Connecting The Probable Dots

Disambiguating strategies for entity oriented search and natural language understanding

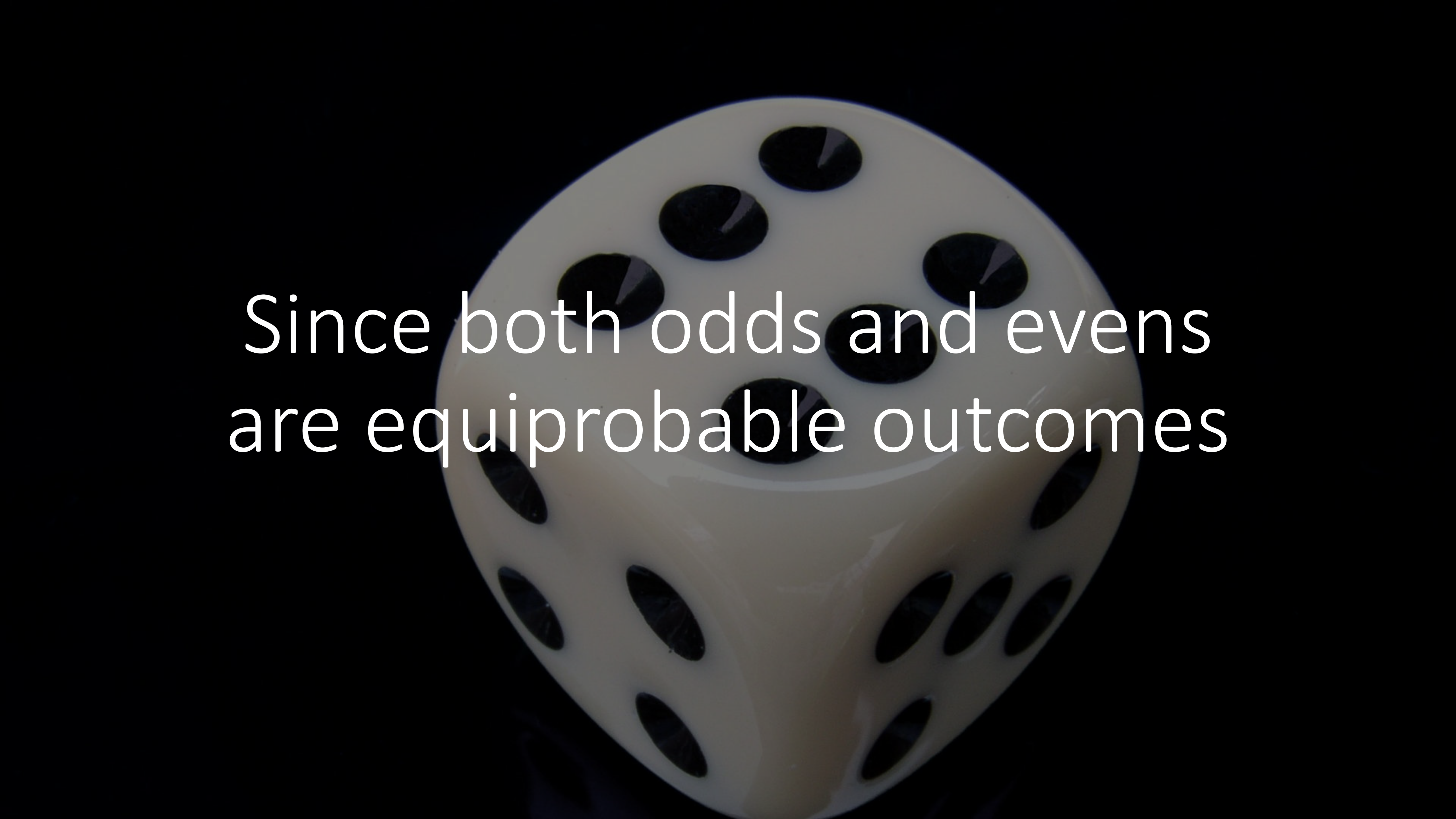
Dawn Anderson

@dawnieando

#FOS20

The probability of rolling an even number or an odd number is 1:2

This is an example of ‘equiprobability’



Since both odds and evens are equiprobable outcomes

“Assigns equal probabilities to outcomes when they are judged to be equipossible” (Wikipedia)

In URLs, content, locations & entities equiprobability can be problematic

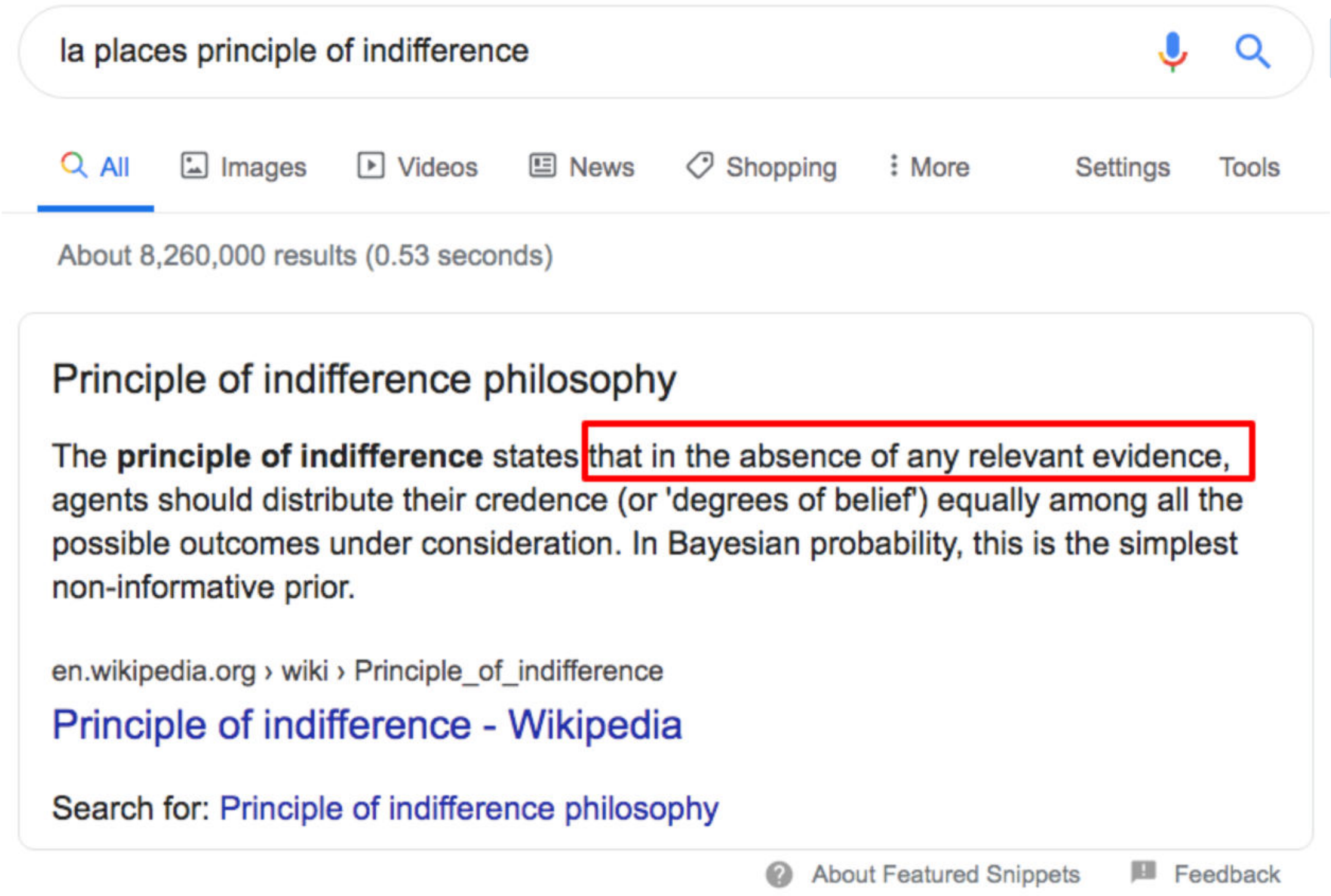
When one of a number of pluralities equally meets an information need

One must beat the other

This could be duplicate content or ‘ambiguity in entities’

Equiprobability is like ‘See saw SEO’

LaPlace’s Principle of Indifference



la places principle of indifference

About 8,260,000 results (0.53 seconds)

Principle of indifference philosophy

The **principle of indifference** states that in the absence of any relevant evidence, agents should distribute their credence (or 'degrees of belief') equally among all the possible outcomes under consideration. In Bayesian probability, this is the simplest non-informative prior.

en.wikipedia.org › wiki › Principle_of_indifference

Principle of indifference - Wikipedia

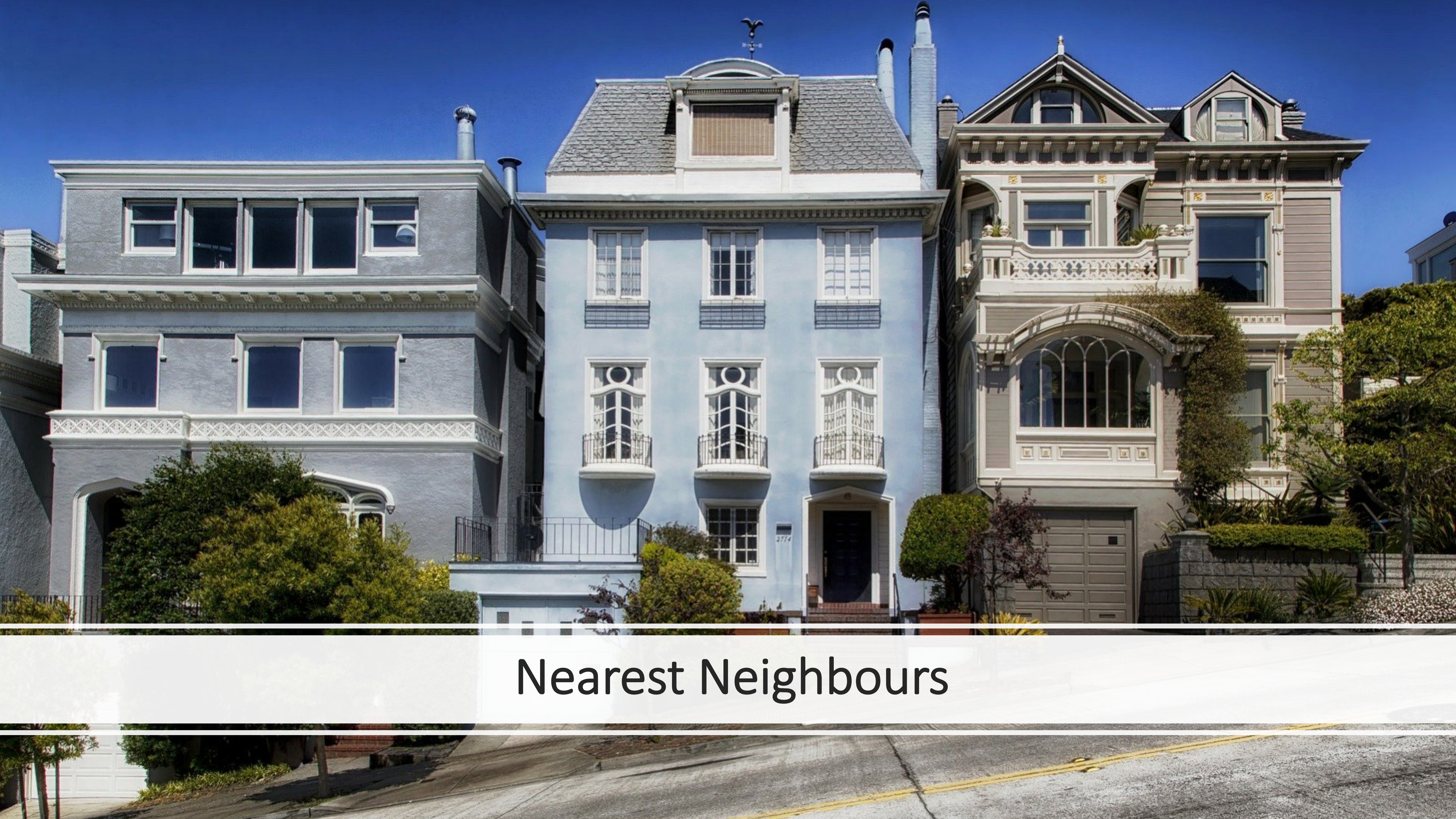
Search for: Principle of indifference philosophy

Where there is equi-possibility there is a need for further confirmations

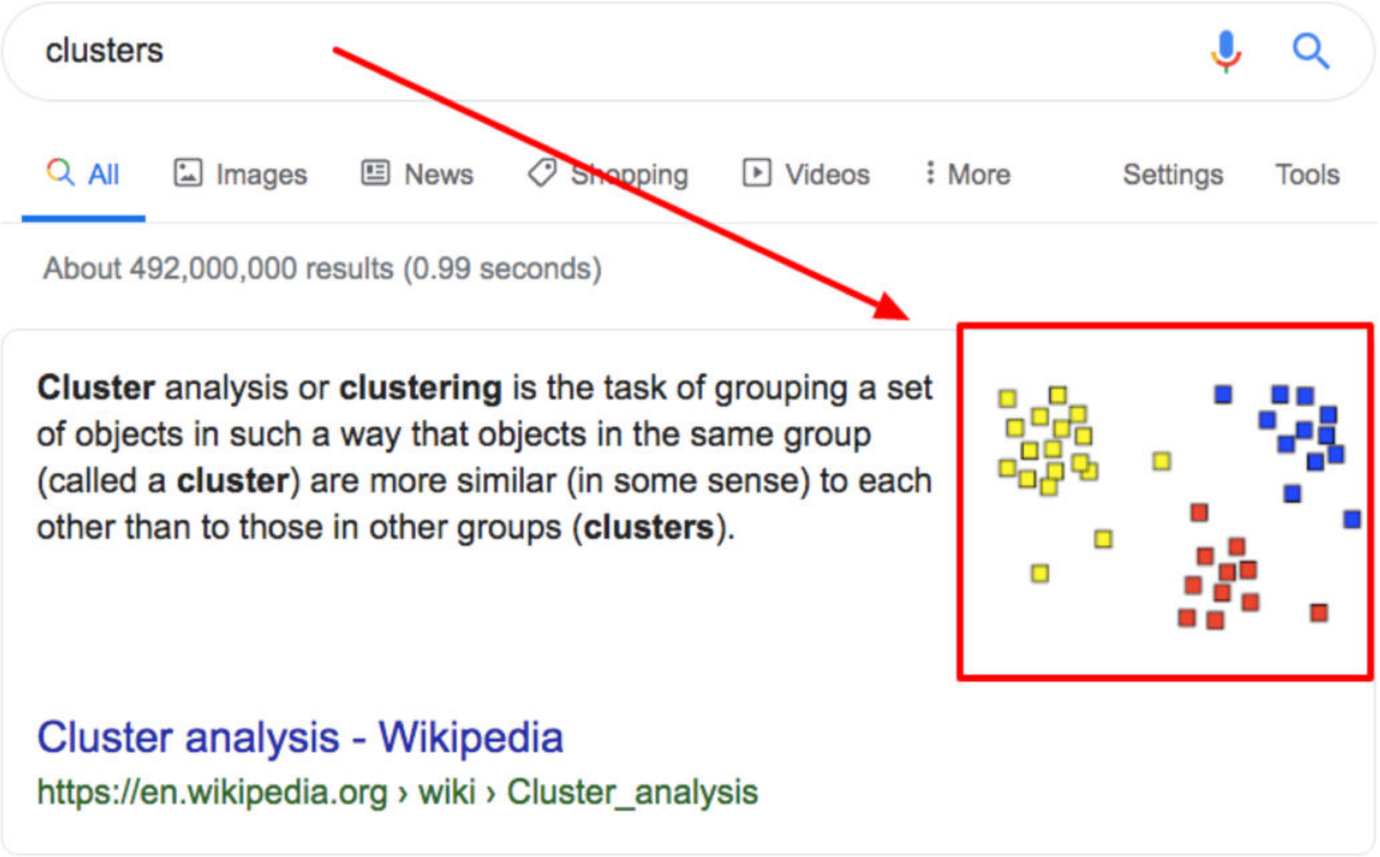
Further ‘confirmations’ beyond the same content or surface form must be sought

Where two or more domain assets, surface forms, or entity determinations are considered equal a representative is ‘the chosen one’ (canonical)

Nearest Neighbours



All the pluralities are clustered together and one is picked



clusters

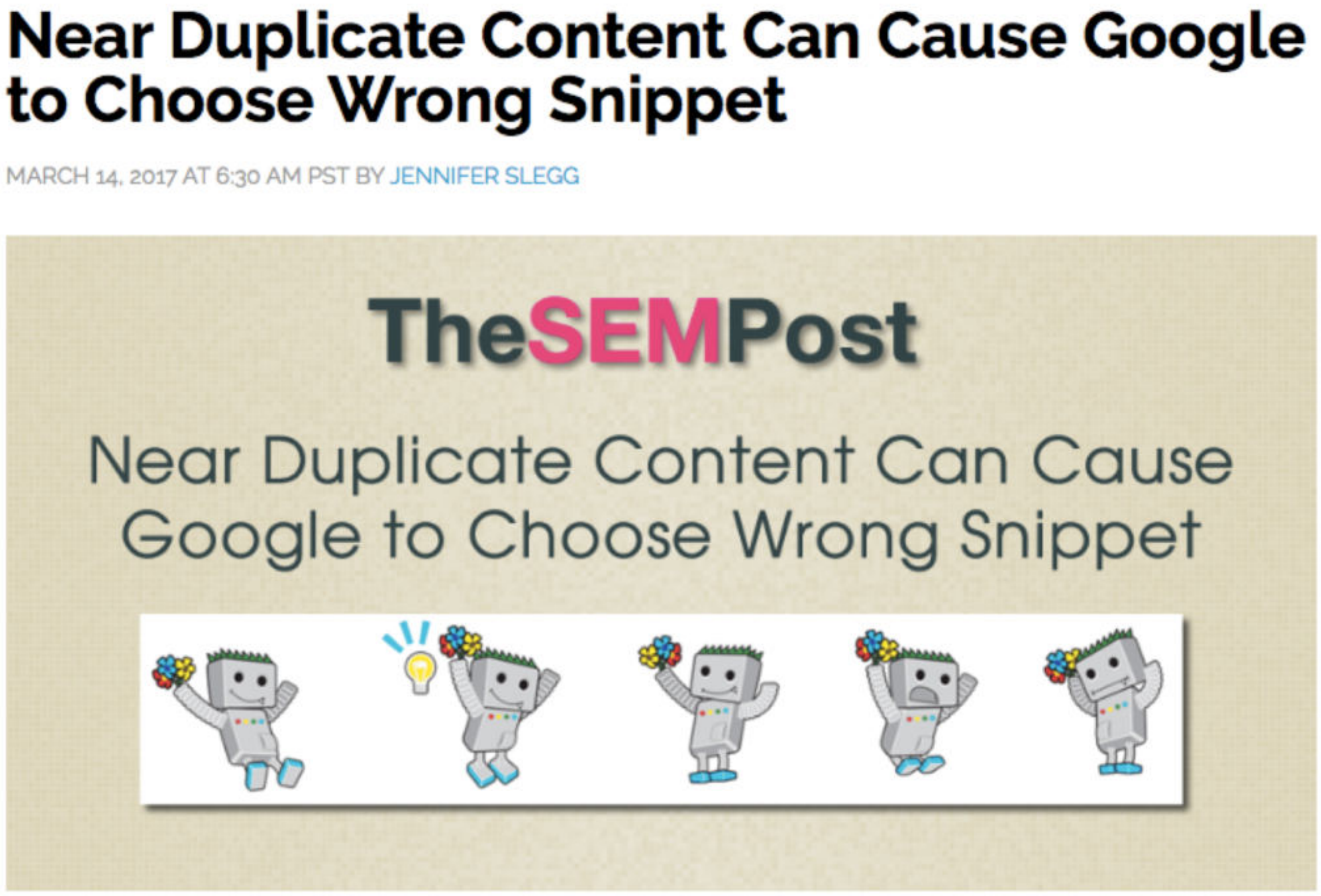
About 492,000,000 results (0.99 seconds)

Cluster analysis or **clustering** is the task of grouping a set of objects in such a way that objects in the same group (called a **cluster**) are more similar (in some sense) to each other than to those in other groups (**clusters**).

Cluster analysis - Wikipedia

https://en.wikipedia.org › wiki › Cluster_analysis

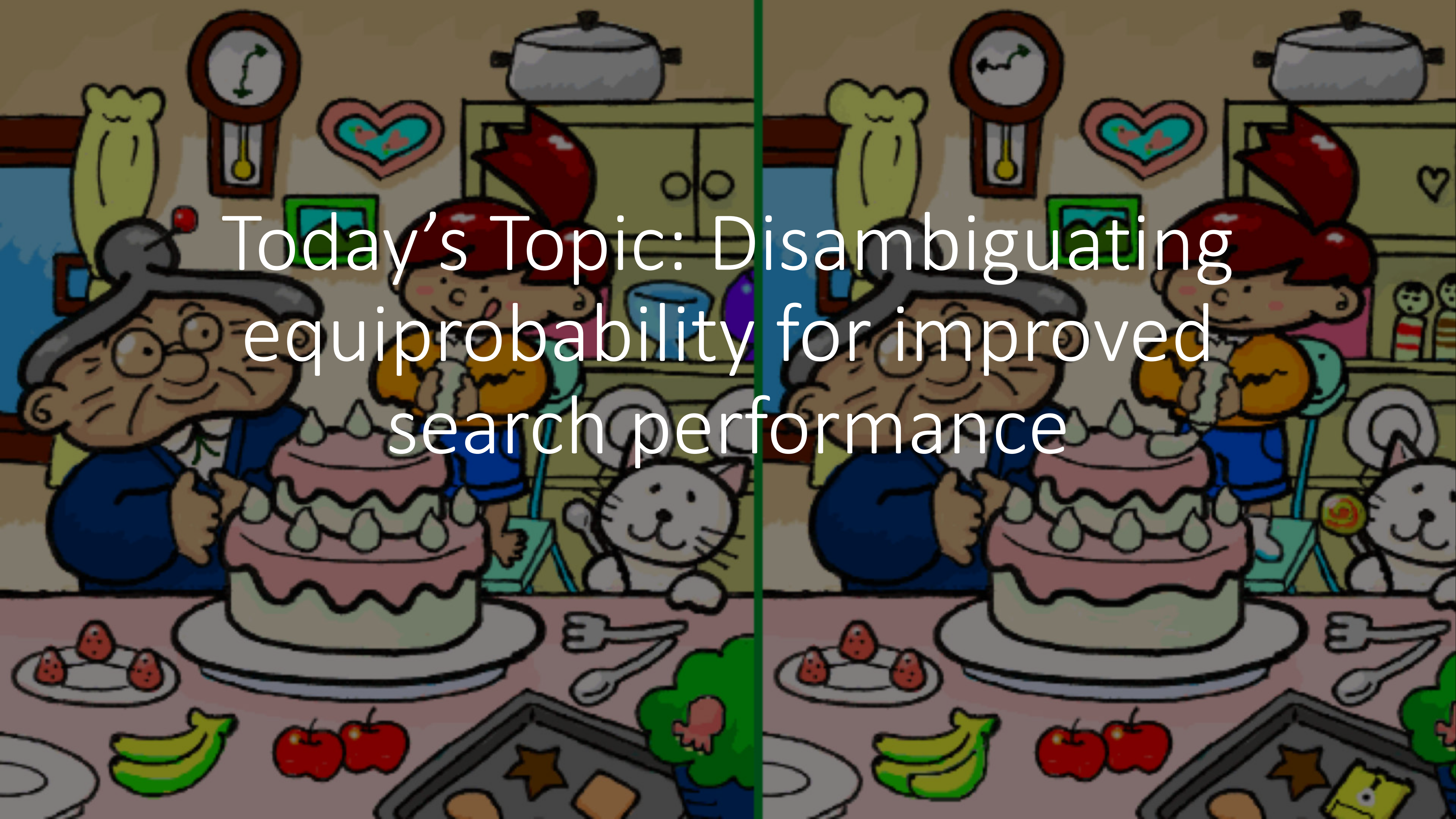
Sometimes 'bits' are picked from multiple 'equiprobable outcomes'



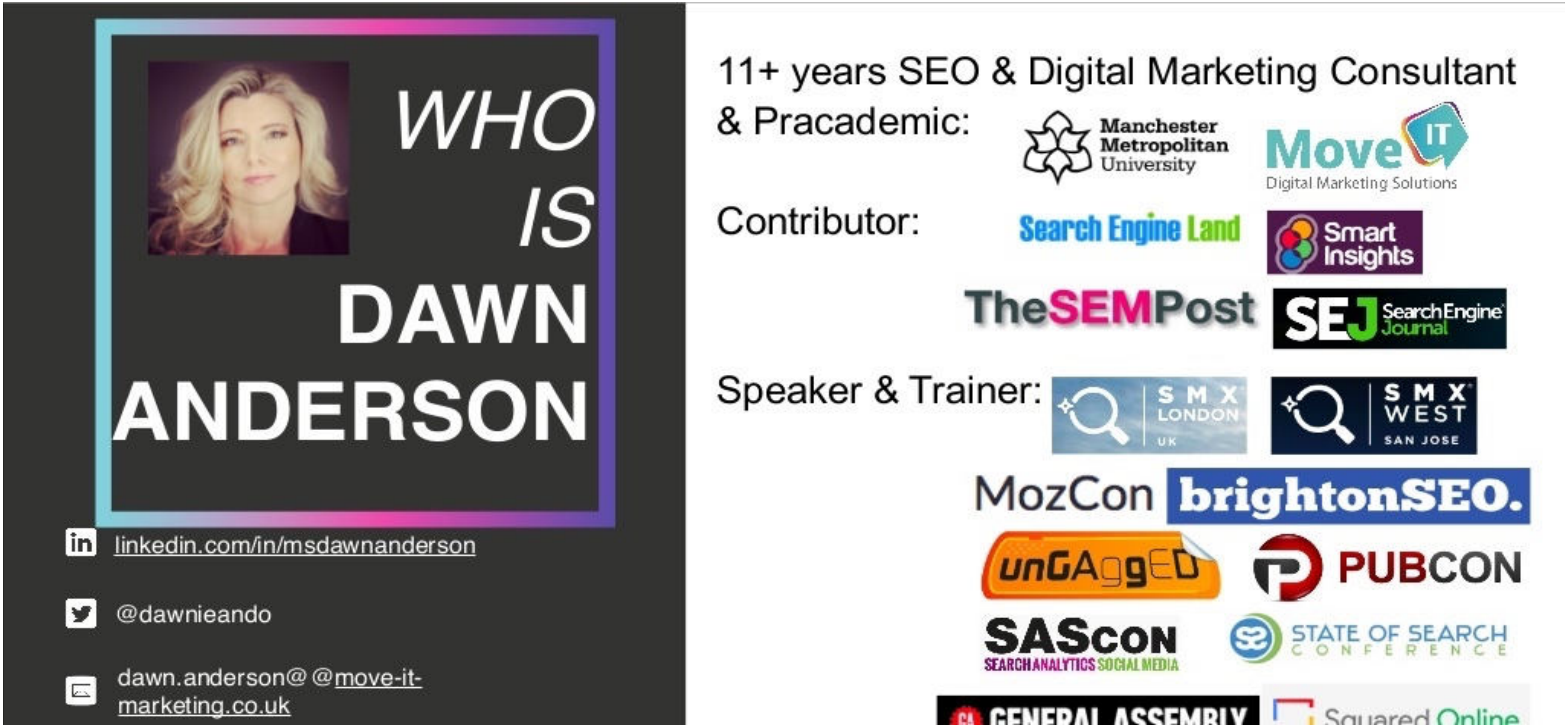
Near Duplicate Content Can Cause Google to Choose Wrong Snippet

MARCH 14, 2017 AT 6:30 AM PST BY JENNIFER SLEGG

TheSEMPost
Near Duplicate Content Can Cause Google to Choose Wrong Snippet



Today’s Topic: Disambiguating equiprobability for improved search performance



WHO IS DAWN ANDERSON

linkedin.com/in/msdawnanderson

@dawnieando

dawn.anderson@ @move-it-marketing.co.uk

11+ years SEO & Digital Marketing Consultant & Pracademic: Manchester Metropolitan University, Move IT Digital Marketing Solutions

Contributor: Search Engine Land, Smart Insights, TheSEMPost, SEJ Search Engine Journal

Speaker & Trainer: SMX London UK, SMX West San Jose, MozCon, brightonSEO., unGAggED, PUBCON, SAScon Search Analytics Social Media, State of Search Conference, General Assembly, Squared Online

WHO IS DAWN ANDERSON

linkedin.com/in/msdawnanderson

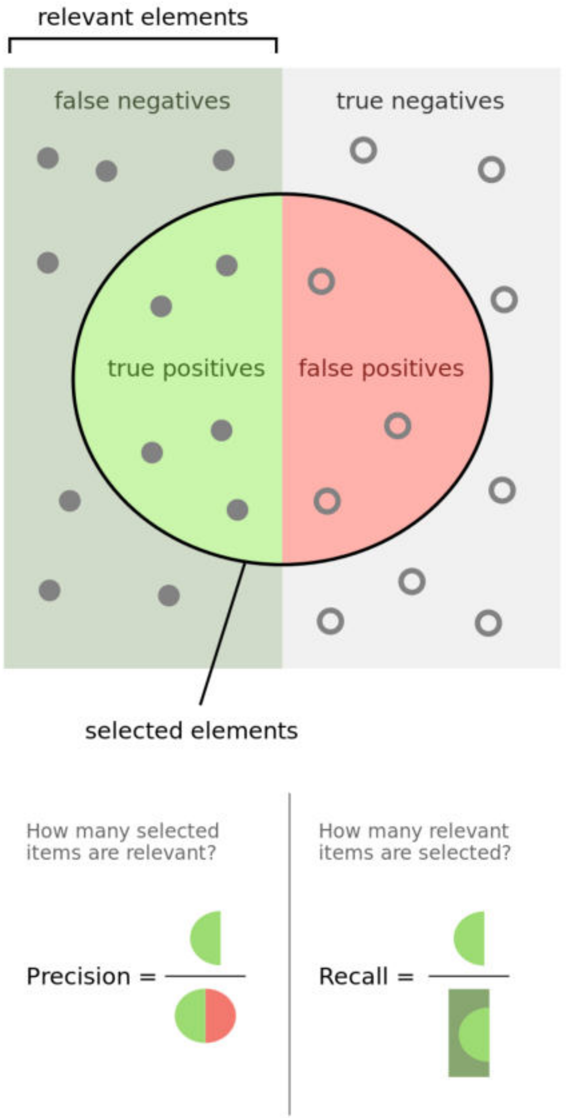
@dawnieando

dawn.anderson@@move-it-marketing.co.uk

move-it-marketing.co.uk

Meet Bert & Tedward

In an ideal world 'precision' and 'recall' would be perfect… but… it's NOT



So search engines must also work off probability determination

Machine learning also now powers some probability determination

Google Uses Machine Learning for Crawling, Indexing & Ranking

Google Uses Machine Learning for Crawling, Indexing & Ranking

MAY 2, 2018 AT 7:20 AM PST BY JENNIFER SLEGG — LEAVE A COMMENT

TheSEMPost
Google Uses Machine Learning for Crawling, Indexing & Ranking

Using 'The Law of Large Numbers'

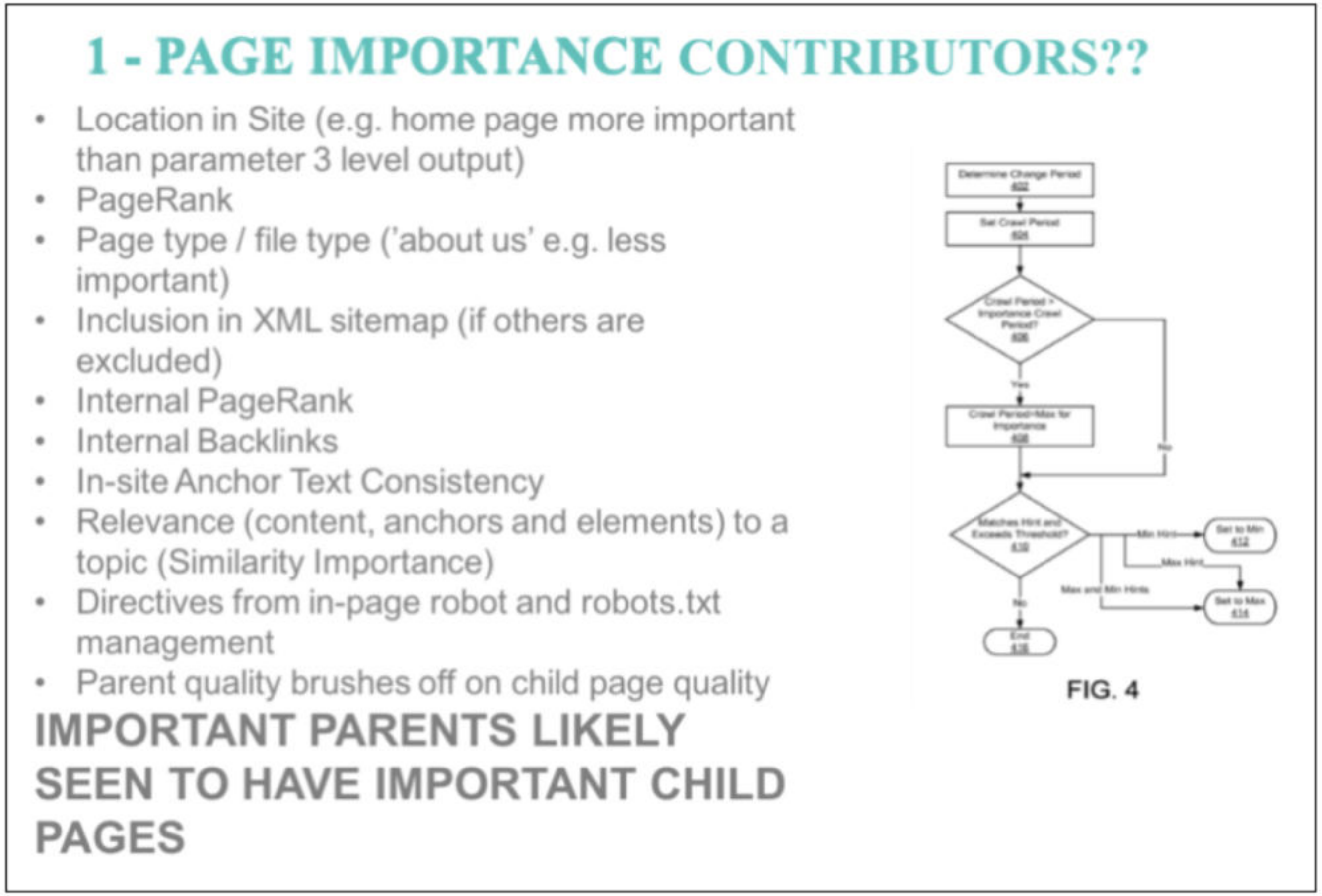
Progressive learning & increasingly ‘educated’ guesses

Sometimes ‘probability’ predictions can be unpredictable

On the ‘long tail’ there’s often very little, or nothing in it

Low to no page-rank pages (but that is also now just one of very many things

There will be lots of contributing factors now



1 - PAGE IMPORTANCE CONTRIBUTORS??

- Location in Site (e.g. home page more important than parameter 3 level output)
- PageRank
- Page type / file type ('about us' e.g. less important)
- Inclusion in XML sitemap (if others are excluded)
- Internal PageRank
- Internal Backlinks
- In-site Anchor Text Consistency
- Relevance (content, anchors and elements) to a topic (Similarity Importance)
- Directives from in-page robot and robots.txt management
- Parent quality brushes off on child page quality

IMPORTANT PARENTS LIKELY SEEN TO HAVE IMPORTANT CHILD PAGES

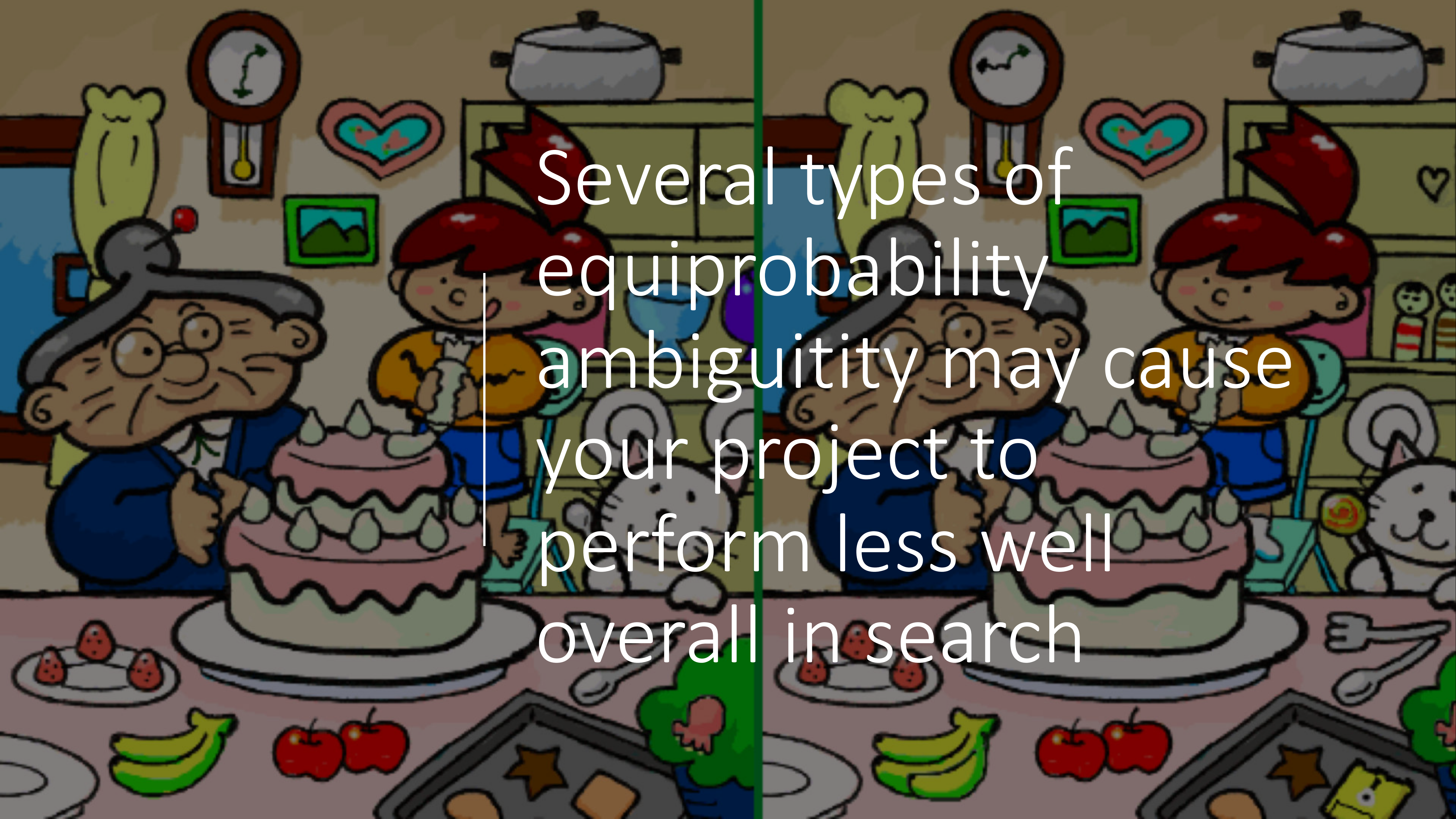
FIG. 4

See-saw SEO

Then the canonical balance Tips From Equiprobability Slightly

Where ambiguity leads to equiprobability intermittently

Your ranking flux might well be transferring equiprobability



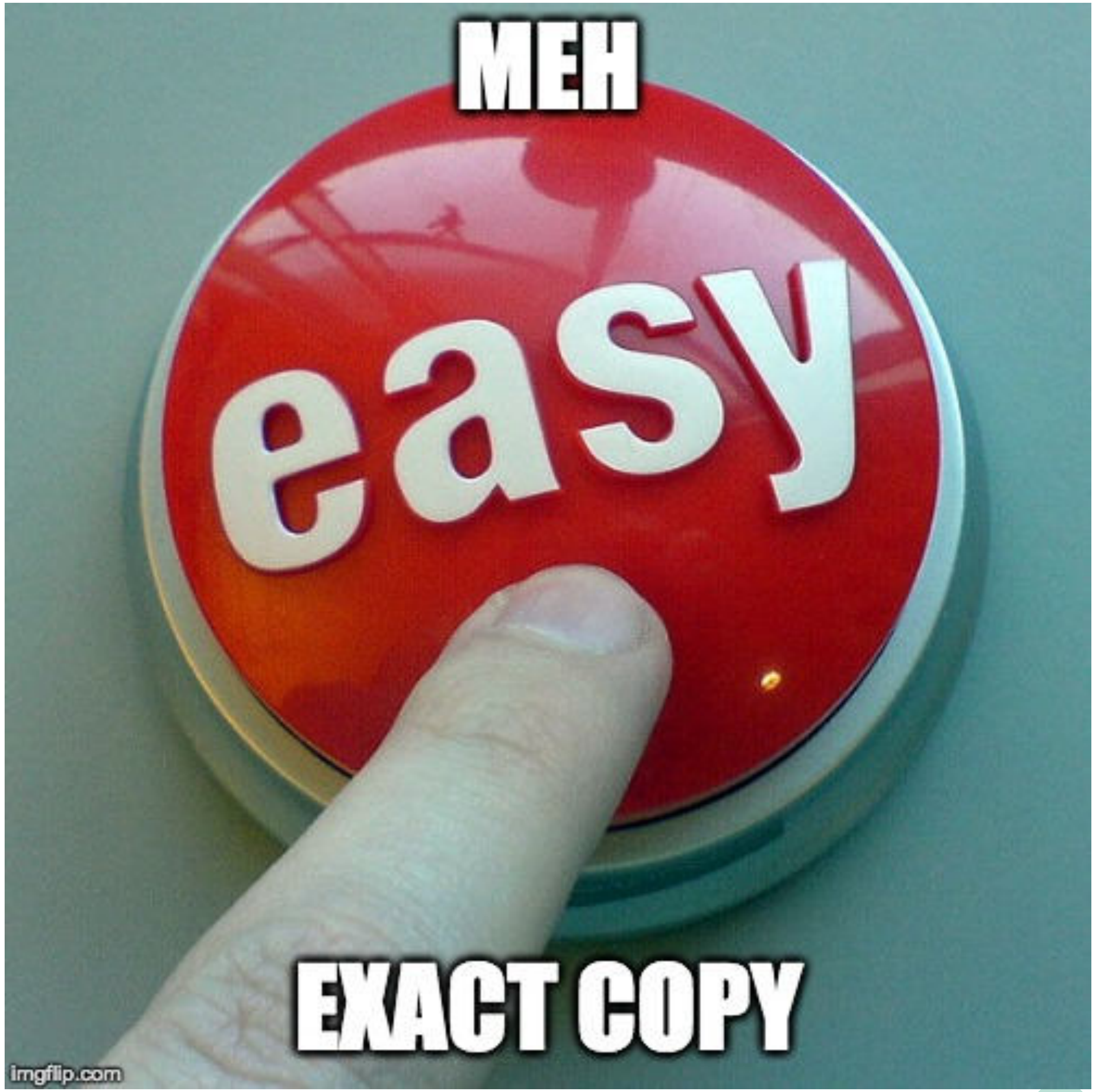
Several types of equiprobability ambiguitity may cause your project to perform less well overall in search

Different types of ambiguity can impact SEO

- Exact duplicate content
- Near duplicate content
- Natural language ambiguity
- Generational cruft based ambiguity
- Machine learned ambiguity ‘lag’
- Dot-to-dot ambiguity
- Semantic heterogeneity
- Location based ambiguity

Exact duplicate content

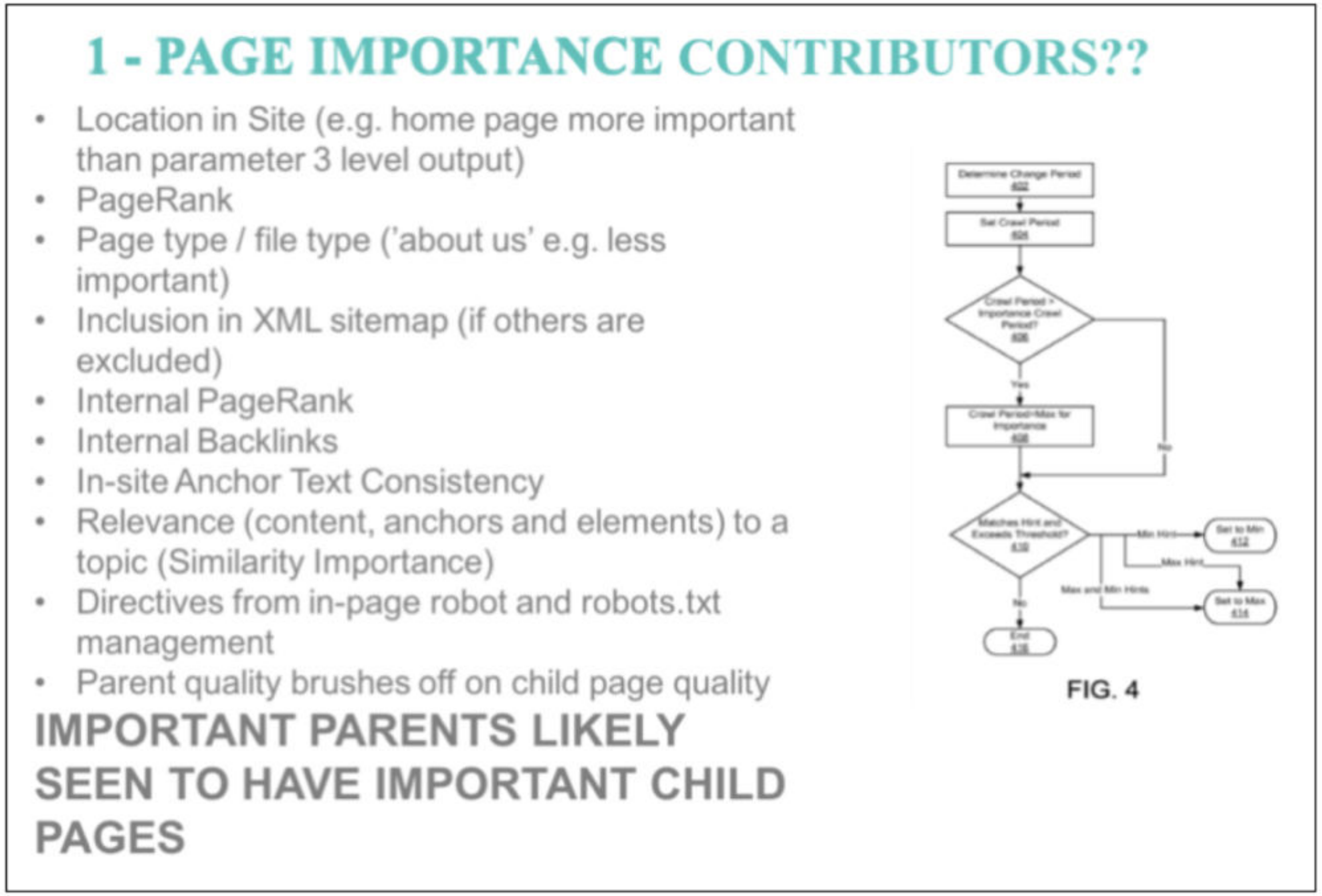
WEB PAGE A
WEB PAGE B
Corporate needs you to find the differences between this picture and this picture.
They're the same picture.
imgflip.com


MEH
easy
EXACT COPY
imgflip.com

Google probably chooses the one with highest ‘probability’

- Most linked to internally
- Most linked to externally
- Included in XML sitemap
- HTTPS
- Prettier URLs

Maybe one or several of many ‘importance’ factors



1 - PAGE IMPORTANCE CONTRIBUTORS??

- Location in Site (e.g. home page more important than parameter 3 level output)
- PageRank
- Page type / file type (’about us’ e.g. less important)
- Inclusion in XML sitemap (if others are excluded)
- Internal PageRank
- Internal Backlinks
- In-site Anchor Text Consistency
- Relevance (content, anchors and elements) to a topic (Similarity Importance)
- Directives from in-page robot and robots.txt management
- Parent quality brushes off on child page quality

IMPORTANT PARENTS LIKELY SEEN TO HAVE IMPORTANT CHILD PAGES

FIG. 4

Google works this out VERY quickly

Near-duplicate content

Easy(ish)

It starts off well enough… but over time

Google realizes these pages are ‘mostly’ the same as others

Some symptoms

- Many matching shingles
- Quilting
- ‘Borrowed’ content at scale
- Feeds
- Big Boilerplate / little main value
- Data driven sites with little value

Sometimes these are sites trying to make a URL footprint well beyond their contribution to a positive ‘Network Effect’

Spoofing the ‘Network Effect’

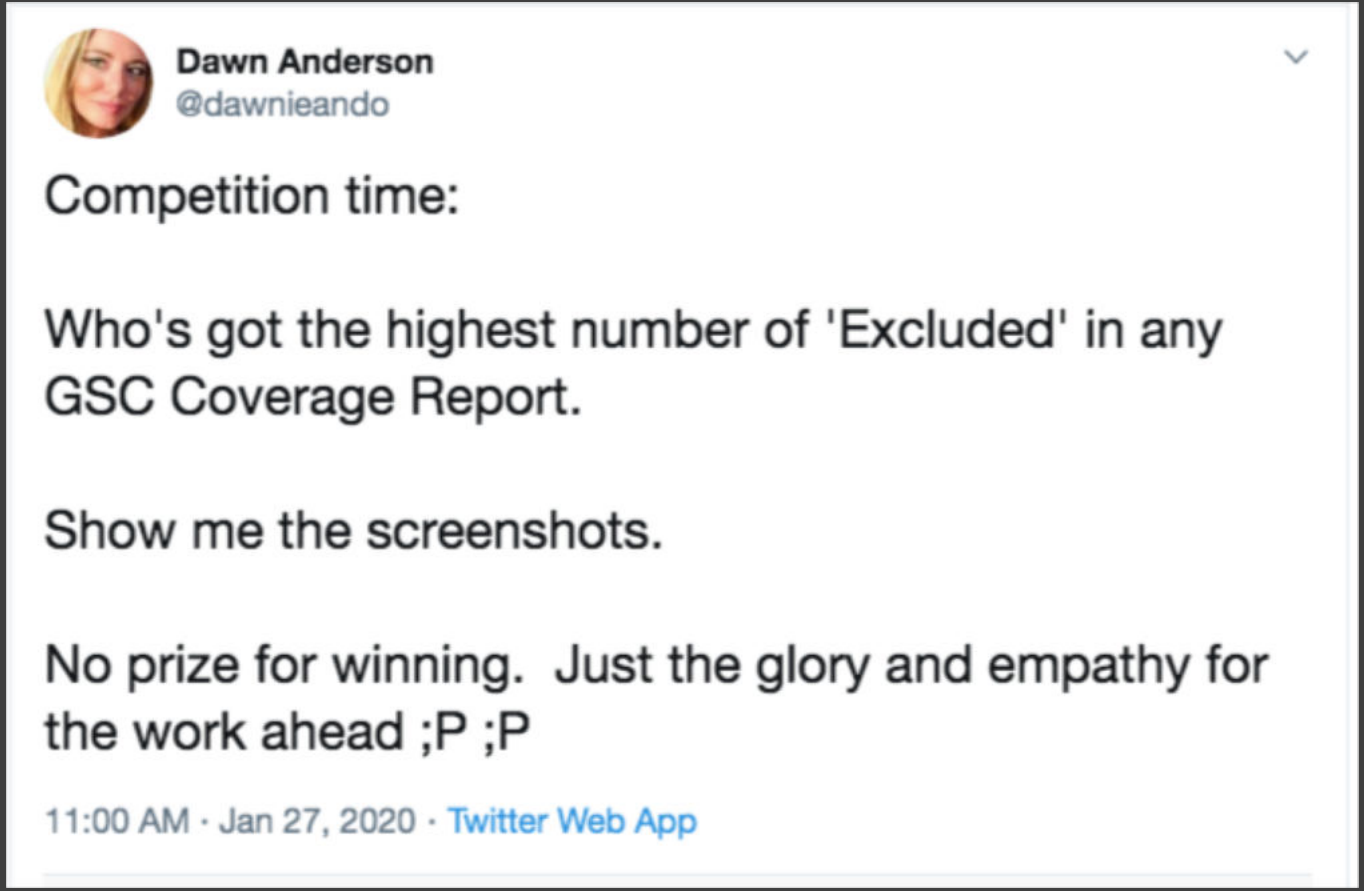
After some crawling & sampling Google puts the pieces of the puzzle together

And begins to exclude pages from the index

Exclusion still happens… A LOT

Particularly on behemoth sites

I ran a little Twitter competition with only empathy as the prize



Dawn Anderson
@dawnieando

Competition time:

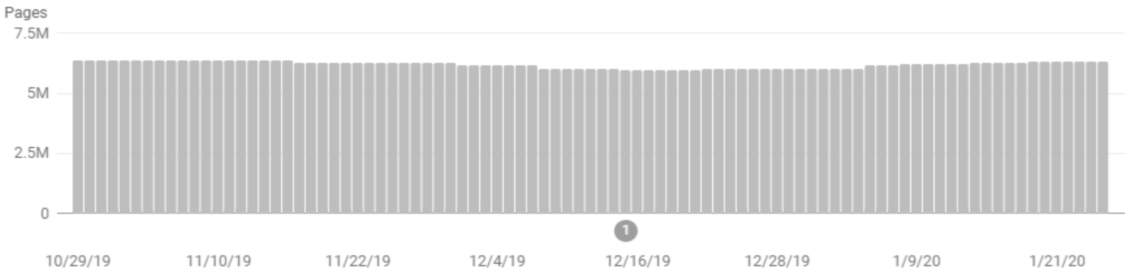
Who's got the highest number of 'Excluded' in any GSC Coverage Report.

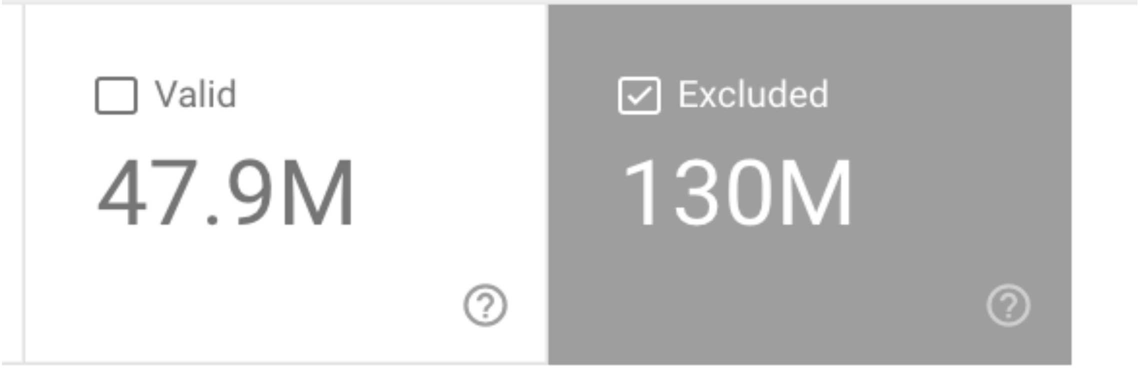
Show me the screenshots.

No prize for winning. Just the glory and empathy for the work ahead ;P ;P

11:00 AM · Jan 27, 2020 · Twitter Web App

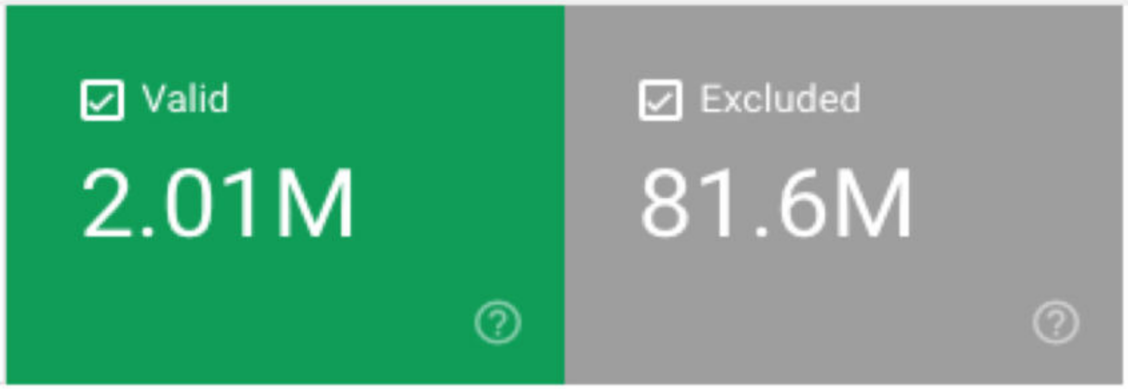
Some of the GSC ‘exclusion’ numbers were huge… millions





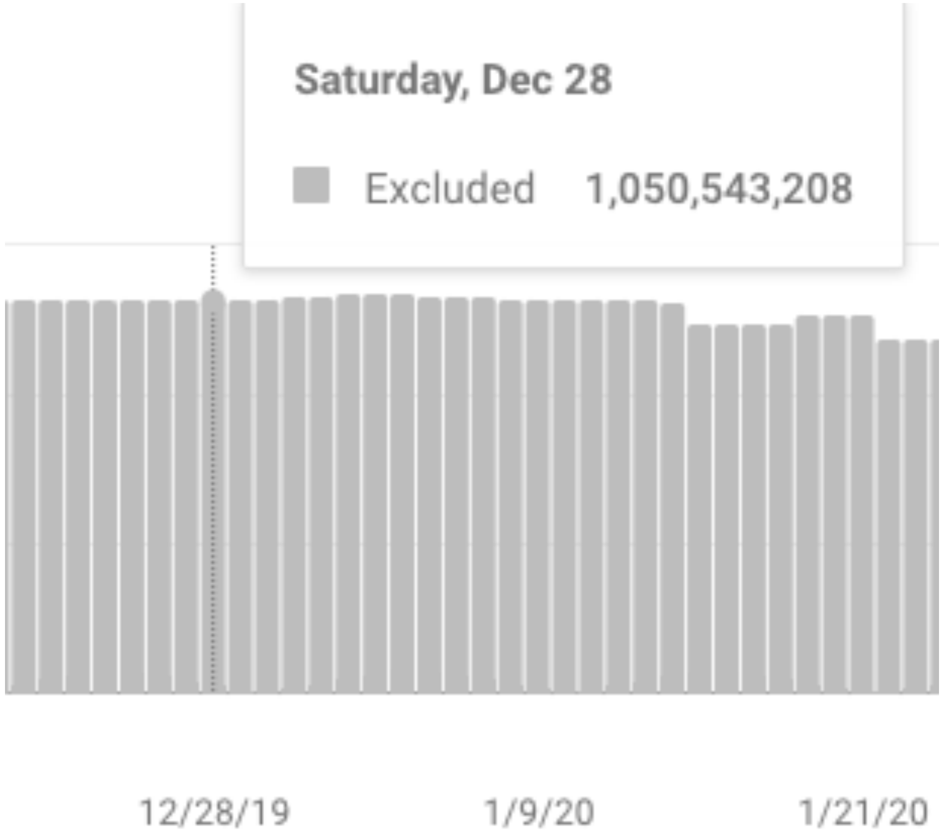
□ Valid
47.9M

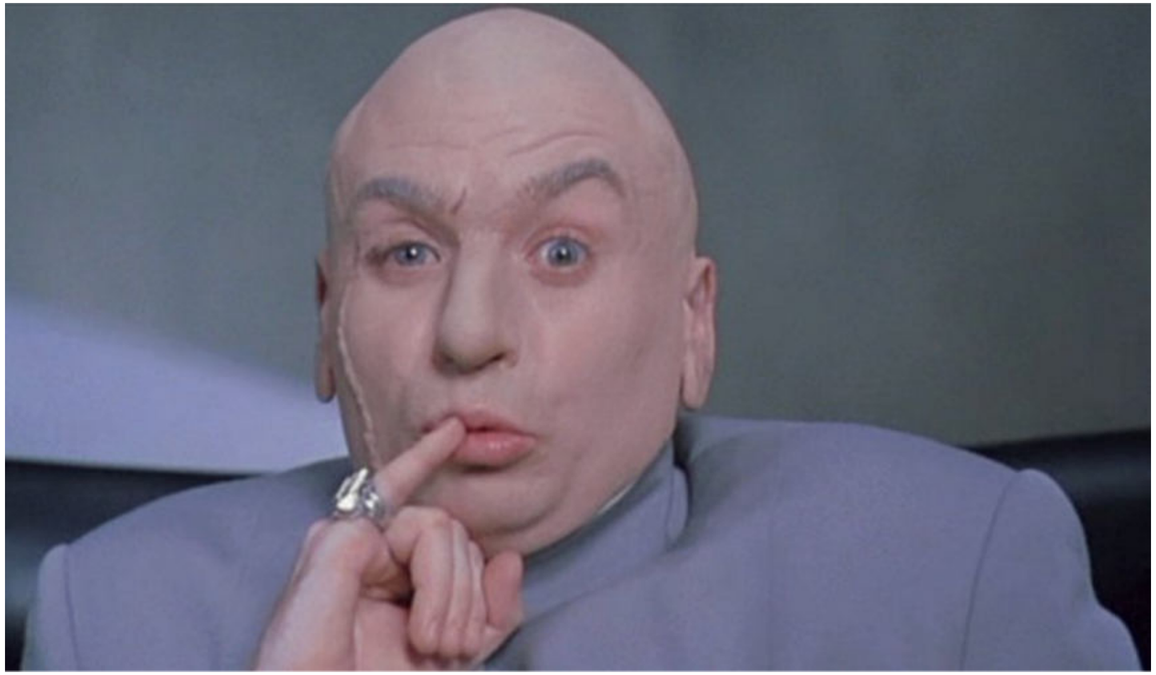
☑ Excluded
130M



☑ Valid
2.01M

☑ Excluded
81.6M

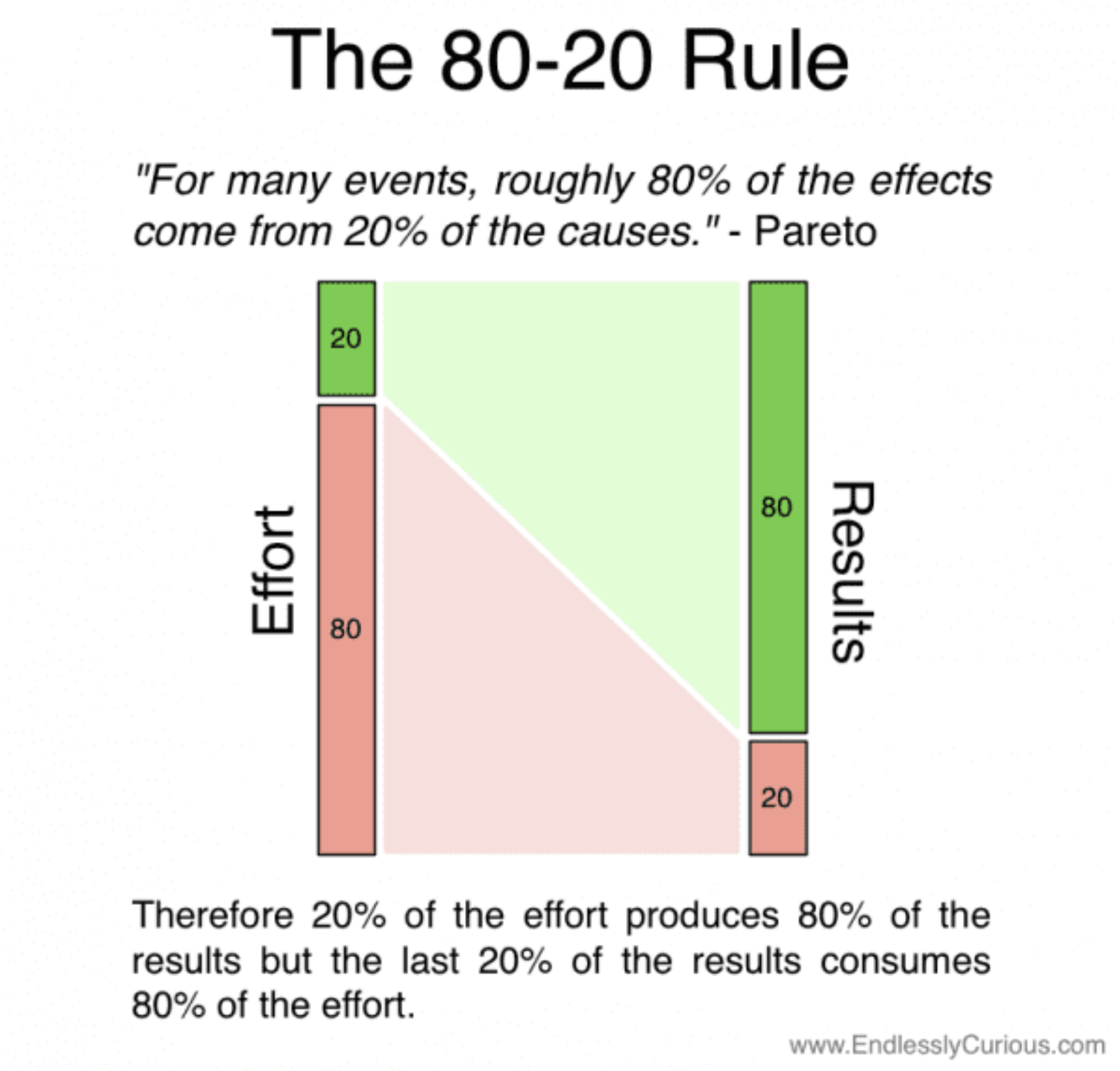




And even over… one billion

I would hazard a guess…The majority of the largest websites in the world have a LOT of near-duplicate ‘transactional’ pages

Crawl demand probably follows the Pareto Principle



The 80-20 Rule

"For many events, roughly 80% of the effects come from 20% of the causes." - Pareto

Therefore 20% of the effort produces 80% of the results but the last 20% of the results consumes 80% of the effort.

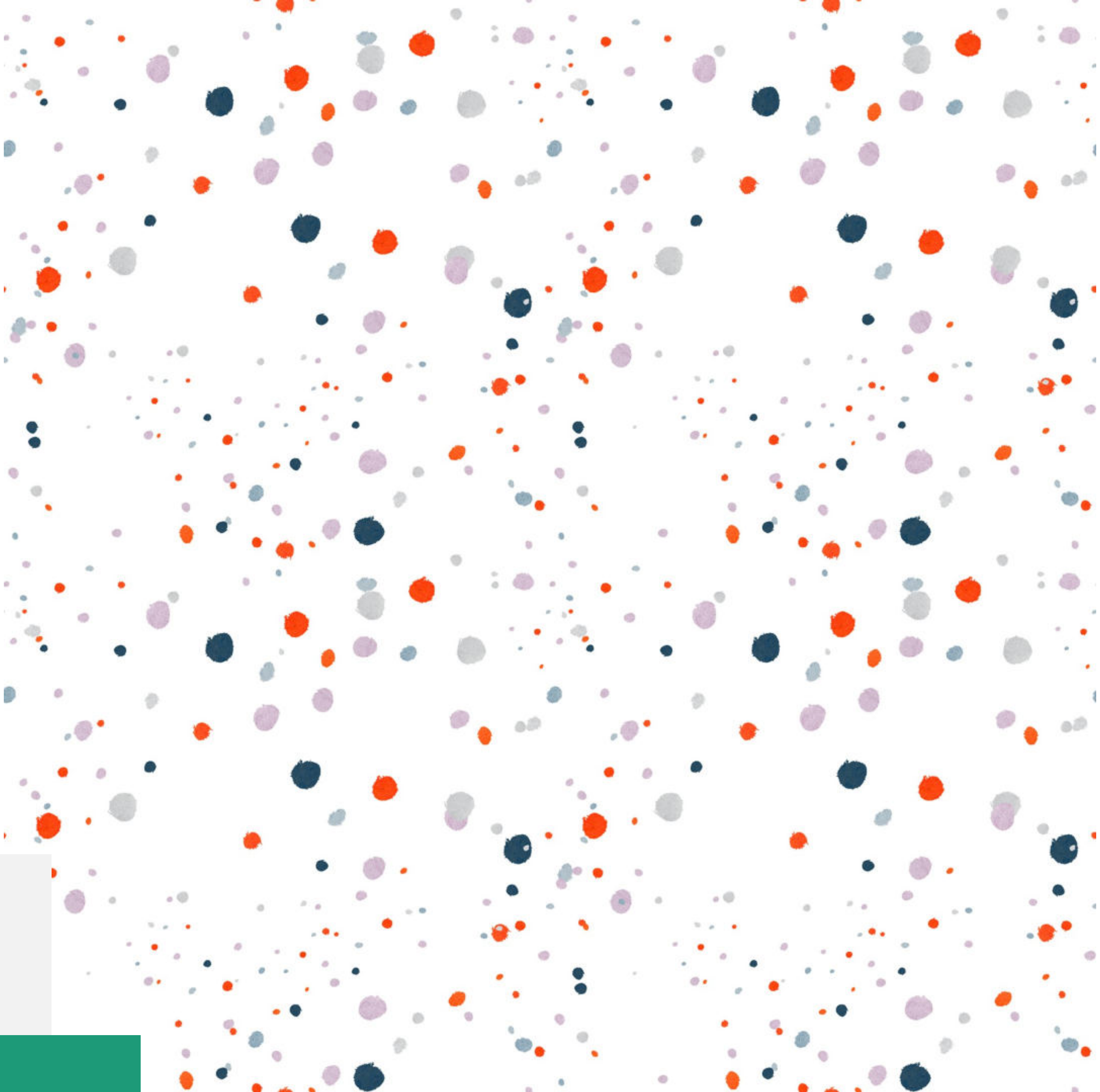
www.EndlesslyCurious.com

Probably 20% of the URLs satisfy 80% of the demand & contribute well to ‘The Network Effect’

Since search engines probably work with positive ‘impact’ value considered

Google will likely use mostly ‘Sampling’ on those heavy ‘excluded URL’ sites

Looking for small samples of content & URL patterns… and… ‘Discovered, not crawled’ will be high



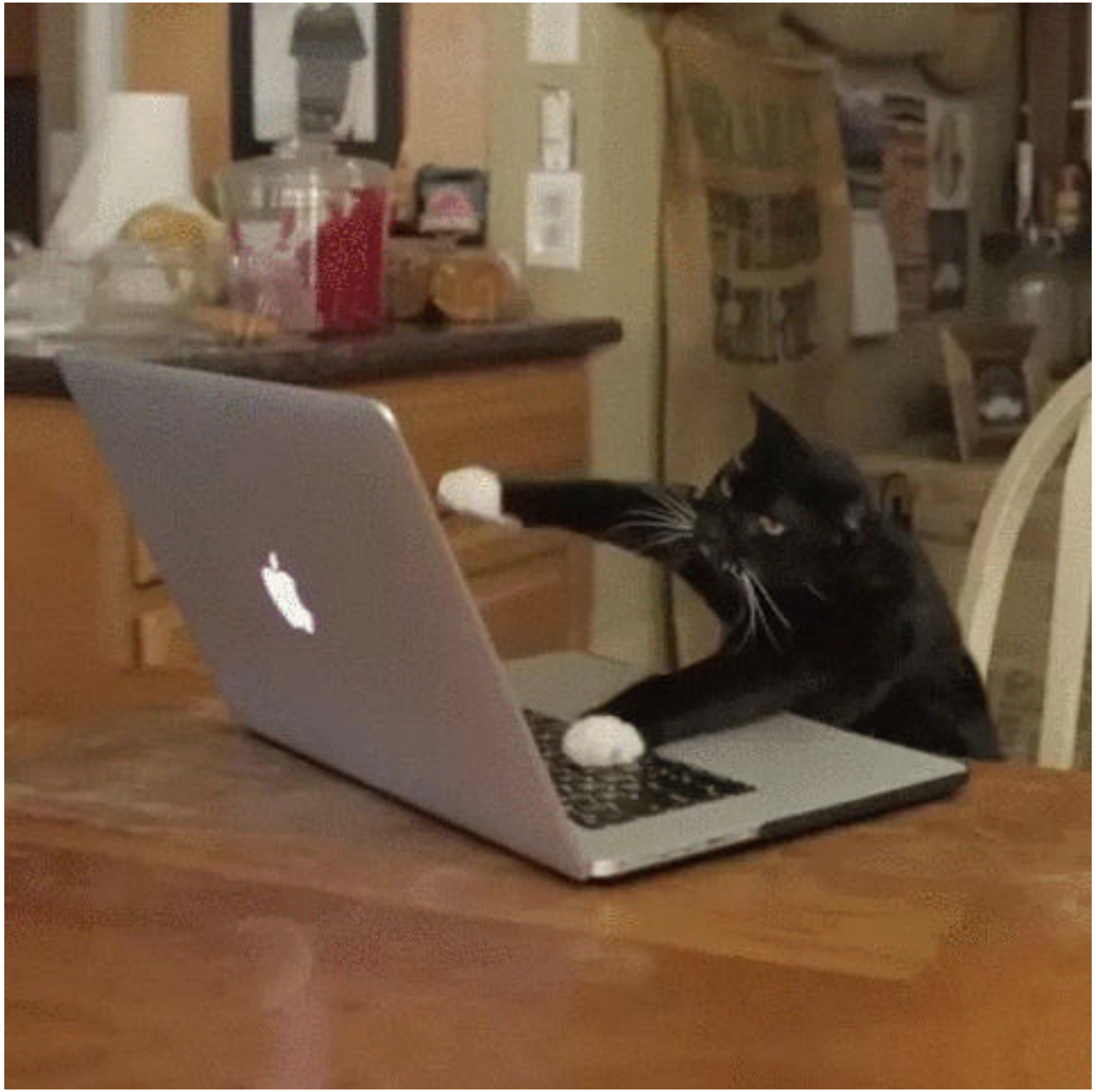
'Discovered, not crawled' in GSC is Google saying…

They know what’s down that site section or URL parameter path

The verdict is in

Those URLs will be ‘starved’ of crawl

However many times you submit and inspect



You have your work cut out once Google puts the pieces of that puzzle together

Natural language ambiguity

For machines words are problematic. Ambiguous… polysemous… synonymous

In spoken word it is even worse because of homophones and prosody

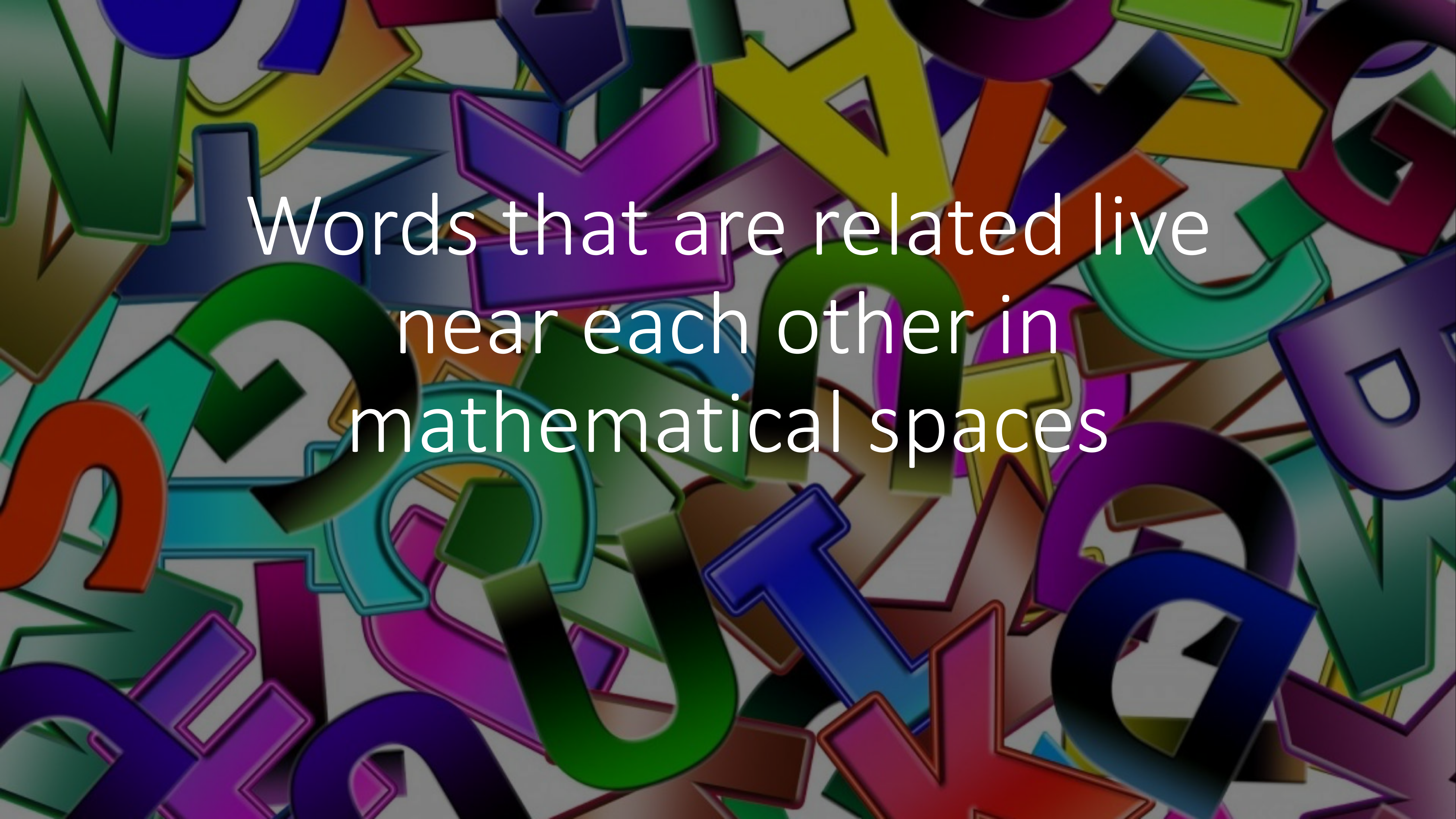
Like “four candles” and “fork handles”

Many words have multiple meanings. Like “like” can be 5 possible parts of speech (POS)

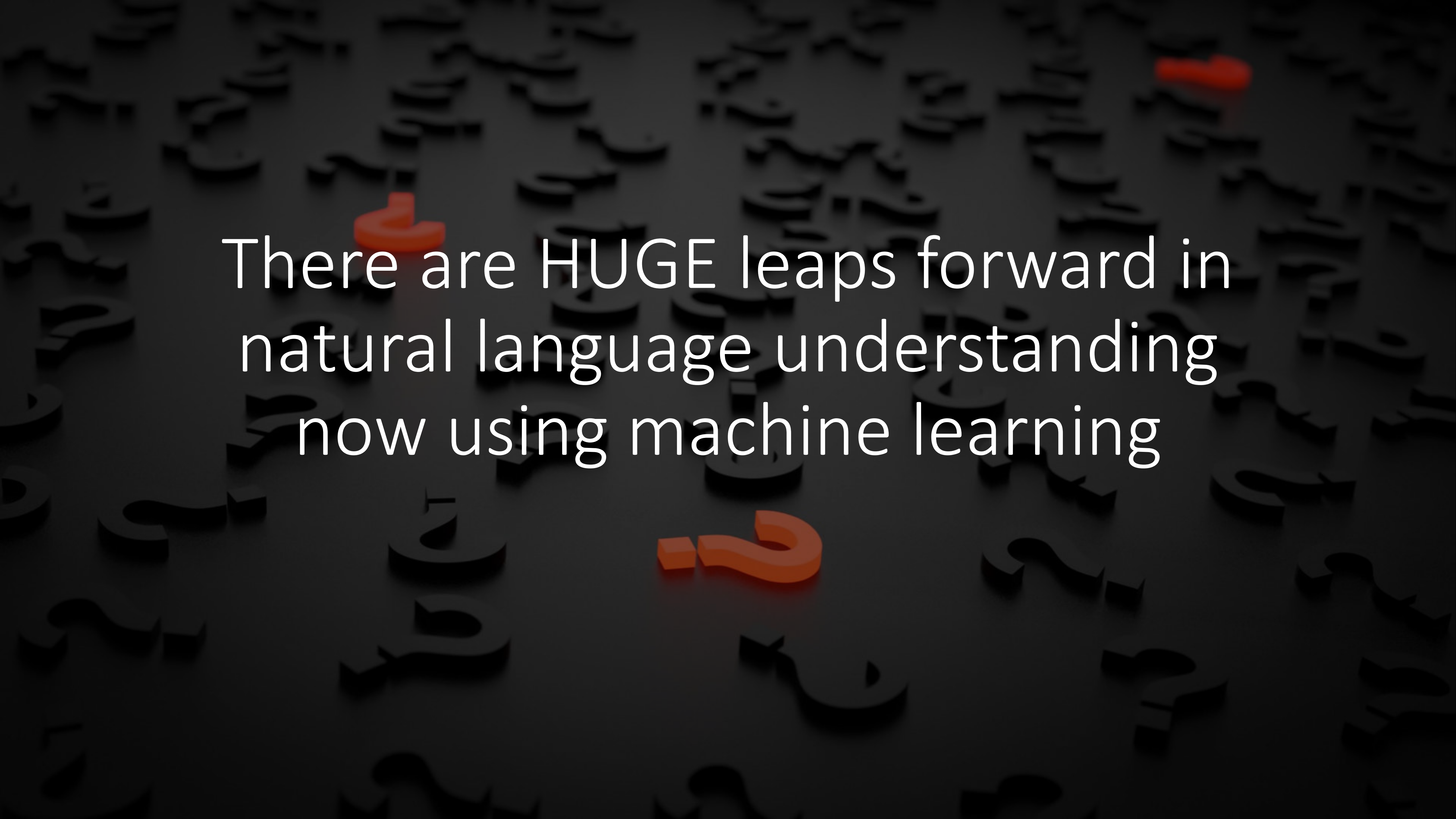
Word’s context helps enormously

And disambiguating words is getting easy(ier)

"You shall know a word by the company it keeps" (Firth, 1957)



Words that are related live near each other in mathematical spaces



There are HUGE leaps forward in natural language understanding now using machine learning

Accelerated mostly by Google’s BERT (Bi-Directional Encoder Representations from Transformers)

Co-occurrence in content helps A LOT

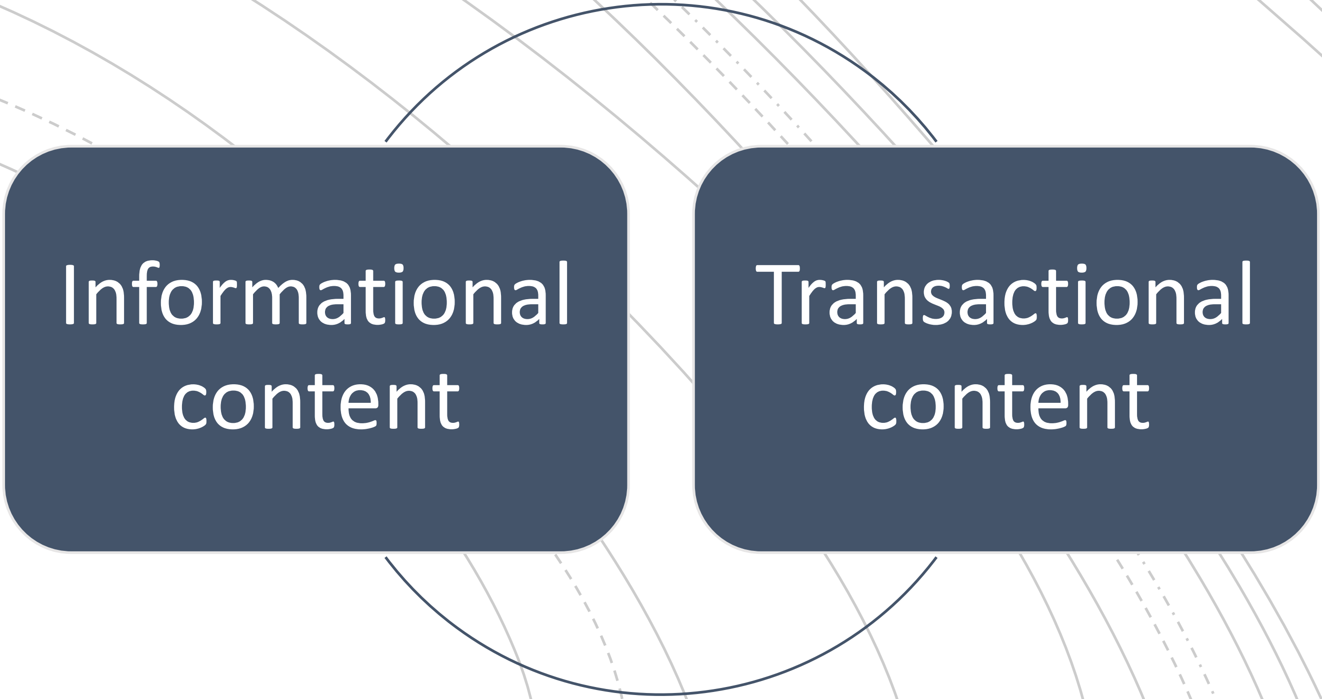
- In the page itself
- In the interconnected pages
- In the subcategorisation
- In the site sections
- In the external relationships
- In the domain ontology as a whole

Relatedness

First level relatedness

Second level relatedness

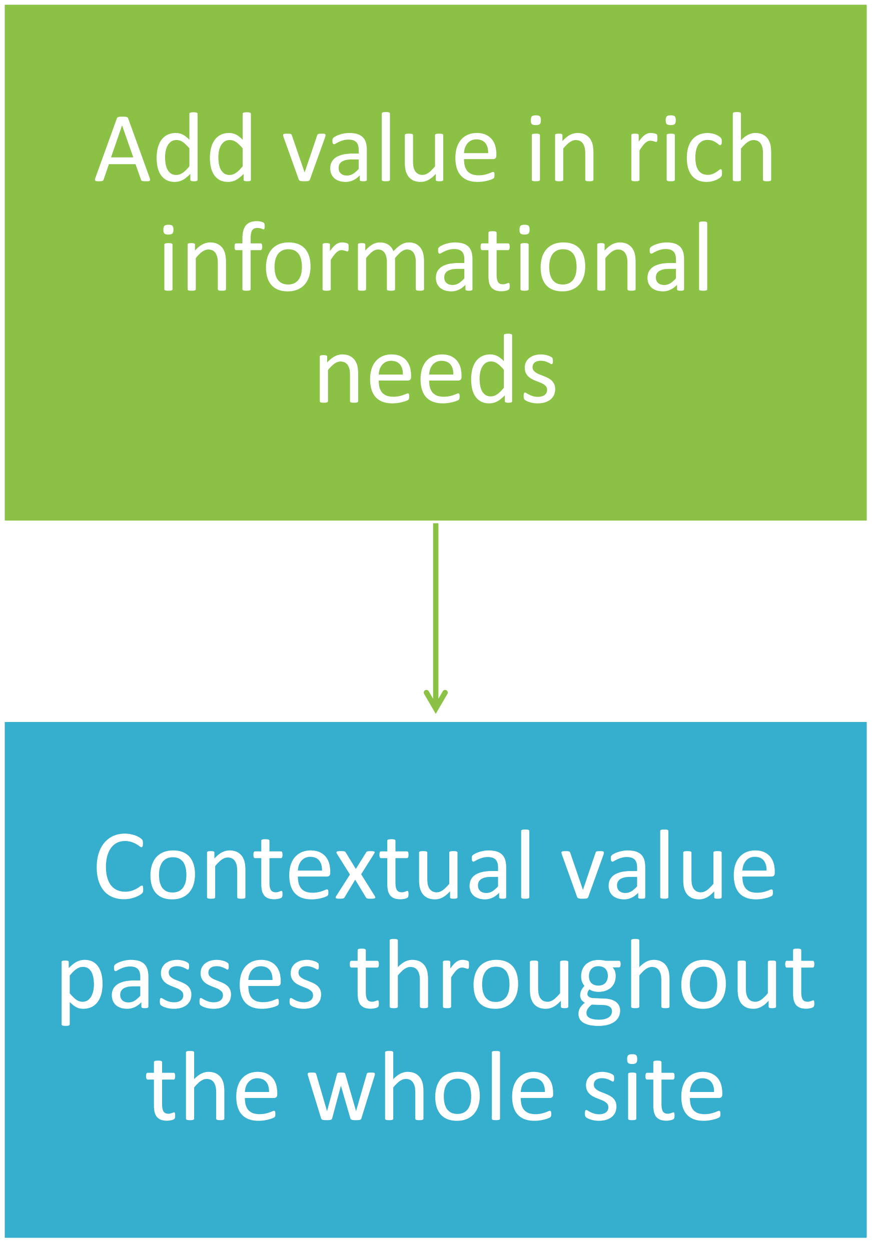
Informational content adds to the contextual 2nd level relatedness of transactional content



Informational content

Transactional content

If you can help with ‘informational needs’ you can ‘probably’ help with transactional needs



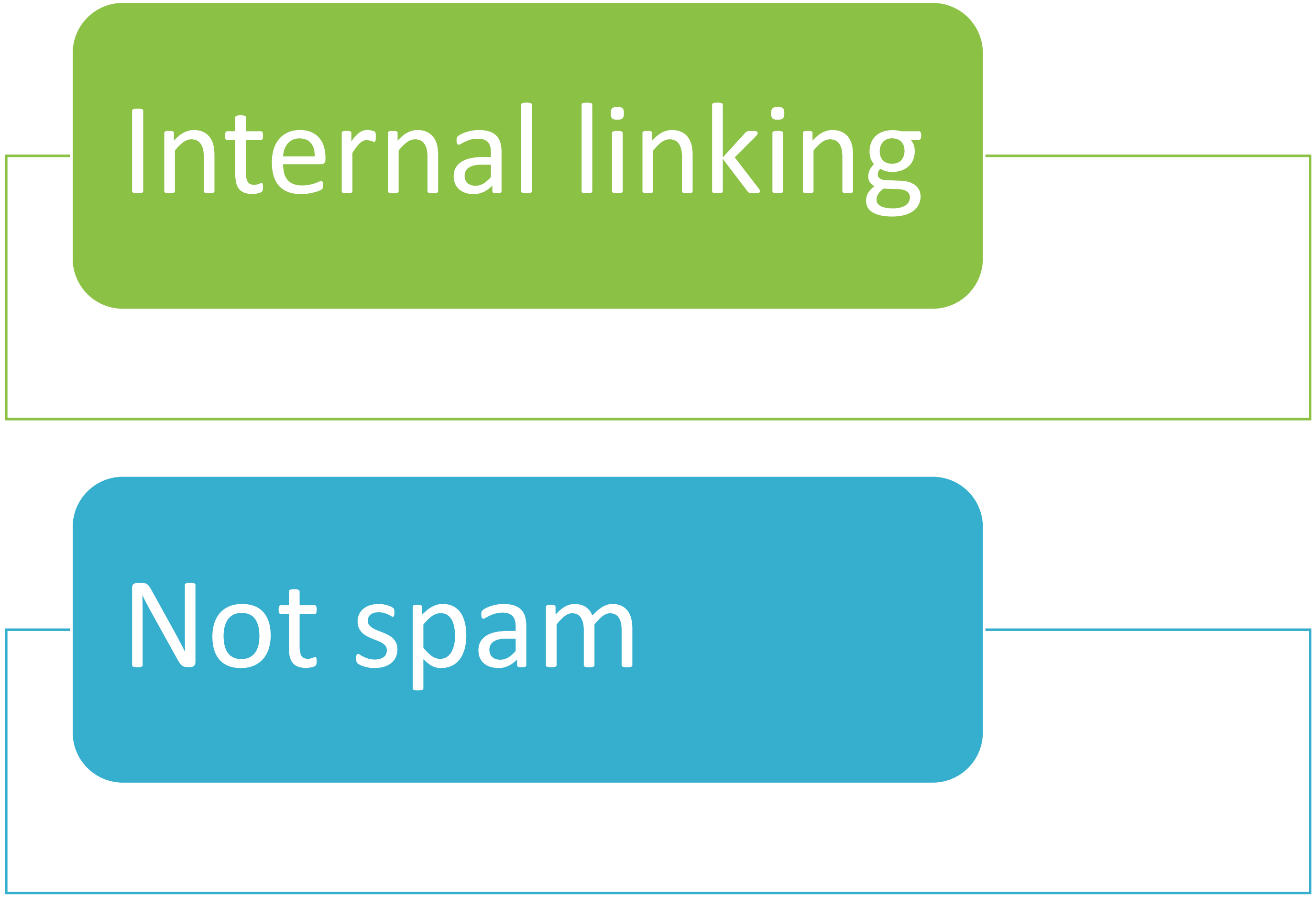
Add value in rich informational needs

Contextual value passes throughout the whole site

Adding Contextual Value Does Not Mean

Adding	Adding loads of content below ecommerce pages
Adding	Adding LSI keywords in ecommerce pages

High topical relatedness & context vectors



Internal linking

Not spam

Utilise conceptual ‘nearest neighbours’ well

Identify content which is ‘very’ closely related

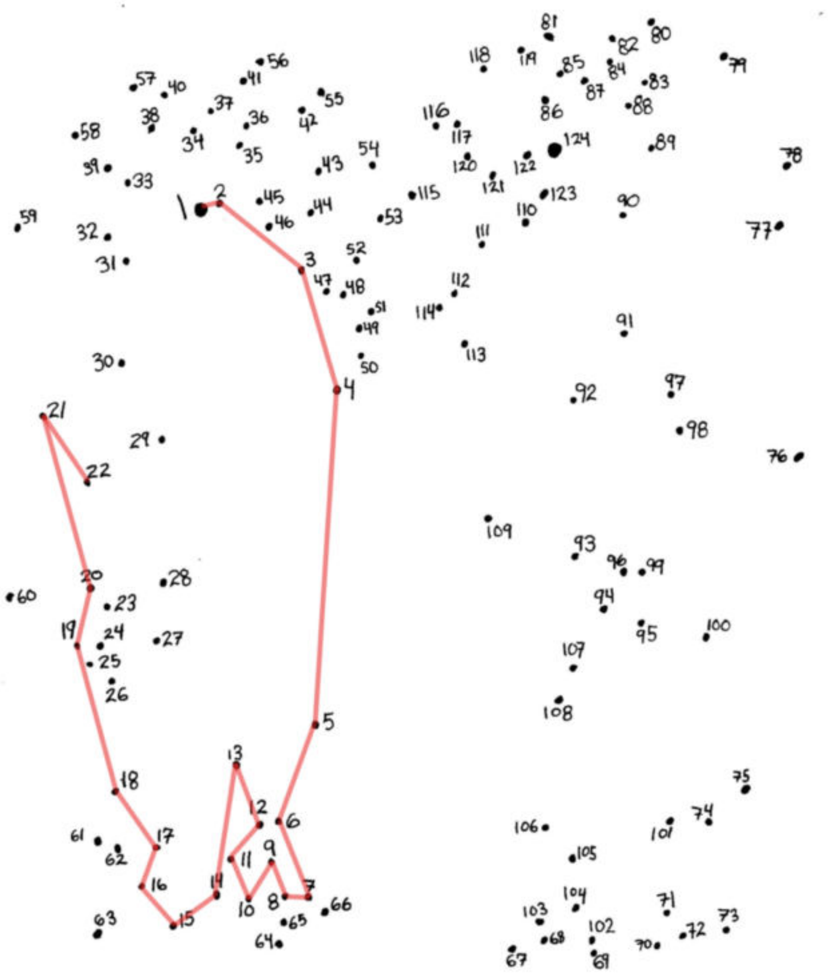
Merge content which is ‘too’ conceptually very similar with no separate demand

Utilise ‘Overflow SEO’ as demand & content naturally grows

Google may realise you can ‘probably’ help with transactional queries

Dot-to-dot ambiguity

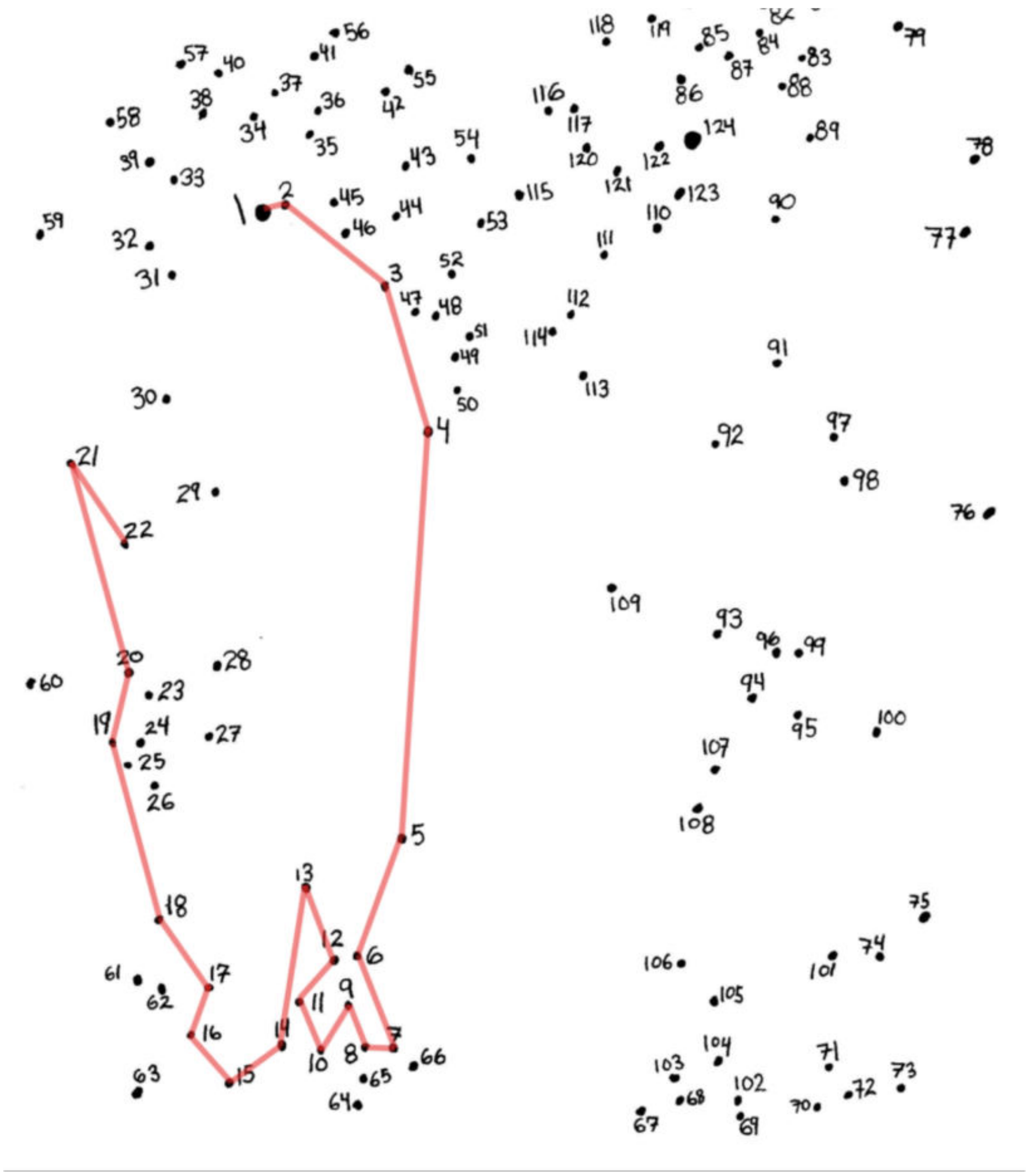
Who knows what this dot to dot puzzle is creating?

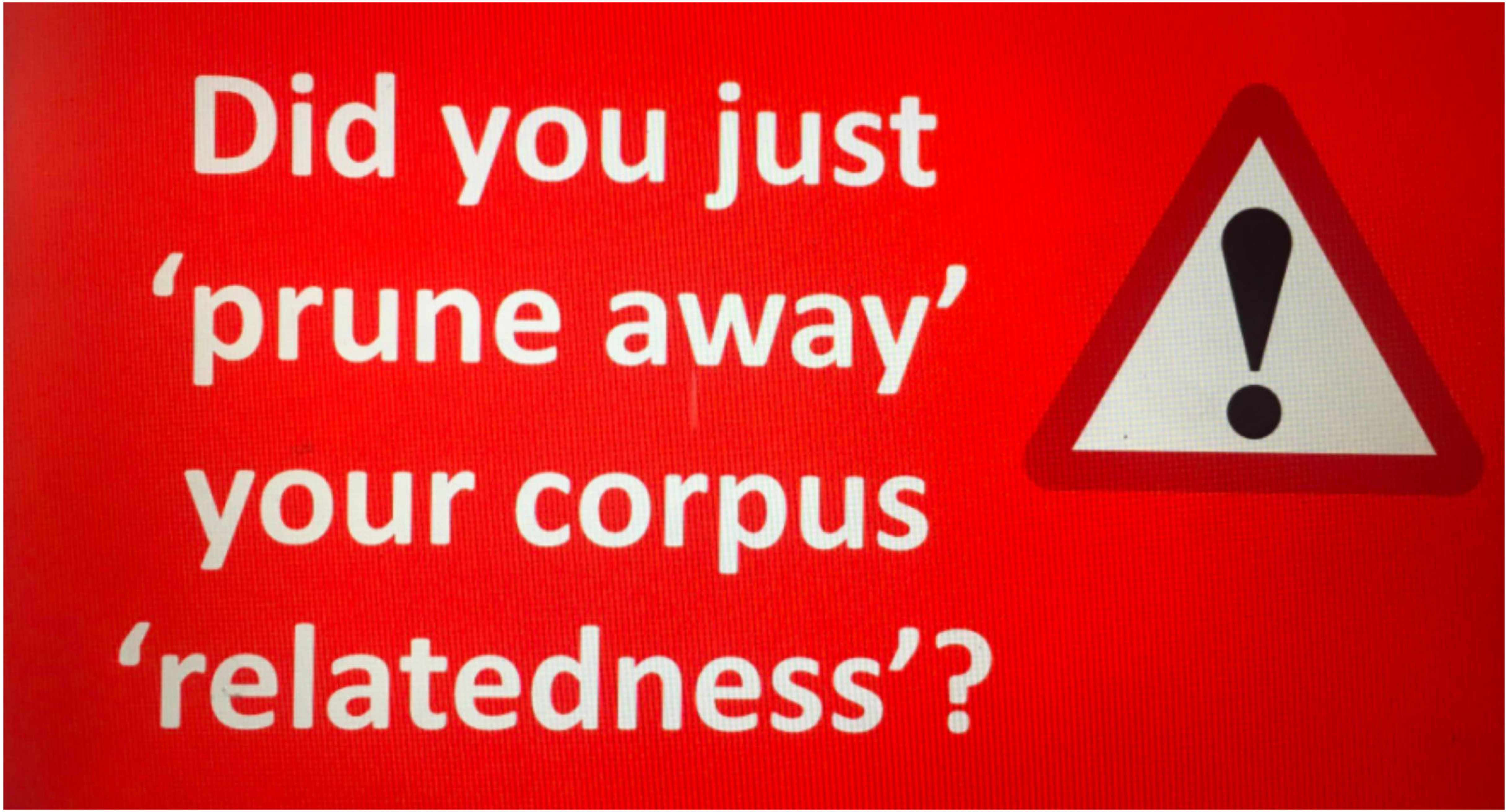


Absolutely no-one

If you chop away or migrate half of your website then equiprobability will likely be a problem

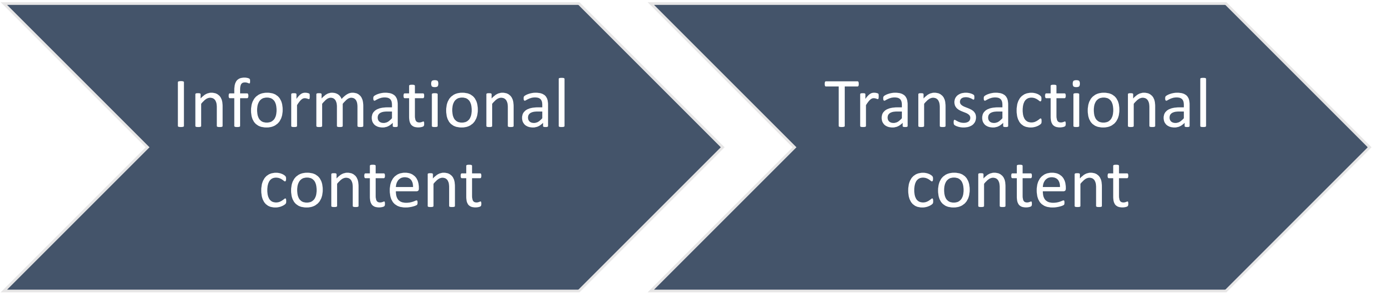
A half drawn website is not the same as a fully drawn previous website?



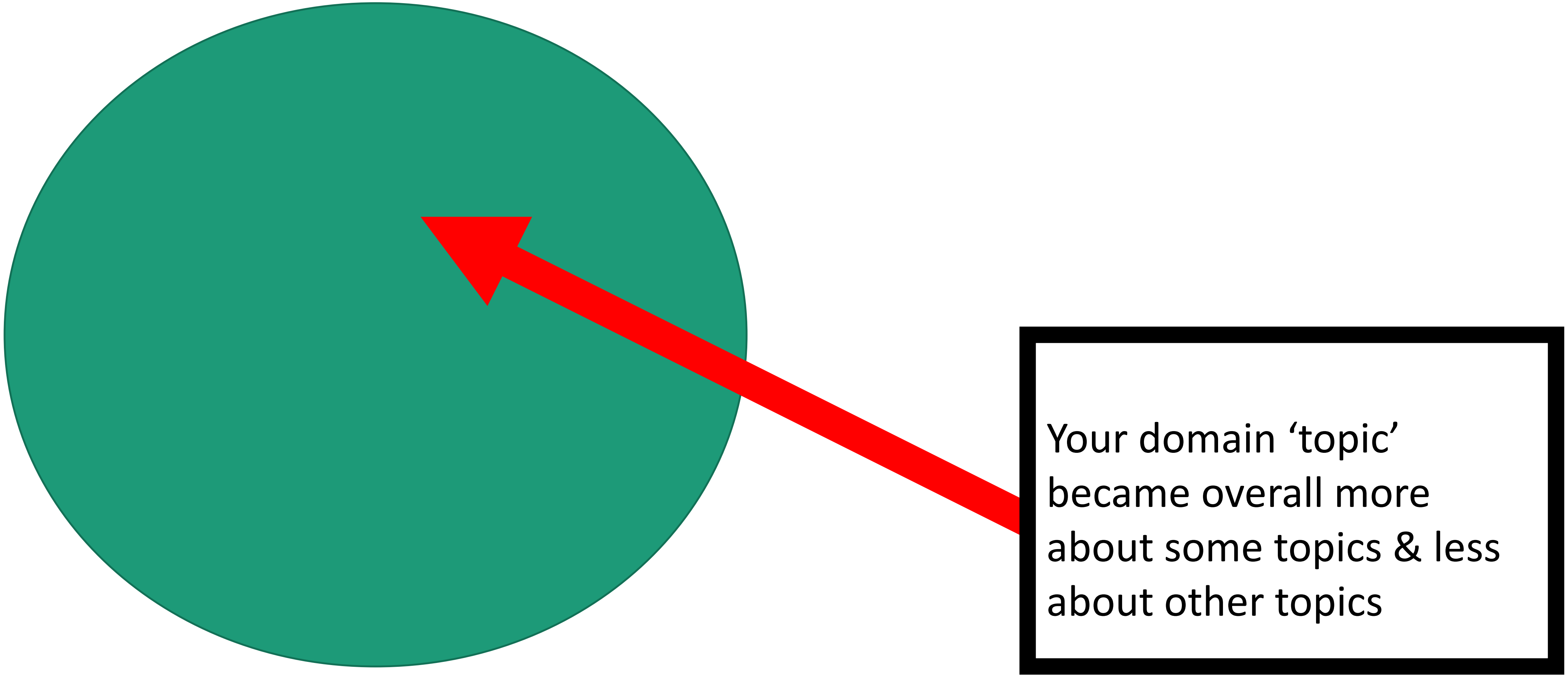


Did you just ‘prune away’ your corpus ‘relatedness’?

Informational content adds to the contextual 2nd level relatedness of transactional content



Although you may rank ‘marginally’ better for fewer things



Your domain ‘topic’ became overall more about some topics & less about other topics

Prune, Merge, Improve, Archive

01

Prune – Only ballast

02

Merge – Highly semantically related content

03

Improve – Evergreen pages worth the effort

04

Archive – Older pieces of temporal content (like a library)

Avoid removing or hiding user generated content (except spam)

Google puts the pieces of a half-site puzzle together (often with bad results)

Cruft based ambiguity

Generational cruft contributes to equiprobability issues

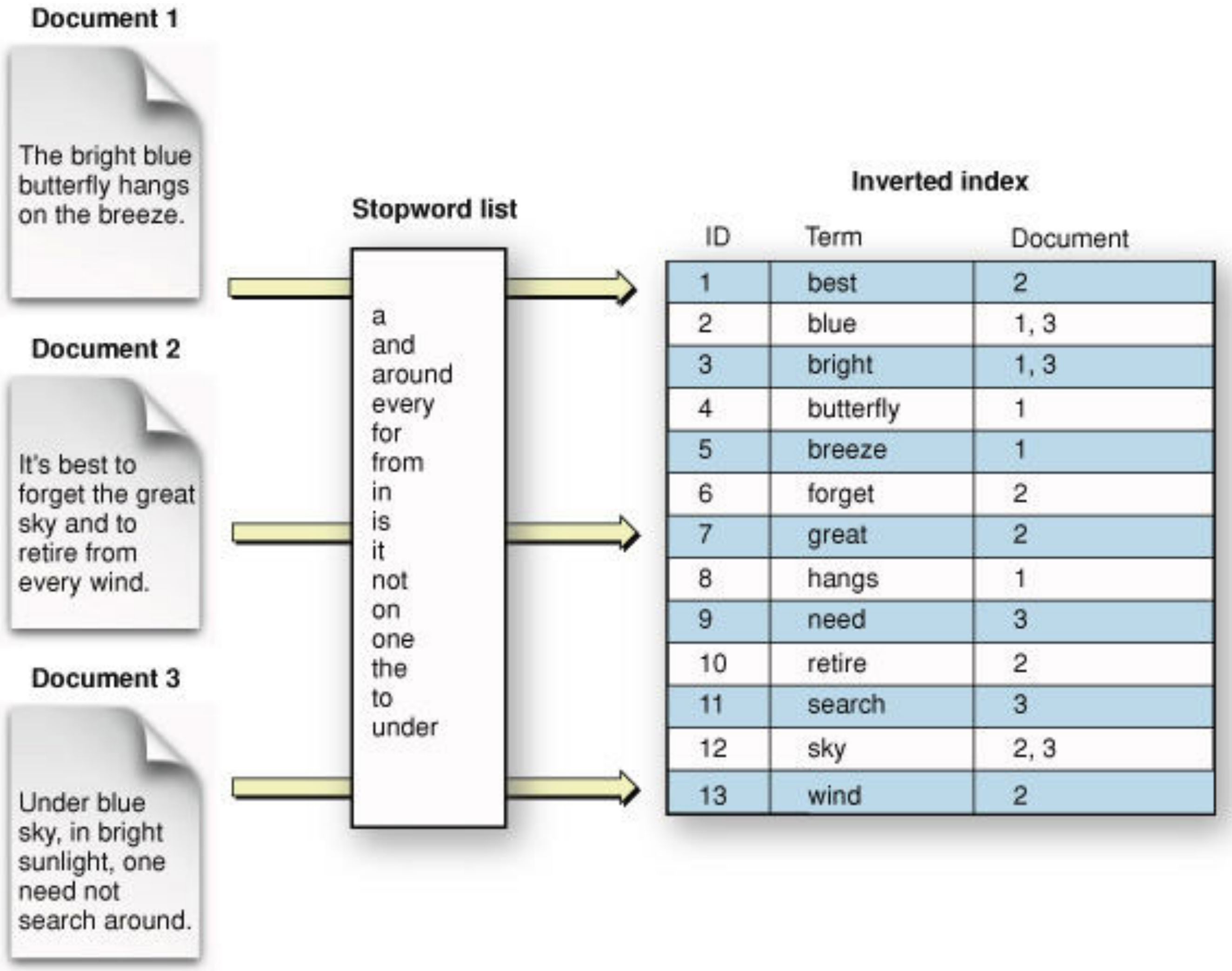
Just a few examples… legacy is a problem

- 301 redirect chains
- Inconsistent 301s
- No 301s at all
- Canonical points to a 301
- Canonical points to a no-index
- No-index points to a canonical

301 Should Mean The Resource Moved

You're supposed to be telling search engines where the words (tokens), topics, concepts & entities went

You’re saying things moved so they can be re-filed



Document 1

The bright blue butterfly hangs on the breeze.

Document 2

It's best to forget the great sky and to retire from every wind.

Document 3

Under blue sky, in bright sunlight, one need not search around.

Stopword list

a
and
around
every
for
from
in
is
it
not
on
one
the
to
under

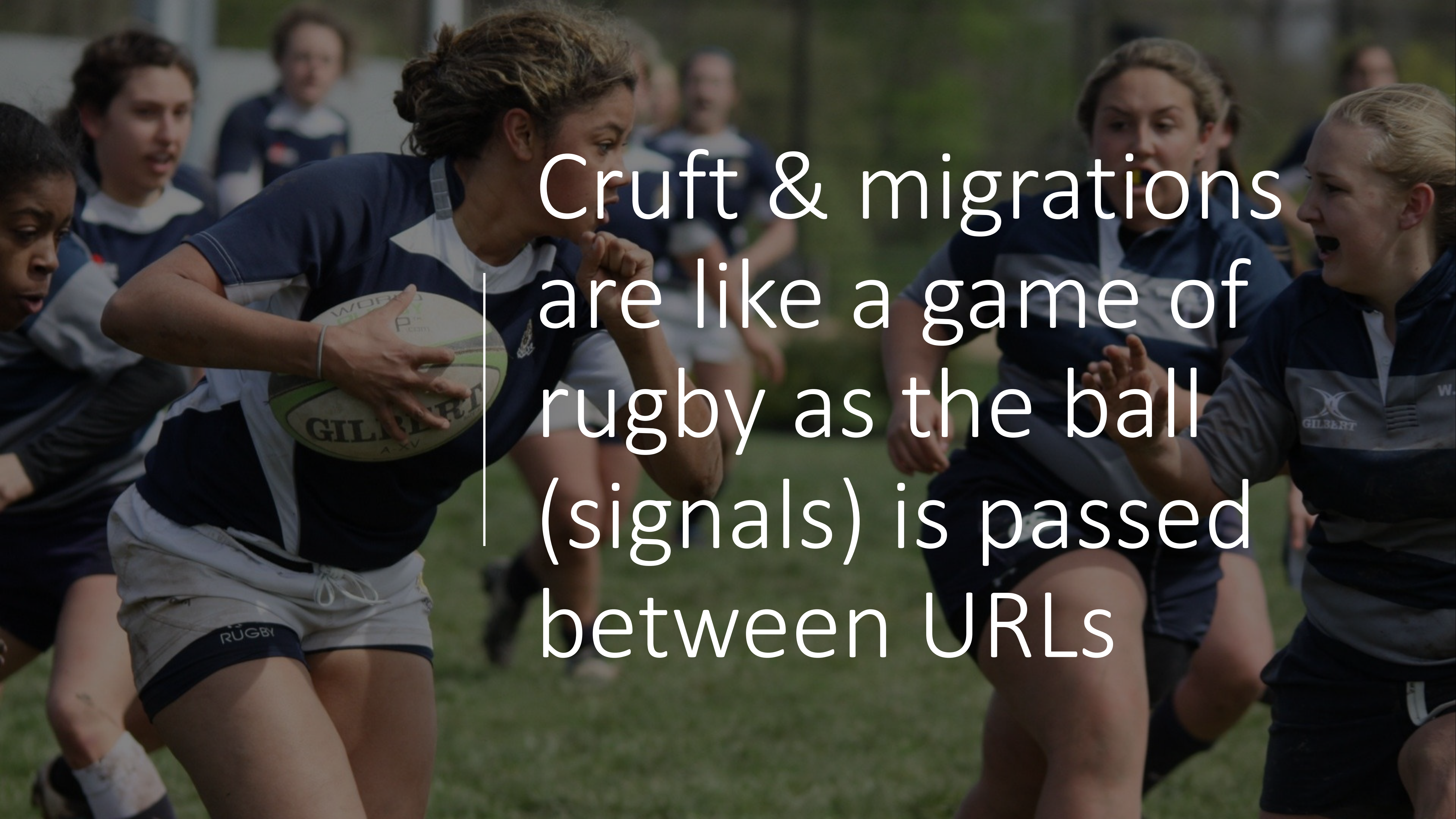
Inverted index

ID	Term	Document
1	best	2
2	blue	1, 3
3	bright	1, 3
4	butterfly	1
5	breeze	1
6	forget	2
7	great	2
8	hangs	1
9	need	3
10	retire	2
11	search	3
12	sky	2, 3
13	wind	2

But often there is little or no match at all

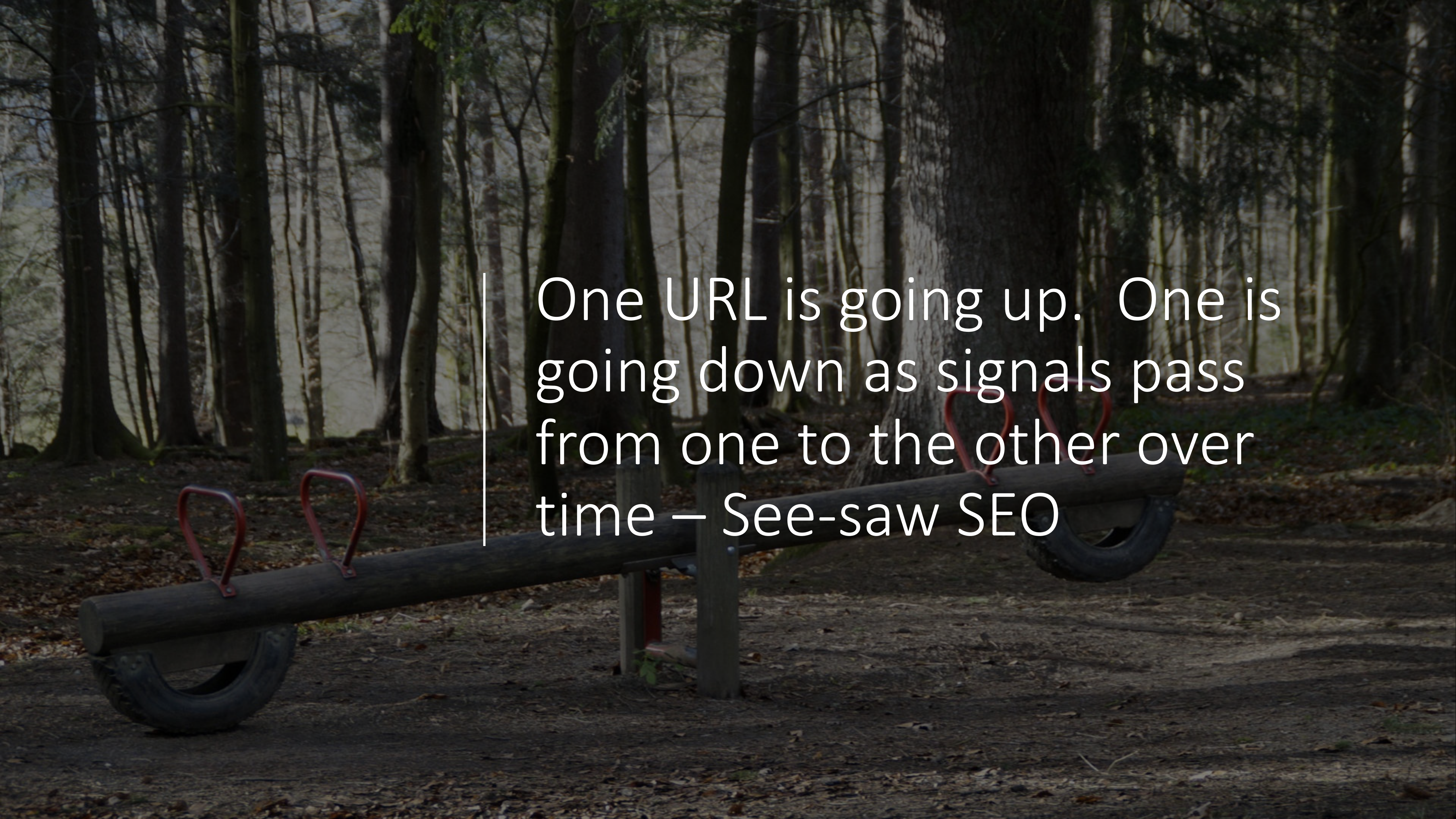
Borderline or true ‘soft 404’

Also, any legacy canonicals are now defunct by the redirect status



Cruft & migrations are like a game of rugby as the ball (signals) is passed between URLs

A sign of this is a temporarily wrong target ranking



One URL is going up. One is going down as signals pass from one to the other over time – See-saw SEO

Google puts the pieces of the puzzle together… eventually (or never)

Machine learned ambiguity ‘lag’

ALL of the above ==

Crawl Budget Woes / Learned Quality Patterns

Your reputation precedes you – You are being judged on the past

I've got a truckload of reputation points
that don't buy you anything
imgflip.com

Your quality & crawl will be ‘machine learned’ by ‘Large Numbers’ over time

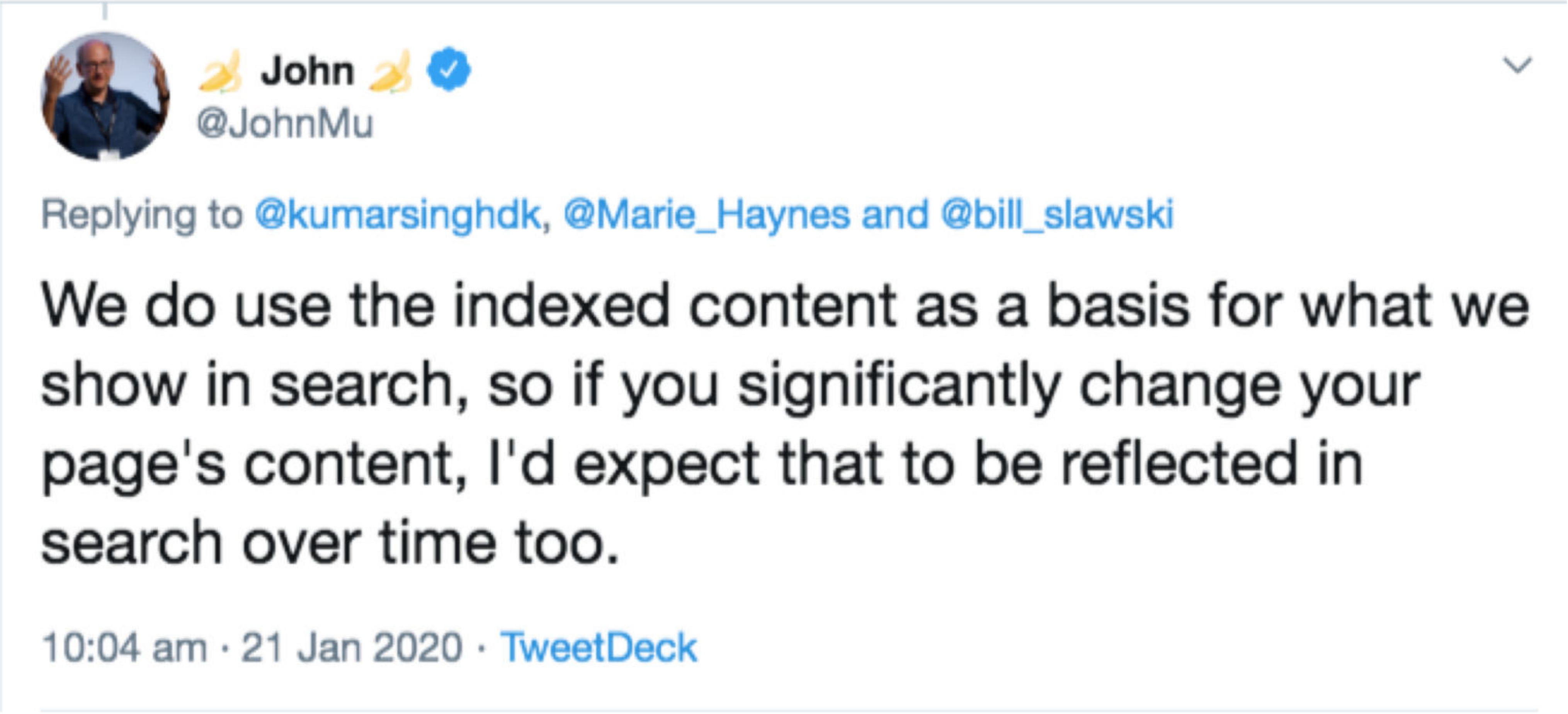
Probably you’ll struggle to get enough crawl for Google to catch up

YOUR SITE IS DOING
Chance
GO DIRECTLY TO JAIL
CRAWL JAIL TIME
DO NOT PASS GO, DO NOT COLLECT $200
imgflip.com

Search engines realise there is no ‘demand’ for your poopy pages’

You’ve probably got URLs which have not been crawled for years

Your performance will be based on what is indexed



🍌 John 🍌 ✓
@JohnMu

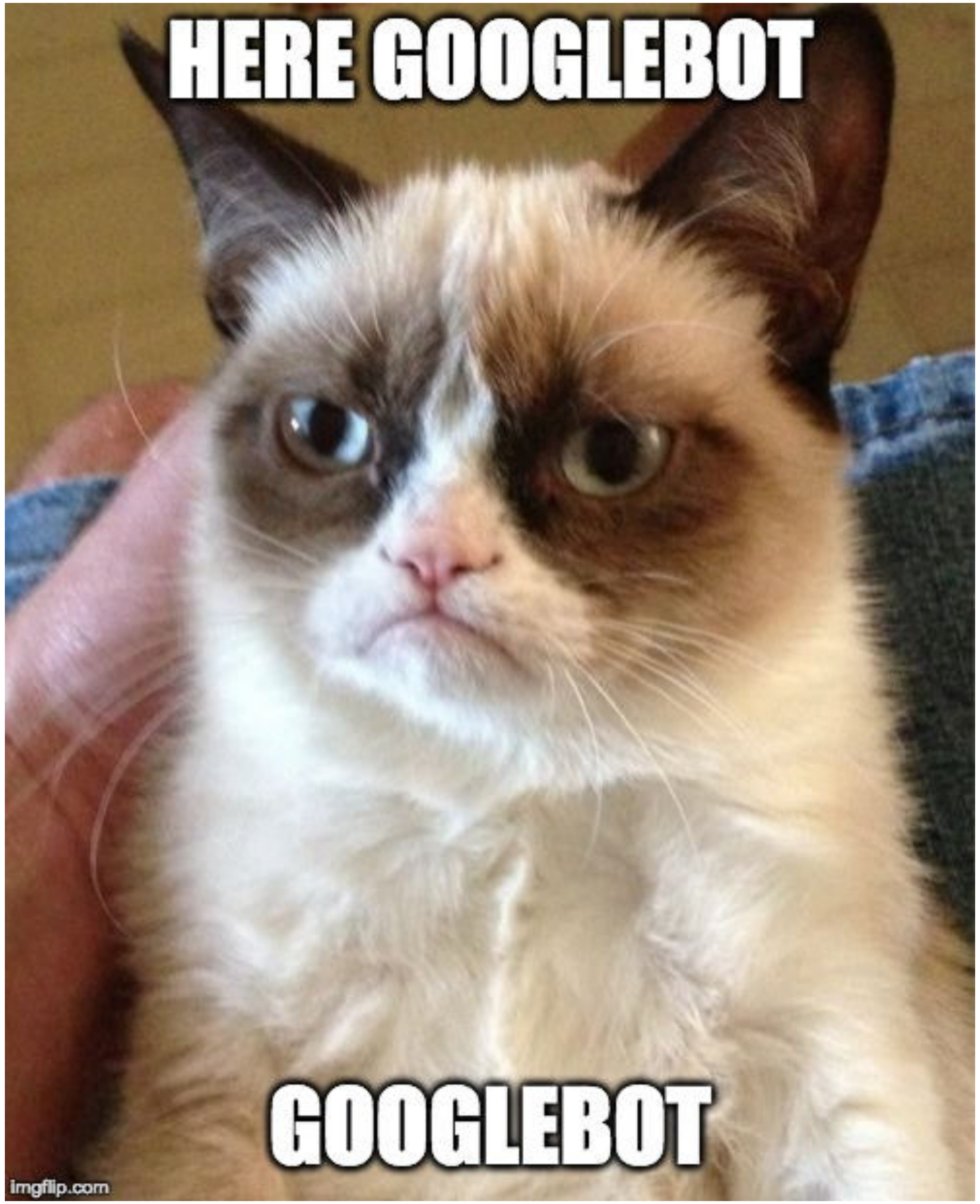
Replying to @kumarsinghdk, @Marie_Haynes and @bill_slawski

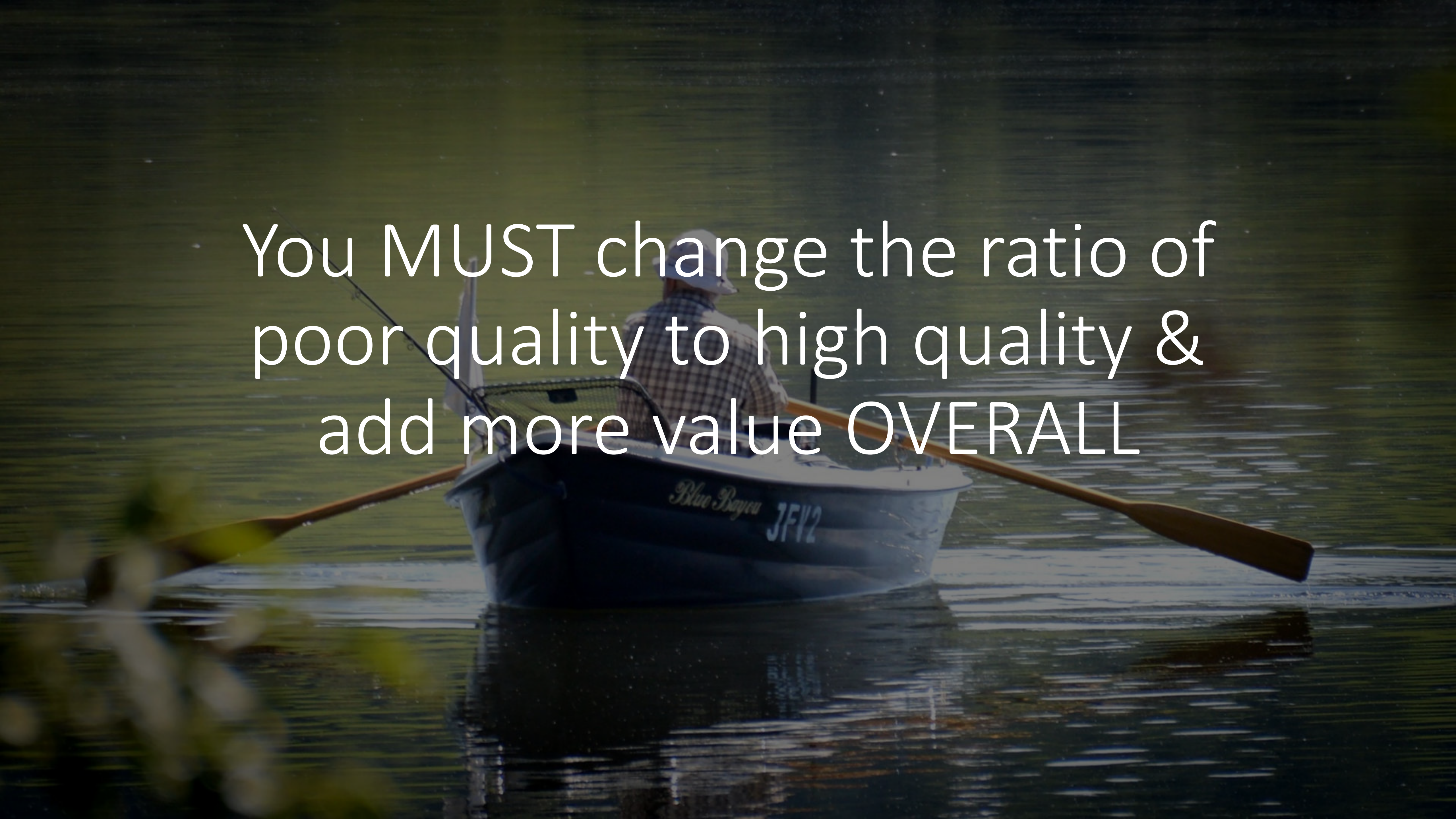
We do use the indexed content as a basis for what we show in search, so if you significantly change your page's content, I'd expect that to be reflected in search over time too.

10:04 am · 21 Jan 2020 · TweetDeck

Fear not… Small Wordpress site

You need to lure a Grumpy Googlebot with tasty quality content morsels





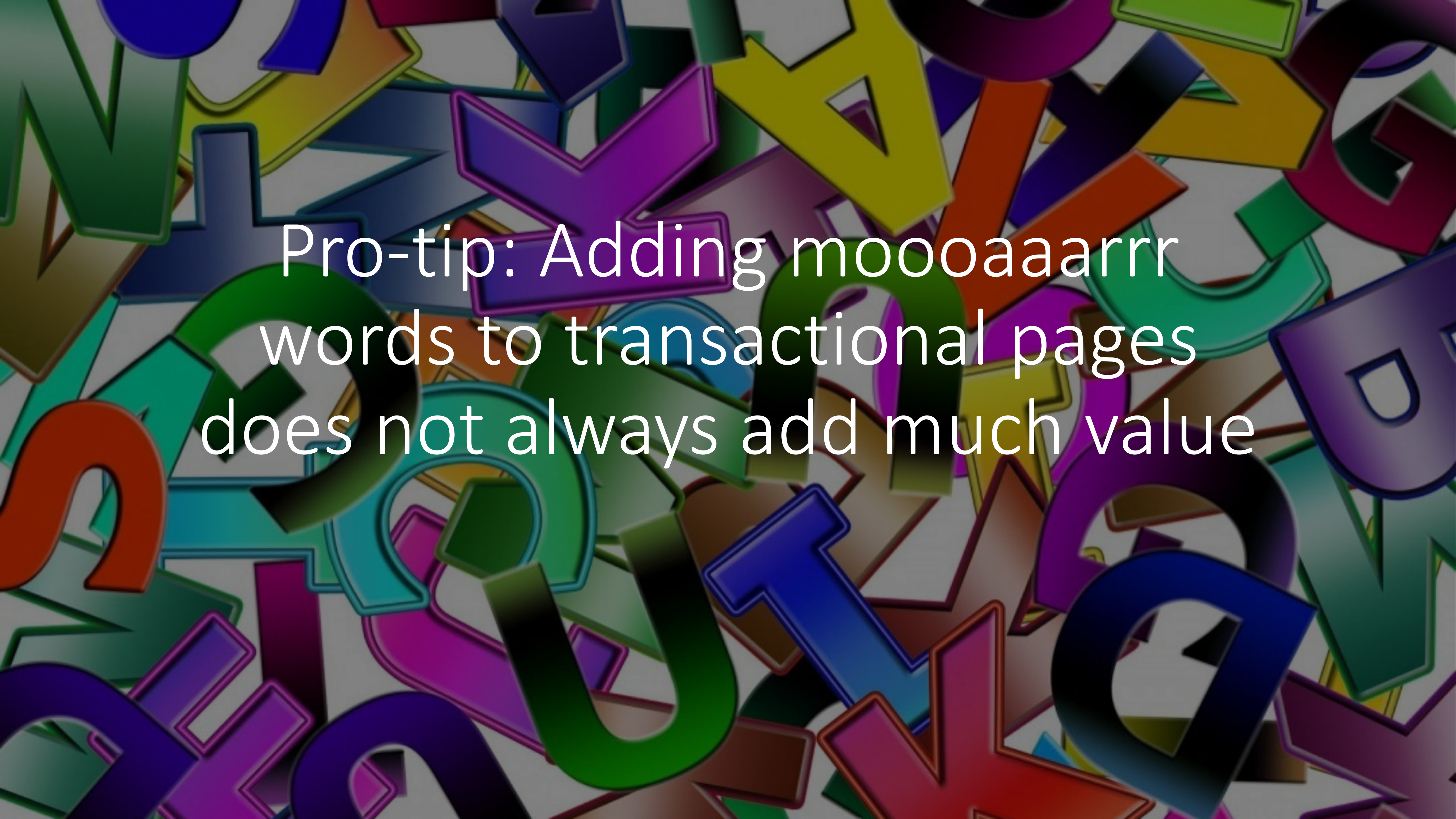
You MUST change the ratio of poor quality to high quality & add more value OVERALL

And contribute positively to ‘The Network Effect’

You need to substantively improve the quality of your pages, regain the crawl and get more demand

Watch for patterns (clues) in GSC Coverage & take a demand-driven approach

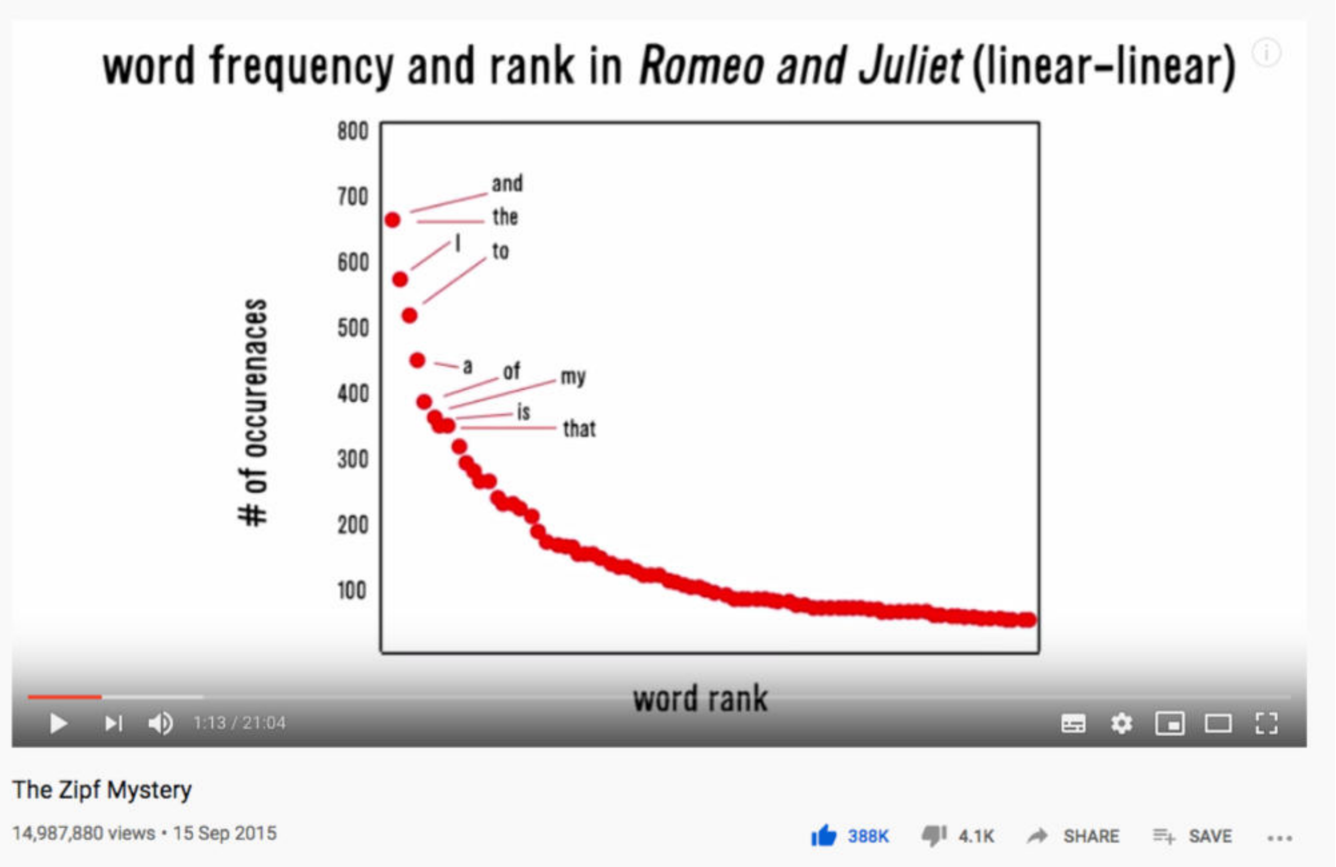
In transactional pages identify valuable content, features & attributes your audience wants



Pro-tip: Adding moooaaarrr words to transactional pages does not always add much value

Plus… There is often a mathematically natural ordered pattern to things ranked by frequency

Like word frequencies many types of 'ordered' popularities will have a Zipfian Distribution



The Zipf Mystery

14,987,880 views • 15 Sep 2015

Where the frequency of x is inversely proportional to its frequency table rank – 1/n

Zipfian Distribution occurs in many other rankings unrelated to language

- Population of cities in a country
- Corporation sizes
- Income rankings
- No. of people watching same TV channel

Nurture internal link graph popularity that mirrors ‘real life popularity

Popularity & Zipfian Distributions

Is it really popular in ‘real life’?

Does it follow ‘The Network Effect’?

Or are you manipulating it for things you want to rank for (which aren’t really popular at all)?

Bow tie of the web & strongly connected components

Building genuine topical hubs (Hub & spoke)

Google puts the pieces of the puzzle together slowly but increases with quality improvements

Equipossibility goes well beyond duplicate content too

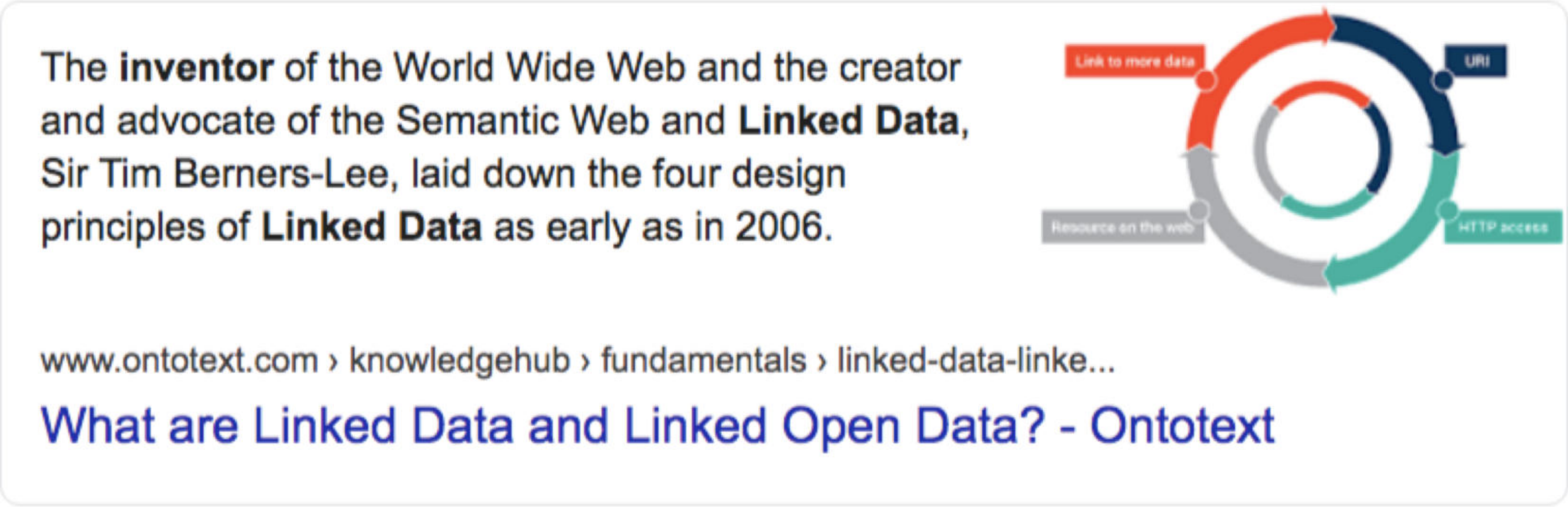
Semantic heterogeneity

The Web of Document Vs The Web of Data

Applies mostly to entity-oriented search

Mis-matching data types or equivalencies in data tables from same or other domains

Linked Data has been around for a long time



The **inventor** of the World Wide Web and the creator and advocate of the Semantic Web and **Linked Data**, Sir Tim Berners-Lee, laid down the four design principles of **Linked Data** as early as in 2006.

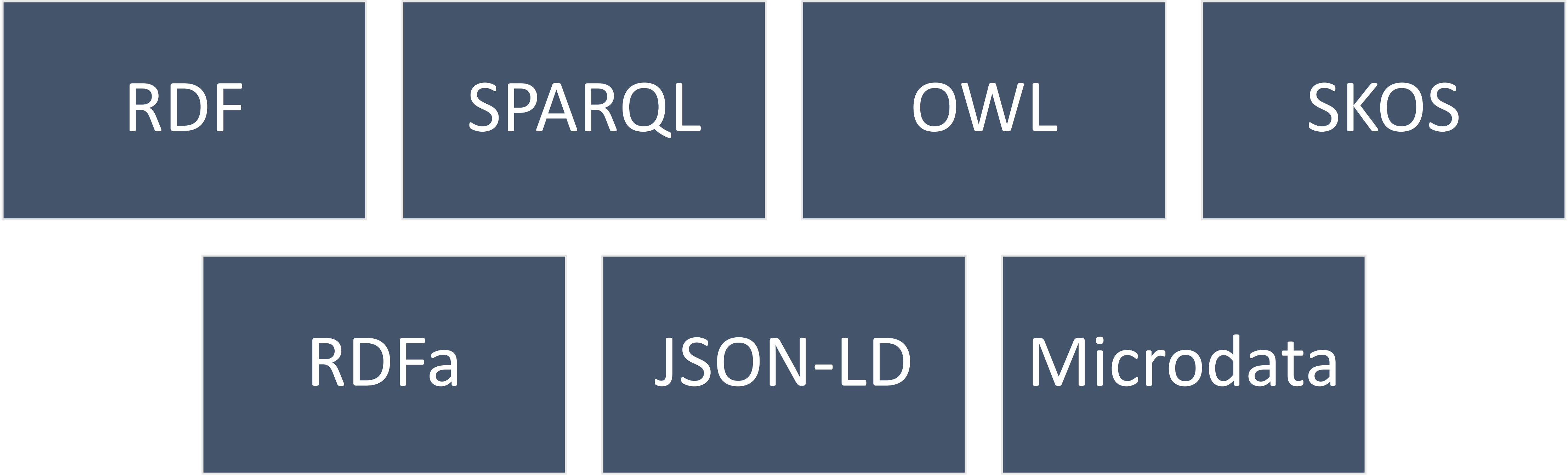
www.ontotext.com › knowledgehub › fundamentals › linked-data-linke...

What are Linked Data and Linked Open Data? - Ontotext

There have historically been several ways to implement linked data

Several linked-data types & web sources come together causing equipossibility

Some Types of Semantic Web Technologies & Their Markups



RDF

SPARQL

OWL

SKOS

RDFa

JSON-LD

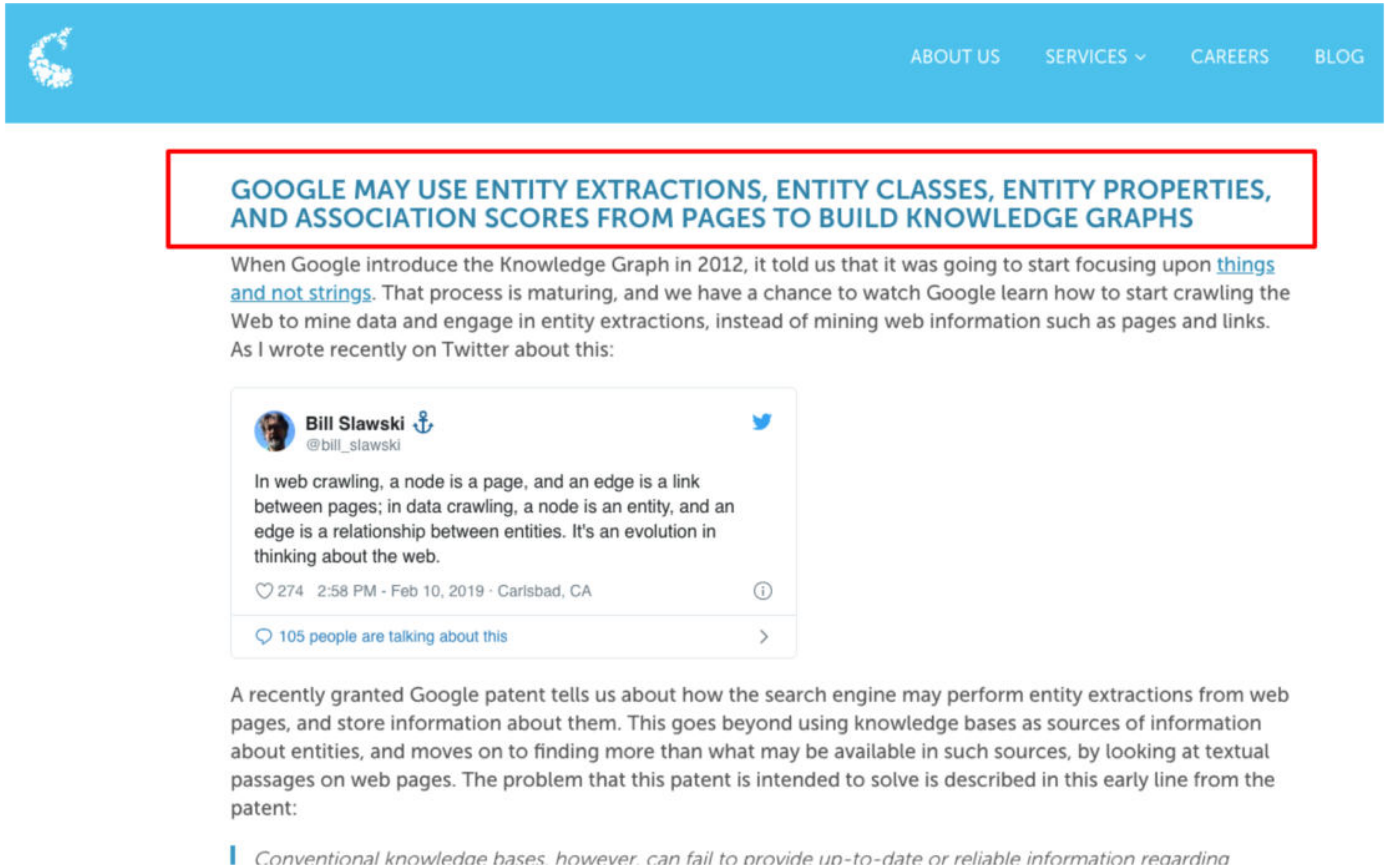
Microdata

Implementation Inconsistencies Prevail (ed)

The Knowledge Graph & Its Data Sources

The majority of Google’s entity types reference the Wikipedia URI or the Knowledge Graph MID

But Now Anything Structured Might Be Used



GOOGLE MAY USE ENTITY EXTRACTIONS, ENTITY CLASSES, ENTITY PROPERTIES, AND ASSOCIATION SCORES FROM PAGES TO BUILD KNOWLEDGE GRAPHS

When Google introduce the Knowledge Graph in 2012, it told us that it was going to start focusing upon things and not strings. That process is maturing, and we have a chance to watch Google learn how to start crawling the Web to mine data and engage in entity extractions, instead of mining web information such as pages and links. As I wrote recently on Twitter about this:

Bill Slawski ⚓
@bill_slawski

In web crawling, a node is a page, and an edge is a link between pages; in data crawling, a node is an entity, and an edge is a relationship between entities. It's an evolution in thinking about the web.

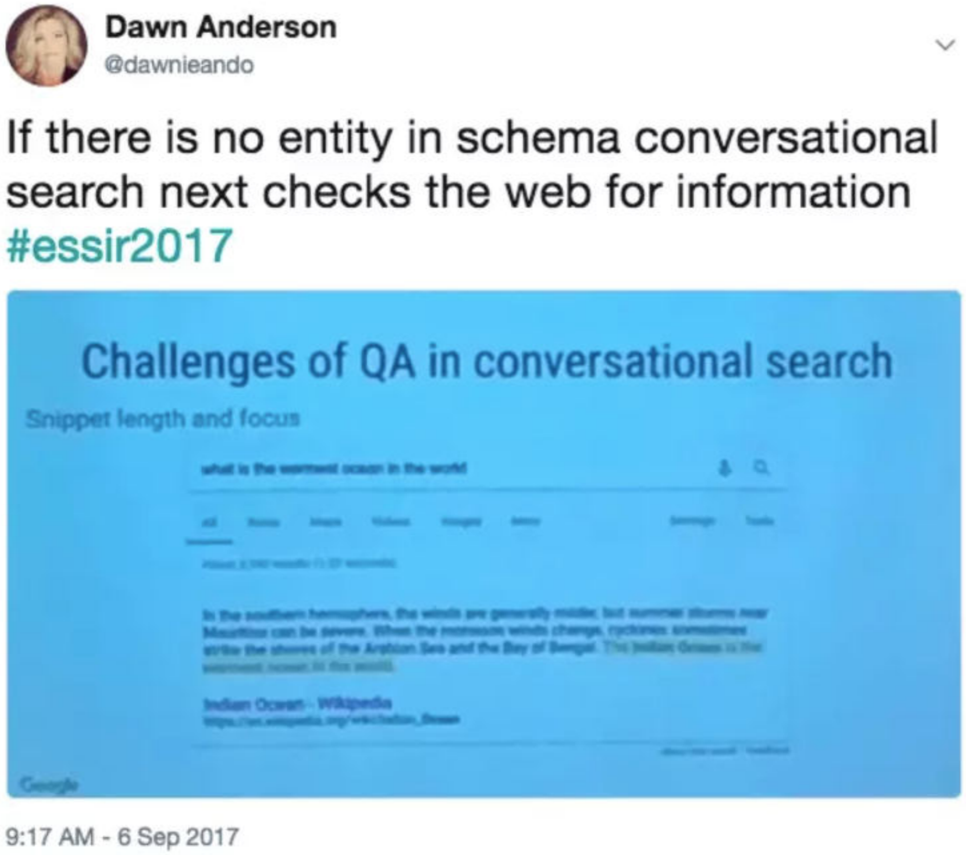
♡ 274 2:58 PM - Feb 10, 2019 · Carlsbad, CA

105 people are talking about this

A recently granted Google patent tells us about how the search engine may perform entity extractions from web pages, and store information about them. This goes beyond using knowledge bases as sources of information about entities, and moves on to finding more than what may be available in such sources, by looking at textual passages on web pages. The problem that this patent is intended to solve is described in this early line from the patent:

> *Conventional knowledge bases, however, can fail to provide up-to-date or reliable information regarding*

We already know conversation search fills gaps via web information



Dawn Anderson
@dawnieando

If there is no entity in schema conversational search next checks the web for information #essir2017

Challenges of QA in conversational search

Snippet length and focus

what is the warmest ocean in the world

In the southern hemisphere, the winds are generally milder, but summer storms near Mauritius can be severe. When the monsoon winds change, cyclones sometimes strike the shores of the Arabian Sea and the Bay of Bengal. The Indian Ocean is the warmest ocean in the world.

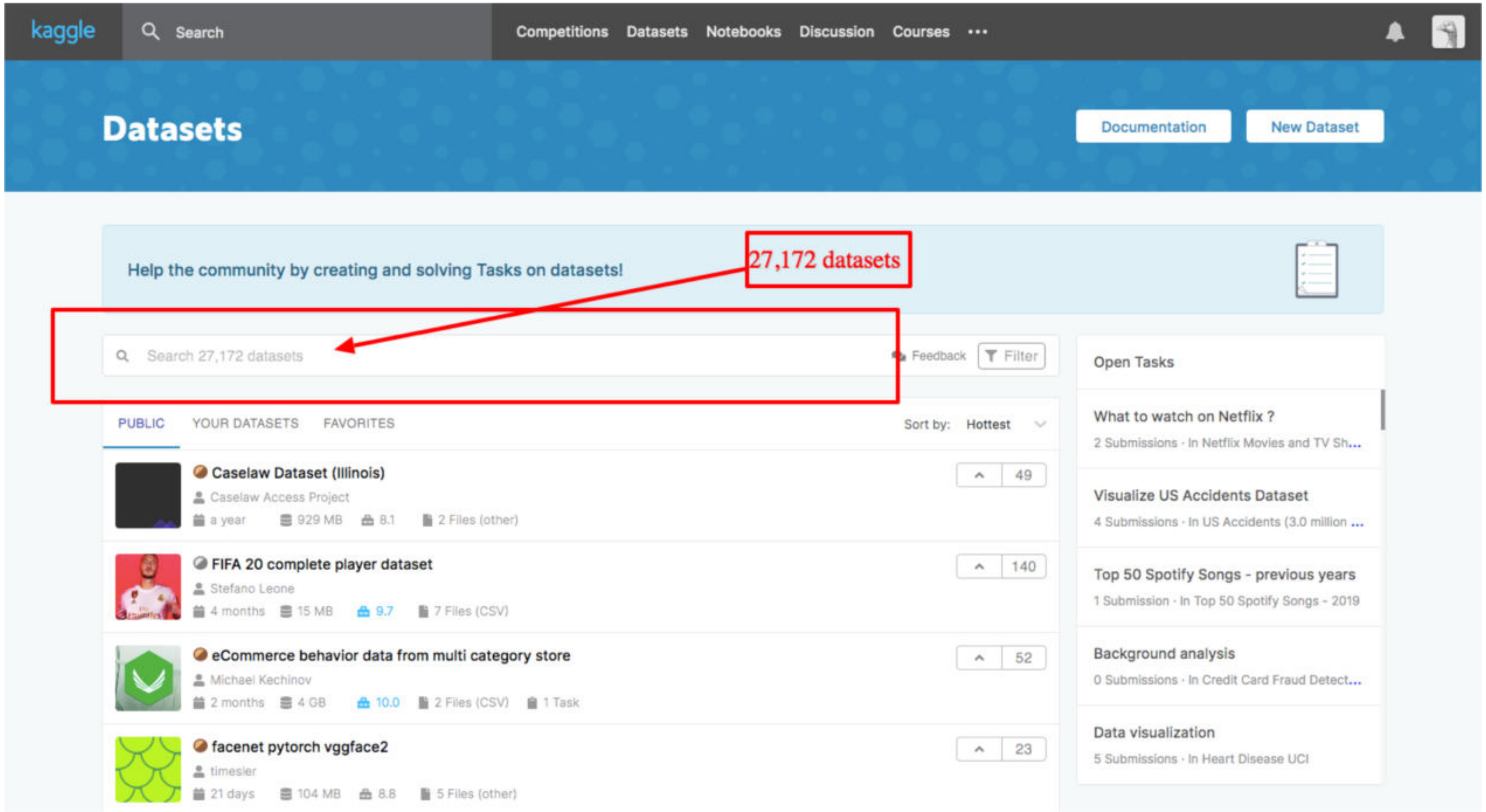
Indian Ocean - Wikipedia
https://en.wikipedia.org/wiki/Indian_Ocean

Google

9:17 AM - 6 Sep 2017

The web is a ‘Data Lake’

Public Datasets (e.g. Kaggle) > 27k datasets



Datasets

Help the community by creating and solving Tasks on datasets!

27,172 datasets

Search 27,172 datasets

Caselaw Dataset (Illinois) — 49
Caselaw Access Project
a year · 929 MB · 8.1 · 2 Files (other)

FIFA 20 complete player dataset — 140
Stefano Leone
4 months · 15 MB · 9.7 · 7 Files (CSV)

eCommerce behavior data from multi category store — 52
Michael Kechinov
2 months · 4 GB · 10.0 · 2 Files (CSV) · 1 Task

facenet pytorch vggface2 — 23
timesler
21 days · 104 MB · 8.8 · 5 Files (other)

Open Tasks

What to watch on Netflix ?
2 Submissions · In Netflix Movies and TV Sh...

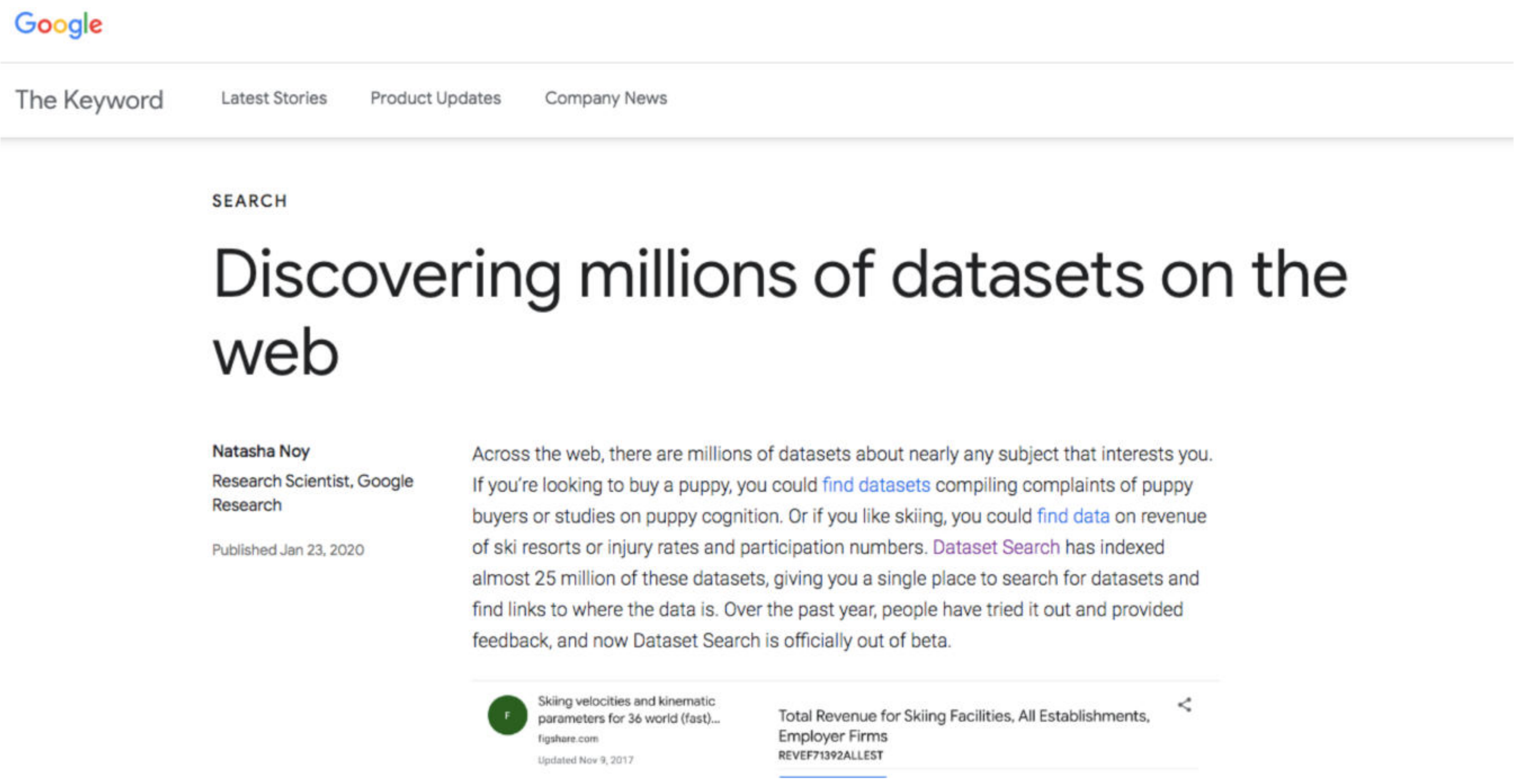
Visualize US Accidents Dataset
4 Submissions · In US Accidents (3.0 million ...

Top 50 Spotify Songs - previous years
1 Submission · In Top 50 Spotify Songs - 2019

Background analysis
0 Submissions · In Credit Card Fraud Detect...

Data visualization
5 Submissions · In Heart Disease UCI

Enter Data Search by Google AI



SEARCH

Discovering millions of datasets on the web

Natasha Noy
Research Scientist, Google Research

Published Jan 23, 2020

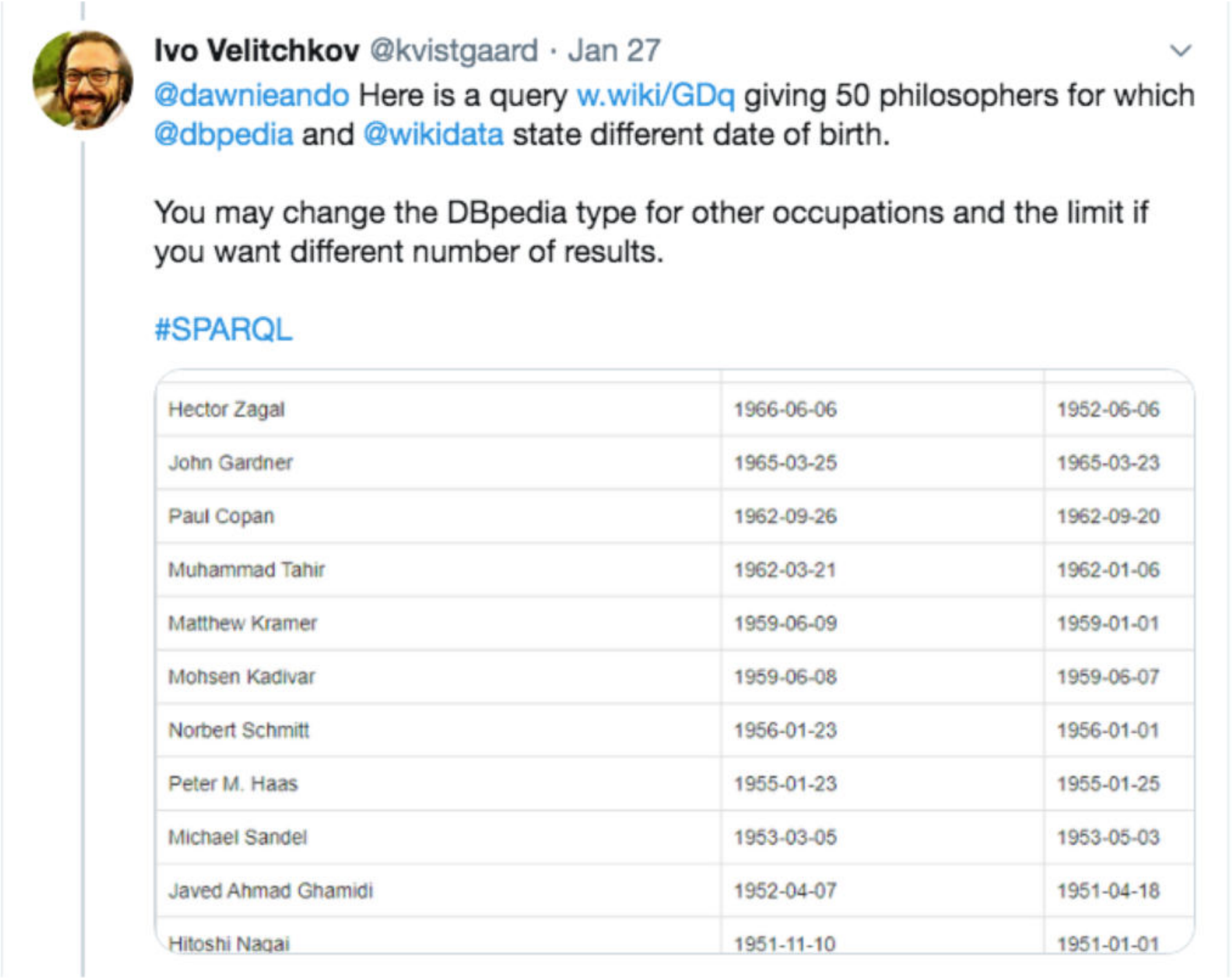
Across the web, there are millions of datasets about nearly any subject that interests you. If you're looking to buy a puppy, you could find datasets compiling complaints of puppy buyers or studies on puppy cognition. Or if you like skiing, you could find data on revenue of ski resorts or injury rates and participation numbers. Dataset Search has indexed almost 25 million of these datasets, giving you a single place to search for datasets and find links to where the data is. Over the past year, people have tried it out and provided feedback, and now Dataset Search is officially out of beta.

F
Skiing velocities and kinematic parameters for 36 world (fast)...
figshare.com
Updated Nov 9, 2017
Total Revenue for Skiing Facilities, All Establishments, Employer Firms
REVEF71392ALLEST

Disambiguating Data Sets (is hard)

Probability Predictions Can Be VERY Wrong

Example: Philosophers Dates of Birth different in Wikidata & DBPedia



Ivo Velitchkov @kvistgaard · Jan 27

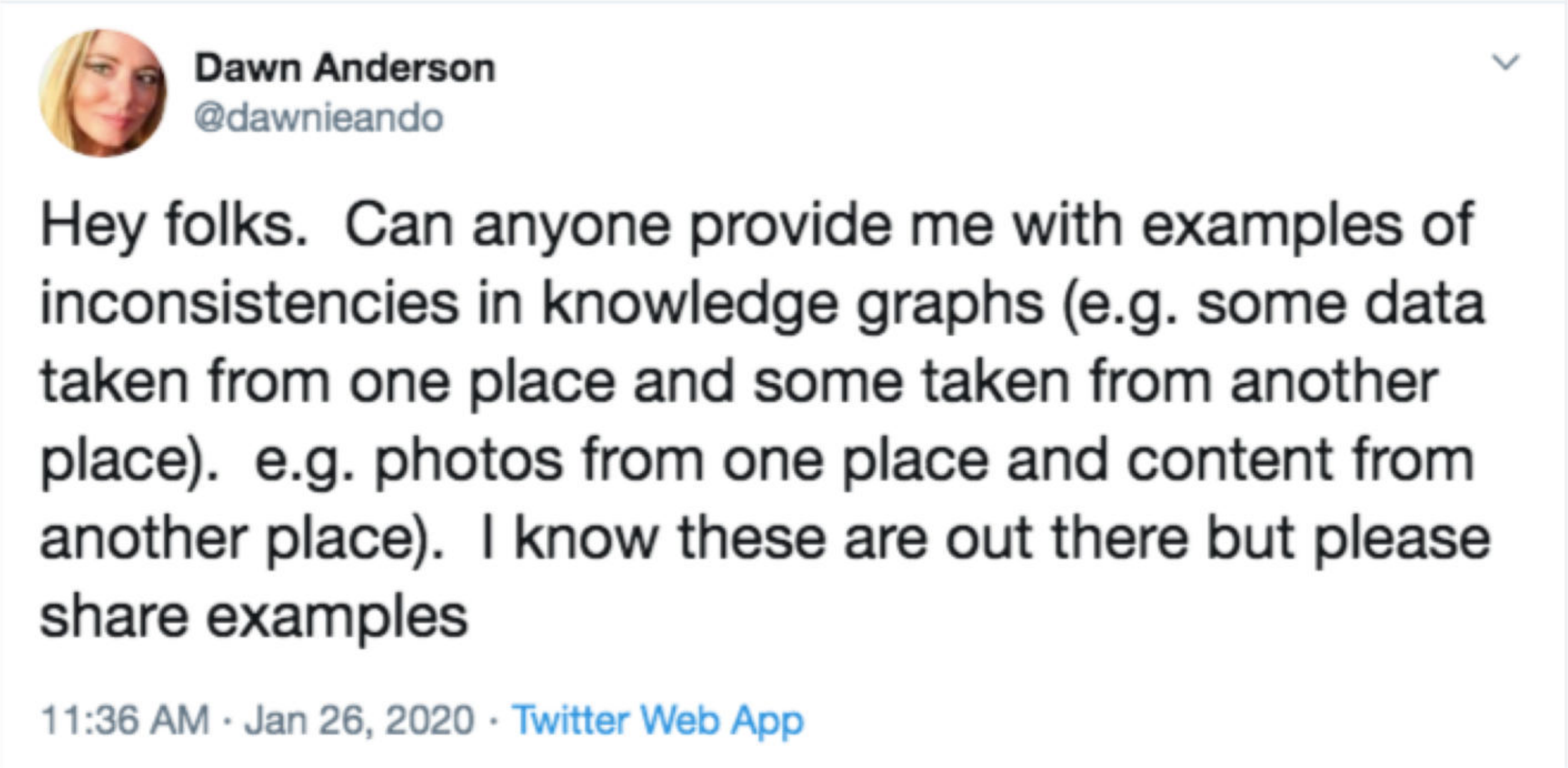
@dawnieando Here is a query w.wiki/GDq giving 50 philosophers for which @dbpedia and @wikidata state different date of birth.

You may change the DBpedia type for other occupations and the limit if you want different number of results.

#SPARQL

Hector Zagal	1966-06-06	1952-06-06
John Gardner	1965-03-25	1965-03-23
Paul Copan	1962-09-26	1962-09-20
Muhammad Tahir	1962-03-21	1962-01-06
Matthew Kramer	1959-06-09	1959-01-01
Mohsen Kadivar	1959-06-08	1959-06-07
Norbert Schmitt	1956-01-23	1956-01-01
Peter M. Haas	1955-01-23	1955-01-25
Michael Sandel	1953-03-05	1953-05-03
Javed Ahmad Ghamidi	1952-04-07	1951-04-18
Hitoshi Nagai	1951-11-10	1951-01-01

I asked for some examples on Twitter

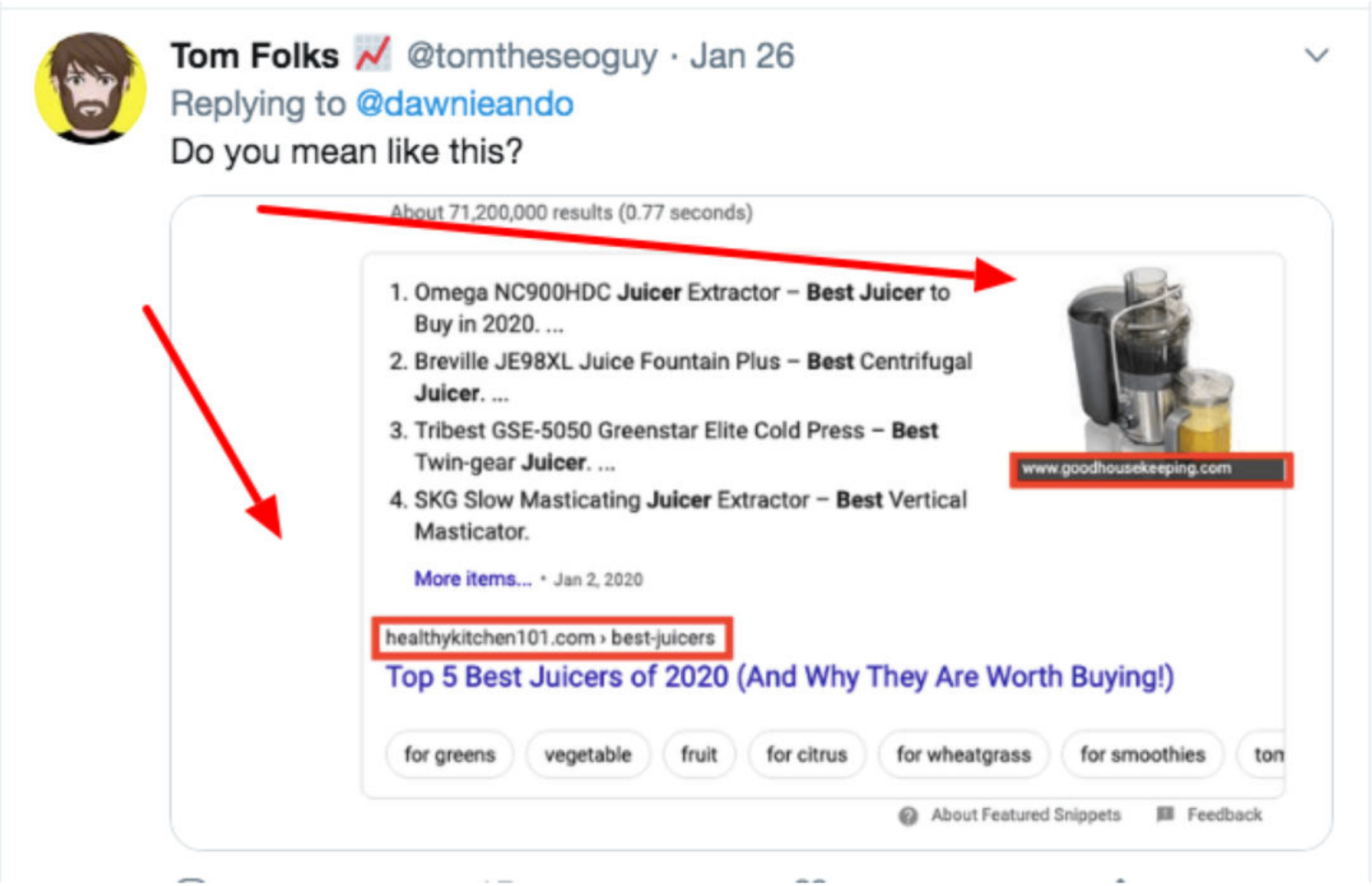


Dawn Anderson
@dawnieando

Hey folks. Can anyone provide me with examples of inconsistencies in knowledge graphs (e.g. some data taken from one place and some taken from another place). e.g. photos from one place and content from another place). I know these are out there but please share examples

11:36 AM · Jan 26, 2020 · Twitter Web App

The community responded



Tom Folks @tomtheseoguy · Jan 26

Replying to @dawnieando

Do you mean like this?

About 71,200,000 results (0.77 seconds)

1. Omega NC900HDC **Juicer** Extractor – **Best Juicer** to Buy in 2020. ...
2. Breville JE98XL Juice Fountain Plus – **Best** Centrifugal **Juicer**. ...
3. Tribest GSE-5050 Greenstar Elite Cold Press – **Best** Twin-gear **Juicer**. ...
4. SKG Slow Masticating **Juicer** Extractor – **Best** Vertical Masticator.

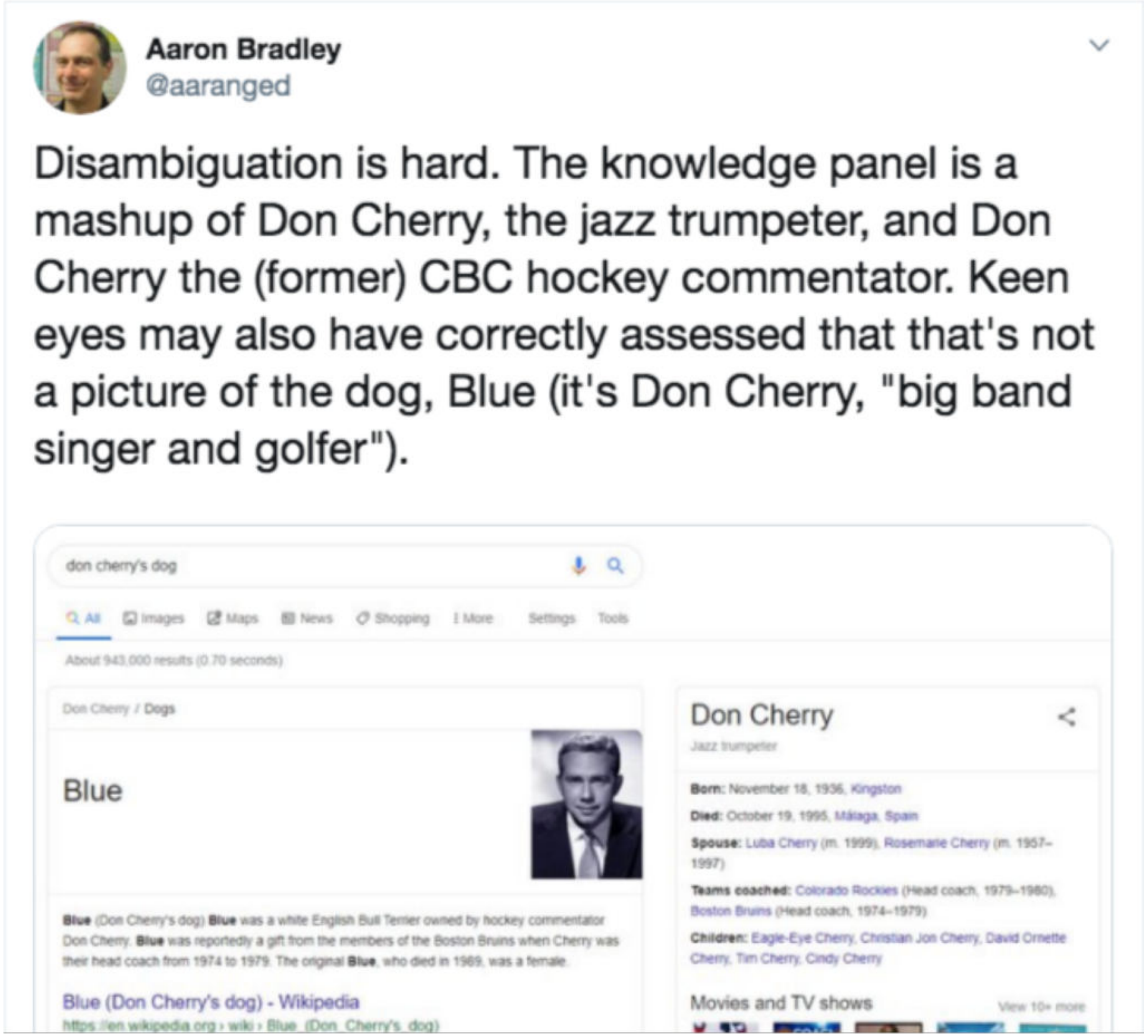
More items... • Jan 2, 2020

www.goodhousekeeping.com

healthykitchen101.com › best-juicers

Top 5 Best Juicers of 2020 (And Why They Are Worth Buying!)

'Mashed-up' Knowledge Graphs (Two Don Cherry's and his dog's name - 'Blue'



Aaron Bradley
@aaranged

Disambiguation is hard. The knowledge panel is a mashup of Don Cherry, the jazz trumpeter, and Don Cherry the (former) CBC hockey commentator. Keen eyes may also have correctly assessed that that's not a picture of the dog, Blue (it's Don Cherry, "big band singer and golfer").

About 943,000 results (0.70 seconds)

Don Cherry / Dogs

Blue

Blue (Don Cherry's dog) **Blue** was a white English Bull Terrier owned by hockey commentator Don Cherry. **Blue** was reportedly a gift from the members of the Boston Bruins when Cherry was their head coach from 1974 to 1979. The original **Blue**, who died in 1989, was a female.

Blue (Don Cherry's dog) - Wikipedia
https://en.wikipedia.org › wiki › Blue_(Don_Cherry's_dog)

Don Cherry
Jazz trumpeter

Born: November 18, 1936, Kingston

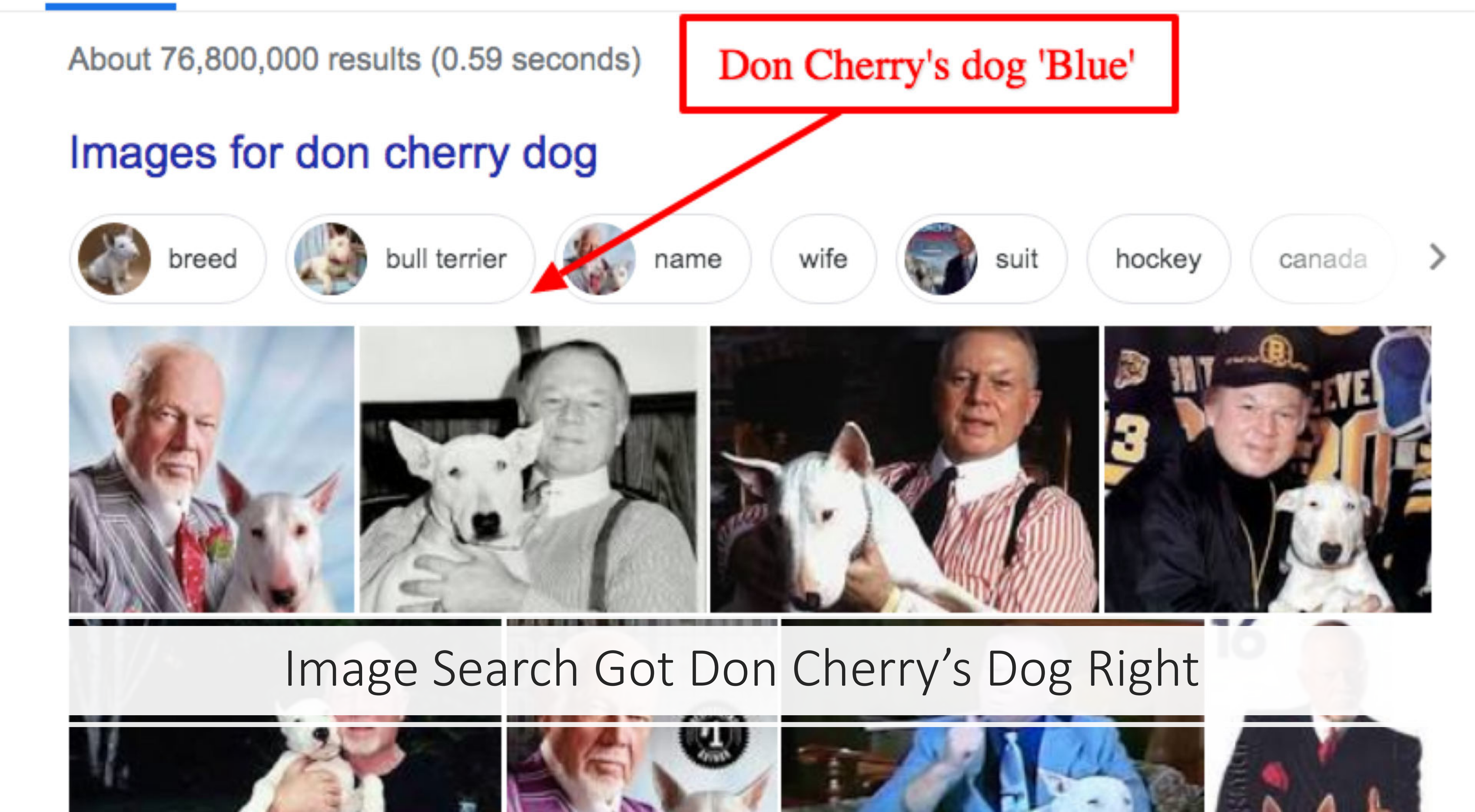
Died: October 19, 1995, Málaga, Spain

Spouse: Luba Cherry (m. 1999), Rosemarie Cherry (m. 1957–1997)

Teams coached: Colorado Rockies (Head coach, 1979–1980), Boston Bruins (Head coach, 1974–1979)

Children: Eagle-Eye Cherry, Christian Jon Cherry, David Ornette Cherry, Tim Cherry, Cindy Cherry

Movies and TV shows



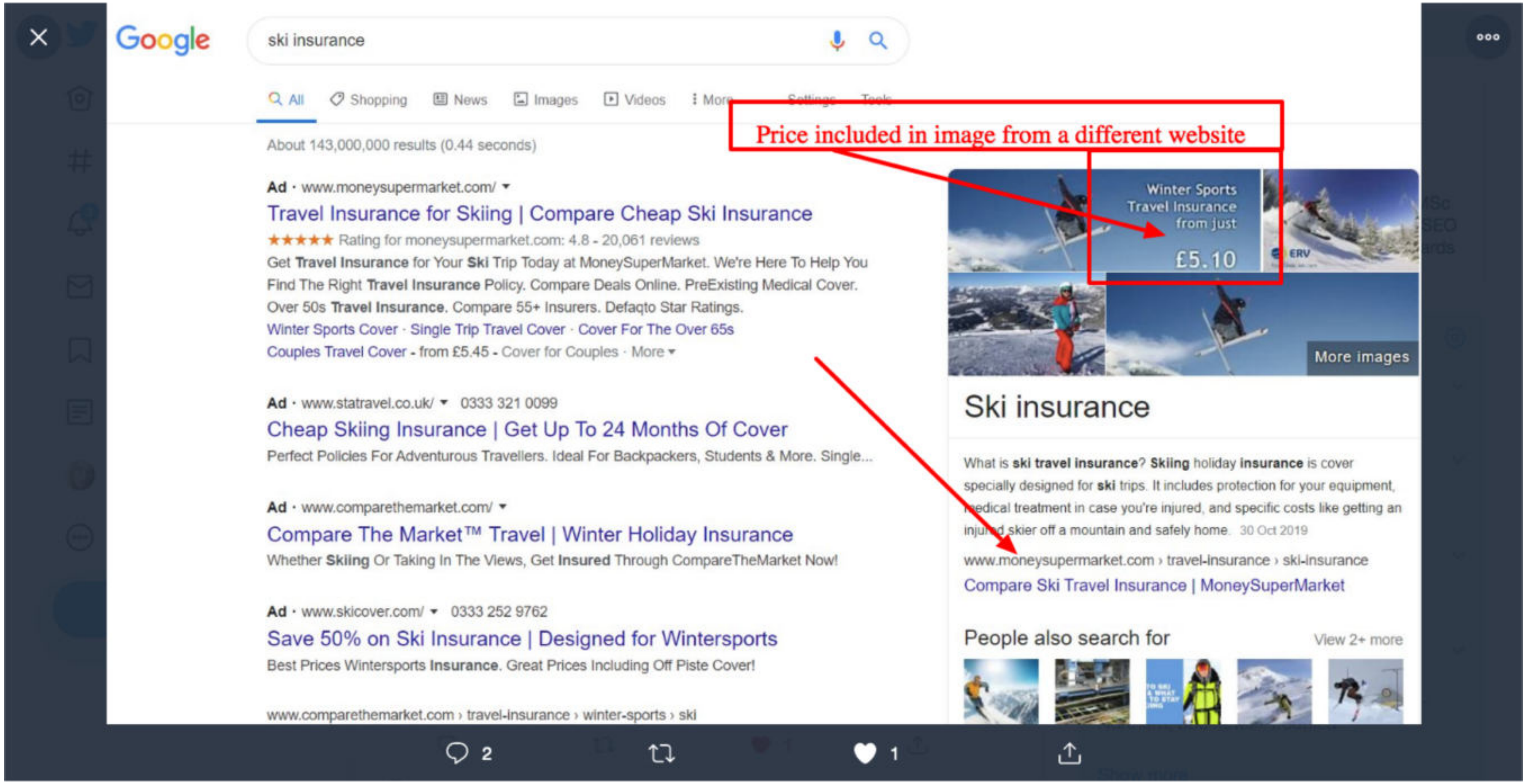
About 76,800,000 results (0.59 seconds)

Images for don cherry dog

Don Cherry's dog 'Blue'

Image Search Got Don Cherry’s Dog Right

Wrong price included in images from other sites



Price included in image from a different website

About 143,000,000 results (0.44 seconds)

Ad · www.moneysupermarket.com/ ▾

Travel Insurance for Skiing | Compare Cheap Ski Insurance

★★★★★ Rating for moneysupermarket.com: 4.8 - 20,061 reviews

Get **Travel Insurance** for Your **Ski** Trip Today at MoneySuperMarket. We're Here To Help You Find The Right **Travel Insurance** Policy. Compare Deals Online. PreExisting Medical Cover. Over 50s **Travel Insurance**. Compare 55+ Insurers. Defaqto Star Ratings.

Winter Sports Cover · Single Trip Travel Cover · Cover For The Over 65s

Couples Travel Cover - from £5.45 - Cover for Couples · More ▾

Ad · www.statravel.co.uk/ ▾ 0333 321 0099

Cheap Skiing Insurance | Get Up To 24 Months Of Cover

Perfect Policies For Adventurous Travellers. Ideal For Backpackers, Students & More. Single...

Ad · www.comparethemarket.com/ ▾

Compare The Market™ Travel | Winter Holiday Insurance

Whether **Skiing** Or Taking In The Views, Get **Insured** Through CompareTheMarket Now!

Ad · www.skicover.com/ ▾ 0333 252 9762

Save 50% on Ski Insurance | Designed for Wintersports

Best Prices Wintersports **Insurance**. Great Prices Including Off Piste Cover!

www.comparethemarket.com › travel-insurance › winter-sports › ski

Ski insurance

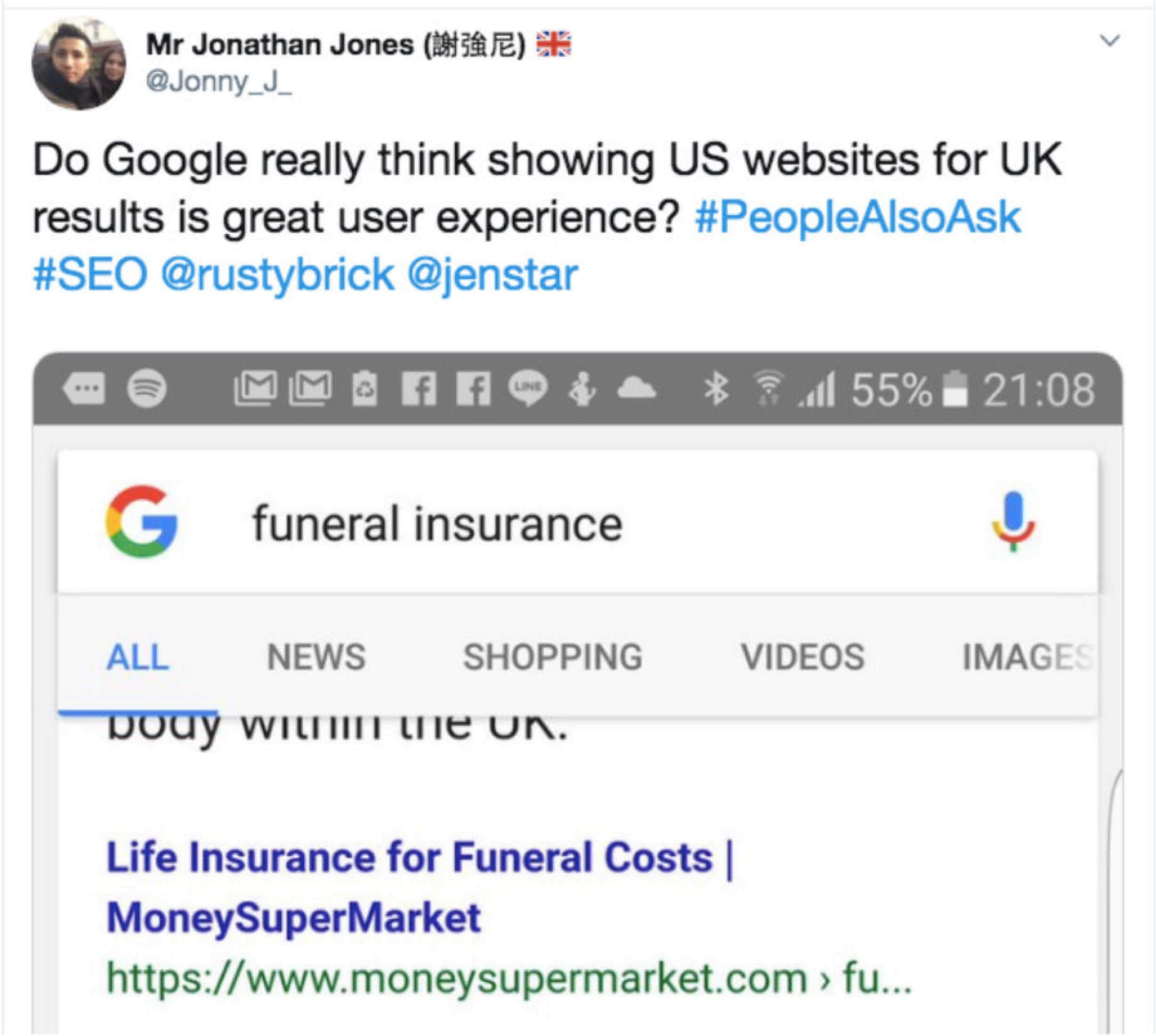
What is **ski travel insurance**? **Skiing** holiday **insurance** is cover specially designed for **ski** trips. It includes protection for your equipment, medical treatment in case you're injured, and specific costs like getting an injured skier off a mountain and safely home. 30 Oct 2019

www.moneysupermarket.com › travel-insurance › ski-insurance

Compare Ski Travel Insurance | MoneySuperMarket

People also search for

View 2+ more



Mr Jonathan Jones (謝強尼) 🇬🇧
@Jonny_J_

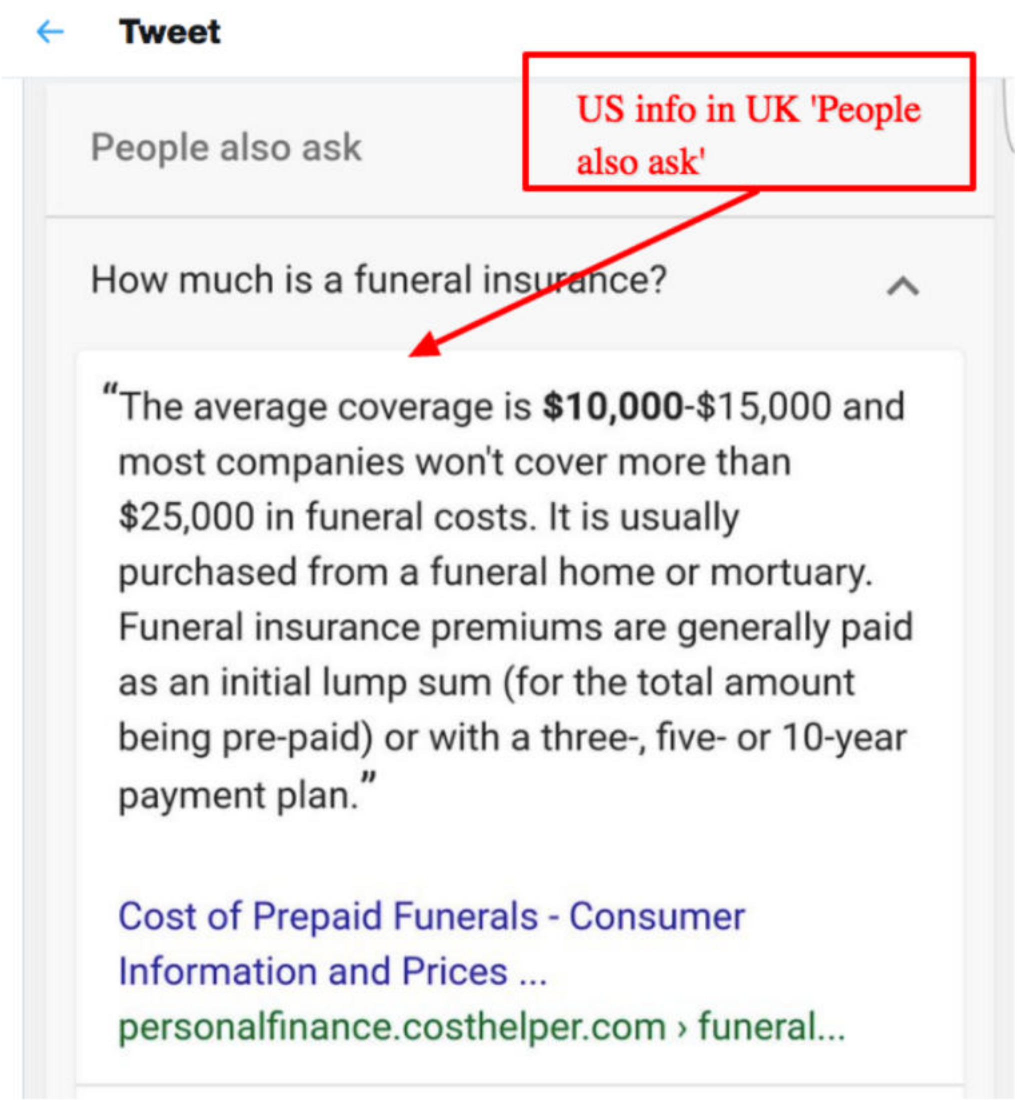
Do Google really think showing US websites for UK results is great user experience? #PeopleAlsoAsk #SEO @rustybrick @jenstar

55% 21:08

funeral insurance

body within the UK.

Life Insurance for Funeral Costs | MoneySuperMarket
https://www.moneysupermarket.com › fu...



US info in UK 'People also ask'

People also ask

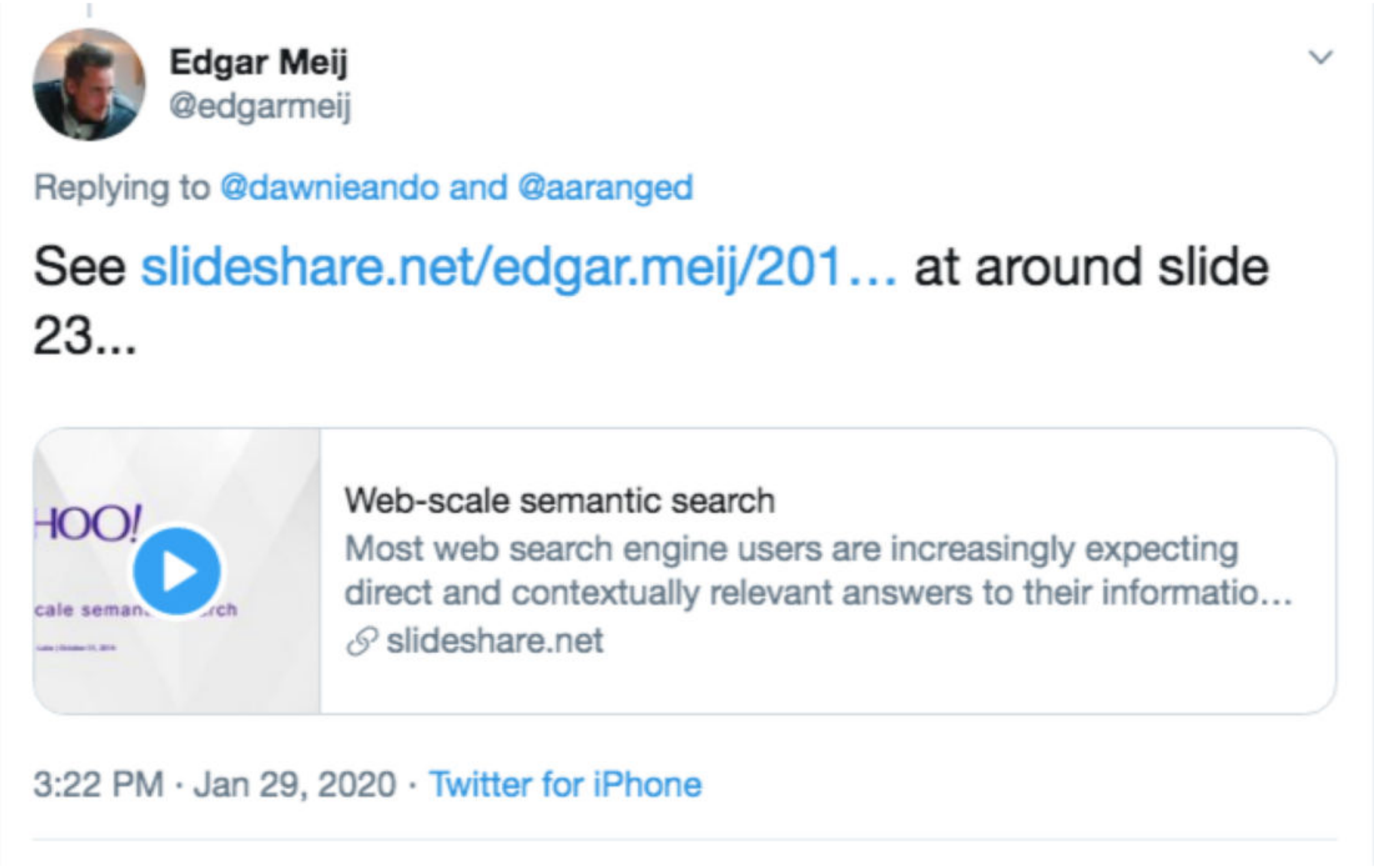
How much is a funeral insurance?

"The average coverage is **$10,000**-$15,000 and most companies won't cover more than $25,000 in funeral costs. It is usually purchased from a funeral home or mortuary. Funeral insurance premiums are generally paid as an initial lump sum (for the total amount being pre-paid) or with a three-, five- or 10-year payment plan."

Cost of Prepaid Funerals - Consumer Information and Prices ...
personalfinance.costhelper.com › funeral...

US Price Info Showing in UK 'People Also Ask'

Edgar’s Were Comical



Edgar Meij
@edgarmeij

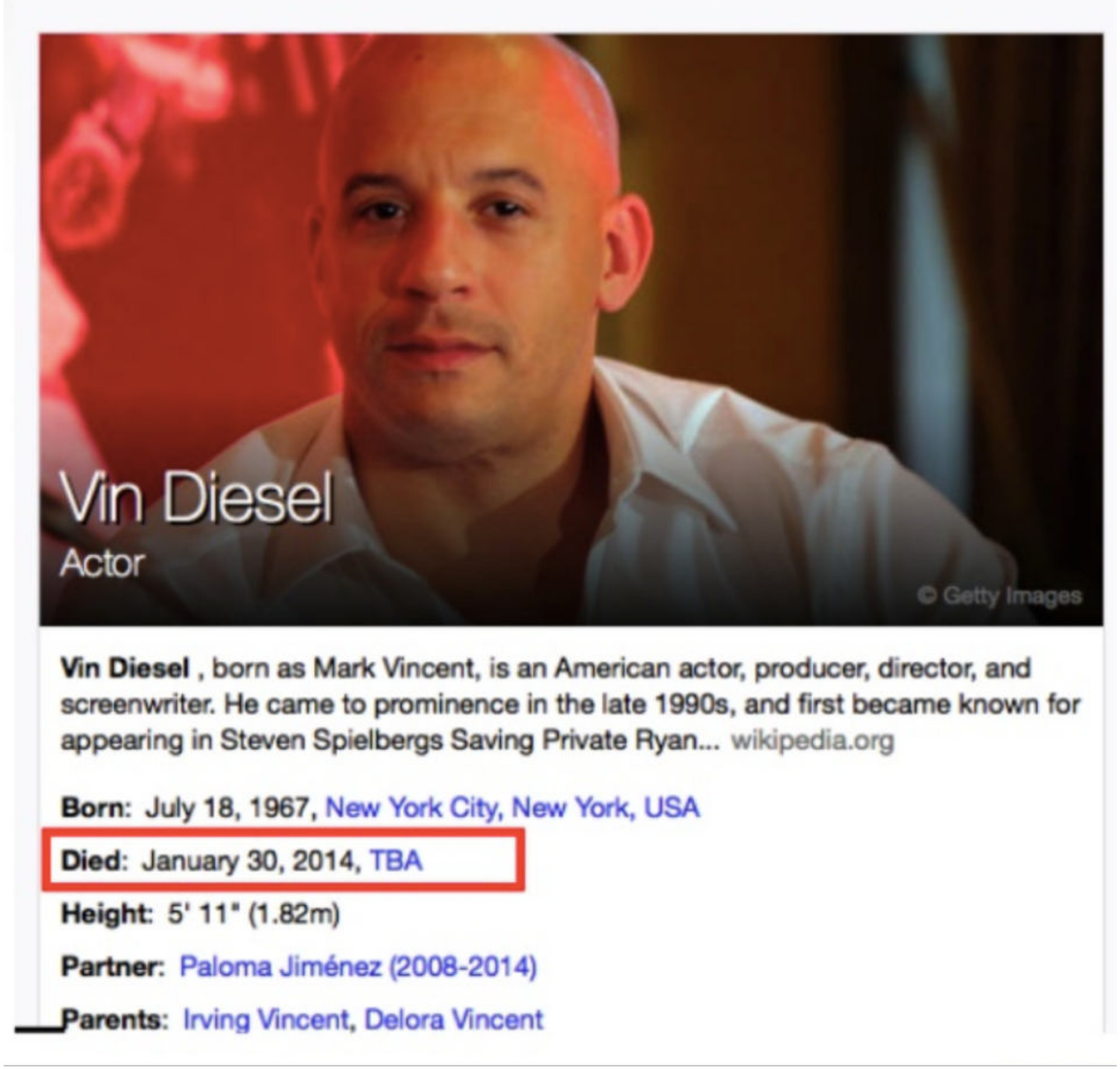
Replying to @dawnieando and @aaranged

See slideshare.net/edgar.meij/201… at around slide 23...

Web-scale semantic search
Most web search engine users are increasingly expecting direct and contextually relevant answers to their informatio…
slideshare.net

3:22 PM · Jan 29, 2020 · Twitter for iPhone

Van Diesel – Death – To Be Advised??



Vin Diesel , born as Mark Vincent, is an American actor, producer, director, and screenwriter. He came to prominence in the late 1990s, and first became known for appearing in Steven Spielbergs Saving Private Ryan... wikipedia.org

Born: July 18, 1967, New York City, New York, USA

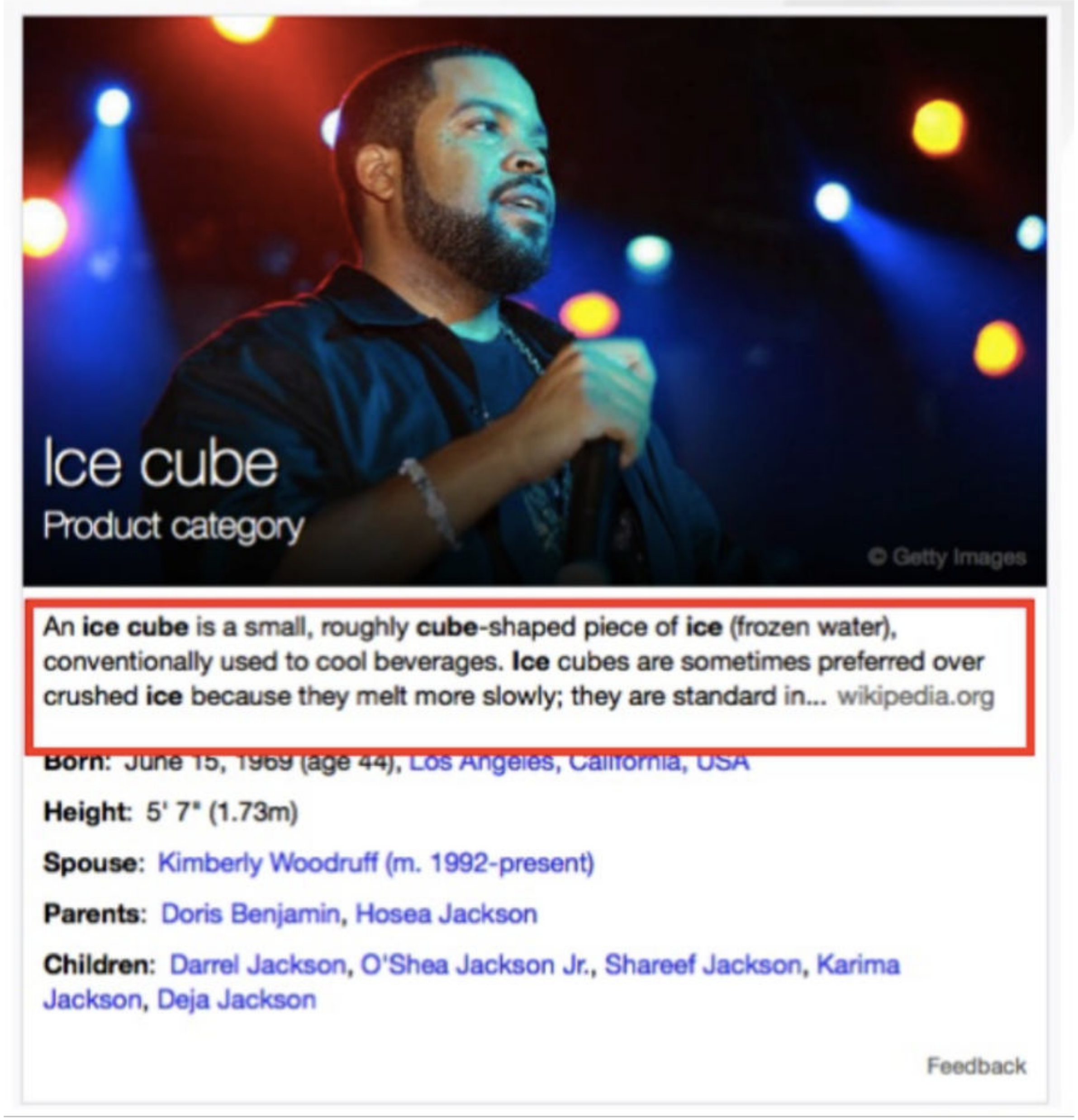
Died: January 30, 2014, TBA

Height: 5' 11" (1.82m)

Partner: Paloma Jiménez (2008-2014)

Parents: Irving Vincent, Delora Vincent

Does An Ice Cube Have Parents?



Ice cube

Product category

© Getty Images

An **ice cube** is a small, roughly **cube**-shaped piece of **ice** (frozen water), conventionally used to cool beverages. **Ice** cubes are sometimes preferred over crushed **ice** because they melt more slowly; they are standard in... wikipedia.org

Born: June 15, 1969 (age 44), Los Angeles, California, USA

Height: 5' 7" (1.73m)

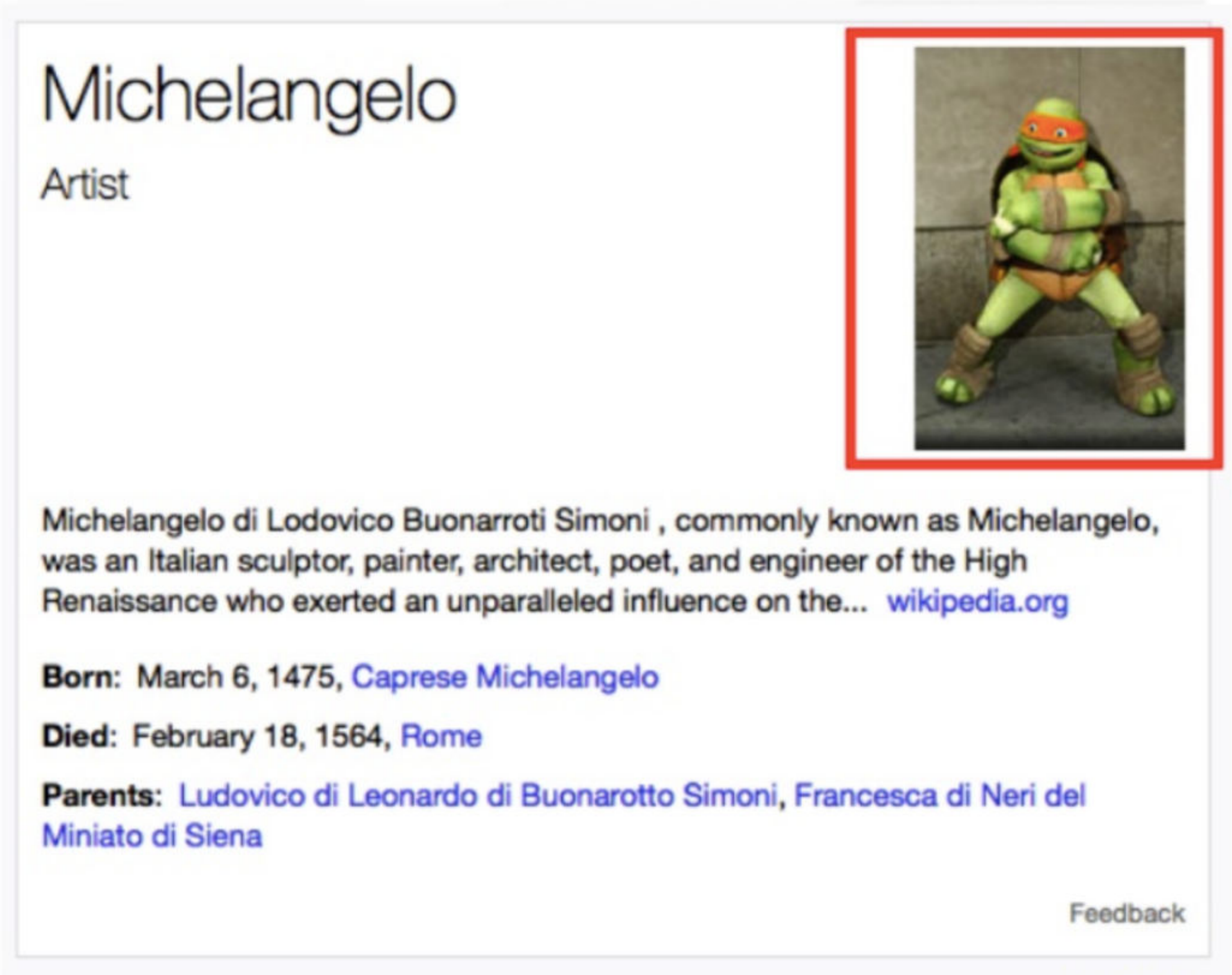
Spouse: Kimberly Woodruff (m. 1992-present)

Parents: Doris Benjamin, Hosea Jackson

Children: Darrel Jackson, O'Shea Jackson Jr., Shareef Jackson, Karima Jackson, Deja Jackson

Feedback

Teenage Mutant Ninja Turtles paint works of art



Michelangelo

Artist

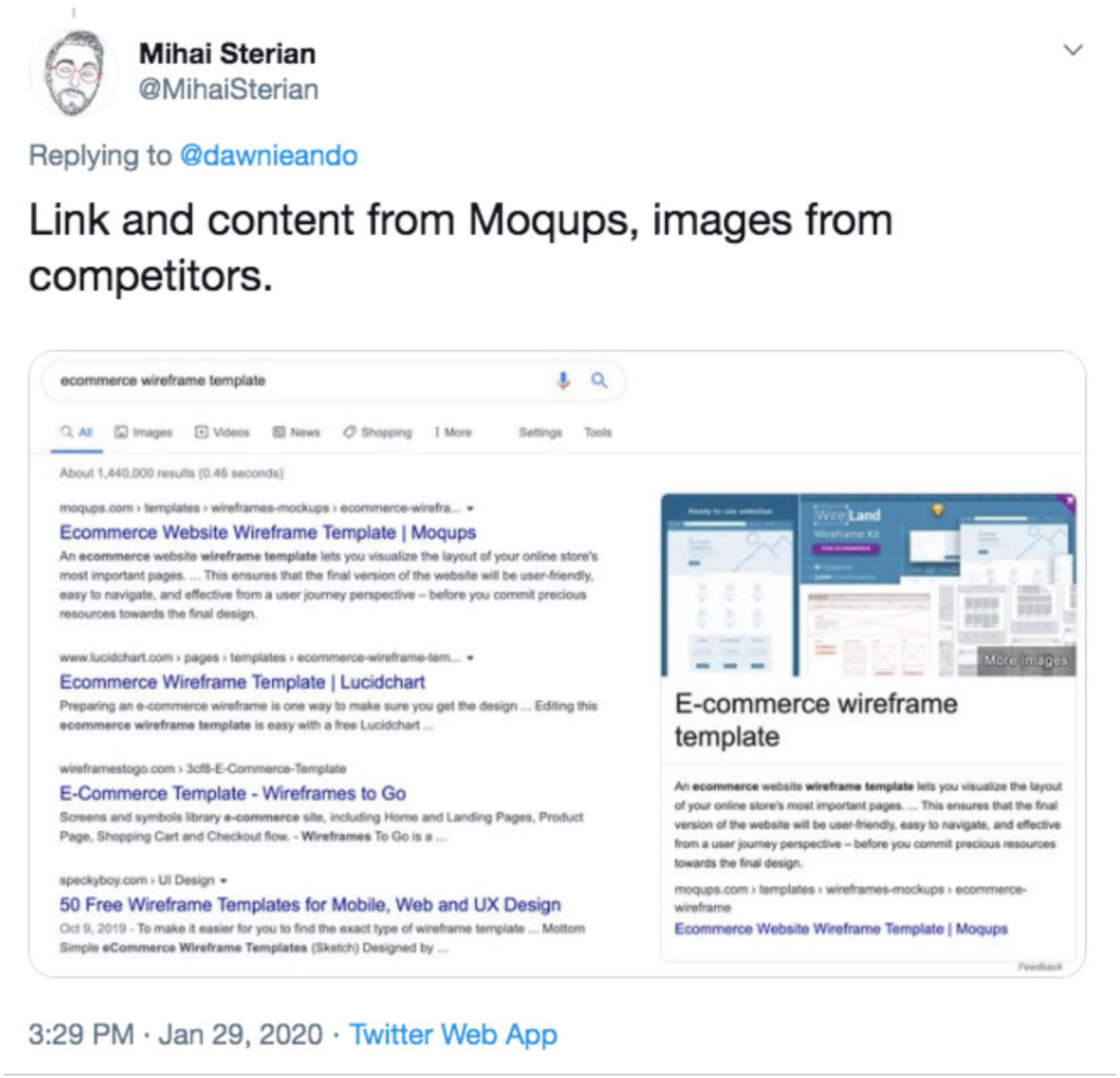
Michelangelo di Lodovico Buonarroti Simoni , commonly known as Michelangelo, was an Italian sculptor, painter, architect, poet, and engineer of the High Renaissance who exerted an unparalleled influence on the... wikipedia.org

Born: March 6, 1475, Caprese Michelangelo

Died: February 18, 1564, Rome

Parents: Ludovico di Leonardo di Buonarotto Simoni, Francesca di Neri del Miniato di Siena

Images from Direct Competitors



Mihai Sterian
@MihaiSterian

Replying to @dawnieando

Link and content from Moqups, images from competitors.

About 1,440,000 results (0.46 seconds)

moqups.com › templates › wireframes-mockups › ecommerce-wirefra... ▾

Ecommerce Website Wireframe Template | Moqups

An **ecommerce** website **wireframe template** lets you visualize the layout of your online store's most important pages. ... This ensures that the final version of the website will be user-friendly, easy to navigate, and effective from a user journey perspective – before you commit precious resources towards the final design.

www.lucidchart.com › pages › templates › ecommerce-wireframe-tem... ▾

Ecommerce Wireframe Template | Lucidchart

Preparing an e-commerce wireframe is one way to make sure you get the design ... Editing this **ecommerce wireframe template** is easy with a free Lucidchart ...

wireframestogo.com › 3cf8-E-Commerce-Template

E-Commerce Template - Wireframes to Go

Screens and symbols library **e-commerce** site, including Home and Landing Pages, Product Page, Shopping Cart and Checkout flow. - **Wireframes** To Go is a ...

speckyboy.com › UI Design ▾

50 Free Wireframe Templates for Mobile, Web and UX Design

Oct 9, 2019 - To make it easier for you to find the exact type of wireframe template ... Mottom Simple **eCommerce Wireframe Templates** (Sketch) Designed by ...

E-commerce wireframe template

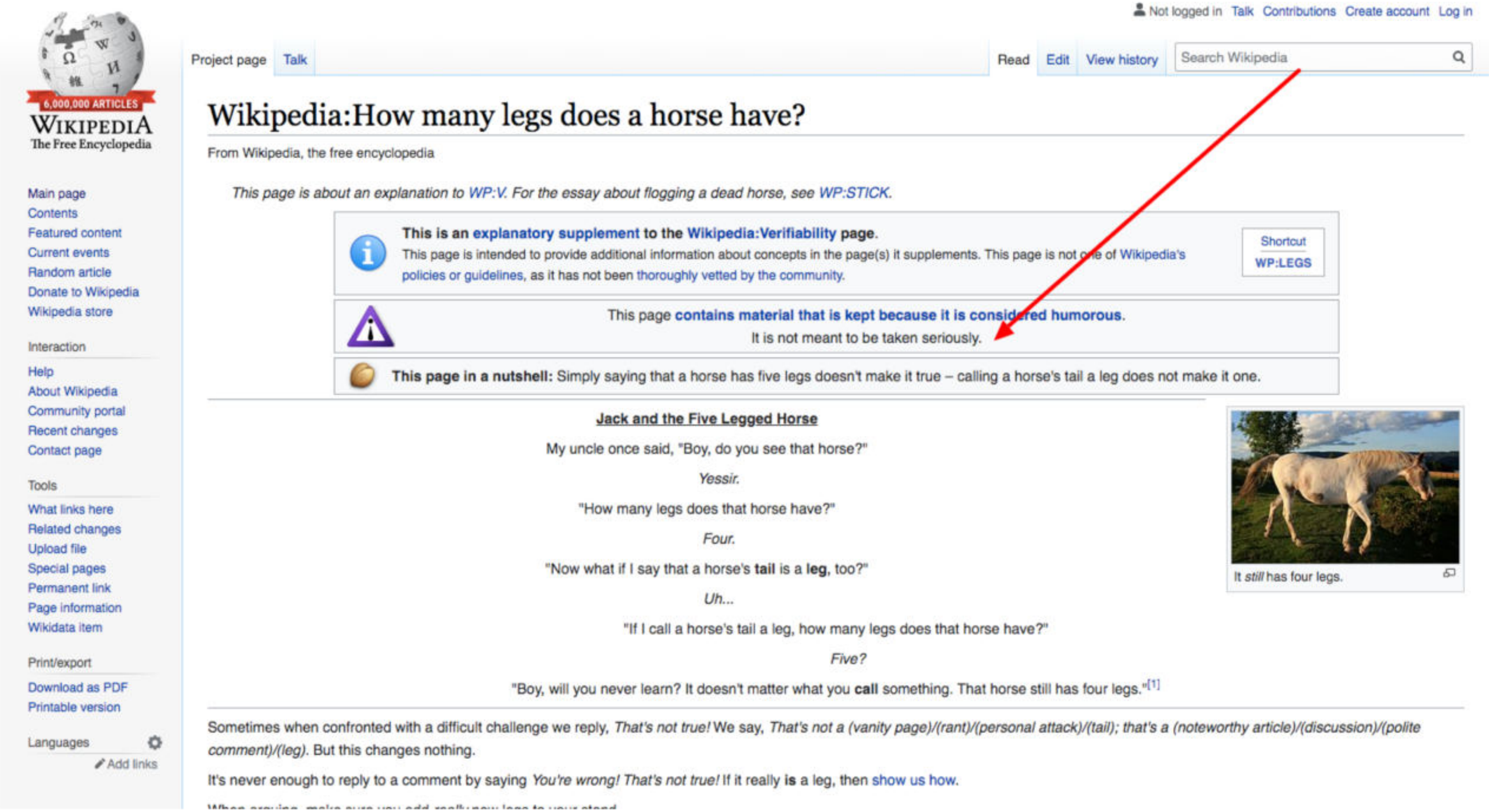
An **ecommerce** website **wireframe template** lets you visualize the layout of your online store's most important pages. ... This ensures that the final version of the website will be user-friendly, easy to navigate, and effective from a user journey perspective – before you commit precious resources towards the final design.

moqups.com › templates › wireframes-mockups › ecommerce-wireframe

Ecommerce Website Wireframe Template | Moqups

3:29 PM · Jan 29, 2020 · Twitter Web App

Wikipedia Made Their Point Well



Wikipedia:How many legs does a horse have?

From Wikipedia, the free encyclopedia

This page is about an explanation to WP:V. For the essay about flogging a dead horse, see WP:STICK.

This is an explanatory supplement to the Wikipedia:Verifiability page.
This page is intended to provide additional information about concepts in the page(s) it supplements. This page is not one of Wikipedia's policies or guidelines, as it has not been thoroughly vetted by the community.

Shortcut
WP:LEGS

This page **contains material that is kept because it is considered humorous.**
It is not meant to be taken seriously.

This page in a nutshell: Simply saying that a horse has five legs doesn't make it true – calling a horse's tail a leg does not make it one.

Jack and the Five Legged Horse

My uncle once said, "Boy, do you see that horse?"

Yessir.

"How many legs does that horse have?"

Four.

"Now what if I say that a horse's **tail** is a **leg**, too?"

Uh...

"If I call a horse's tail a leg, how many legs does that horse have?"

Five?

"Boy, will you never learn? It doesn't matter what you **call** something. That horse still has four legs."[1]

It *still* has four legs.

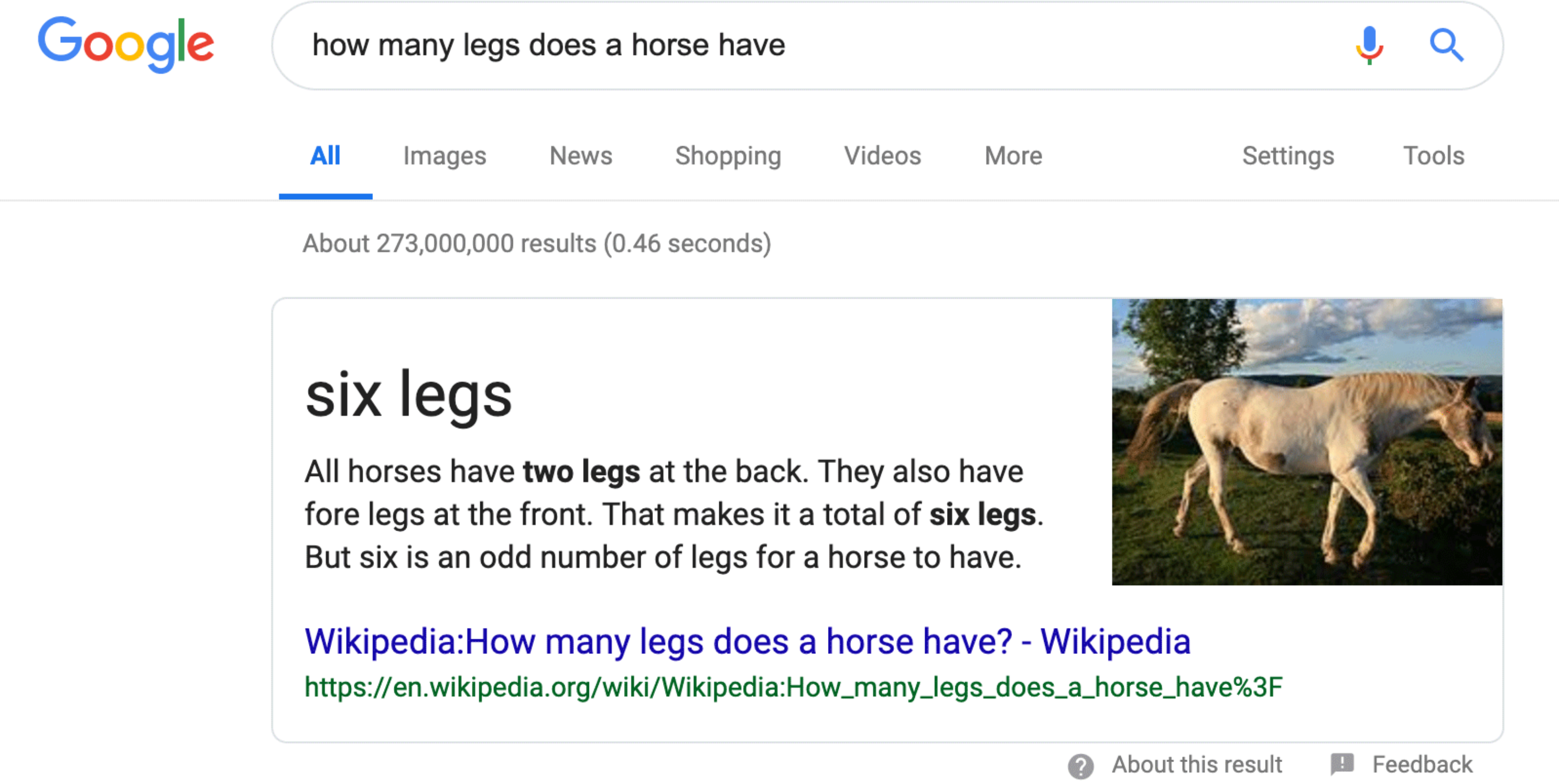
Sometimes when confronted with a difficult challenge we reply, *That's not true!* We say, *That's not a (vanity page)/(rant)/(personal attack)/(tail); that's a (noteworthy article)/(discussion)/(polite comment)/(leg)*. But this changes nothing.

It's never enough to reply to a comment by saying *You're wrong! That's not true!* If it really **is** a leg, then show us how.

When arguing, make sure you add really new legs to your stand

“Simply saying that a horse has five legs doesn't make it true – calling a horse's tail a leg does not make it one.” (Wikipedia, 2019)

A six legged horse?



how many legs does a horse have

About 273,000,000 results (0.46 seconds)

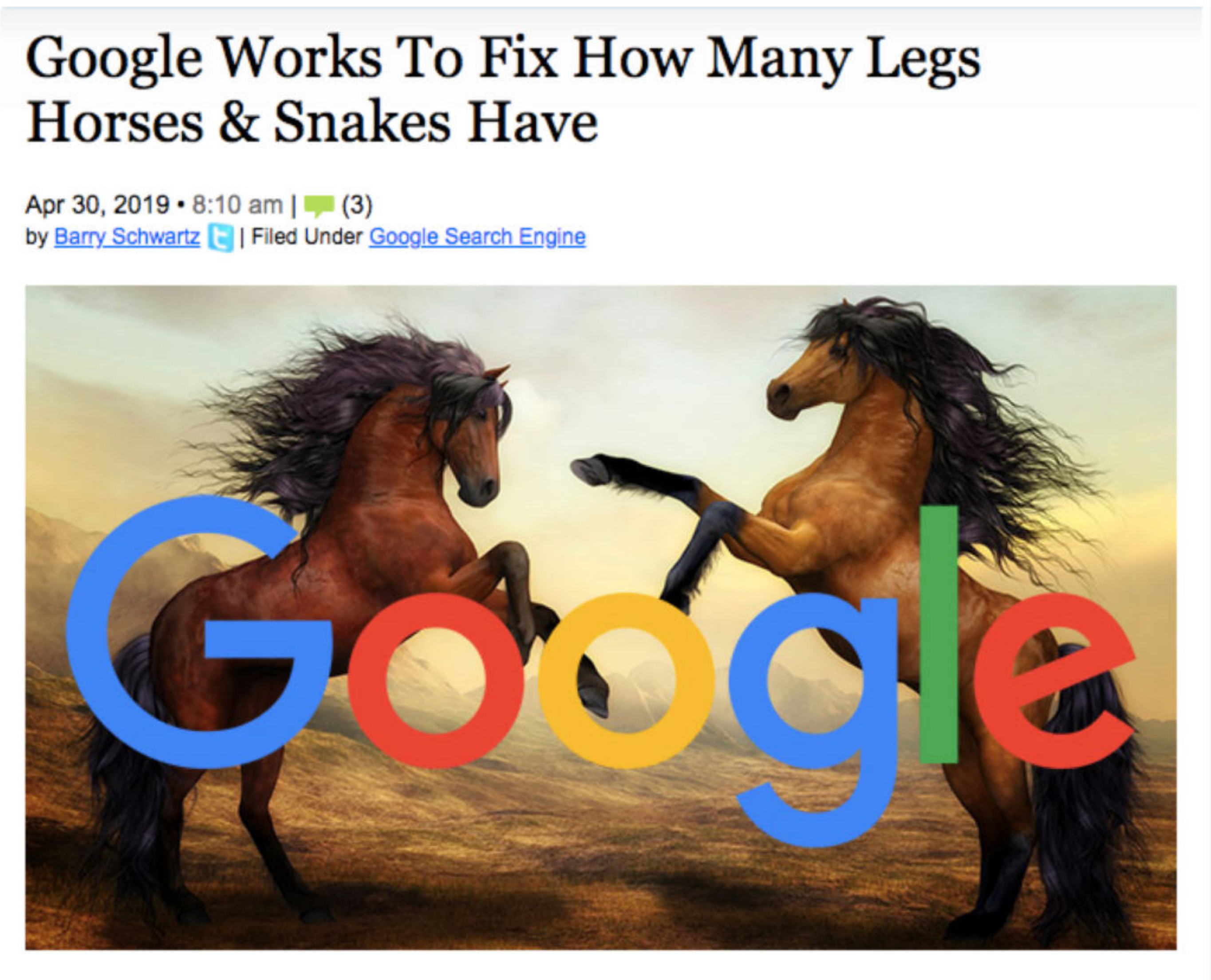
six legs

All horses have **two legs** at the back. They also have fore legs at the front. That makes it a total of **six legs**. But six is an odd number of legs for a horse to have.

Wikipedia:How many legs does a horse have? - Wikipedia

https://en.wikipedia.org/wiki/Wikipedia:How_many_legs_does_a_horse_have%3F

Somes Reputable Knowledge Repositories are Just Plain ‘Wrong’



Google Works To Fix How Many Legs Horses & Snakes Have

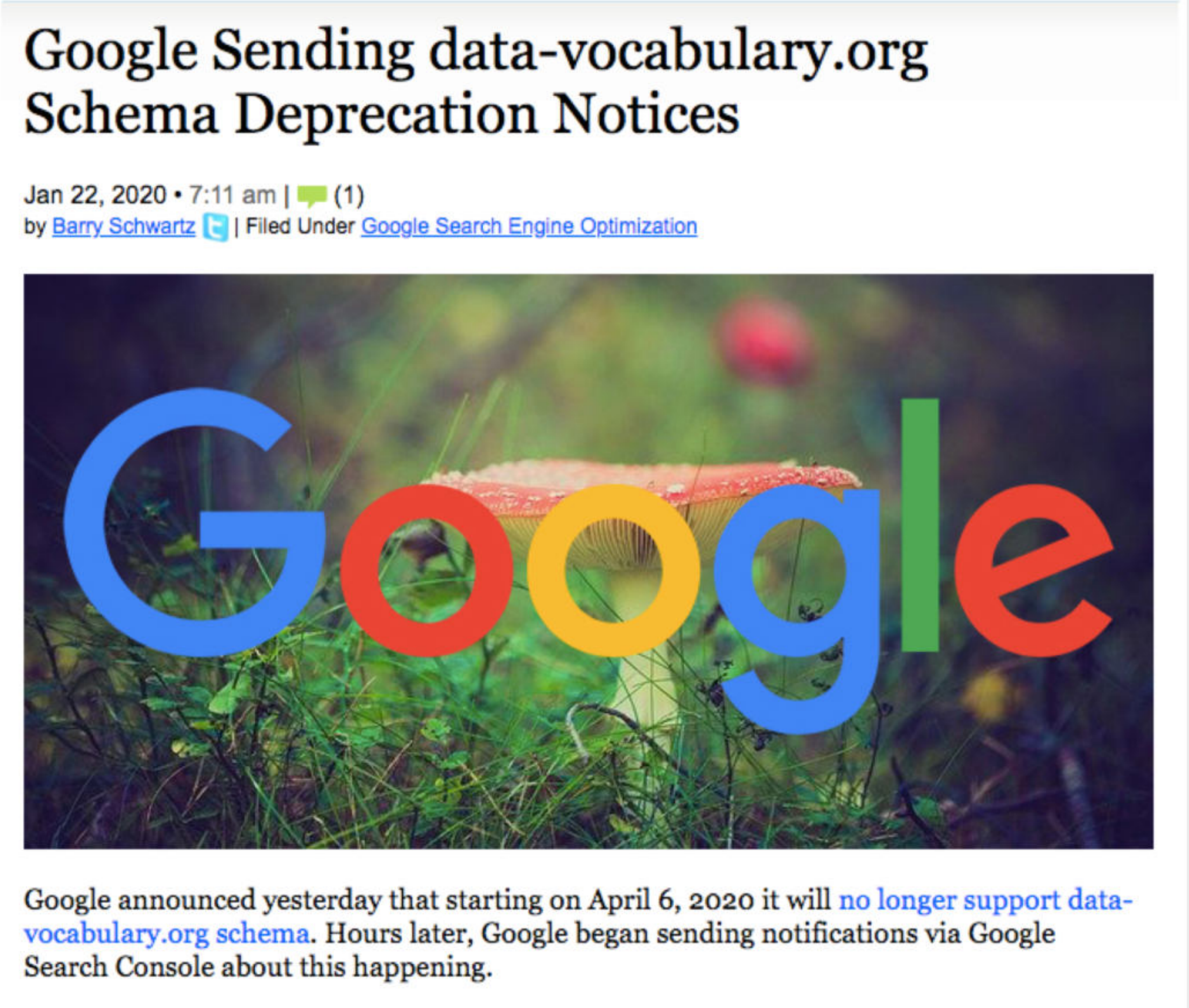
Apr 30, 2019 • 8:10 am | 💬 (3)
by Barry Schwartz | Filed Under Google Search Engine

Equivalency between data sources is sought

Efforts Are Underway to Resolve

Since 2011 major search engines have been focusing on Schema commonly

Google Deprecating Data-vocabulary Schema

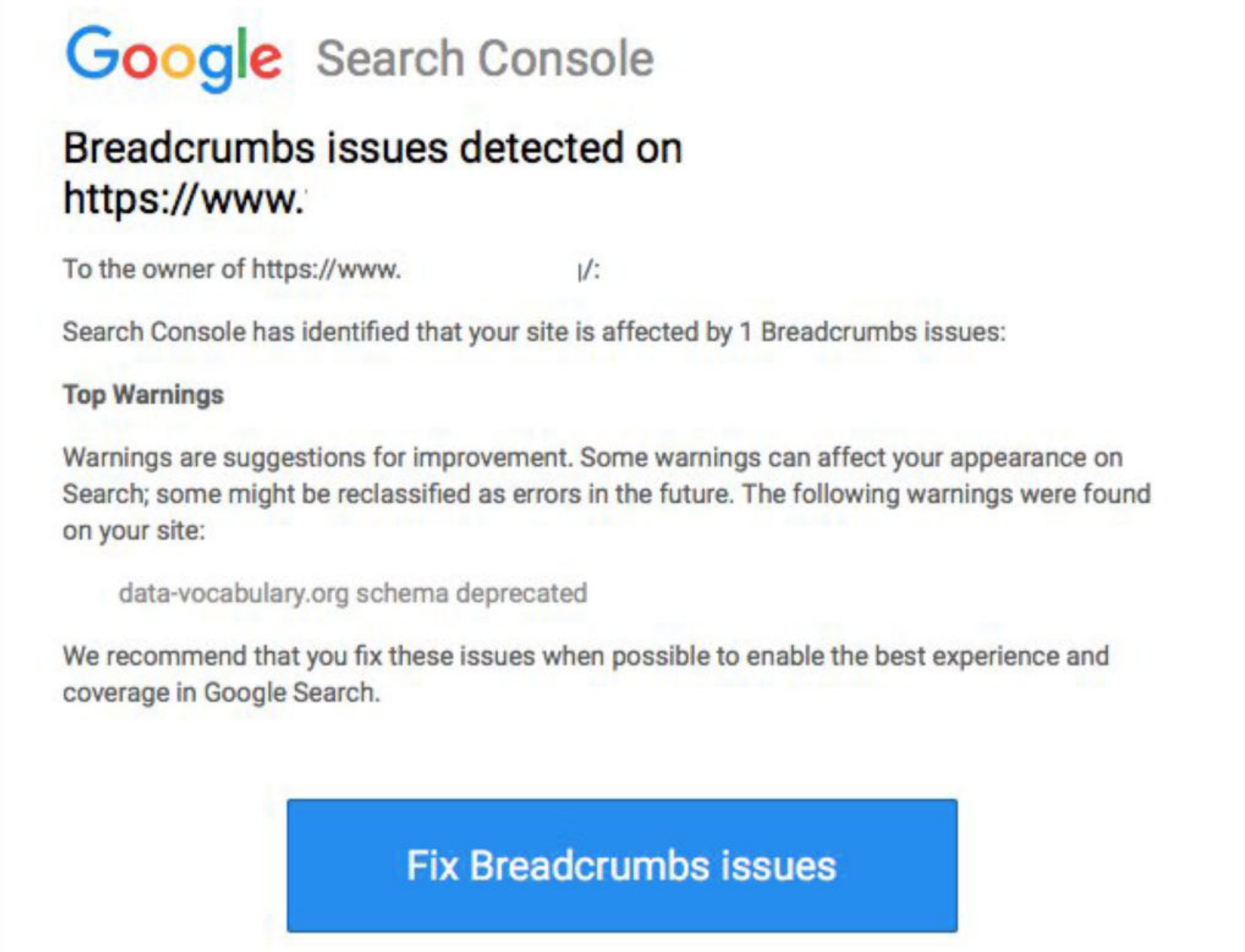


Google Sending data-vocabulary.org Schema Deprecation Notices

Jan 22, 2020 • 7:11 am | (1)
by Barry Schwartz | Filed Under Google Search Engine Optimization

Google announced yesterday that starting on April 6, 2020 it will no longer support data-vocabulary.org schema. Hours later, Google began sending notifications via Google Search Console about this happening.

Your Breadcrumbs Markup May Be Impacted



Google Search Console

Breadcrumbs issues detected on https://www.

To the owner of https://www. /:

Search Console has identified that your site is affected by 1 Breadcrumbs issues:

Top Warnings

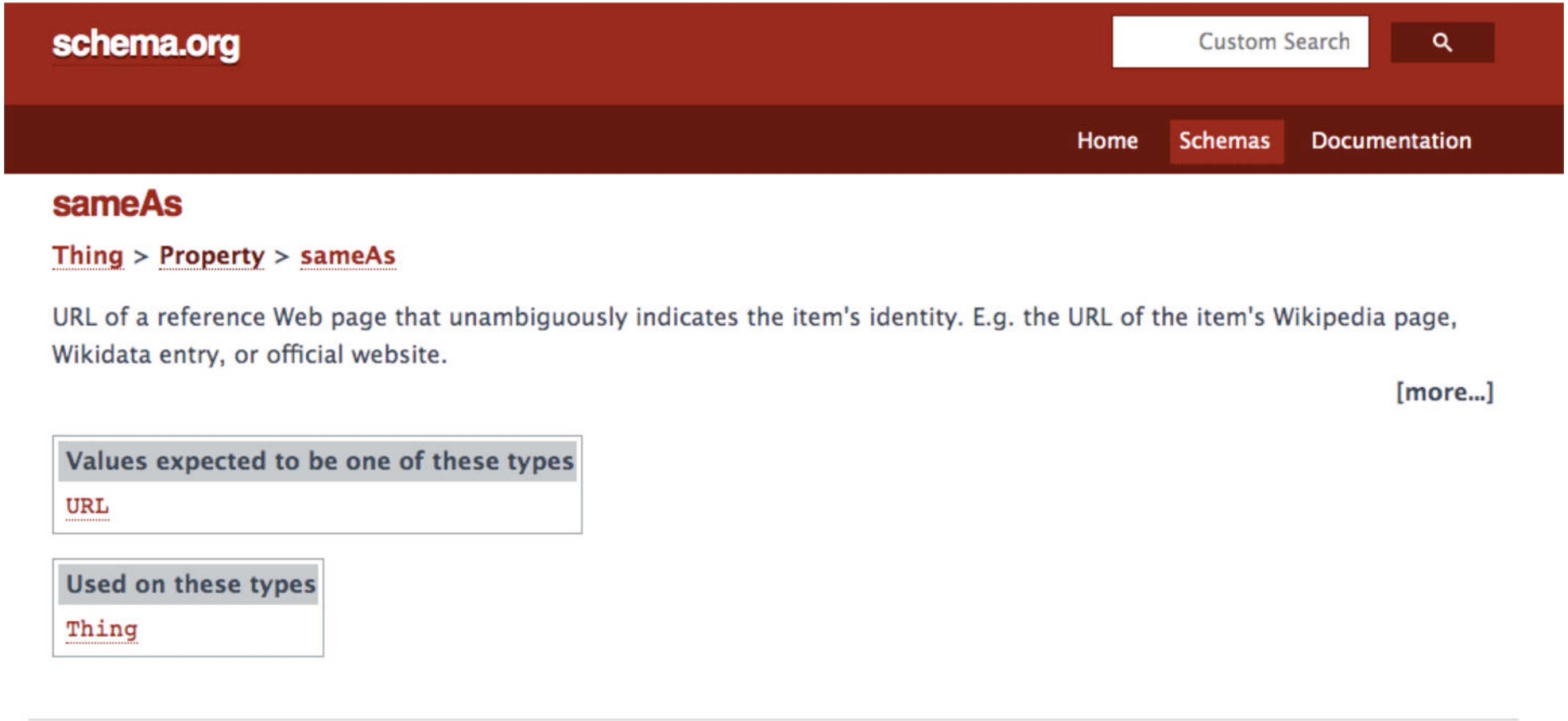
Warnings are suggestions for improvement. Some warnings can affect your appearance on Search; some might be reclassified as errors in the future. The following warnings were found on your site:

data-vocabulary.org schema deprecated

We recommend that you fix these issues when possible to enable the best experience and coverage in Google Search.

Entities often have many surface forms

Schema:SameAs and OWL:SameAs



sameAs

Thing > Property > sameAs

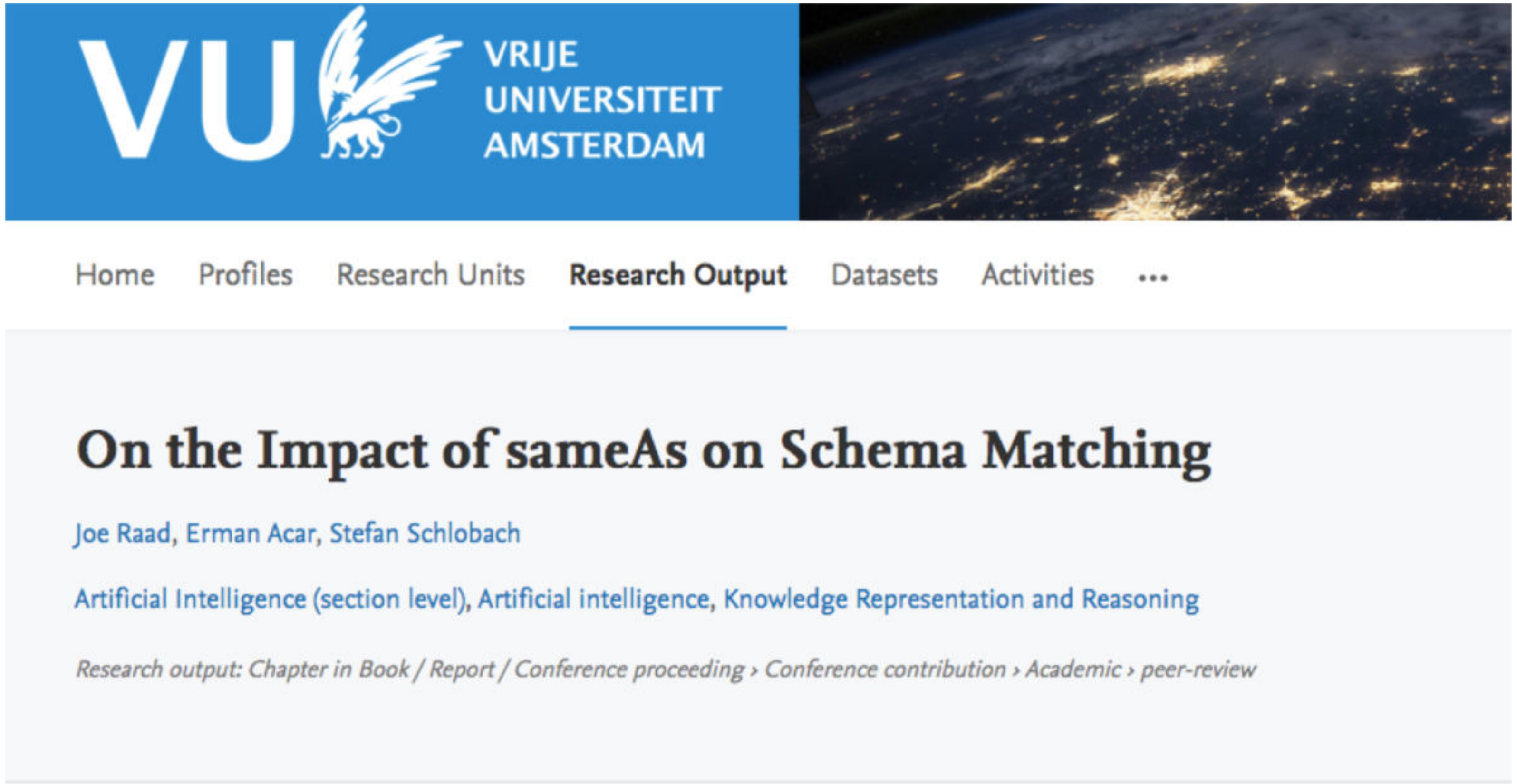
URL of a reference Web page that unambiguously indicates the item's identity. E.g. the URL of the item's Wikipedia page, Wikidata entry, or official website.

[more...]

Values expected to be one of these types
URL

Used on these types
Thing

SameAs Seems To Be Effective



On the Impact of sameAs on Schema Matching

Joe Raad, Erman Acar, Stefan Schlobach

Artificial Intelligence (section level), Artificial intelligence, Knowledge Representation and Reasoning

Research output: Chapter in Book / Report / Conference proceeding › Conference contribution › Academic › peer-review

BE consistent in ALL mentions of your brand name & ALL THINGS GENERALLY

Remember: Consistency is one of the Kings

Utilise ‘known’ (popular) public datasets & knowledge repositories

Avoid conflict with ‘known’ (popular) public datasets & ‘knowledge repositories’

Realise that well known knowledge repositories may be wrong

But they may be conceptually still be ‘right’

They’ll likely win in an equipossibility face-off

Leave no room for ambiguity where possible

Google puts the pieces of the puzzle together often badly

Location based ambiguity

Semantic Heterogeneity is a problem locally

Entity Address (location) is a VERY Big Problem with entity-oriented search

street number

Locality - city or town

street/route name, if detected

postal code, if detected

country, if detected

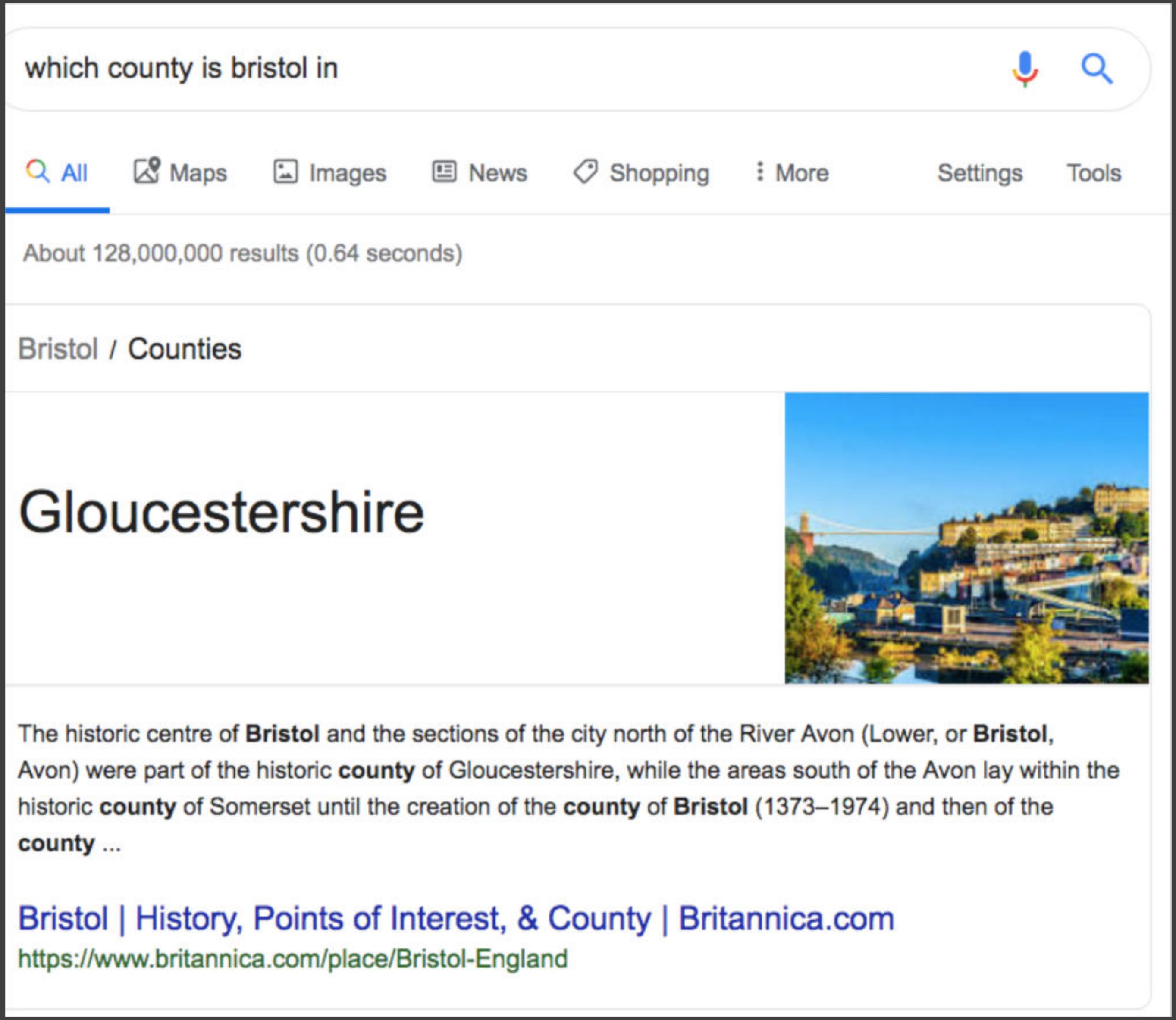
broad_region - administrative area, such as the state, if detected

Narrow_region - smaller administrative area, such as county, if detected

Narrow_Region is VERY Problematic

- Inconsistent data from knowledge repositories
- Inconsistent data from websites
- Inconsistent data from searchers
- HUGELY SPLIT EQUIPROBABILITY

Which County is Bristol in?



About 128,000,000 results (0.64 seconds)

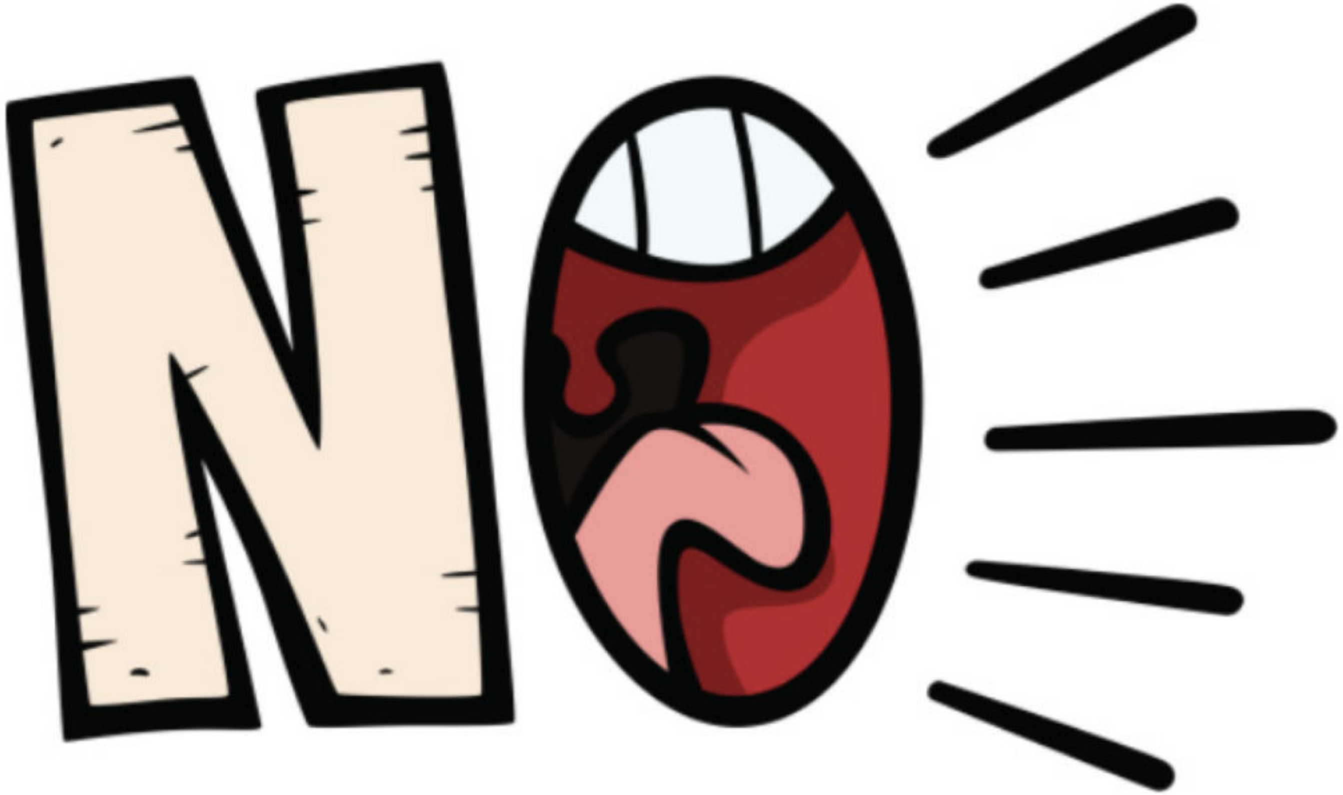
Bristol / Counties

Gloucestershire

The historic centre of **Bristol** and the sections of the city north of the River Avon (Lower, or **Bristol**, Avon) were part of the historic **county** of Gloucestershire, while the areas south of the Avon lay within the historic **county** of Somerset until the creation of the **county** of **Bristol** (1373–1974) and then of the **county** ...

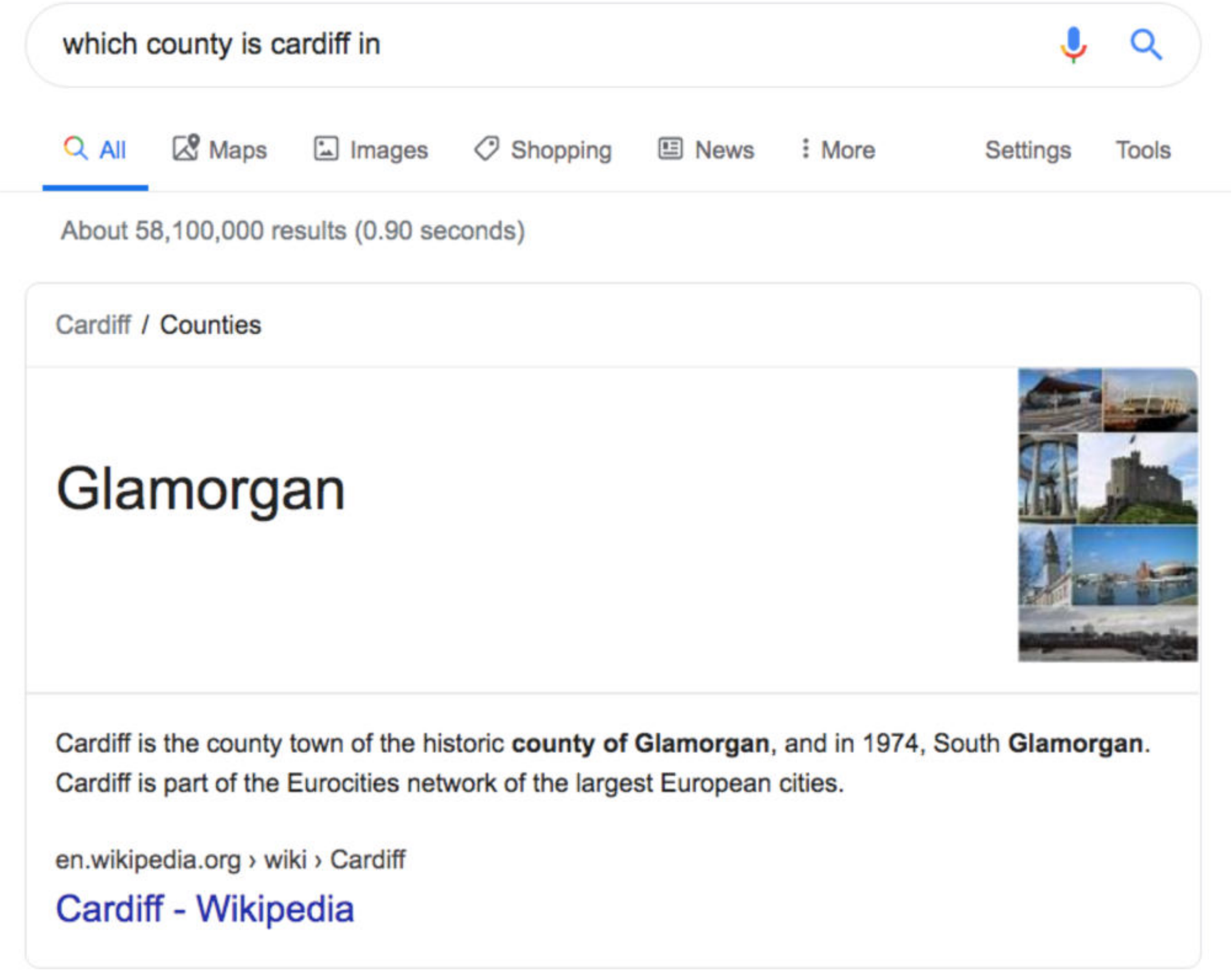
Bristol | History, Points of Interest, & County | Britannica.com
https://www.britannica.com/place/Bristol-England

Bristol is BOTH a county and a city



In Scotland and Wales it’s worse

Which County is Cardiff in?



About 58,100,000 results (0.90 seconds)

Cardiff / Counties

Glamorgan

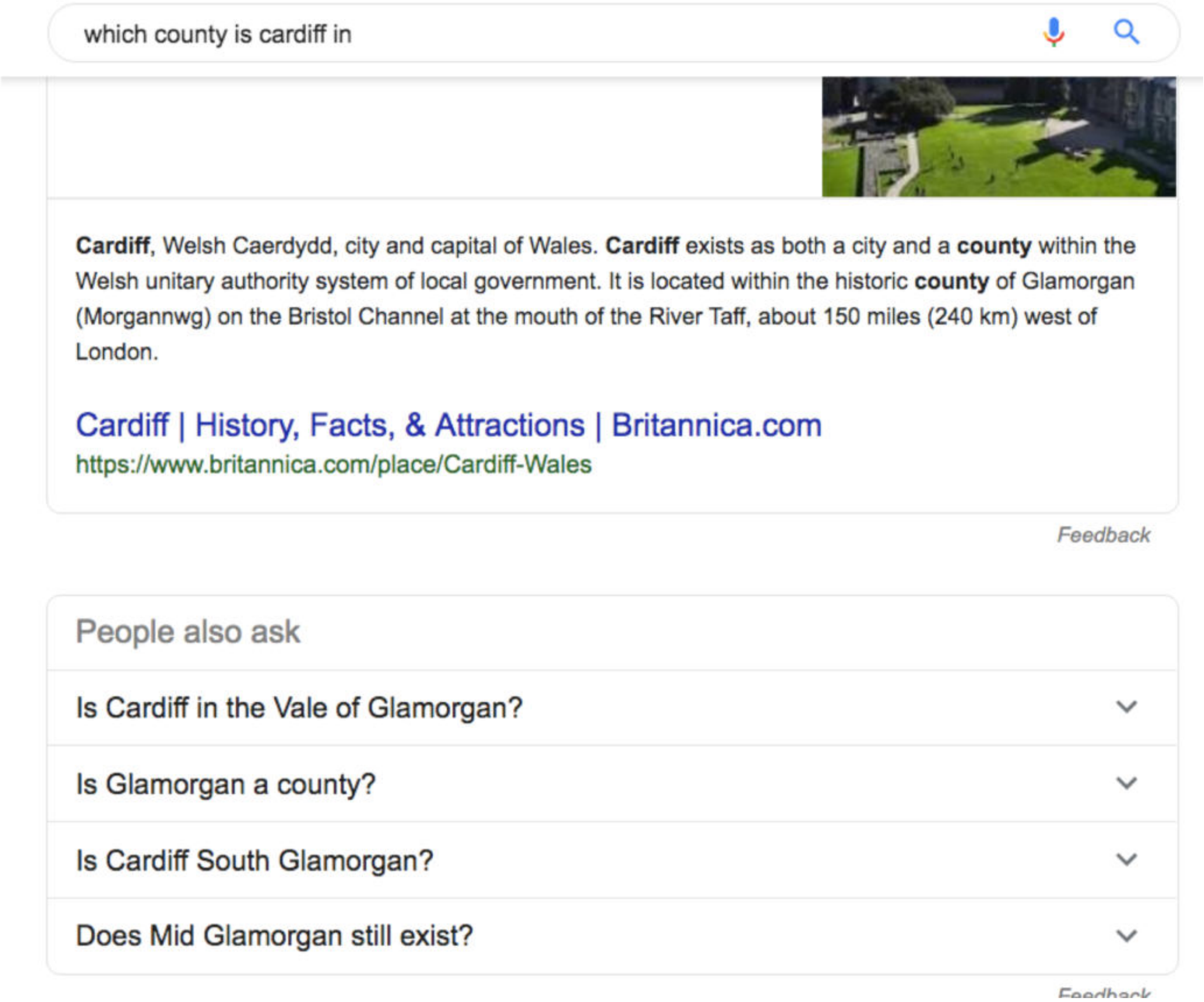
Cardiff is the county town of the historic **county of Glamorgan**, and in 1974, South **Glamorgan**. Cardiff is part of the Eurocities network of the largest European cities.

en.wikipedia.org › wiki › Cardiff

Cardiff - Wikipedia

Cardiff is in South Glamorgan NOT Glamorgan

But even humans are not so sure



which county is cardiff in

Cardiff, Welsh Caerdydd, city and capital of Wales. **Cardiff** exists as both a city and a **county** within the Welsh unitary authority system of local government. It is located within the historic **county** of Glamorgan (Morgannwg) on the Bristol Channel at the mouth of the River Taff, about 150 miles (240 km) west of London.

Cardiff | History, Facts, & Attractions | Britannica.com
https://www.britannica.com/place/Cardiff-Wales

People also ask

Is Cardiff in the Vale of Glamorgan?

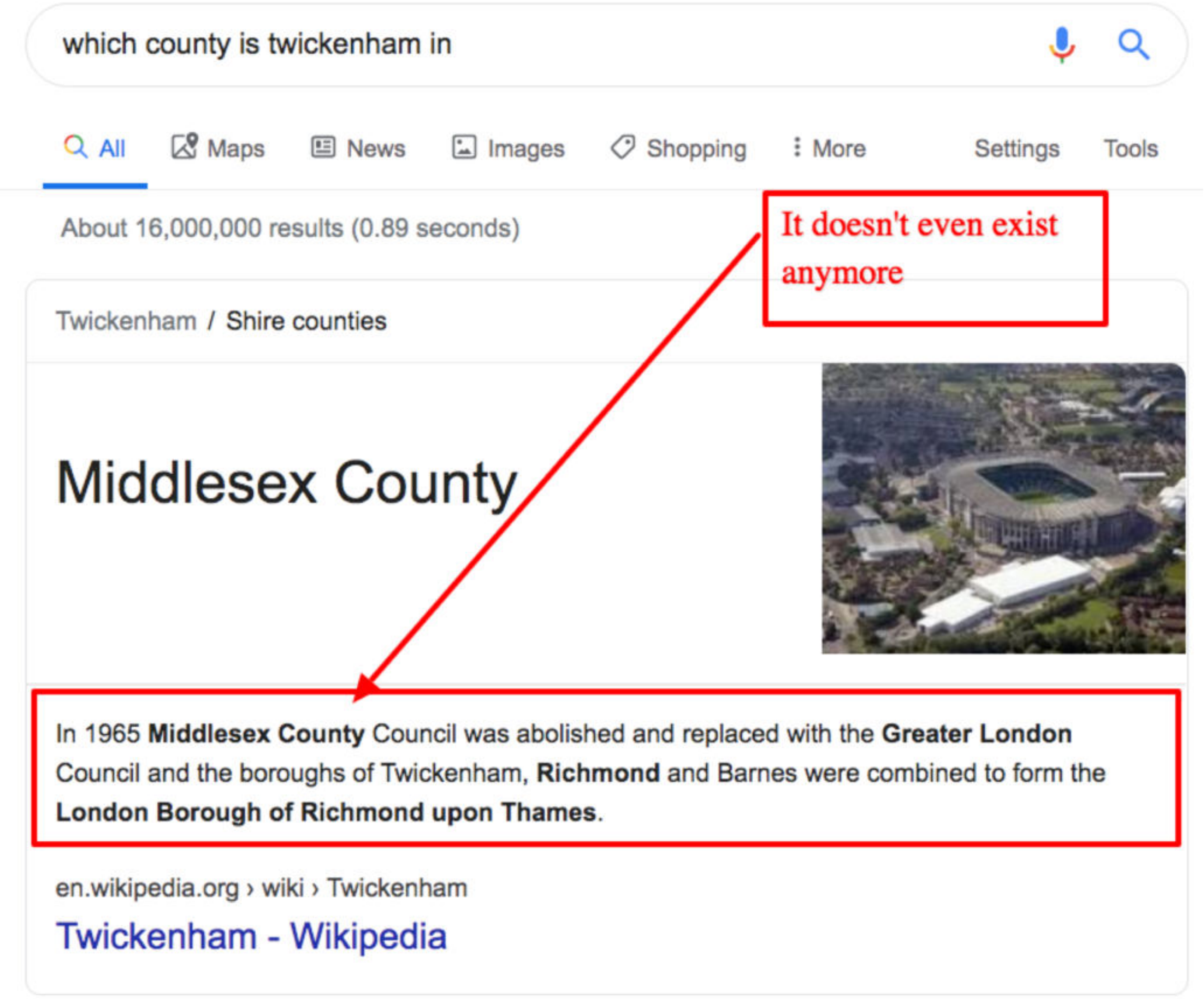
Is Glamorgan a county?

Is Cardiff South Glamorgan?

Does Mid Glamorgan still exist?

Google Does Not Understand the Word ‘Historic'

What about counties that don’t even exist any more?



which county is twickenham in

About 16,000,000 results (0.89 seconds)

It doesn't even exist anymore

Twickenham / Shire counties

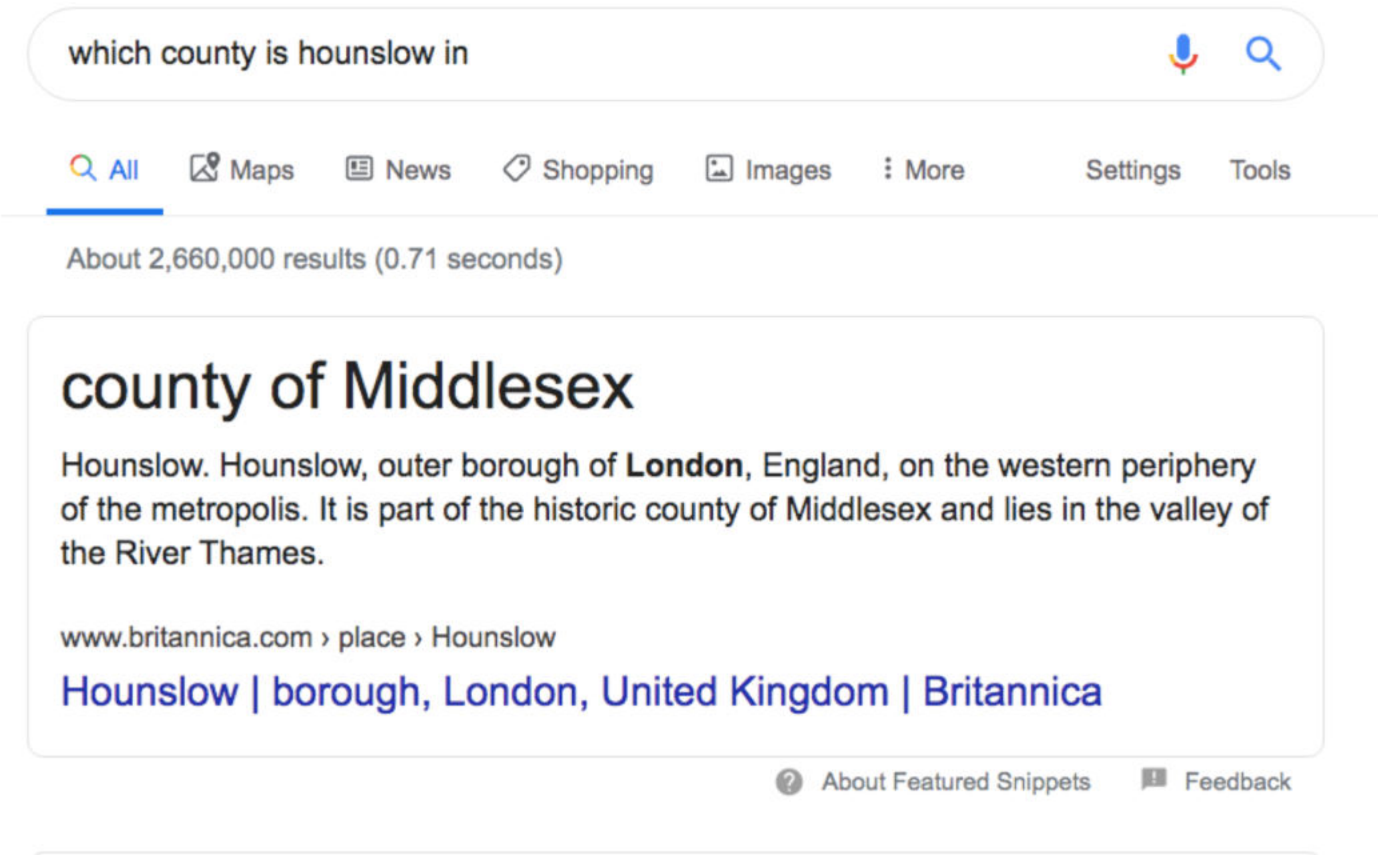
Middlesex County

In 1965 **Middlesex County** Council was abolished and replaced with the **Greater London** Council and the boroughs of Twickenham, **Richmond** and Barnes were combined to form the **London Borough of Richmond upon Thames**.

en.wikipedia.org › wiki › Twickenham

Twickenham - Wikipedia

At least this one mentions London as well as Middlesex



About 2,660,000 results (0.71 seconds)

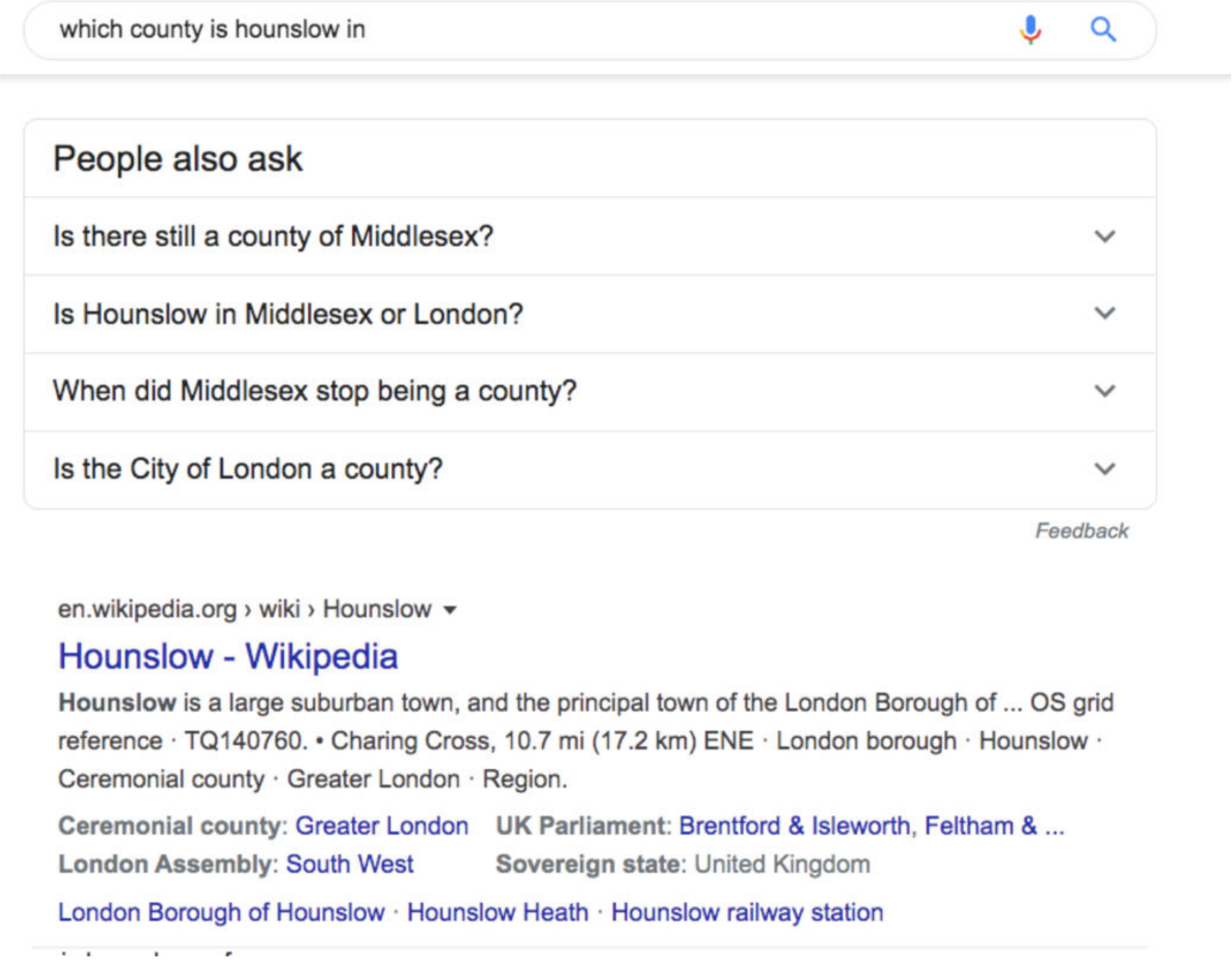
county of Middlesex

Hounslow. Hounslow, outer borough of **London**, England, on the western periphery of the metropolis. It is part of the historic county of Middlesex and lies in the valley of the River Thames.

www.britannica.com › place › Hounslow

Hounslow | borough, London, United Kingdom | Britannica

Everyone... Is confused



which county is hounslow in

People also ask

Is there still a county of Middlesex?

Is Hounslow in Middlesex or London?

When did Middlesex stop being a county?

Is the City of London a county?

Feedback

en.wikipedia.org › wiki › Hounslow

Hounslow - Wikipedia

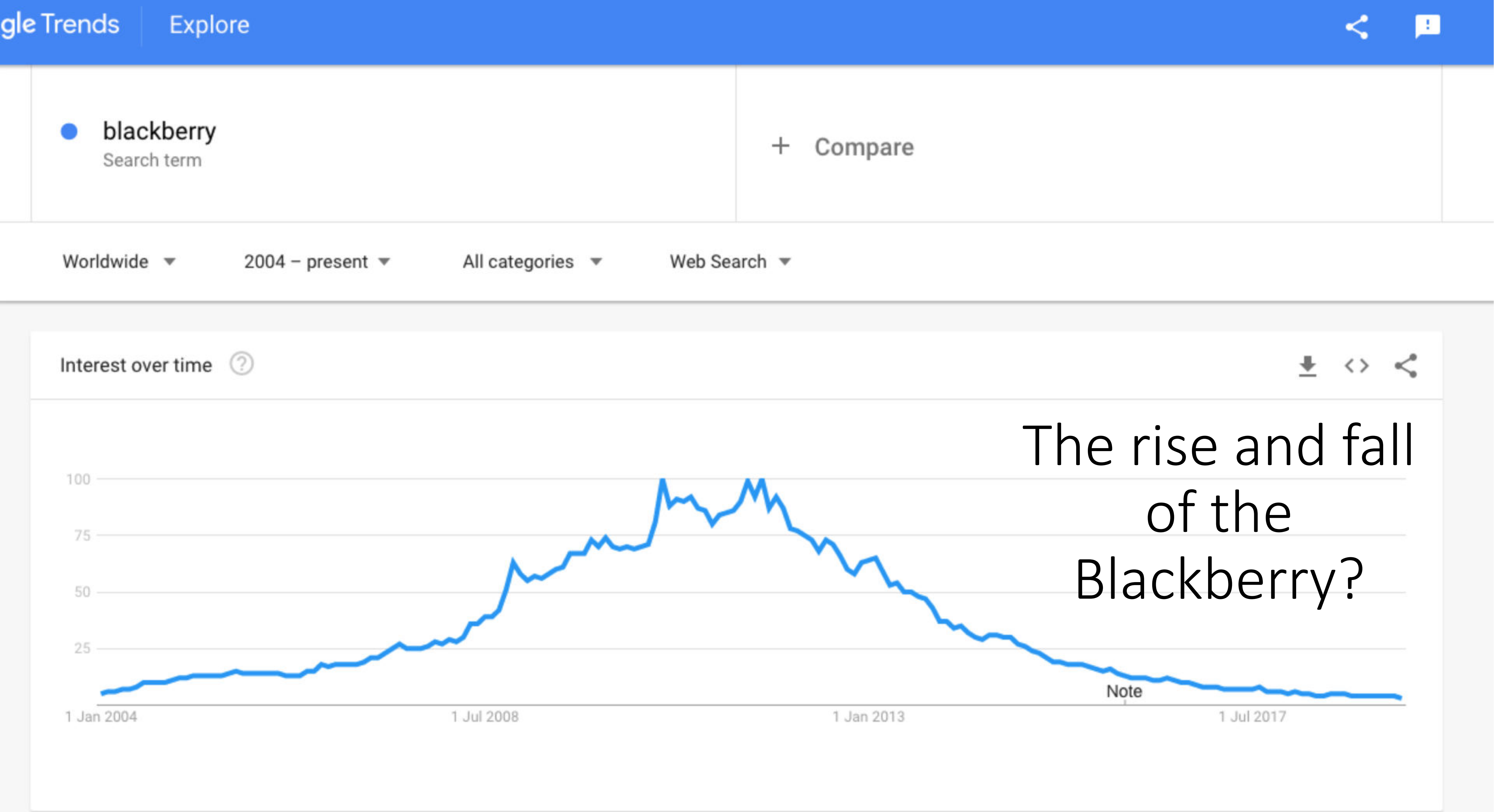
Hounslow is a large suburban town, and the principal town of the London Borough of ... OS grid reference · TQ140760. • Charing Cross, 10.7 mi (17.2 km) ENE · London borough · Hounslow · Ceremonial county · Greater London · Region.

Ceremonial county: Greater London
UK Parliament: Brentford & Isleworth, Feltham & ...
London Assembly: South West
Sovereign state: United Kingdom

London Borough of Hounslow · Hounslow Heath · Hounslow railway station

The passage of time adds new meaning to queries sometimes too

The rise and fall of the Blackberry?



gle Trends | Explore

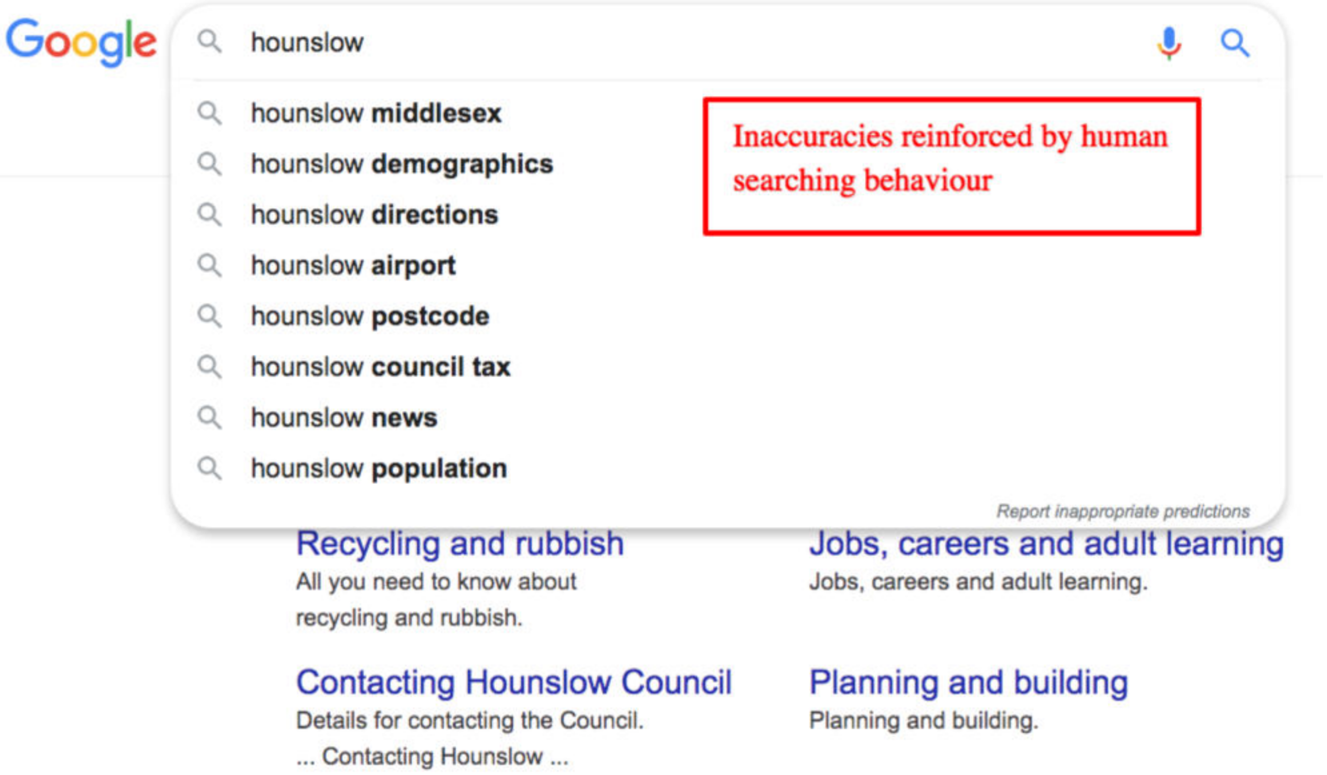
● blackberry
Search term

+ Compare

Worldwide ▾ | 2004 – present ▾ | All categories ▾ | Web Search ▾

Interest over time

Older humans probably keep ambiguous county + town combinations alive



Google

hounslow

- hounslow **middlesex**
- hounslow **demographics**
- hounslow **directions**
- hounslow **airport**
- hounslow **postcode**
- hounslow **council tax**
- hounslow **news**
- hounslow **population**

Report inappropriate predictions

Inaccuracies reinforced by human searching behaviour

Recycling and rubbish
All you need to know about recycling and rubbish.

Jobs, careers and adult learning
Jobs, careers and adult learning.

Contacting Hounslow Council
Details for contacting the Council. ... Contacting Hounslow ...

Planning and building
Planning and building.

Maximise the ‘Local’ Opportunity

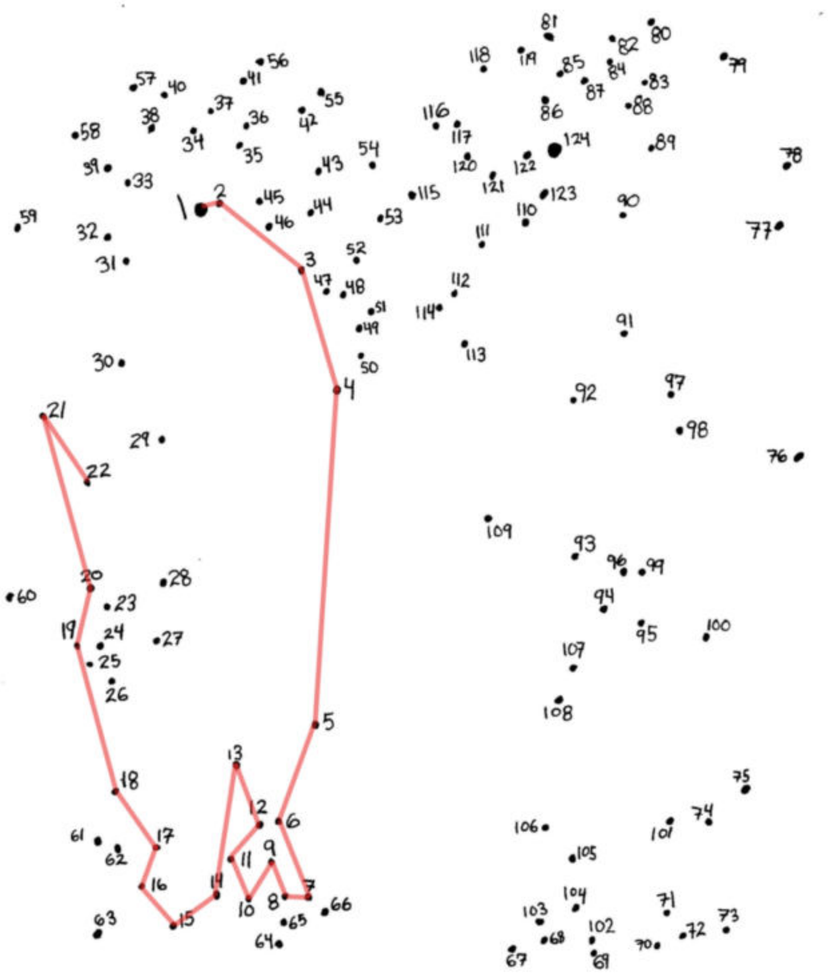
In location focused pages adding a postcode could be much more valuable than adding lots of words

Contextual relatedness in local spaces is proximity-based

AreaServed Schema & Internal linking by Proximity (e.g. lat & long) is powerful

Google sometimes puts the pieces of the puzzle together sporadically

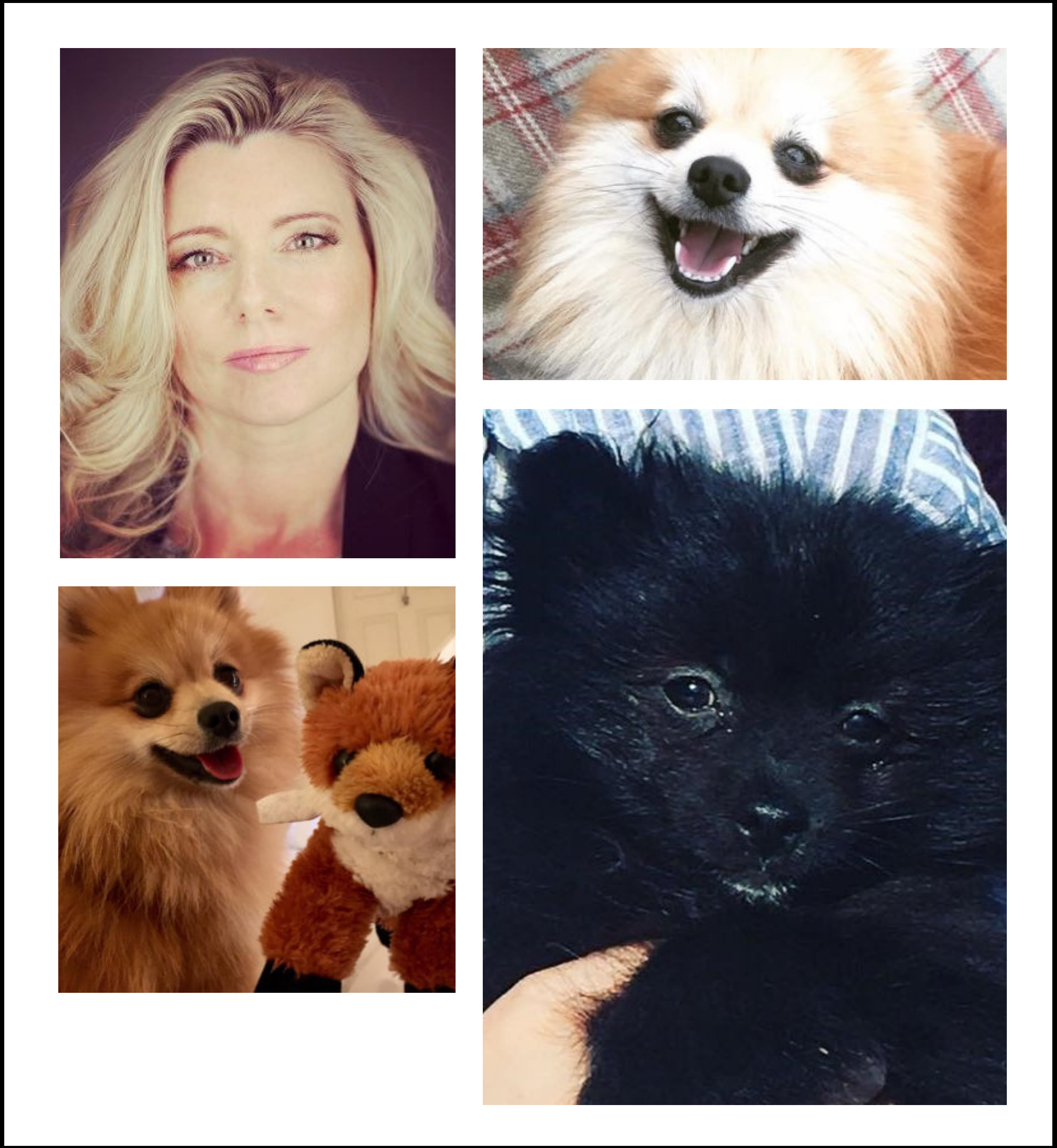
Align ALL the dots



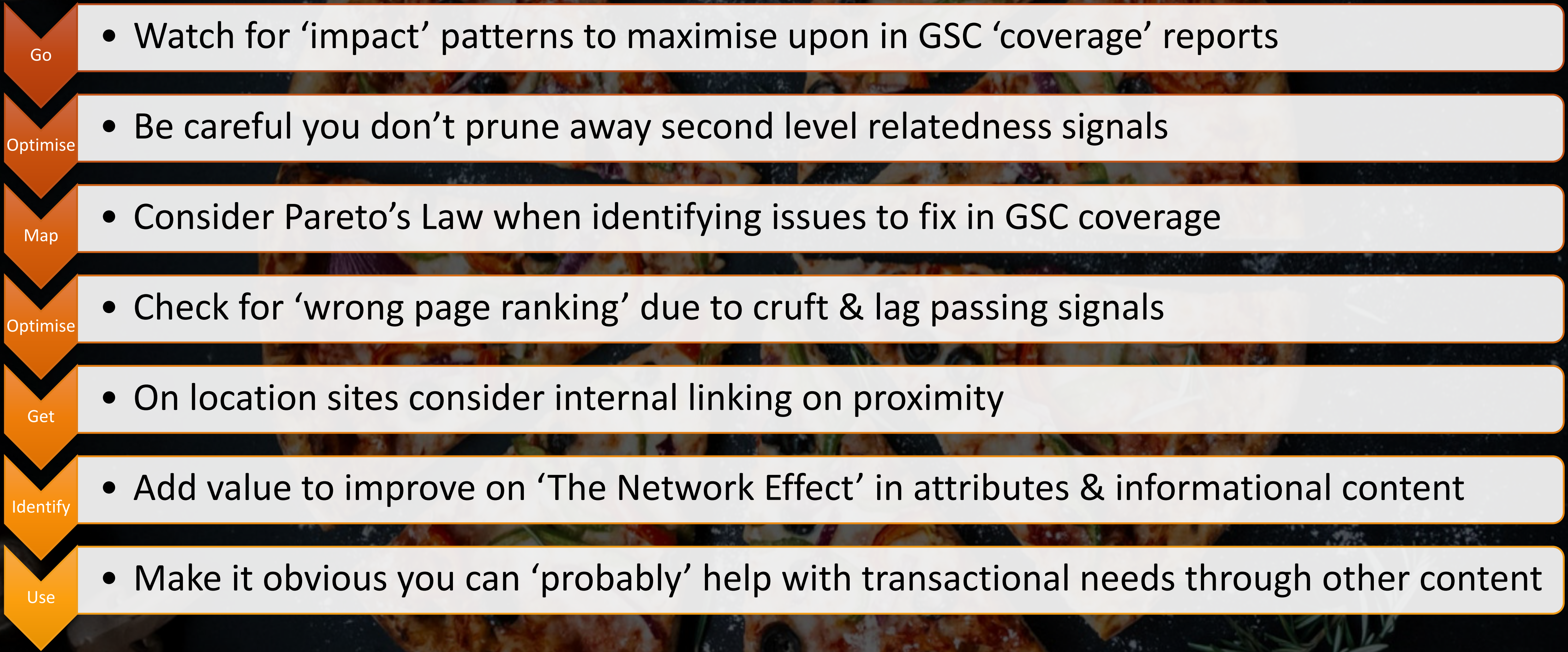
Most Importantly - Be Consistent

Keep in Touch

- @dawnieando
- @BeBertey



Some Take Aways



Go

- Watch for ‘impact’ patterns to maximise upon in GSC ‘coverage’ reports

Optimise

- Be careful you don’t prune away second level relatedness signals

Map

- Consider Pareto’s Law when identifying issues to fix in GSC coverage

Optimise

- Check for ‘wrong page ranking’ due to cruft & lag passing signals

Get

- On location sites consider internal linking on proximity

Identify

- Add value to improve on ‘The Network Effect’ in attributes & informational content

Use

- Make it obvious you can ‘probably’ help with transactional needs through other content