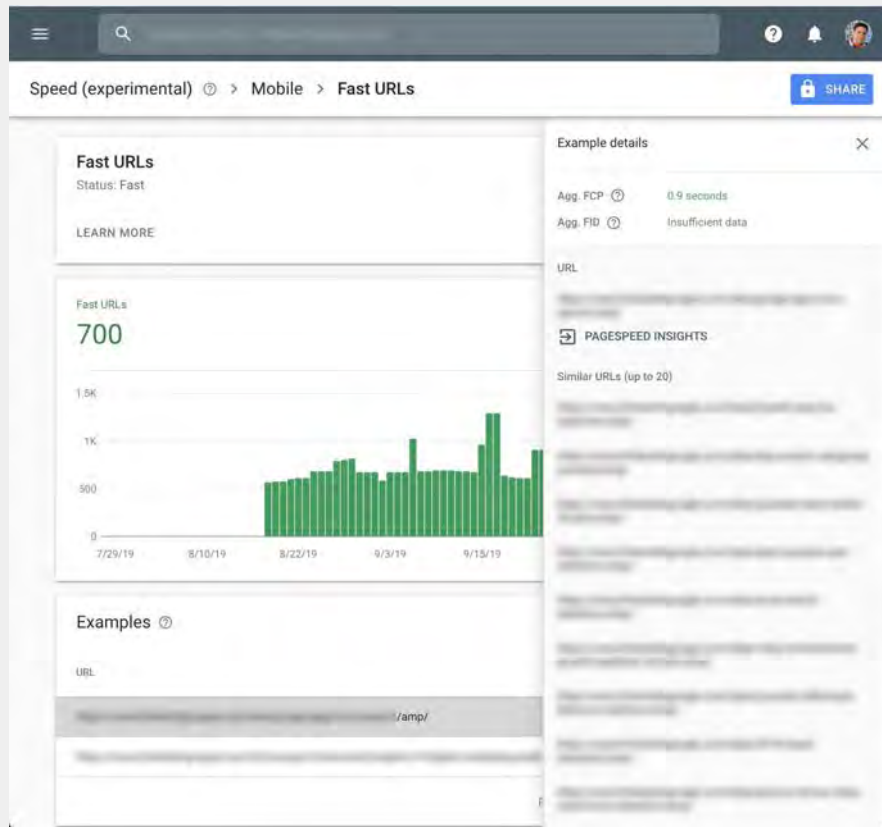
Validate speed fixes resulted
in improved user experience



Get positive examples

Identify what type of pages are performing well

Apply similar implementation to other sections of your site

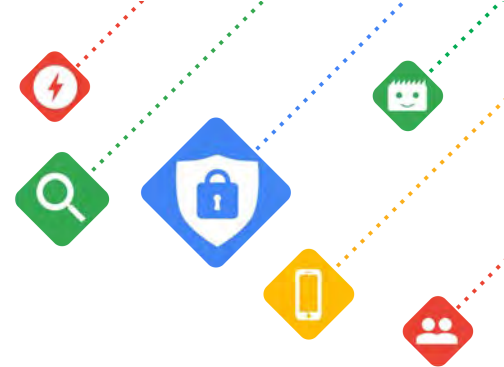


Thank you

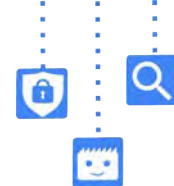
Web Deduplication



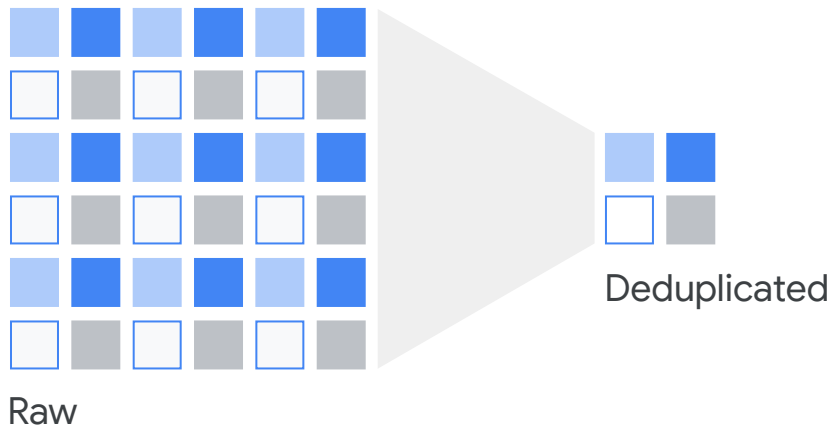
Allan Scott
Software Engineer



Deduplication

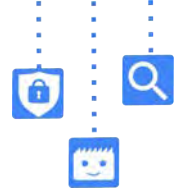


What is it?



1. Identify and cluster duplicate web pages.
2. Pick representative URLs, index unique pages.
3. Forward signals to representative URLs.

Deduplication



Why deduplicate?

1

Users don't want the "same" page repeated.

2

More room for distinct content in the index.

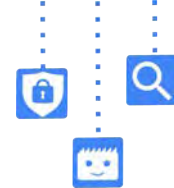
3

You retain signals when redesigning your site.

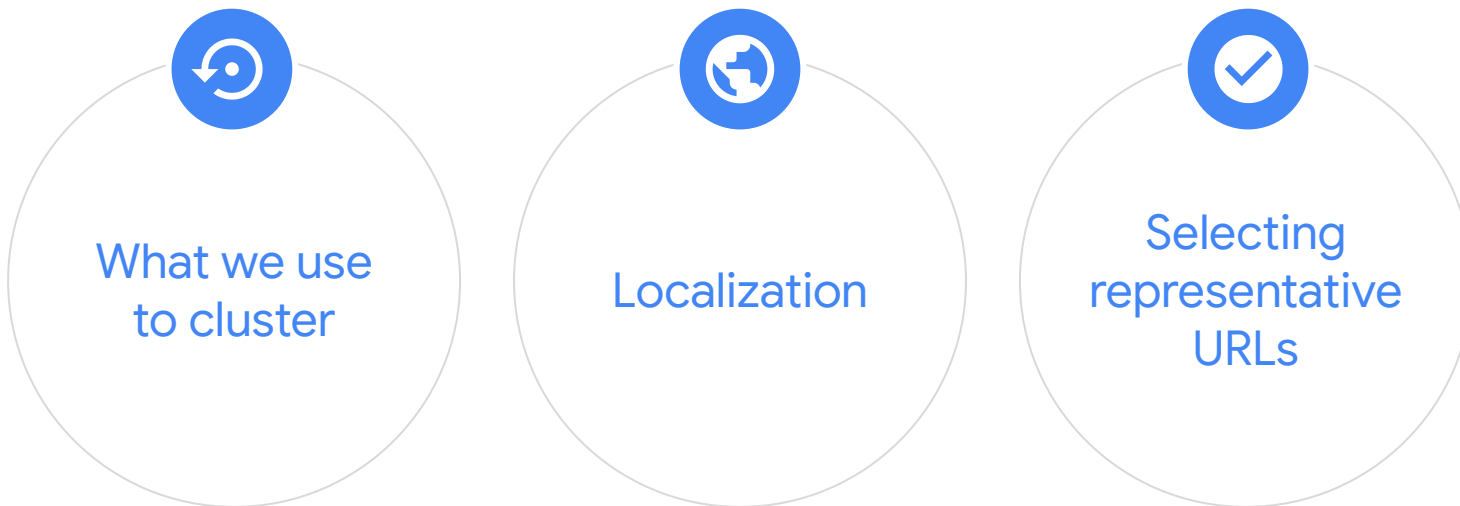
4

We can also find alternate names.

Deduplication



Overview

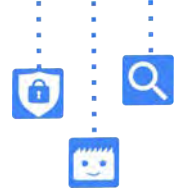


Clustering

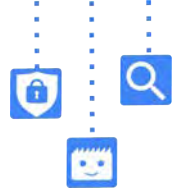


Signals we use

1. Redirects.
2. Content.
3. Rel=Canonical.
4. Etc...



Clustering

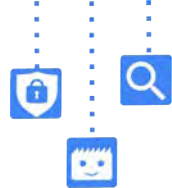


Redirects

We largely trust redirects, as they are almost perfectly predictive of duplicate content.

This is part of why we recommend using them when you move/redesign your sites.

Clustering



Content

We also use content checksums.

We make efforts to ignore boilerplate.

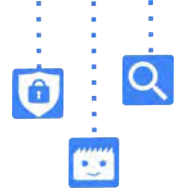
Catches many soft error pages.

This is why we prefer getting an HTTP error.

Clustering

Rel=Canonical

We use your annotations to cluster.
These tend to get more verification.
Thresholds are still intentionally loose.
Had more than a few “broken script” clusters.

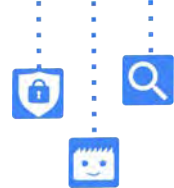


Localization



tl;dr version

1. When main content is the same, pages cluster.
2. This can include boilerplate-only localization.
3. Clever geo-redirecting often also clusters.
4. We use hreflang alternates to bridge this gap.

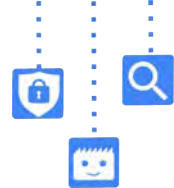


Canonicalization



Representative
URLs

1. To deduplicate we must pick a representative.
2. Candidates compete on weighted criterion.
3. Weights are trained by ML.

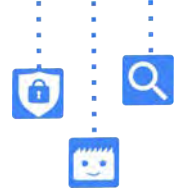


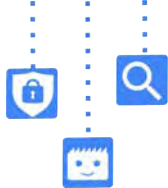
Canonicalization



Signals

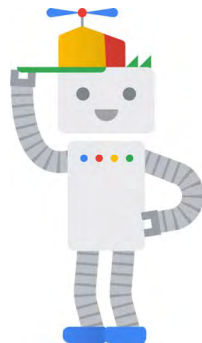
1. Hijacking is our overriding concern here.
 - Escalations via WTA forums are valuable.
2. Second concern is “user experience”:
Security, slow meta refresh, etc.
3. Third, webmaster signals:
Redirects, rel=canonical, sitemaps.





Some Suggestions

1. Use redirects to clue us into your site redesign.
2. Send us meaningful HTTP result codes.
3. Check your rel=canonical links.
4. Use hreflang links to help us localize.
5. Keep reporting hijacking cases to the forums.
6. Secure dependencies for secure pages.
7. Keep canonical signals unambiguous.



Thank you

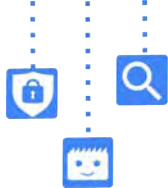
Google Images



Francois Spies

Product Manager: Google Images

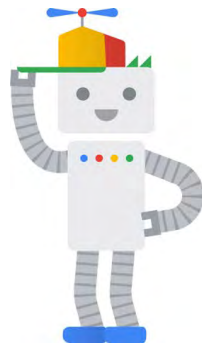


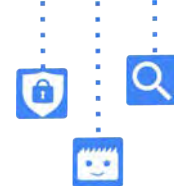


01 Product overview

02 Product changes

03 SEO best practices





Event planning



DIY



Work



School



Shopping



Visual Definition



Interior design



Medical



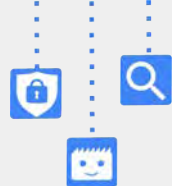
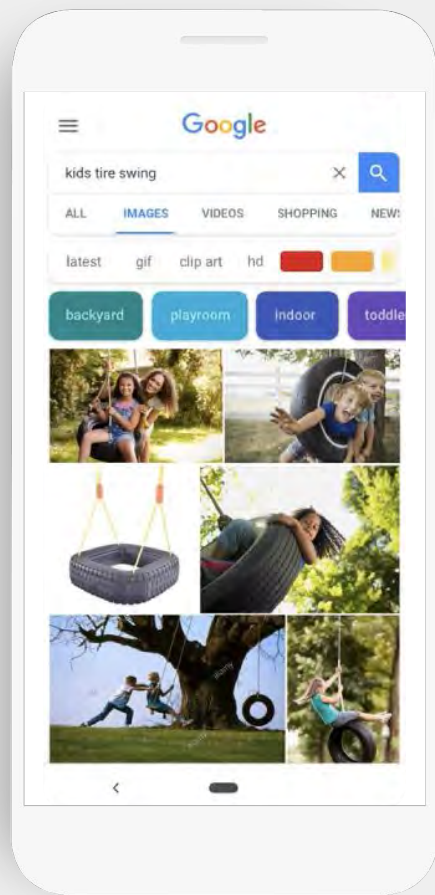
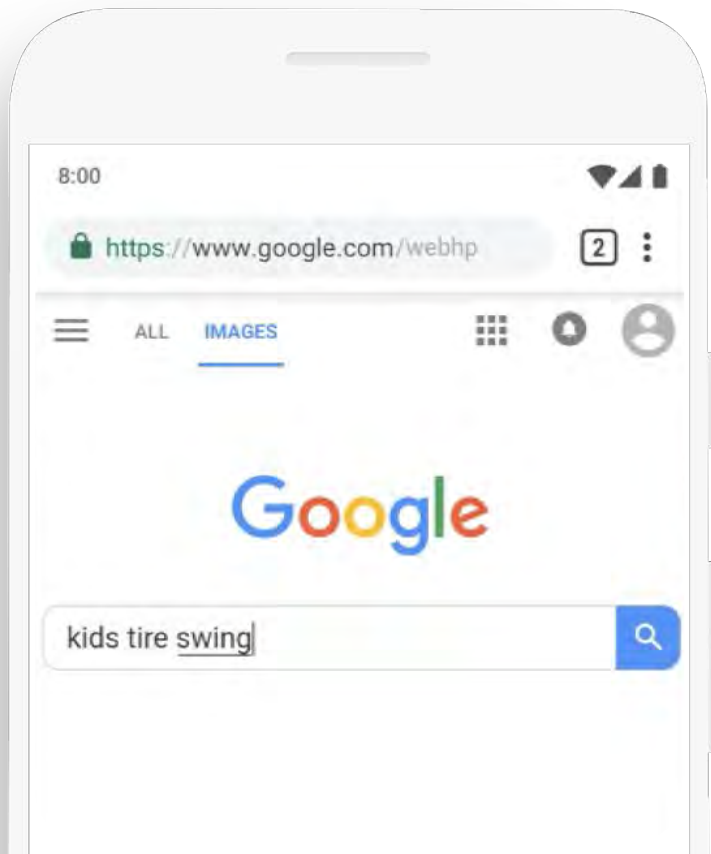
Communication

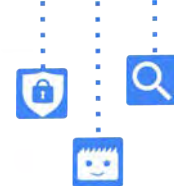
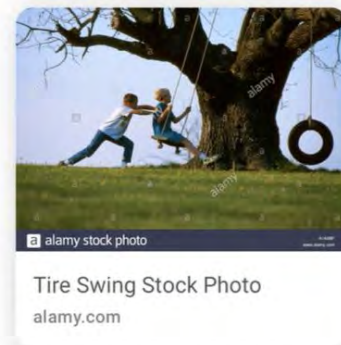
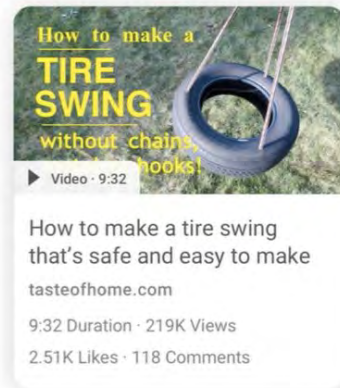
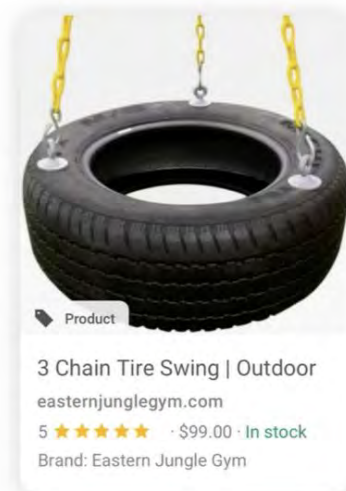
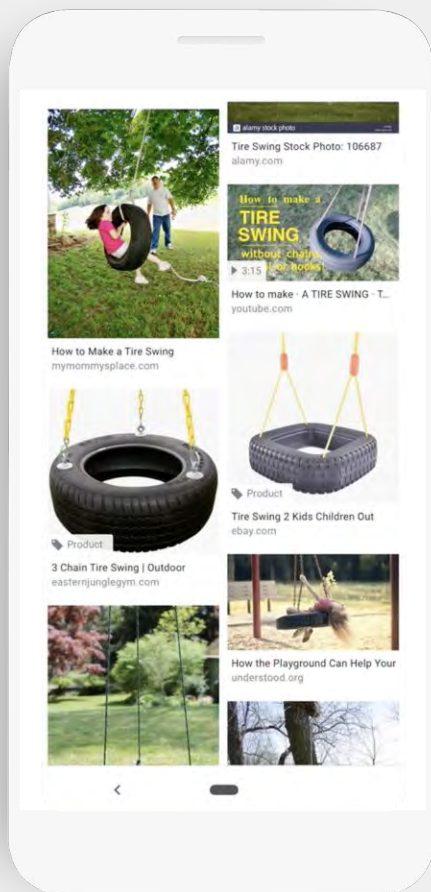


Visual Interest

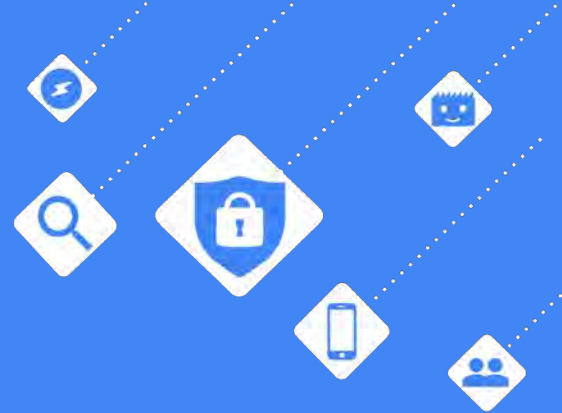


Fashion





Recent Product Changes



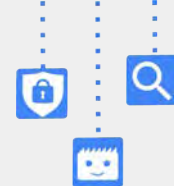
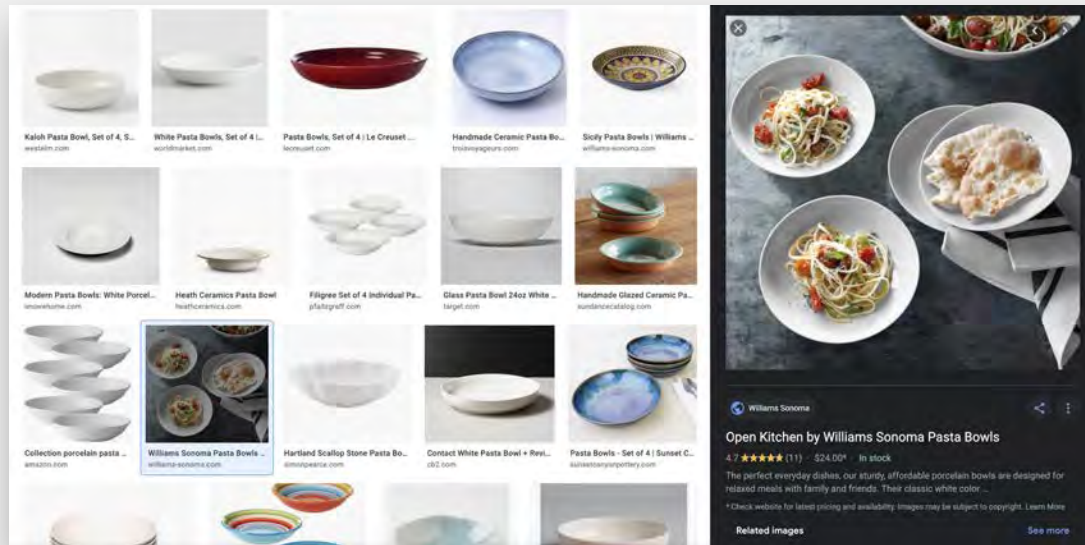
Product Updates

Images on Desktop

Viewer in side panel

More prominent
publisher context
alongside image

Align viewer format
across surfaces

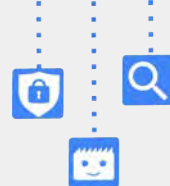
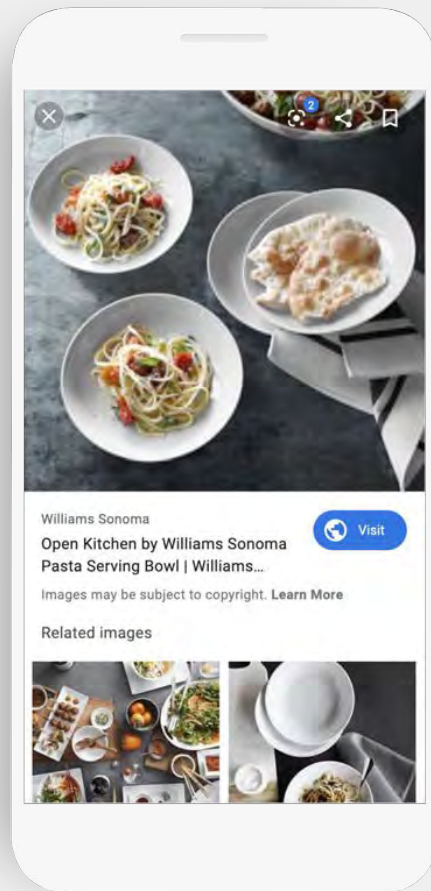


Product Updates

Images viewer

Aesthetic changes
to improve legibility

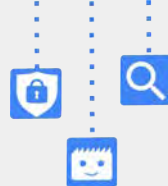
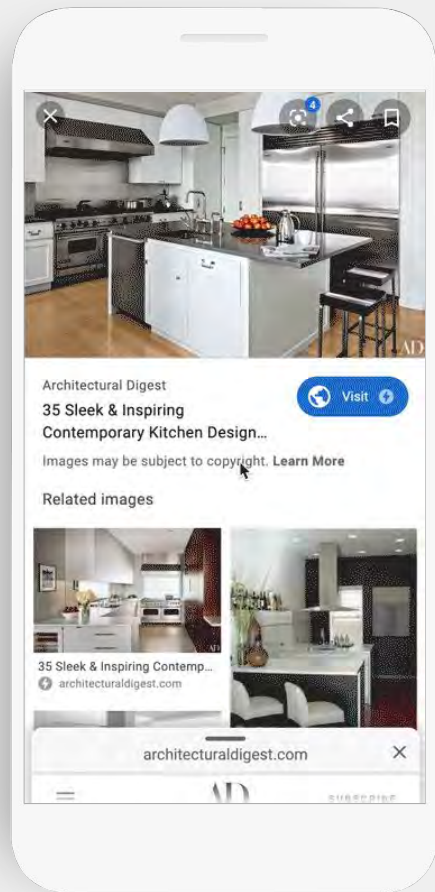
More prominent button
to visit page



Product Updates

Swipe to Visit

Reduce friction for users
to visit AMP web pages



Optimization Best Practices



Use structured data for rich results (especially for products, videos, recipes)



The Flavor Bender

Sri Lankan Chicken Curry

4.9 ★★★★★ (34) · 55 min · Yield: 6 people


Learn how to make authentic Sri Lankan chicken curry with these tips and variations for the best ...

Ingredients

- 2 - 3 tbsp coconut oil
- ½ onion (yellow or red)
- 4 garlic cloves (minced)
- 1 inch ginger (minced)
- 6 - 7 curry leaves or 2 bay leaves

Images may be subject to copyright. [Learn More](#)

Related images



Amazon.com

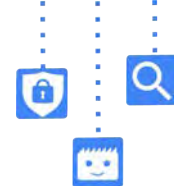

MICHELANGELO 8 Quart Stock Pot with 8 Quart Induction Pot, 8 Quart Stock Pot, Lid

4.5 ★★★★★ (21)

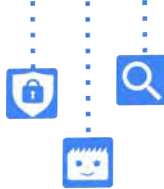
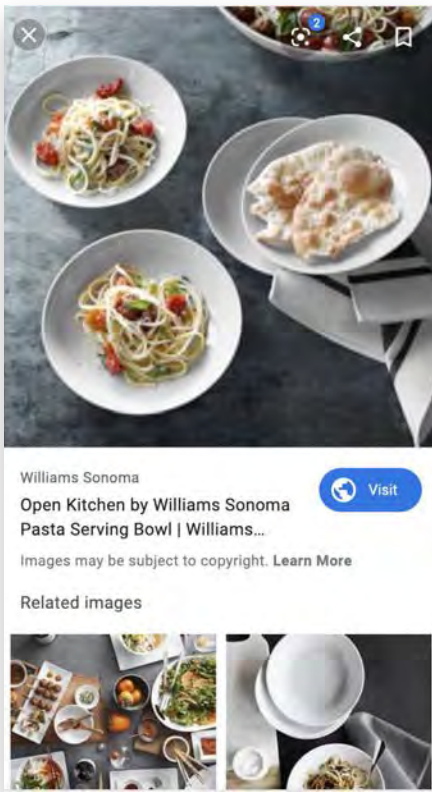
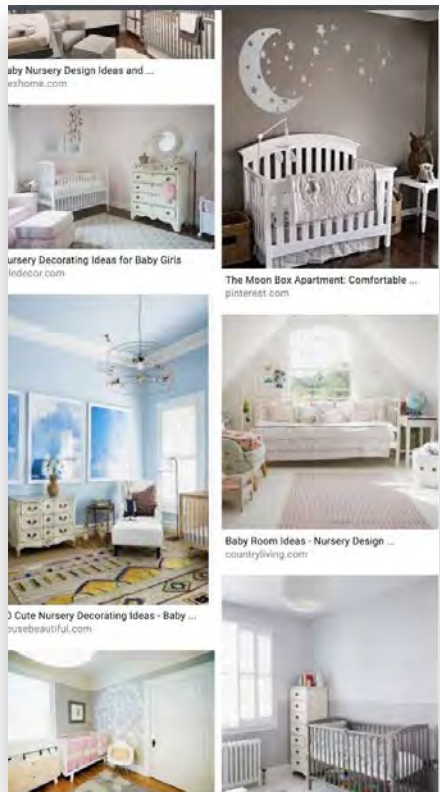
Stainless steel 8 Quart stock pot with lid history of the MICHELANGELO kitchenware: today, combined with superb craftsmanship, innovative design and ...

Images may be subject to copyright. [Learn More](#)

Related images

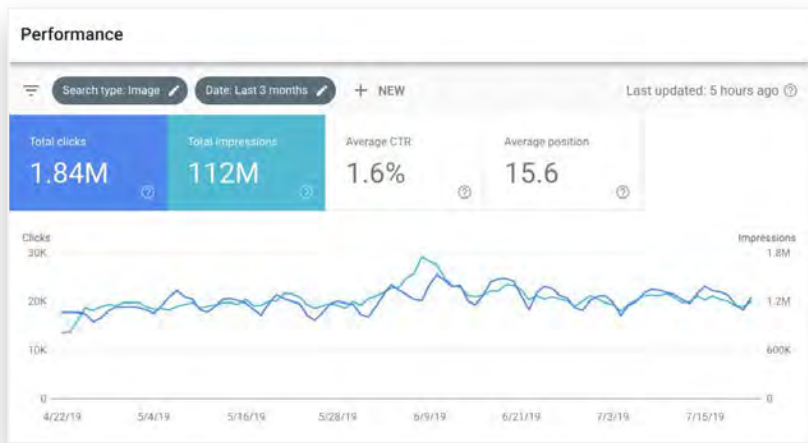


Use descriptive
titles, captions,
and filenames



Use high-quality
& optimized
images, placed
next to text,
on mobile-
friendly pages





In closing, a reminder that you are able to **check traffic** from Google Images within Google Search Console



Thank you

Rendering



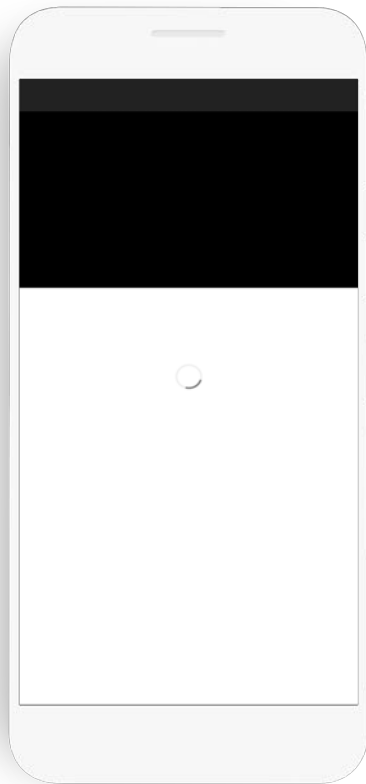
Erik Hendriks
Software Engineer



Rendering allows
Googlebot to see the
same web as users do.



Rendering

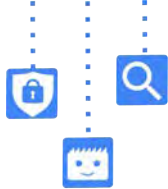


Before



After



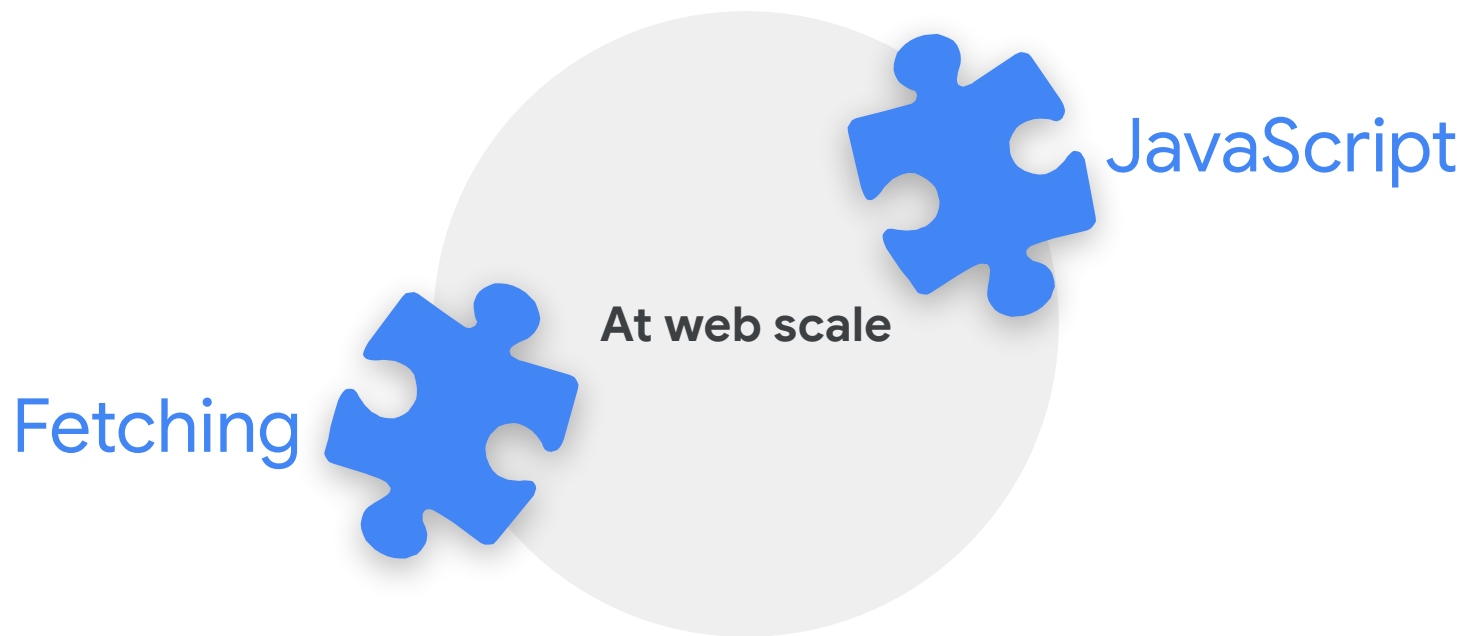
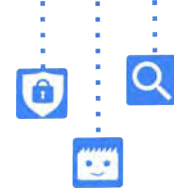


Googlebot renders with **Chrome**

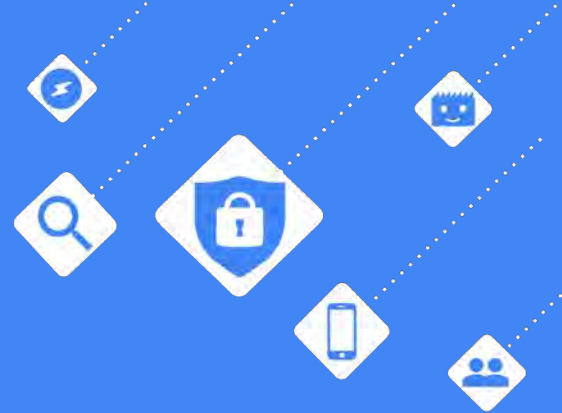
Challenges



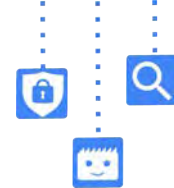
Challenges



Fetching



Fetching



Limited access
(robots.txt)

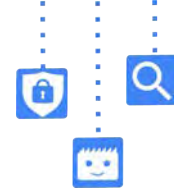
1

Differences

2

Limited
crawl volume

Fetching



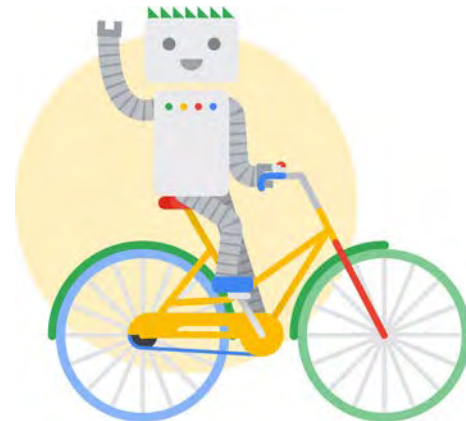
50-60

Resource fetches
per page

(obeying robots.txt)

60-70%

Cache Hit Rate

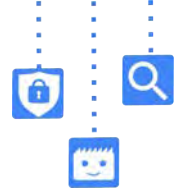


Fetching

20x



Fetching



Corners will be **cut**

1

We do not obey HTTP caching rules

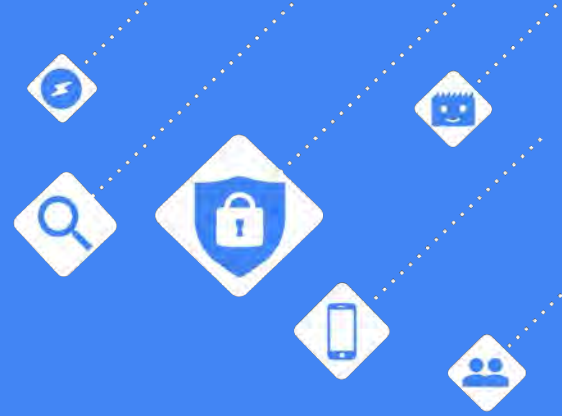
- Don't rely on clever caching

2

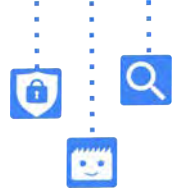
We might not fetch everything

- Minimize fetches
- Be resilient

JavaScript



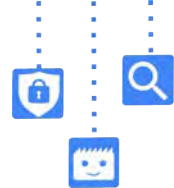
JavaScript



Performance matters!

- ✓ We limit CPU consumption
- ✓ We will interrupt script
- ✓ Excessive CPU consumption negatively impacts indexing

JavaScript



Popular ways to fail

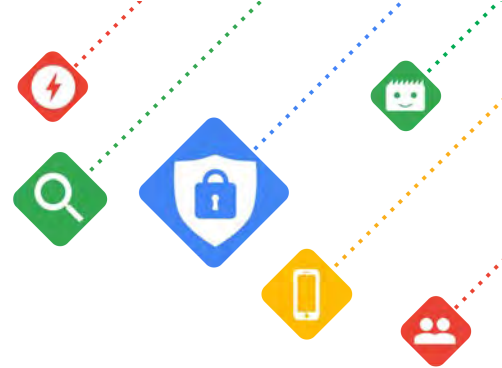
- ✗ Error loops (robots.txt, missing features)
- ✗ Cloaking
- ✗ Cryptocurrency miners

Thank you

Titles & Snippets/ Result Previews



Phiroze Parakh
Software Engineer



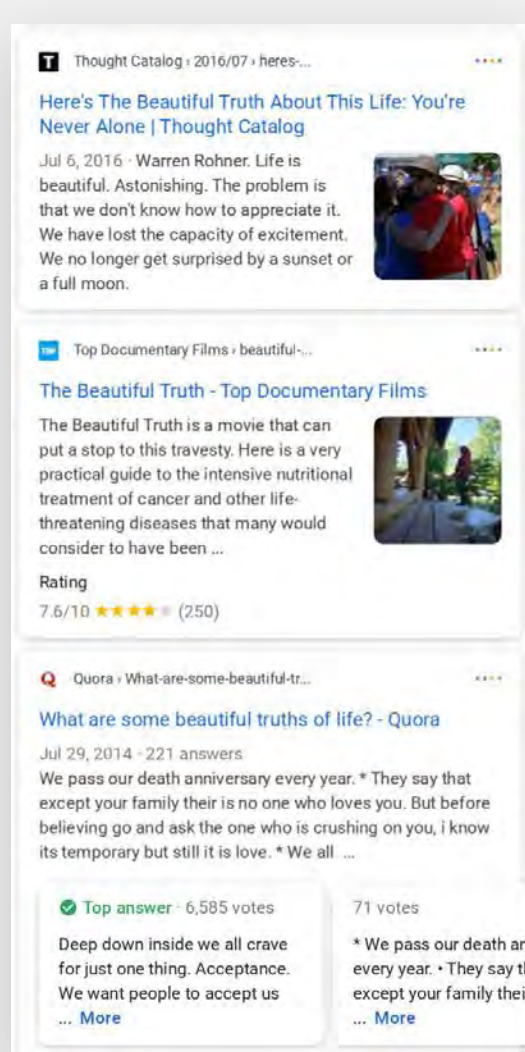
Result Preview Pillars

Relevance to user need

Help users choose between results

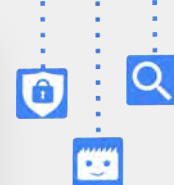
Illuminate depth of content
within results

Express diversity of content in
the ecosystem

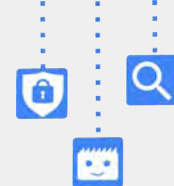


The screenshot displays three search results for the query "The Beautiful Truth".

- Result 1 (Thought Catalog):** A text-based result titled "Here's The Beautiful Truth About This Life: You're Never Alone | Thought Catalog". It includes a quote by Warren Rohner dated July 6, 2016, and a small image of two people.
- Result 2 (Top Documentary Films):** A video-based result titled "The Beautiful Truth - Top Documentary Films". It provides a synopsis of the film as a practical guide to cancer treatment and includes a rating of 7.6/10 based on 250 reviews.
- Result 3 (Quora):** A discussion-based result titled "What are some beautiful truths of life? - Quora". It shows a question asked on July 29, 2014, with 221 answers. The top answer, which has 6,585 votes, discusses the concept of death anniversaries and acceptance.



Web Result Architecture



Attribution

 New York Times 

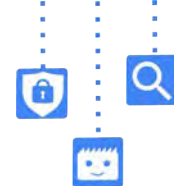
Review: 'Ocean's 8' Women Walk Away With
a Male Franchise. Sort Of ...

Title

The party gets started early in "Ocean's 8," a
frothy female-driven caper. Stuffed with talented,
beautiful women playing naughty, this is the late...

Content
Preview

Relevance in Snippets - an Example



[queries in dbms]

What Is the Definition of a Database Query? - Lifewire

<https://www.lifewire.com> › How To › Software › Databases ▼

Sep 12, 2019 - A database **query** extracts data from a database and formats it in a readable form. A **query** must be written in the language the database requires. ... What a **Database Management System (DBMS)** Is and How It Works.

[database queries]

What Is the Definition of a Database Query? - Lifewire

<https://www.lifewire.com> › How To › Software › Databases ▼

Sep 12, 2019 - A **database query** extracts data from a database and formats it in a readable form. A **query** must be written in the language the database ...

Images in Previews

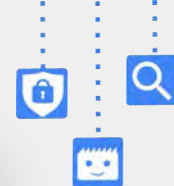
Images are relevant to the query

Placement on right

- Images are secondary to title/snippet

Galleries support contentful pages

Users visit a greater diversity of sites

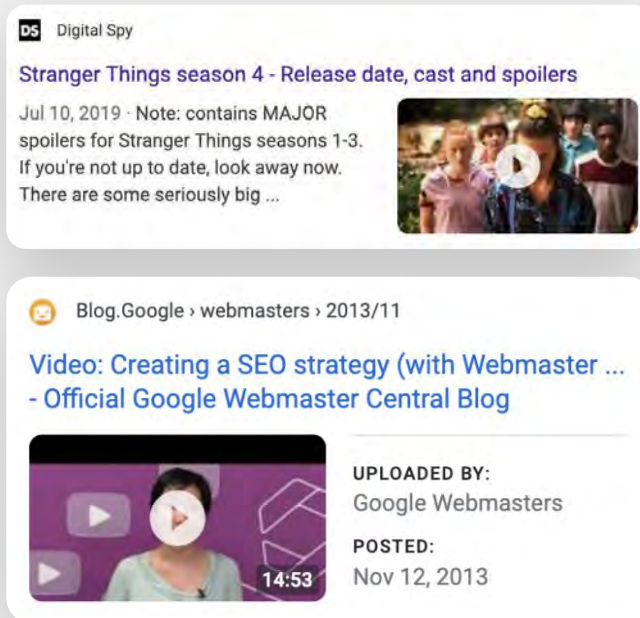
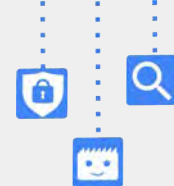


Video in Previews

Video is relevant to the query

Preview indicates if video is dominant or supportive

Video metadata informs user experience



Sitelinks in Previews

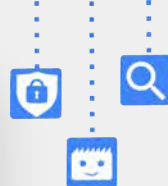
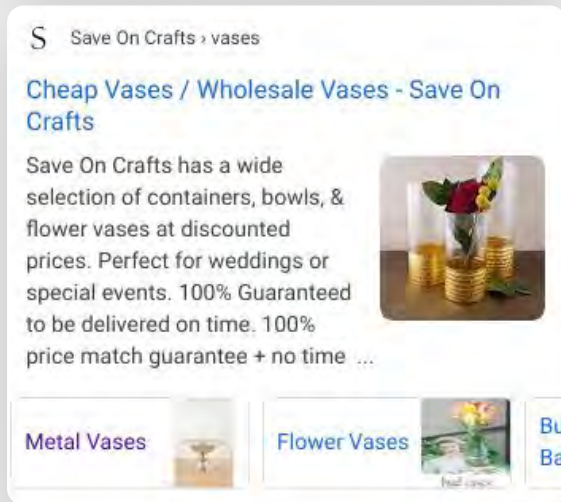
Links are relevant to query

Links extracted algorithmically

- Menus, Site-structure
- Drives traffic into a diverse set of sites

Sitelink-images help users

- Pithy links are better understood



Entity Facts in Previews

Relevance to needs around
the entity

Facts extracted algorithmically

- Tables, Lists

Official SSC Multi Tasking Staff Recruitment Notification 2017 - Testbook Blog

[Testbook](#) > [Blog](#) > [Notifications](#)

🕒 AMP - Dec 31, 2016 - Here's the complete SSC Multi Tasking Staff Recruitment notification. Check eligibility & apply now for the exam to grab 1 of 8300 vacancies!

Date of Common Written Exam: 16.04.2017, 30.04.2017 & 07.05.2017

Last Date to fill Application: 30.01.2017

Start Date to fill Application: 31.12.2016

Date of Descriptive Test: To be announced later

Kilian Jornet | Kilian Jornet Burgada

[Kilian Jornet Burgada](#) > [kilian-jornet](#)

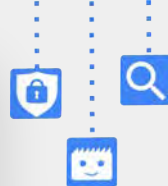
I'm (semi) nomadic, and live somewhere between Chamonix, Tromsø, and the Pyrenees. My life is spent pursuing and fighting for my dreams. Card. Weight: 58-59kg. Height: 171cm. FC max: 205. FC rest: 34 ...

VO2 max: 85-90 ml / min / kg

Weight: 58-59kg

Height: 171cm

Fat: 8%



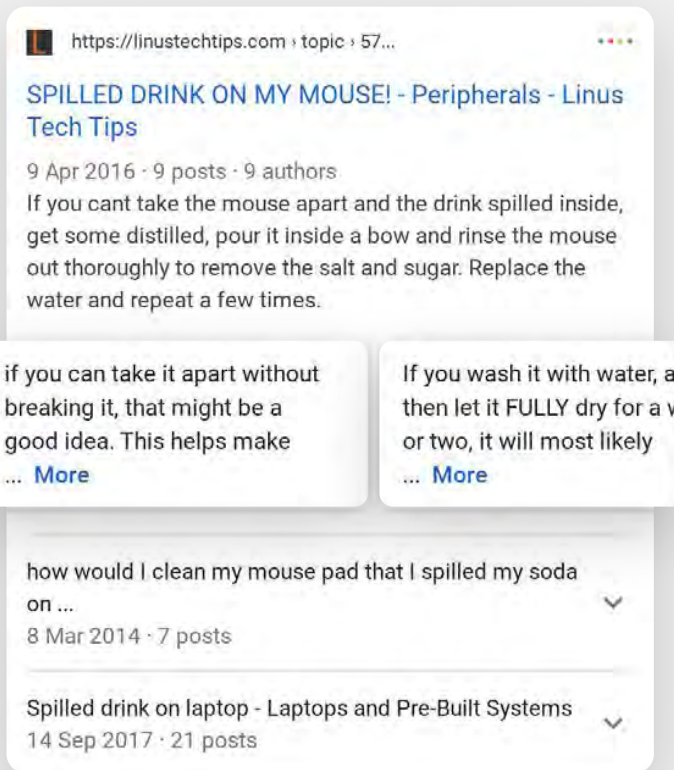
Forums as Previews

Posts extracted algorithmically

Forum cluster is relevant to query

Users find value in forum sites

Markup (QnA) takes priority



Tables & Lists as Previews

Pages with dominant Tables/Lists


Helps users contrast content

Structure and position on the page guides the preview

1718 Roster - University of Central Florida
www.knightswrestling.net › roster

NAME	HOMETOWN	CLASS	WEIGHT
Thomas Agresta	Orlando, FL	Junior	197
Ethan Ahrens	Mims, FL	Freshman	174
Luttre Eglaus	Fort Pierce, FL	Grad	197

→ [View more rows \(19\)](#)

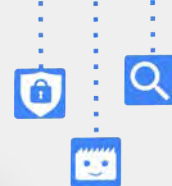
 The Active Times › adventure › mo...

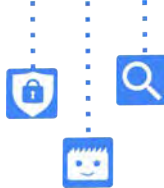
The Most Dangerous Sea Creatures We Never Want to Encounter

List includes:

Great White Shark
Sea Snakes
Textile Cone Snail
Tiger Shark
Box Jellyfish
Blue-Ringed Octo...
Flower Urchin

[View full list](#)





Summary

Relevance infuses everything

Attribution helps user choose the source

Express the **depth** of content

- Sitelinks drive traffic into the site
- Forum previews show related content

Preview **diversity** supports the ecosystem

- Imagery, Video, Tables, Lists, Facts

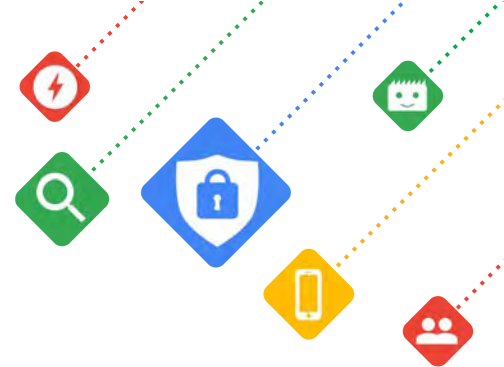


Thank you

Googlebot & Web Hosting



Jin Liang
Software Engineer

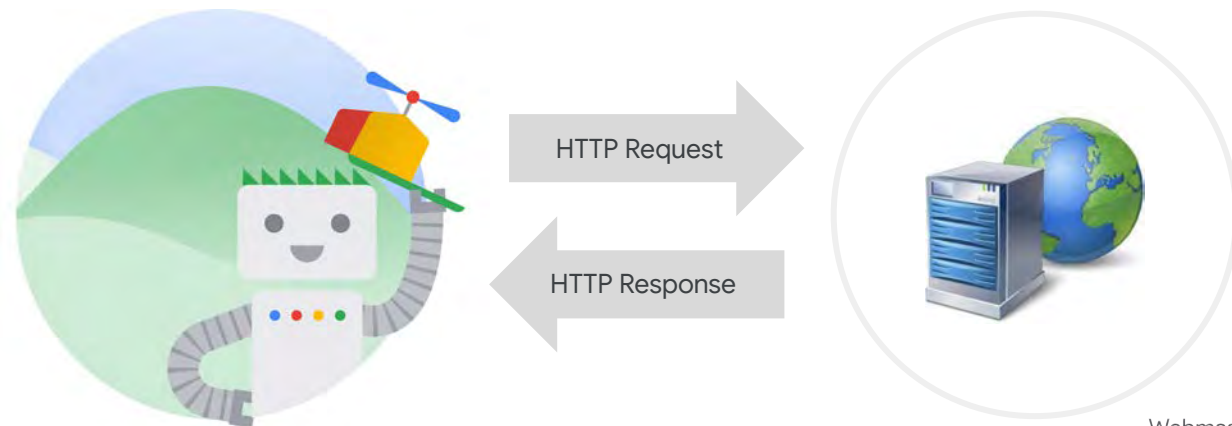
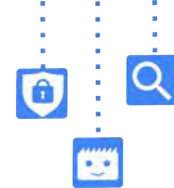


Googlebot and Web Hosting

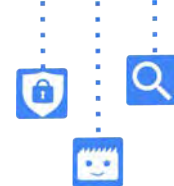
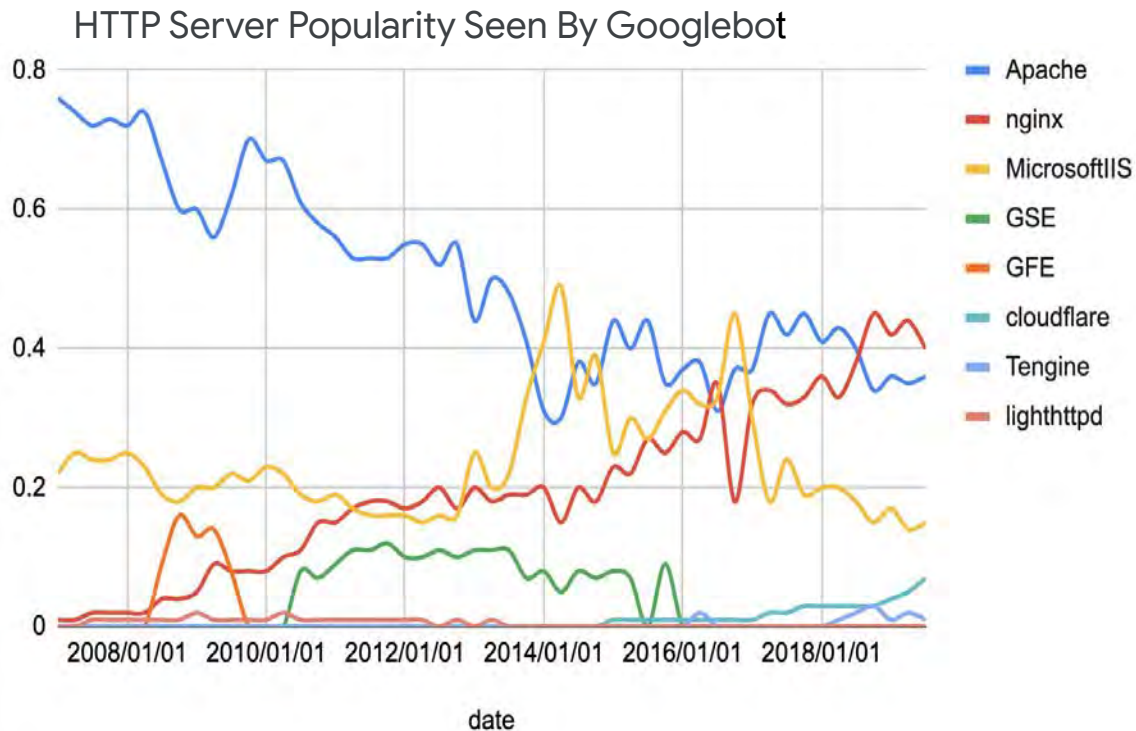
01 How has HTTP serving evolved

02 Robots.txt explained

03 Google crawl rate

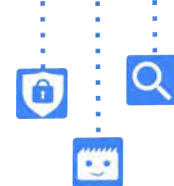
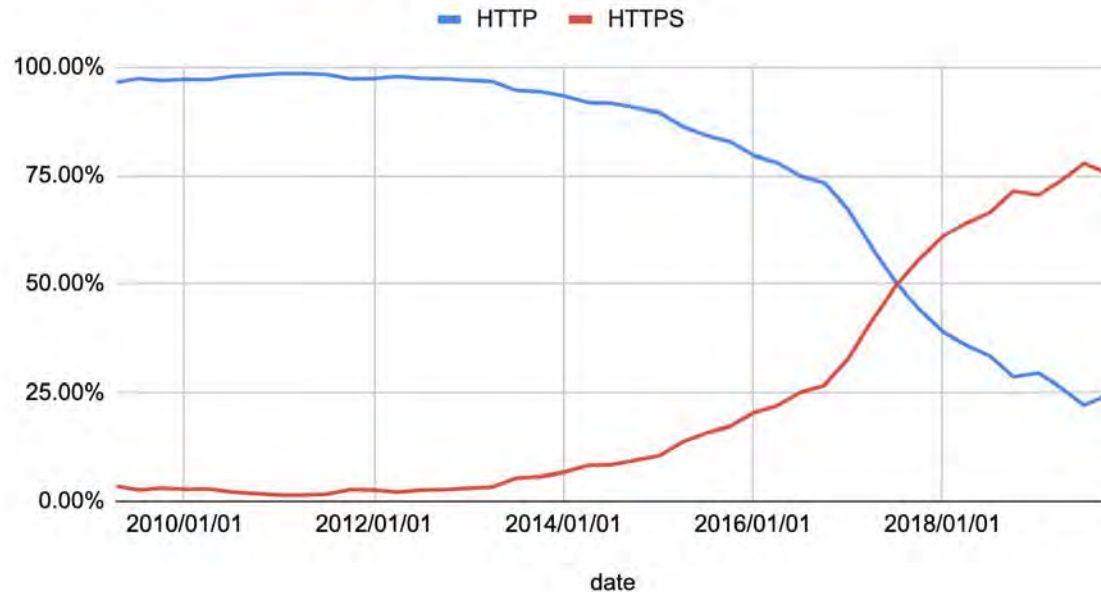


HTTP Server Popularity Changes



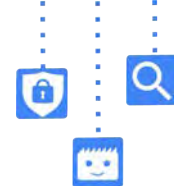
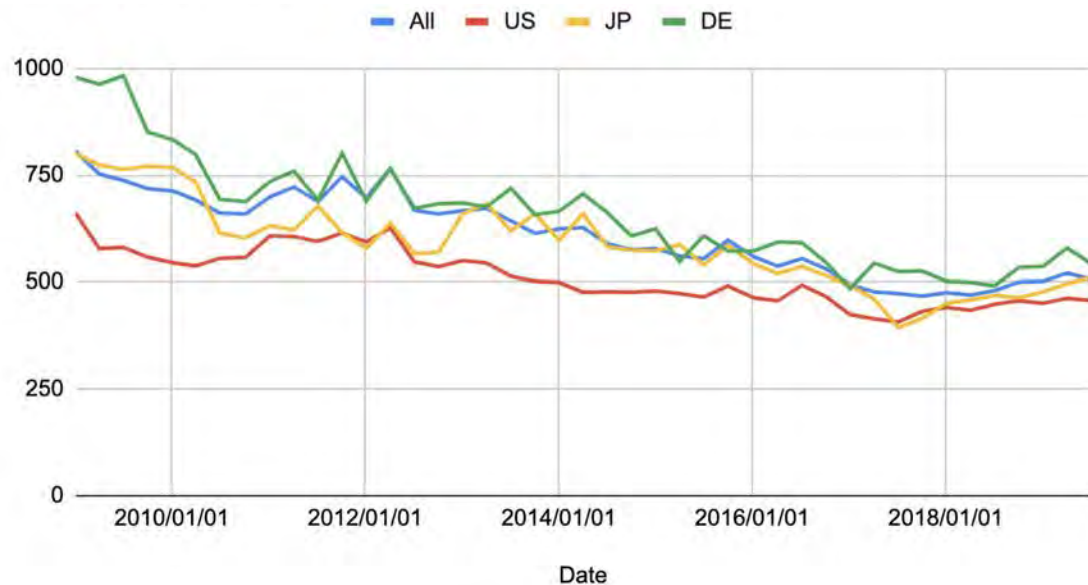
Traffic Shifting To HTTPS

HTTP vs HTTPS Traffic

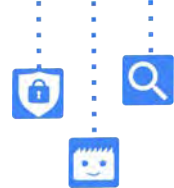


HTTP Fetching Getting Faster

Average Download Time (ms)



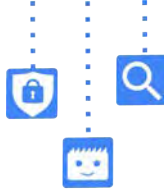
Robots.txt Explained



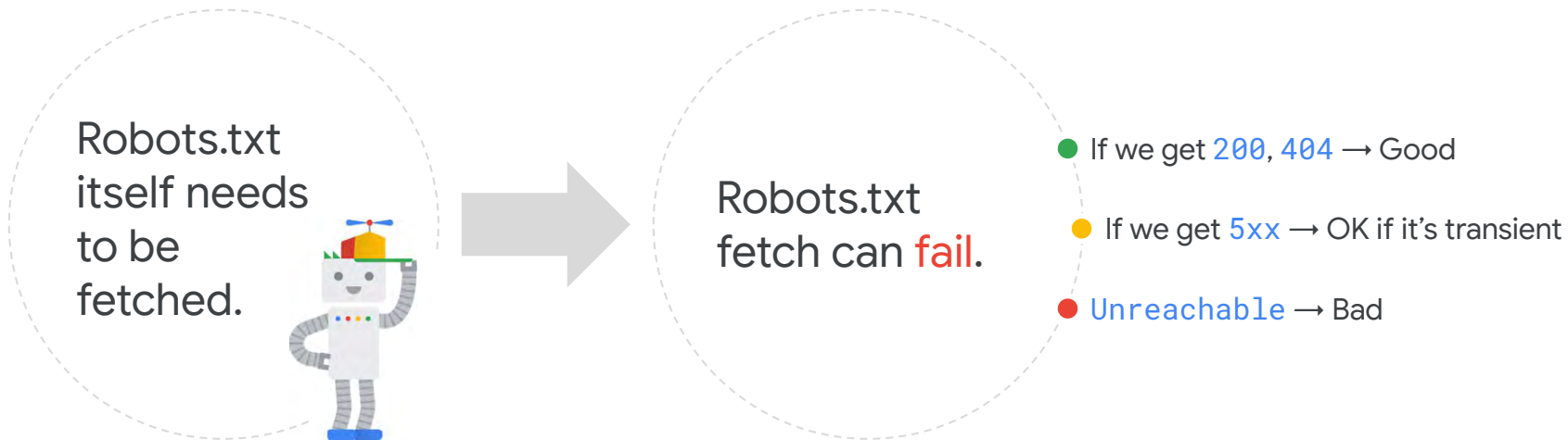
```
User-Agent: Googlebot  
Allow: /useful-content  
Disallow: /do-not-crawl  
...
```

Common way for webmasters to specify access to their websites
Not a standard, and no effort to make it so...
... until we proposed to IETF in July!

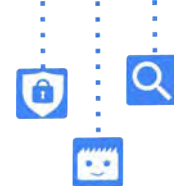
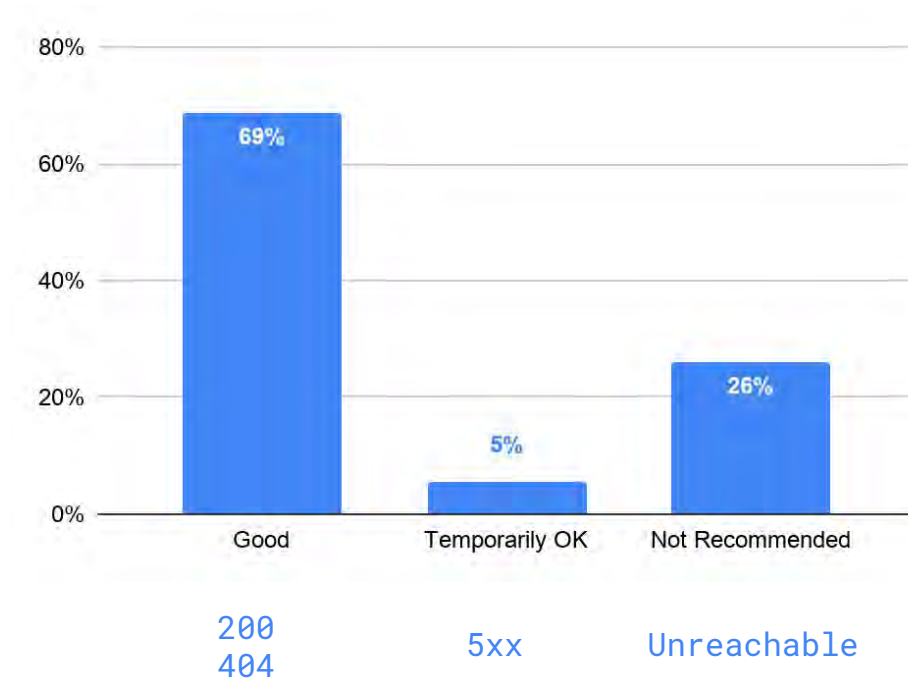
Robots.txt fetching issues



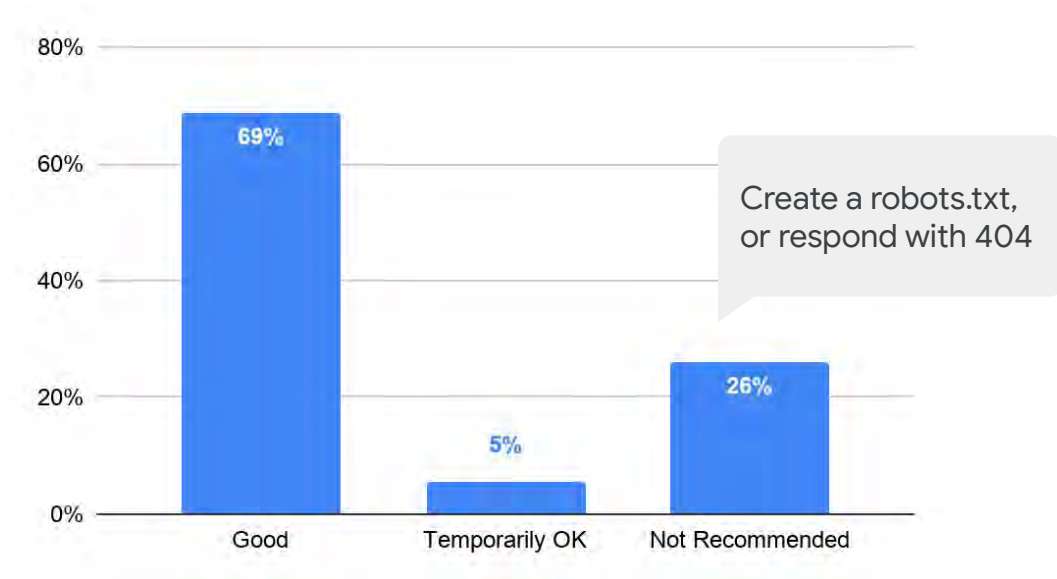
For every URL we crawl, we must first check the robots.txt.



Robots.txt fetch status



Robots.txt usage recommendation



Use robots to control access
Return 404 for open access

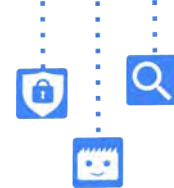
200
404

5xx

Unreachable

Do not always return 5xx

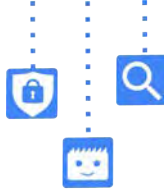
Crawl Rate Limiting



Googlebot has sophisticated way to determine **how fast** to crawl a site

Webmaster can set a **custom crawl rate** in search console if needed





Custom Crawl Rate Recommendations

Custom crawl rate does not trigger more crawl.

It may accidentally reduce our crawl rate.

Our recommendation: “Leave it to Google” unless your site is being overloaded



Thank you

Claiming your knowledge panel



Lauren Clark
Product Manager



Alex Hsu
Product Manager



Billions of searches
take place on Google
every year, and
many of those are
about people
and organizations.

When they find you, how do
you want to introduce yourself?



Google

Technology company



google.com

Google LLC is an American multinational technology company that specializes in Internet-related services and products, which include online advertising technologies, search engine, cloud computing, software, and hardware. It is considered one of the Big Four technology companies, alongside Amazon, Apple, and Facebook. [Wikipedia](#)

CEO: Sundar Pichai (Oct 2, 2015–) [Trending](#)

Founded: September 4, 1998, [Menlo Park, CA](#)

Headquarters: [Mountain View, CA](#)

Subsidiaries: [YouTube](#), [Google.org](#), [DoubleClick](#), [Google AdMob](#), [MORE](#)

Founders: [Larry Page](#), [Sergey Brin](#)

Did you know: Google is the world's second-largest internet company by revenue. [wikipedia.org](#)

Profiles



YouTube



Twitter



Instagram



Pinterest



LinkedIn

People also search for

[View 15+ more](#)



YouTube



Apple



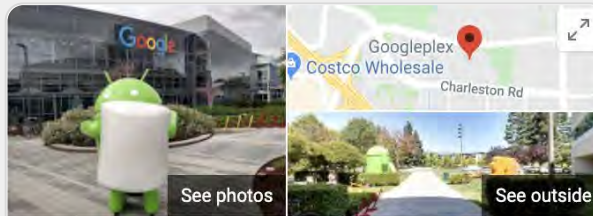
Amazon.c...



Yahoo!



Microsoft Corporati...



Googleplex

[Website](#)

[Directions](#)

[Save](#)

4.3 ★★★★★ [8,311 Google reviews](#)

Building complex in Mountain View, California

The Googleplex is the corporate headquarters complex of Google and its parent company Alphabet Inc. It is located at 1600 Amphitheatre Parkway in Mountain View, California, United States. [Wikipedia](#)

Address: 1600 Amphitheatre Pkwy, Mountain View, CA 94043

Phone: (650) 253-0000

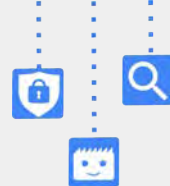
Owner: [Google](#)

Built: July 2004; 15 years ago

Did you know: Google's Mountain View headquarters – known as the Googleplex – have long been lauded as the cream of the crop when it comes to tech campuses. [aol.com](#)

[Suggest an edit](#)

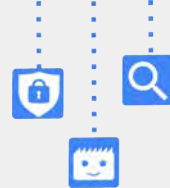
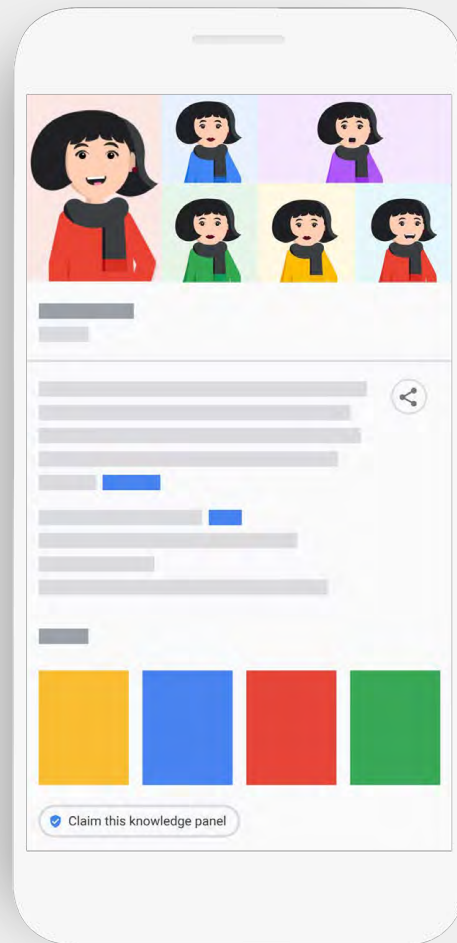
Know this place? [Answer quick questions](#)



Claim your knowledge panel



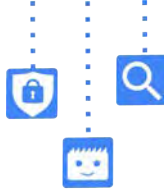
Find your person
or business and
look for a prompt
to **claim**



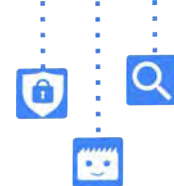
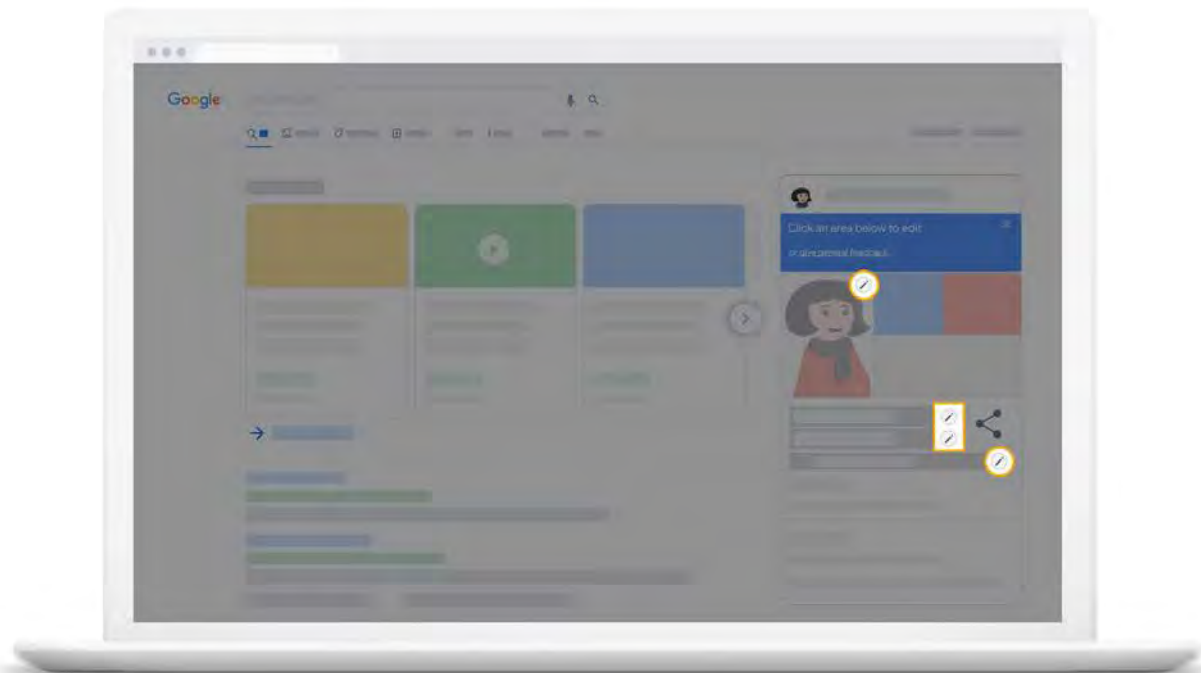
Update your info



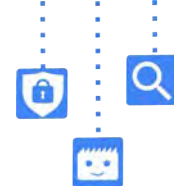
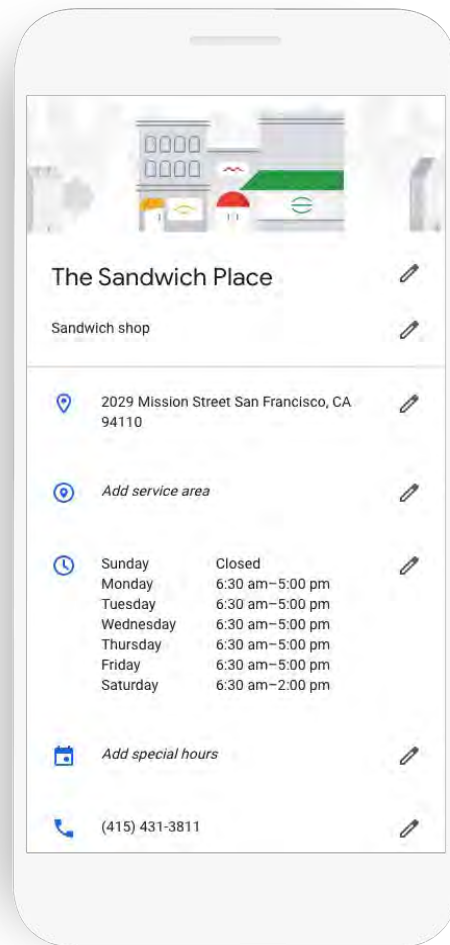
Businesses that complete their information are **twice as likely** to be considered reputable by consumers.



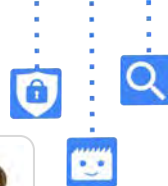
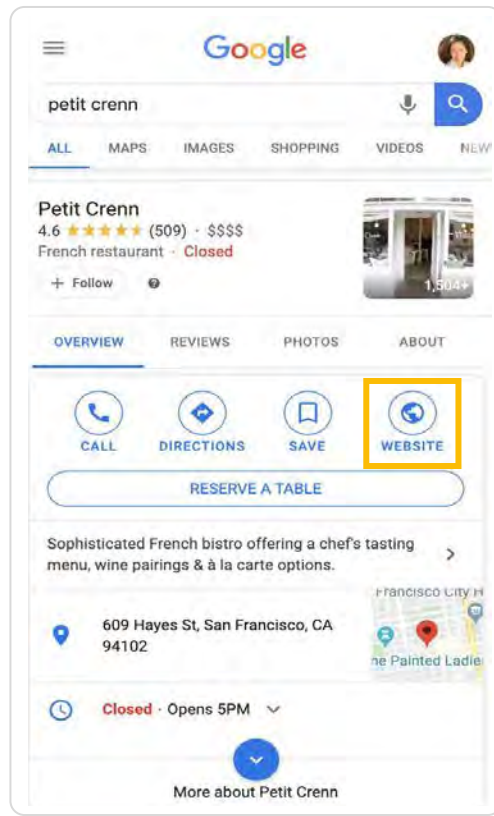
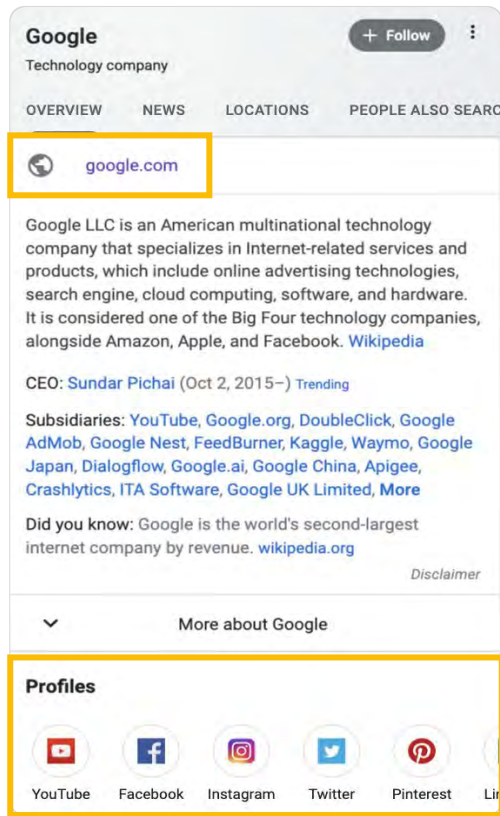
Suggest updates
to inaccurate
or outdated
information in
your knowledge
panel



Add any missing information -
like your hours,
customer
service number,
and street
address.



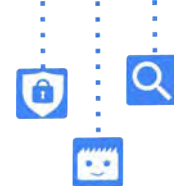
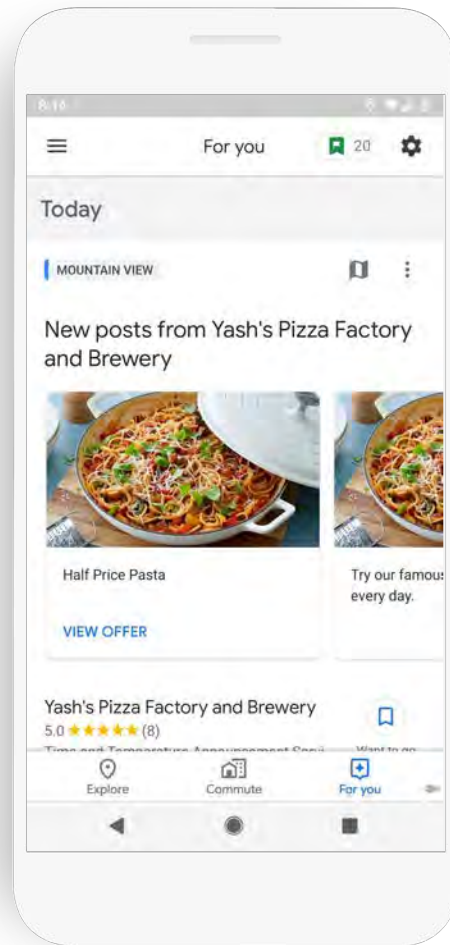
Make sure
your website
and other
web profiles are
up to date



Engage searchers

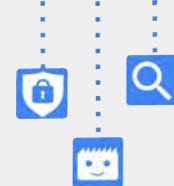


Create content
to engage
and inform
searchers
on Google

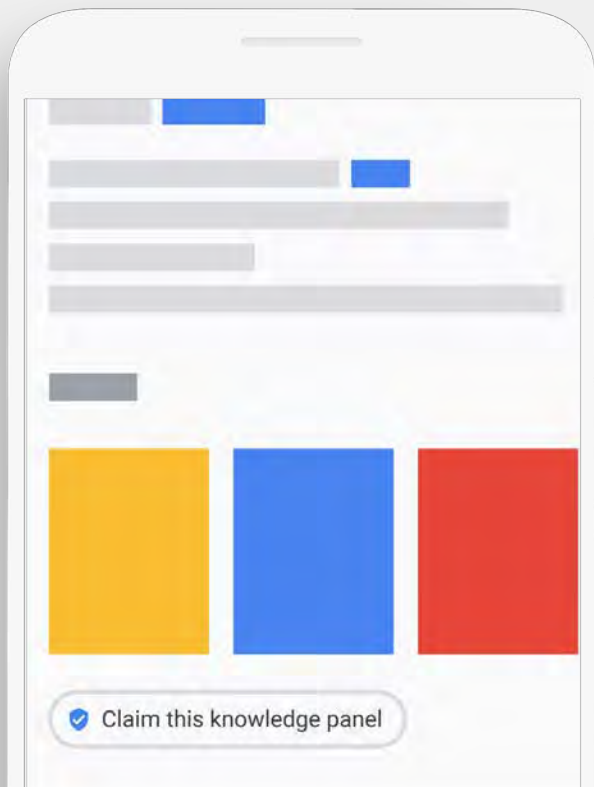


Get Started

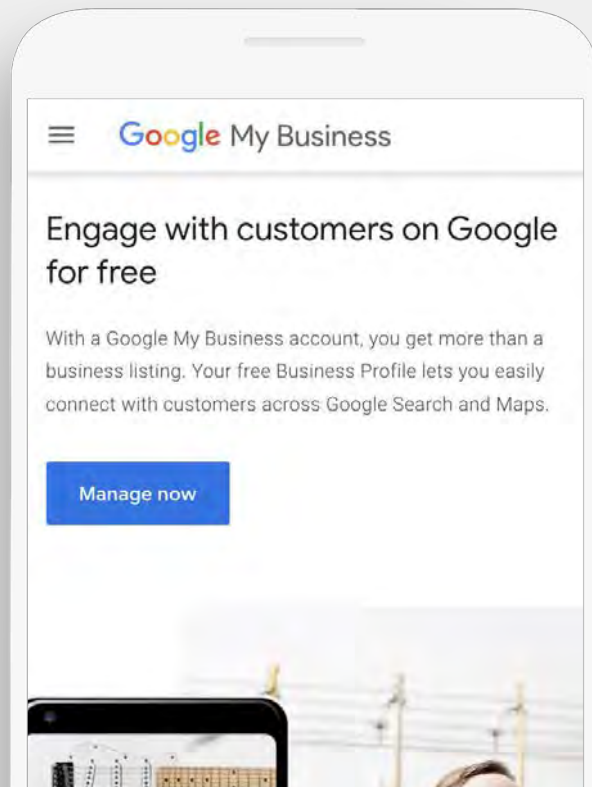




People and Brands



Businesses



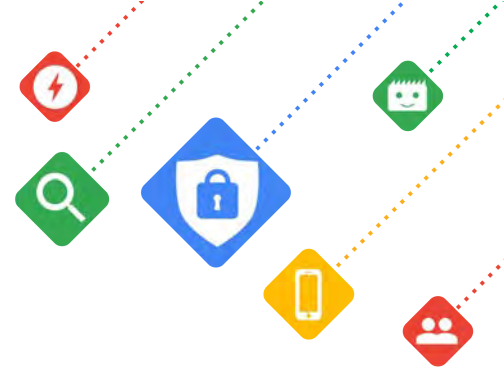
Thank you

Improving Search Over the Years



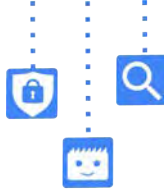
Paul Haahr

Distinguished Engineer, Search



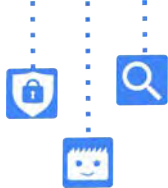
Ranking Case Studies



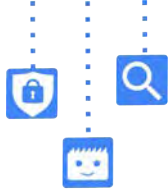


How do Google Search
engineers think about
ranking problems?

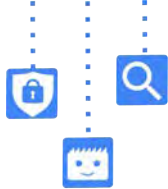




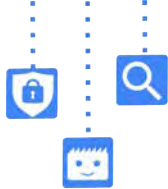
I could try to explain a
step-by-step methodology...



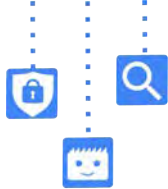
But there isn't one.



There are **many**.



And they involve **a lot** of
debugging, experimentation,
evaluation, guesswork, research,
and (often) luck.



Instead, here are
some **examples...**

Synonyms and Siblings



Google's Synonyms System

- User vocabulary \neq Document vocabulary
- System tries to bridge the gap by automatically adding alternative words
- Similar to using OR, but usually less important than original terms
- One of Google Search's most important ranking components



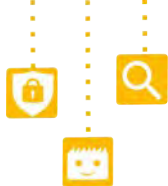


An example...

[cycling tours in italy]

⇒

[cycling OR cycle OR bicycle OR bike OR biking
tours OR tour OR holidays OR vacation
in
italy OR italian]



Contextual:
Synonyms
depend
on other
query words

[**gm** truck] ⇒ “general motors”

[**gm** barley] ⇒ “genetically modified”

[baseball **gm** salary] ⇒ “general manager”

Not the same as English Synonyms

- Designed to find good search results
- Hidden behind the scenes (mostly)
- Unimportant whether they're actually synonyms to a human reader



But...



For a short time in 2005, Google's top result for **[united airlines]** was continental.com

(The two companies did merge in 2010, but it wasn't our fault.)

(We hope.)





Why? Synonyms

(And a couple
of unrelated bugs
that I'm not going
to talk about.)

[united airlines]

⇒

[united OR continental
airlines OR air OR airline]

How do we fix things?

- We want algorithmic solutions
- Don't just manually block the problems
- Look for patterns of failures



Synonyms sometimes finds siblings

- We can learn pairs of words that serve similar roles but aren't interchangeable
- Consider pairs of searches:
 [**united** *reservations*]
 [**continental** *reservations*]

 [**united** *newark airport*]
 [**continental** *newark airport*]
 ...
- “Siblings” (often rival siblings!)



Can we
distinguish
siblings
from useful
synonyms?

- Again, look to searches people do
- People compare siblings to each other:
[united vs continental]
[canon vs nikon]
[beatles vs stones]
[godzilla vs king kong]
...
- Look for [X vs Y] queries from logs,
use as a negative signal for
 $X \Rightarrow Y$ and $Y \Rightarrow X$ synonyms



Then comes
the hard part

- Process logs, build data, run experiments, evaluate the results, tune, repeat...

- Eventually, find many other synonym failures:

cat⇒dog

part time⇒full time

- But we also lost some good synonyms:

sign in⇒sign on

address⇒contact



Lessons



Understanding
patterns of
failures can
reveal solutions



By not patching
over algorithmic
problems manually,
we get more
general solutions



Every
change
has wins
and losses



Non-Compositional Compounds



Information Retrieval

Information Retrieval is mostly about matching and counting words

- Including title vs body, links, frequency, etc

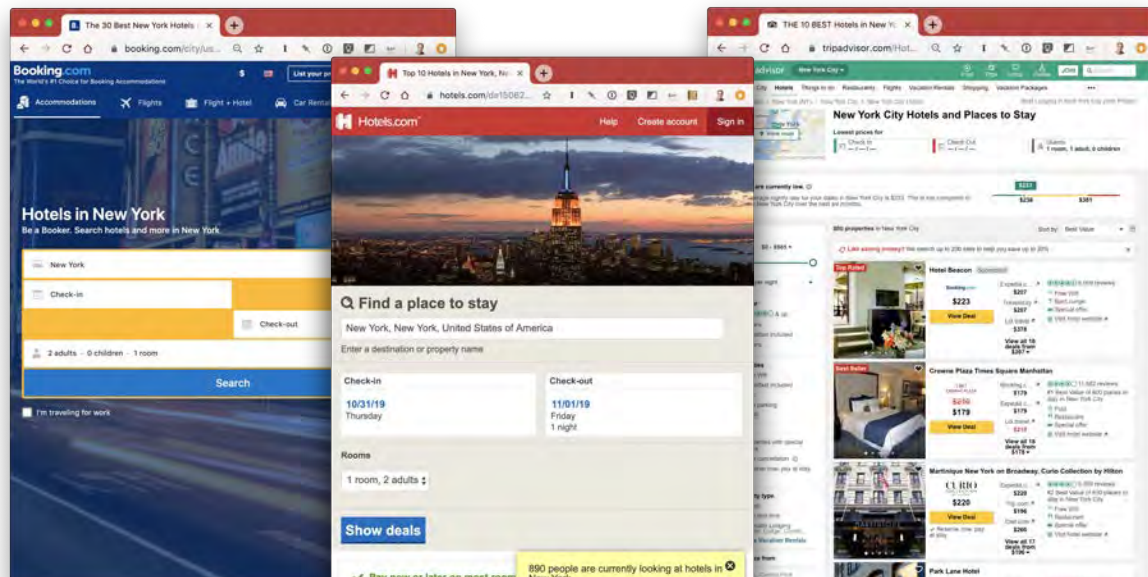
This is the basic underpinning of Search.



Relevance
comes from
matching words

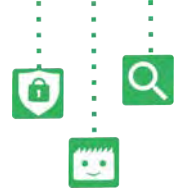
Consider these pages, which are
good matches for [new york hotels]

- Title, body, links, etc



But sometimes
it's a bit too
simplistic

Are they good matches for `[york hotels]`?



Compounds

A **compositional compound** is “a phrase of two or more words where the words composing the phrase have the same meanings in the compound as their conventional meanings”.

A **non-compositional compound** is one where the meanings differ.



New York

“New York” is **non-compositional**.

Even though it is formed by compounding “New” and “York,” there’s nothing York-related now.

Not all place names follow the same rule

- “York” is not “New York”
- “Vegas” is “Las Vegas”



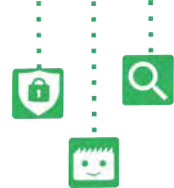
Can we
identify non-
compositional
compounds?

Algorithm:

- Start with a set of “X Y” phrases
- Look at pages where “X Y” occurs
- If “X” or “Y” only appears in “X Y” on most of those pages, guess that it’s non-compositional

For “new york”:

- “new” appears alone on many
- “york” appears alone on very few





Matching NCCs

Now that we have non-compositional compounds, what do we do?

Specialized matching code:

```
[york hotels]
```

⇒

```
[(ignore_left:new york) hotels]
```

- Meaning: “Don’t match ‘york’ if the word to the left is ‘new’.”



Lessons



Edge case: would be very hard to predict in advance, but obvious to the first person who tries this query



Once seen, it's obvious there is a general pattern here

[fantasy game] is not “final fantasy”

[view office] is not “mountain view”

...



Hard work is done offline, ahead of time



Small change in matching code



Language Evolves Over Time

If you received “🤪” in a text in 1996, when Google launched at Stanford, would you have known it meant “rolling on the floor, laughing”?

Probably not. [The first emoji appeared in 1997.](#)



Why Search for Emoji?

- People use emoji all the time
- But, often, they're not sure what they mean exactly
- So, they search for them



Emoji in Search

Unfortunately, for a long time, Search ignored emoji and other “special characters”

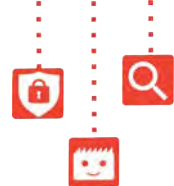
- “Nobody searches for them”
- Expensive to index if they’re not used

What happened?

[😄] didn’t find anything

[smiley face 😄] sort-of worked

[😄 meaning] found dictionaries



Index/Query Alignment

Changes to what Search indexes are complicated, because they need to go in the right **order**:

- First, update indexing to allow emoji
- Wait for documents to be reindexed
- Then, change query parsing

But first:

- Prove the cost is worth it!
- Even though they didn't work, people were using emoji in **>1 million searches** per day



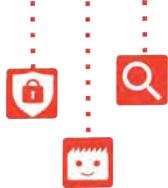
But...

After data was in the index and query parsing was fixed, we ran evaluations, with **very negative results**.

Lots of other systems and models that needed to be updated before launch:

- Link processing
- Spelling
- Autocomplete
- ...





While we
were at it...

Also added **math** and other symbols

- $[\infty]$
- $[\Sigma x]$
- $[P \neq NP]$

And someone else did (some) **punctuation**

- $[+=]$
- $[== \text{ vs } ===]$
- $[P! = NP]$

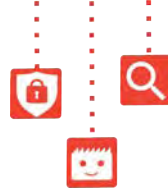
Lessons



Things which look easy
from the outside can be
a lot of work to implement



All the assumptions
you bake into your code
can change over 20 years





The Search Emoji Team

