

# Preliminary Thoughts 1/21

## **Core Idea:**

- 1) Google wants to give you the "highest quality"
- 2) High quality is subjective - we need to ask who is determining the quality?
- 3) Google is "playing god" by defining high quality - which defaults to advertising and bias

## **Solution:**

Transactional, localized → to create an unbiased result and how would we do

- By developing a more comprehensive and unbiased search directory, we can help people to more easily find the information they need and make informed decisions, free from the constraints of traditional search algorithms that have rapidly become commercially "gamed" and biased.
- [Telephone directory - Wikipedia](#)      [Library of Congress](#)

The screenshot shows an email from Google Location History to Sam. The subject is "Reminder: You have Location History turned on". The email body contains a message from Google explaining what Location History is and how it works. It also includes links for managing location history settings and a link to the Help Center.

**Reminder: You have Location History turned on**

**Science & Futuristic Technology**

**CNET Is Quietly Publishing Entire Articles Generated By AI (2 minute read)**

CNET has been using AI to generate its financial explainer articles. Articles published under 'CNET Money Staff' are generated using automation technology and then edited and fact-checked by an editor. CNET has so far put out around 73 AI-generated articles. Despite Google claiming to penalize AI-generated content, some of the articles have received large amounts of traffic. It appears that the company is trying to keep the experiment as lowkey as possible.

**Hi Sam,**

To protect your privacy, this is a reminder that you have Location History turned on.

Location History is a Google Account setting that creates Timeline, a personal map that helps you remember places you've been, and routes and trips you've taken. It can also give you more personalized experiences across Google, such as place recommendations, based on where you've been.

**You're in control**

You can review your Location History data and change your settings anytime. If you'd like to turn off Location History, review and delete your data, or auto-delete your data, visit [myactivity.google.com/activitycontrols](#).

For step-by-step instructions, visit our [Help Center](#).

8:46 AM (2 hours ago) ☆ ⓘ

# Week 1 - Readings

## The 3 Types of Search Queries & How You Should Target Them

- Navigational Search Queries
  - A navigational query is a search query entered with the intent of finding a particular website or webpage
  - "Youtube" or "facebook"
  - Have to own the main site basically
- Informational Searches
  - Queries that cover a broad topic (e.g., colorado or trucks) for which there may be thousands of relevant results.
  - Not a specific site, but they are looking for an answer to a question or learn how to do something
  - Google creates "Knowledge Graph" to answer it themselves
  - Their solution: The goal is to position yourself as a trustworthy, authoritative source of information
- Transactional Searches
  - **A transactional search query is a query that indicates an intent to complete a transaction, such as making a purchase**
  - May include exact brand and product names or be generic or actually include terms like buy/purchase/order
    - There is business end of the conversion funnel
  - Many local searches are transactional as well → "Dever wine shop"
    - Vertical searches are a subset of transactional search queries, and they represent people looking to make a transaction in a specific industry
      - These include local searches, restaurant searches, hotel searches, flight searches, etc
  - *Article proceeds list various "tactics" to game/cater to the algorithm*

- These are exactly the kinds of queries that are mostly likely to deliver ROI in paid search. If people are looking for a specific type of product to buy, a sponsored ad is just as likely as an organic result to deliver what they need.
  - Sponsored results take up a lot of the available space on the SERP for commercial/transactional queries. If you want visibility above the fold for transactional keywords, you should consider PPC.
  - Google offers lots of bells and whistles for sponsored ads and product listings. For example, you can include a picture of your product. Your options in organic results are more limited and less controllable.
  - In one study, we found that people click on paid results over organic results 2 to 1 for queries with high commercial intent. This is probably because sponsored results take up so much of the above-the-fold real estate on these types of searches, because the new ad formats are so eye-catchingly clickable, and because lots of search engine users can't tell the difference between ads and non-ads. (NB: Commercial search queries are just a small percentage of total search query volume overall, so organic results still take the lion's share of overall clicks. [More on that here.](#))
- 

## Is local search inherently transactional?

- Example: When does the Post Office open?
  - There's no featured snippet or answer box in this case, because Google is answering the question directly in the local pack (which appears first in my results). How about a search for "Where is the Space Needle in Seattle?"
- Am I interested in the history of tacos at this moment? No, Google has rightly determined that I want to find conveniently located tacos and put them in my mouth.
- **Miriam raises a critical point – even if searcher intent isn't always transactional, Google has a vested interest in driving local transactions, and SERP features will always reflect Google's interests.**
- <https://pointy.withgoogle.com/>

## What is Local SEO, and How Does It Work? | Mailchimp

- Any business that has a physical location or serves a geographic area can benefit from local SEO
- Today Google looks at hundreds of factors both on and off your website to determine if it's relevant to display as a search result.
- Most people don't realize that Google isn't doing a live search of the entire internet every time someone types something into the search

bar. It's actually searching a stored copy of all of the sites that Google has discovered. This copy is called the Google Index.

- **Google figured out that people seeking certain types of businesses need results from their immediate area. That's why Google's local search algorithm includes a proximity factor, which is a fancy way of saying that Google takes your location into account when you search for a local keyword**
- Article proceeds to explain tactics local businesses can use to rank higher - in an already very limited local panel

## PhantomBuster

- Lead generation

## Ranking Results – How Google Search Works

- Google's ranking systems are designed to do just that: sort through hundreds of billions of webpages and other content in our Search index to present the most relevant, useful results in a fraction of a second
- To give you the **most useful information, Search algorithms look at many factors and signals**, including the words of your query, relevance and usability of pages, expertise of sources, and your location and settings. The weight applied to each factor varies depending on the nature of your query.
- Meaning
  - *This involves steps as seemingly simple as recognizing and correcting spelling mistakes, and extends to trying to our sophisticated synonym system that allows us to find relevant documents even if they don't contain the exact words you used*
    - “change laptop brightness” and finds “adjust laptop brightness”
  - The term “change” could have different meanings: replace, exchange, adjust
- Relevance
  - The most basic signal that information is relevant is when content contains the same keywords as your search query
- Quality of content
  - expertise, authoritativeness, and trustworthiness.
- Usability of web pages

- For example, our systems would look at page experience aspects, such as if content is mobile-friendly, so that those on mobile devices can easily view it.
- Similarly, they look to see if content loads quickly, also important to mobile users.
- **This is a point we could attack**
- Context and Settings
  - Information such as your location, past Search history, and Search settings all help us to ensure your results are what is most useful and relevant for you in that moment.
  - *We use your country and location to deliver content relevant for your area. For instance, if you're in Chicago and you search "football", Google will most likely show you results about American football and the Chicago Bears first. Whereas if you search "football" in London, Google will show results about soccer and the Premier League*
- **Some sort of verification system is needed**

## Our Approach – How Google Search Works

- "That's why Search makes it easy to discover a broad range of information from a wide variety of sources." - Google
  - **NOT true when it comes to transactional searches! Biased against local businesses**
- Statement on "Sell ads, nothing more"
  - Ads are how Search remains accessible to everyone. But ads are the only thing we sell.
  - We don't charge anyone to appear in our search index. Whether a business, individual or organization buys ads is not a factor in our search algorithms. We never provide special treatment to advertisers in how our search algorithms rank their websites, and nobody can pay us to do so.

## 2022 Google Algorithm Recap and Looking Forward to 2023 – Vizion Interactive

- Big factor in decisions is page speed, load performance, responsiveness to user input
- August 2022 Update

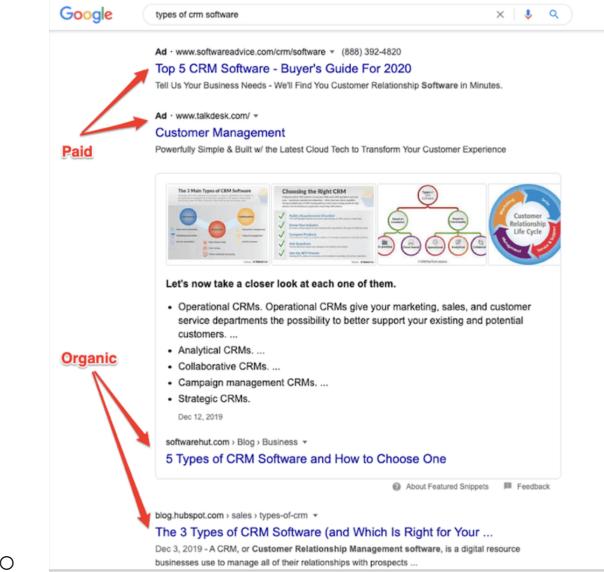
- It's designed to look for AI and system generated content that's built solely for search engines and not for people. The automated ranking systems of Google identify low-value content and classifies that site, or page, as having unhelpful content.
- Key Takeaway for your content strategy
  - Prove ownership
- Google has thirty-two guides for the most popular types of schema in their help document: Understand How Structured Data Markup Works.

## Optimize your site for search engines (for beginners)

- "Google doesn't accept payment to rank a website higher, or rank websites higher if they advertise on Google. If anyone tells you otherwise, they're wrong."
- Be complete: Mention everything that you have to offer. Google is smart, but we can't guess what you don't tell us.
- **Keep your site up to date:** If you haven't updated your site since 1925, it's time to remove the references to Calvin Coolidge in your current events blog.
  - **Small businesses can't do this since they are busy  
RUNNING THEIR SMALL BUSINESS**
- Get referrals: Have your site mentioned in useful places online. See if you can get into your local online news sites, or in any other appropriate resources.
  - **All these things are catering to Google and don't actually help people get what they need**
- Want your site to do better, but don't have the time? **If you think your site needs more help than you can provide, consider hiring a professional.**

## SERP 101: All About Search Engine Results Pages

- SERPs



- Every SERP is unique, even for search queries performed on the same search engine using the same keywords or search queries.
  - **We want to give one universal SERP for a given area**
- Two types of content
  - Organic and paid
  - Knowledge Box → Google
- **Commercial intent** is high for transactional searches
- Ranking signals → the search algorithm used by Google features hundreds of ranking factors, and while nobody outside of Google knows precisely what they are, some are thought to be more important than others
- Gaining: Common off-page SEO techniques include link building and exchange, social bookmarking, content marketing, submissions to directories and search engine indexes, and the creation of online communities on social media
- Paid search functions as an auction. Advertisers bid on keywords that are relevant to their business that can trigger the display of their ads when users search for those terms
- **Google Business Profile - Get Listed on Google**

## A Complete Guide to the Google RankBrain Algorithm

- At its core, RankBrain is a machine learning system that builds off Hummingbird, which took Google from a “strings” to “things” environment.

- Took it from “reading” literal characters, and instead “seeing” the entity they represented
- Example/Hummingbird
  - Jason barnard - Until Hummingbird, Google saw those characters and a collection of 2 words and 13 characters which, arranged in that order, and used on a page often enough, would make it relevant for the search string “jason barnard”.
  - Each core entity is represented by a Machine ID.
  - ***In the context of RankBrain, is that with Hummingbird, Google (/m/045c7b) now understood the world as a collection of things and not strings***
- RankBrain is a pre-screening system
  - Google hadn’t seen 15% of queries and had no context for them - so RankBrain would also consider environmental contexts
- How it works
  - “Victoria bc” is not two things - but one
    - Can look to the meaning
  - For example, if Google sees the ID for Jason and the ID for me often co-occurring, they would be able to connect the two.
  - Remember, this system launched to address the problem of queries that Google had never encountered before, there are going to often be no-or-low volume words that can’t be monitored themselves.
- [\*\*https://cloud.google.com/natural-language/#natural-language-e-api-demo\*\*](https://cloud.google.com/natural-language/#natural-language-e-api-demo)
- “Optimize for RankBrain”
  - Review the entities with the NLP API.
  - Deeper dive into entity Knowledge Graph Explorer.
  - Take all of this and put it in the context of you (what device and location you have).
  - Run some tests from other locations and devices using a tool like Mobile Moxie to compare the results and understand how things skew.

## Do Unbiased Search Engines Exist? Top Alternative Search Engines to Consider - Neeva

- We have to remember than search engines SELECT which information to show us - which could be biased as they decide on relevance
- Search engines want to show results that people will click on, which means that they are biased towards engagement
  - Engagement isn't always a signal of quality content
  - **EXAMPLE:** Consider, for example, the top ranking results for personal health queries. Advertising driven sites like WebMD often outrank government sites like National Institutes of Health. This is because WebMD is better at getting people to read through their content and click around.
  - **EXAMPLE:** The same goes for sites that review products. You might be looking for the best sneakers, but the top results are likely sponsored by the very products they recommend. These sites have a big incentive to get you to interact by clicking on a link to buy something quickly. This helps them rank high, get more views, and pull in advertising money—without providing real service to the consumer
- Other types of bias
  - **Location** - Many people enjoy this feature, but it does mean that search results differ depending on where you are in the world.
  - **Search History** - Using your search history to personalize results can sometimes create the phenomenon of a "filter bubble"
  - **Offensive content** - Google, for example, altered its algorithms after it received complaints about the anti semitic and racist results that appeared when searching for "jews" and "black girls"
- They are inherently biased since they give priority to result that fit certain criteria
  - Keywords - Google's algorithms prioritize sites that share detailed, useful information; defines useful by key words
  - Get backlinks
- Ads power search engines
  - *When it comes to bias, ad-powered search poses two main issues. The first is conflict of interest. Ad-supported search engines allow advertisers to buy their way to the top of the*

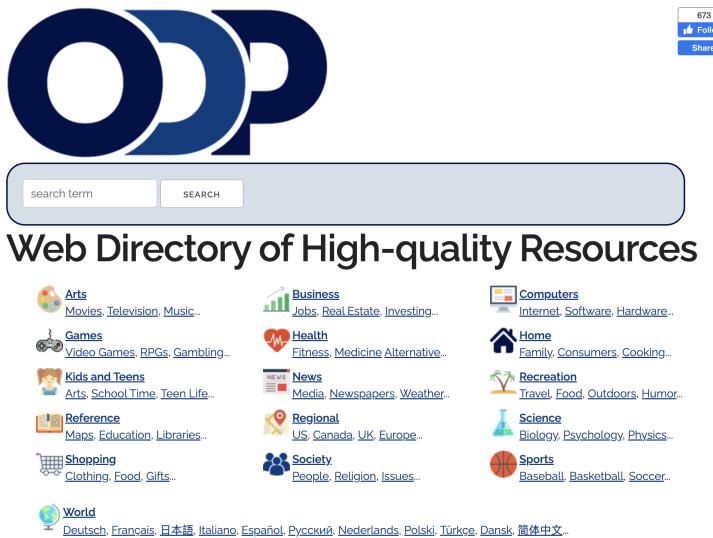
*results page for any given search term. This is especially true for retail queries—that is, queries for which the intent is to buy something.*

- **Just because a website ranks high on search results does not make it a reliable source**
- Top alternative search engines
  - Anonymous search engines
    - *Since they don't know anything about you, anonymous search engines eliminate the possibility of location or search history bias. What they don't eliminate is the conflict of interest that comes from having an ad-based model. Their results are also subject to engagement bias—you'll find this one nearly impossible to avoid.*
  - Metasearch Engines
    - Metasearch engines, like Searx and Metager, pull search results from hundreds of different sources, so they reflect a different set of biases from Google and Bing
  - Private Search engines
  - ^^^Lists various options

## 19 - Web search basics [not linked]

- Web is unprecedented in scale and lack of coordination in its creation
- <https://www.stanford.edu/home/atoz/contact.html>
  - Parts of the url
- The mass publishing of information on the Web is essentially useless unless this wealth of information can be discovered and consumed by other users
  - Two categories: (1) full-text index search engines such as Altavista, Excite and Infoseek and (2) taxonomies populated with web pages in categories, such as Yahoo!
    - *presented the user with a keyword search interface supported by inverted indexes and ranking mechanisms*
    - *Vs*
    - *allowed the user to browse through a hierarchical tree of category labels. While this is at first blush a convenient and intuitive metaphor for finding web pages, it has a number of drawbacks: first, accurately classifying web pages into taxonomy tree nodes is for the most part a*

*manual editorial process, which is difficult to scale with the size of the Web*



- - <http://www.odp.org/homepage.php>
  - High effort on human part
  - Sub-tree's need to be clear in the user's mind: *Furthermore, in order for a user to effectively discover web pages classified into the nodes of the taxonomy tree, the user's idea of what sub-tree(s) to seek for a particular topic should match that of the editors performing the classification*
- The essential feature that led to the explosive growth of the web – decentralized content publishing with essentially no central control of authorship – turned out to be the biggest challenge for web search engines in their quest to index and retrieve this content
  - **Businesses are finite with some barrier to entry (having the product)**
- Spam
  - Hiding words in the background to get better crawls
  - *At its root, spam stems from the heterogeneity of motives in content creation on the Web. In particular, many web content creators have commercial motives and therefore stand to gain from manipulating search engine results. You might argue that this is no different from a company that uses large fonts to list its phone numbers in the yellow pages; but this generally costs the company more and is thus a fairer mechanism*
- Link analysis/Link Spam

- Cost per mil (CPM) basis: the cost to the company of having its banner advertisement displayed 1000 times
  - Also cost per click model
  - Goto (then Overture) was the original bid for query terms
- Duplicates
  - The simplest approach to detecting duplicates is to compute, for each web page, a fingerprint that is a succinct (say 64-bit) digest of the characters on that page
  - Shingles
- Math equations/proofs (?)

## SEARCH NEUTRALITY AS AN ANTITRUST PRINCIPLE [not linked]

- "Objective metric of relevance"
- Arguing against a need to be neutral
- Ten Blue Links
  - Websites and search engines
  - Websites are information wells, search engines provide ultimate information
  - Vertical integration as SE's began operating websites
  - Inaccurate model of the internet
    - The one monopoly profit argument only holds if the complementary goods are consumed in fixed proportions, which search services and sites are not
  - ARGUMENT:
    - *Thus, a major flaw in the monopoly leverage story is that even if a particular search engine were dominant as a search vehicle, search engines are not necessarily dominant when it comes to reaching websites. In most cases, a critical mass of users knows where they want to go without conducting a search. Manipulation of a search engine to favor particular sites might induce more traffic to visit the site, but it seems unlikely that it could foreclose customers from reaching competitive sites*
- Consequently, search engines are no longer just focusing on document retrieval. Instead, they are working towards direct question answering." By "figur[ing] out the 'intent' of the person conducting the search" and then displaying all the related content that he might want

to see, search engines are shifting away from the paradigmatic ten blue links towards a world of richer results

- Search engines today are more than search engines - google shopping, google flights, knowledge panel, YouTube
- Pro search engine bias:
  - *Unless the search engine is to remain stuck in the ten blue links paradigm, search engine companies must have the freedom to make strategic choices about the design of their services, including the decision to embed proprietary functions traditionally performed by websites in the engine's search properties*
- Dominant firms may sometimes have special antitrust obligations not shared by weaker rivals, but those obligations should never stand in the way of the firm's ability to innovate
- *Acts of "pure malice" that do not result in distortions of the competitive market belong to the world of tort, not antitrust*

## Are Search Engines Biased? Detecting and Reducing Bias using Meta Search Engines

- *The main objectives are to develop an interpretable model of search engine behaviors and to design statistical tests pointing out suspicious instances as possible bias, without knowing the detailed ranking algorithms implemented by search engines*
- Google, Bing, Yahoo, Duckduckgo, Qwant, Startpage
- Organic Links
- This has led to the so-called and still vivid search neutrality debate, asking whether SEs should be neutral in their organic rankings—i.e., base their rankings on relevance only—and if so, how that can be monitored
  - Connected to Net Neutrality - ISPs should treat equally all content flowing through their networks, regardless of its kind, source or destination
- **But bias might not be easy to detect biases and switch SEs when unfairness is just for a limited number of queries, and the SE market is not as competitive as it seems**
- *We will define the relevance of a page/link as 84 the average visibility of its positions among SEs*

The rest of the paper is organized as follows. Section 2 presents our model, more precisely how we compute scores of links and SEs for given search terms based on average visibility. Section 3 introduces the statistical tests for potential bias, each of them based on Dixon’s Q test for outlier detection. Given the several SEs available, Section 4 proposes the two meta-engines aiming at reducing bias. The whole methodology is implemented on a publicly available web site to detect and reduce bias (thanks to our meta engines) as described in Section 5. Extensive tests are conducted and analyzed in Section 6, on detected bias but also to compare all the search engine outputs. Finally we conclude and describe our next research directions in Section 7.

- 
- **Section 2**
- **Relevance**

“how closely the content on the landing page matches the intent behind the search query. This includes matching terms directly on the page as well as terms used in links referring to the page. Bing also considers semantic equivalents, including synonyms or abbreviations, which may not be exact matches of the query terms but are understood to have the same meaning.”

- 
- Dixon’s Q test [36, 37] is designed to quickly identify outliers in datasets that only contain a small number of observations
- **MATH/SKIMMED**

 PA Explain to a 5th grader this: In case of equality in a given subset, a vote equal to  $r_{i,k}$  is removed for all 398 competitors  $i$ , and the algorithm repeated with this excluded value (for this 399 subset of competitors) by computing the new majority judgment values. This 400 is repeated up to a differentiation is made, which happens except if some 401 candidates have exactly the same set of rankings.

 Imagine there is a group of friends trying to decide which restaurant to eat at. They each have their own favorite restaurant and they write it down. If there are two or more restaurants with the same number of votes, they will take away one vote for each restaurant and try again. They keep doing this until there is only one restaurant with the most votes. This is done so that they can make a fair decision on where to eat.

- 
- Meta Search engine built: <https://snide.irisa.fr/search>
  - **Includes scores**
  - Note that we ran our tool on the 1000 most-searched terms according to the mentioned website, but we encountered some data collection issues for 233 of them
- **GRAPHS ARE CONFUSING?**
  - *This 577 metric allows to see which SEs tend to give similar results. In particular, 578 the distance between AOL and Yahoo is very small, which is consistent with 579 our observations of Figure 6*

- for having a low score (test “SE score”, with tested hypothesis “*the score of that SE is not an outlier*”),
- for not showing high enough the most consensual page (test “Slot of top-consensus page”, with tested hypothesis “*the visibility given by that SE to the top-consensus page is not an outlier*”),
- for displaying in their top slot a page that is not deemed relevant by the other SEs (test “Top-SE page vs others”, with tested hypothesis “*the visibility given by that SE to its top-ranked page is not an outlier (among all visibilities given by SEs for that page)*”),
- for displaying in their top slot a page much less relevant than the others (test “Top-SE page scores”, with tested hypothesis “*the score of that SE’s top-ranked page is not an outlier (among all SE top-ranked-page scores)*”).

We note that Ask is the SE performing worst in all 4 test categories, followed closely by DirectHit and Duckduckgo, with Google a bit behind.

Looking at those 4 metrics, the best-performing SEs appear to be AOL, Bing, Lilo and Yahoo. However, note that this is likely to be a consequence of those SEs returning quite similar results, as shown before in Table 3: since the tests are based on outlier detection, a biased behavior will not be detected if implemented by enough SEs. Hence again, those results do not show that AOL, Bing, Lilo and Yahoo are better than the other SEs, but rather that their behavior do not individually stand out when compared to the 12 others. [To deal with that issue of redundancy among SEs, a possibility](#)

To correlate with what we did in [32], we re-identify in Table 4 for each SE the (ordered) 10 queries with lowest relative score with respect to the consensus SE. The table seems to show less sensitive terms than what we

AllTheInternet	AOL	Ask	Bing	DirectHit	Duckduckgo	Ecosia
communao	focus	aright+move	coco	ebay+kleinanzeigen	ynet	mac
make+my+trip	paris	sears	sing	bbc+football	rambler	movistar
quot	allegro	9+gag	cars+game	american+online+inc	goggle+news	chromecast
bidmate	gossip	verizon+wireless	camera	sears	att	lazada
bee+bee	united+airlines	live+cricket+score	leso	9+gag	man+united	car+game
otto	anual	mealdeal	verizon+wireless	verizon+wireless	zimbio	zimbi
film+training	quot	rottentomatoes:captain+america+civil+war+ive+critic+score	skies	arsenal	citybank	terris
camera+360	gaps	free+mobile	happy+birthday	rotten+tomatoes	lowes	cheap+flight
mail+online	bonprix	lowes	fitbit	gamework	gamework	odnoklassnik
bookers	fifa+16	national+hockey+league	free+mobile	fanango	fanango	chatro

Google	Lilo	Lycon	Tourma	Yahoo	Vippy
focus	car+g	video+one	restauran+near+me	paris	mac
restaurant+near+me	falsehseit	anime+gong	pizza	focus	internet
chrome	otto	east	tests	allegro	cinepolis
tesla	american+online+inc	coco	enterprise	gossip	email
fitbit	emoji	mail+online	western+union	united+airlines	east
currency+converter	fitbit	cameras	traductor+d+googlemp3+music+d+downloading	eme	gospip
united+airlines	fitbit	zimbi	avacuca	quot	video+one
pizza	shatoe	happy+birthday	lego	gaps	apps
mega	movistar	fitbit	app+lock	bouprix	nissan
happy+wheels	happy+wheels	applock	cars+game		

Table 4: 10 lowest-scores terms (relative to the consensus) for each SE.

had observed in [32], suggesting either that the discrepancies among SEs are not economically motivated or that such biases are not detected since implemented by several SEs. One can nevertheless notice the search term “chrome”, in the list of the terms for which Google’s score is the lowest relative to the consensus. For that term, one can expect that Google tends to refer more to its Chrome web browser rather than to the metal plating when compared with the other SEs; that specific search term is analyzed

- It 651 seems that Google makes an extensive use of the geographical data of the 652 user, and refers to the French-speaking version significantly more often than 653 the other SEs; Google is also on the lower-end side for the visibility it gives 654 to the English-speaking version

- 6.3. Comparison of meta engines
  - Re-skim?
- Conclusion
  - The basis of our model is the interpretation of the positions in search engines ranking in terms of visibility given to webpages

## Axiom.ai

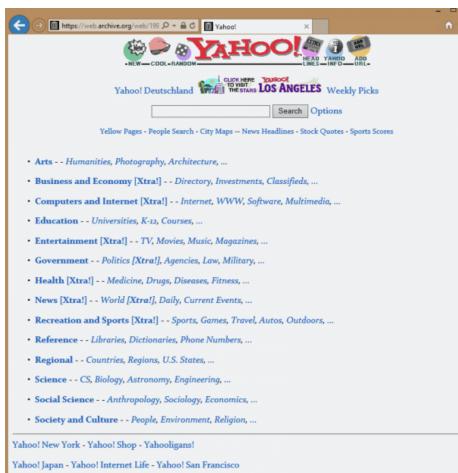
### Potential Names

- Spot-A-Shop
- Shop Around Town
- BuyNearby
- TryNBuy

## Week 2 - Readings

### Yahoo killing off Yahoo after 20 years of hierarchical organization | Ars Technica

- 2014 - retiring its namesake product
  - In the early days of the Web, these categorized, human-curated Web listings were all the rage
  - Search engines existed but bad quality
  - **When web was smaller - directories were a useful alternative way of finding sites of interest**
- *As the Web grew, directories became less useful—there was no way they could ever hope to be exhaustive—and Google, in particular, made search engines useful*



## The Yahoo Directory -- Once The Internet's Most Important Search Engine -- Is To Close

- For some, it may seem like there has always been Google, and it's always been the main way people have found things on the web. But before Google, there was the Yahoo Directory and its hand-compiled list of websites, organized into topics
- Yahoo was started nearly 20 years ago as a directory of websites that helped users explore the Internet. While we are still committed to connecting users with the information they're passionate about, our business has evolved and at the end of 2014 (December 31), we will retire the Yahoo Directory
- A "directory" relies on humans to review websites, summarize them with short descriptions and organize them into categories. When Yahoo started, this system was effective, because there weren't that many pages on the web (relatively speaking) and automated search technology to organize websites wasn't very good.
- NOT TRUE FOR LOCALIZED TRANSACTIONAL SEARCHES
  - First, Google. Google came along in 1998 and quickly proved that you could have search results generated by automation, rather than through human effort, and have better relevancy.
- **Yahoo worked because internet was smaller and automation was bad**

## 20 Web crawling and indexes

- Web crawling is the process by which we gather pages from the Web, in order to index them and support a search engine. The objective of crawling is to quickly and efficiently gather as many useful web pages as possible, together with the link structure that interconnects them
- Features it **MUST** provide
  - Robustness - overcome traps, not such traps are malicious
  - Politeness - web servers have both implicit and explicit policies on the speed of a crawler
- **SHOULD** provide
  - Distributed, Scalable, Performance/efficiency, Quality, Freshness, Extensible
- Crawling
  - *The process*

- Difficult at scale while keeping all the factors in mind
- Frontier
- IP Address, DNS serve
- A common heuristic is to insert a gap between successive fetch requests to a host that is an order of magnitude larger than the time taken for the most recent fetch from that host
- Distributing indexes
- Connectivity serves
  - 1. Similarity between lists: Many rows of the table have many entries in common. Thus, if we explicitly represent a prototype row for several similar rows, the remainder can be succinctly expressed in terms of the prototypical row.
  - 2. Locality: many links from a page go to “nearby” pages – pages on the same host, for instance. This suggests that in encoding the destination of a link, we can often use small integers and thereby save space.
  - 3. We use gap encodings in sorted lists: rather than store the destination of each link, we store the offset from the previous entry in the row.
- ***VERY CONFUSING AND TECHNICAL***

### Hierarchical Navigation: An Exploration of Yahoo! Directories

- . Several issues emerged from the students’ reporting of their search process and a comparison between hierarchical navigation and keyword searching: citation order of facets, precision vs. recall, and other factors influencing searchers’ successes and preferences
- Yahoo began hierarchical browning structure exclusively → *Why do successful enterprises like Yahoo! and Google invest in something akin to the classified catalogs that American libraries abandoned over a century ago*
- Hierarchical navigation may use traditional classifications or it may use classification-like schemes such as Yahoo! directories or the structure of files within folders within folders on one’s computer
- A classificatory structure also facilitates navigation up and down a hierarchy
  - Do searchers consider classificatory structures a viable alternative or supplemental approach to information retrieval?
- **Exploratory study**
  - *Research emphasis on natural language has drawn attention away from controlled vocabularies, including classification*
- Previous Research on the topic

- Nancy Williamson (1997, 24) noted that: "classification aids in retrieval through the use of logical structure and helpful order to facilitate browsing and filtering of large quantities of data
- **hierarchical browsing "improves precision by first defining and narrowing the domain for searching" It does so by collocating (inclusion) and partitioning (exclusion)**
- Research Method
  - Convenience sample
  - Testing ability to use classification
  - Started with "Science"
- Results
  - Hierarchical Paths Taken
    - "Science" to "Biology" to "Microbiology" to "Organizations" (Path A)
    - once they successfully completed the task required in the assignment there was no reason for them to explore other options

Path Taken	Number of Students
A. Science→Biology→Microbiology→Organizations	21
B. Science→Biology→Microbiology→Web Directories	1
C. Science→Biology→Parasitology→Organizations	1
D. Science→Biology→Organizations	1
E. Science→Biology→Organizations→Microbiology	3
F. Science→Biology→Organizations→Professional	1
G. Science→Medicine	1
H. Science→Medicine→Microbiology and virology→Organizations	1
I. Science→Organizations→Biology→Microbiology	3
J. Science→Research	1
■ Not indicated	1

- Three students believed that this type of search was useful for people who were able to follow easily the hierarchical path because of their familiarity with the general area of the search topic

Advantage	# of Students	Disadvantage	# of Students
Quick, easy, and not time-consuming	13	Requires more steps	3
High precision	11	Low recall	11
Showing relationships between topics	8	Limited to the structure developed by Yahoo	3
Useful for experts or someone familiar with the topic	3	Requires knowledge of a particular subject hierarchy	8
No need to generate search terms	2	Subjectivity in categories and assignment of sites to categories	1
No need to know advanced search techniques	1	Requires knowledge of hierarchical search	4
		Updated slowly	1

- 

Table 5. Advantages and disadvantages of hierarchical navigation.

- **common complaint about keyword searching given by those students (10) was the overwhelming number of sites retrieved, a large percentage of which were irrelevant (i.e., low precision rate). Six students pointed out that users needed to come up with useful keywords on their own and not many users would be able to do so**
- Discussion
  - the hierarchical structure itself, including issues of specificity and citation order; the influence of these two approaches on precision and recall; time consumed in retrieval; and knowledge and skill required for retrieval
  - When the two search methods were compared, navigating through Yahoo! directories yielded high precision but low recall and keyword searching resulted in low precision but high recall.
  - Time
    - In other words, hierarchical navigation took more steps and was more time consuming in the searching stage; but keyword searching required more time spent on evaluating the appropriateness of the retrieved sites individually, many of which did not meet the search criteria. In this way, the precision and recall of search results were clearly linked to time taken
  - The hierarchy follows the conventional notion common in bibliographic classifications of classification by discipline with form as the last facet
- Conclusion
  - advantage of showing relationships between topics → relationships more visible
  - The second implication for classification is that several students referred to the structure of the hierarchy as a limitation, requiring knowledge of the particular subject hierarchy of hierarchical searching and limiting what might be searched
  - **"It points to hierarchical navigation as a viable alternative to the ubiquitous keyword searching."**

## PROPOSED REMEDIES FOR SEARCH BIAS:

- 2012
- "Search bias" → FTC going after google, article is proposed remedies

- Microsoft, Expedia, Foundem
- This article concludes that the cures proposed by the competitors are worse than
- Antitrust Law → Promote consumer welfare, competition and innovation
- 5 Categories
  - Search “Neutrality”
  - Ten Blue Links, not Universal Search
  - Google-Specific Fair-Use Limit
  - No Bidding
  - Continuous Disclosure
- Background
  - held a hearing on Google that focused almost exclusively on one topic: whether Google's search engine preferences other Google products or deliberately harms Google's competitors
  - Does it have monopoly power - and does it abuse that power to exclude rivals on a basis other than efficiency or innovation
  - Before 2007, Google displayed results generally as 10 links to Web pages on each results page. In 2007, Google introduced Universal Search, which displays results not only from Web sites, but also, from images, videos, news, → THIS favors Google's specialized search properties (Maps, Places, Products)
    - Leading competitors use this style too
    - Chomp for App Store Search
  - Prediction of future: they will be mobile and increasingly reliant on artificial intelligence processing of concepts, not keywords
  - ***According to Google, its users understandably go to Google for answers, not for links to sites consisting of additional links to answers***
- Remedies in Antitrust Law
  - Terminate a defendant's unlawful conduct, to prevent that conduct's recurrence, and to reestablish the opportunity for competition → generally neglected
  - Consumer welfare (NOT competitor welfare) → A remedy should not slow down the most efficient and innovative leaders merely to help weaker competitors
  - Technological innovation and dynamic efficiency

- Don't want to interfere with an industry - nonintervention unless the court is confident that it has identified anti-competitive conduct and can apply an effective remedy
- *such as a price squeeze remedy, would require constant monitoring and the matching of two moving targets in different markets: the quality of Web sites ranked in search results and the consumer preferences for their relevance. For these reasons, the proposed remedies are flawed and should not be considered*
- Remedies Proposed
  - "Neutral" display of search results
    - The FairSearch White Paper proposes 'prohibiting Google from reducing the ranking of any site because it competes with a Google Site'
    - "Technical committee" to implement
    - → **Hard to define "neutrality"**
      - **Search results right now are defined as opinions**
    - Flight comparison → *CRS displays results for a specific class of queries (flights) based on a few simple criteria, such as listed price, stops, destination, date, and time. On the other hand, search engines must respond to far more complex (and often incomplete or suggestive) queries*
  - Technical Committee Proposal
    - Didn't work in Microsoft case
  - Ten Blue Links - Forbidding universal Search
    - Browser swapping - Giving users an option to choose default map service, default news, etc
    - Features vs products
    - Google not a pure play at anything like its competitors so shouldn't limit it
  - Transparency
    - Users may or may not consider Google News to be an entirely separate product, but there is no evidence of consumer confusion regarding which company (Google) has organized news results into a news box on Google's search page
  - Foundem complains that its Quality Score was very low, but Google is at least transparent about why - Google values the

Quality Scores “feature original content that can’t be found on another site”

- ***GOOGLE IS DECIDING WHAT TO VALUE***

- CONTENT RULES: FAIR USE AND SHARING
  - Snippets, copyright/fair-use claims
  - Saying that also competitors should be held to the same standards as Google if Google is to be limited
- NO BIDDING: ACQUISITIONS AND EXCLUSIVE AGREEMENTS
  - ***Need to see a benefit to consumers***

- Conclusion

- Argues that the proposed remedies cause more harm than good for consumers/competition/high-tech innovations

complaints are not only irremediable, but also, lack merit under US antitrust laws. As a writer at the nation’s top technology blog wrote:

Even if Google is abusing its monopoly powers, then what? Any remedies imposed on Google could be worse for consumers than the uncertain consequences of keeping Google unchecked. If Google had to pass every change to its search engine through an antitrust filter that could really screw up search, something which most of us depend on every single day.<sup>182</sup>

In short, the FTC has better uses for its scarce resources than pursuing an inquiry or investigation that would merely raise the cost of one rival, Google, for the benefit of other rivals that have proposed unworkable and detrimental remedies for their novel theories of antitrust harm that is effected through “search bias.”

- 

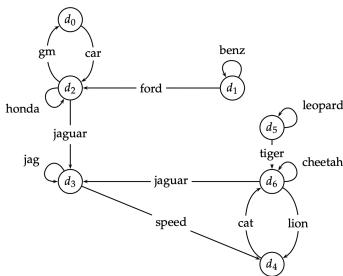
## IDEA FOR SEARCH TERM:

- Beret
- Track Spikes

# Week 3/Bonus - Readings

## Manning IR textbook Ch. 21 'Link Analysis'

- We begin by reviewing some basics of the Web as a graph in Section 21.1, then proceed to the technical development of the elements of link analysis for ranking
- Link Spam
- Web Assumptions:
  - 1) The anchor text pointing to page B is a good description of page B
  - 2) The hyperlink from A to B represents an endorsement of page B, by the creator of page A.
- Page Rank
  - . Our first technique for link analysis assigns to every node in the web graph a numerical score between 0 and 1, known as its PageRank. The PageRank of a node will depend on the link structure of the web graph
  - Random walk → . The idea behind PageRank is that pages visited more often in this walk are more important
- A Markov chain is a discrete-time stochastic process: a process that occurs in a series of time-steps in each of which a random choice is made. A Markov chain consists of N states. Each web page will correspond to a state in the Markov chain we will formulate.
  - MATH FORMULAS



- ► Figure 21.4 A small web graph. Arcs are annotated with the word that occurs in the anchor text of the corresponding link.
- Topic Specific Page Rank
  - We now consider teleporting to a random web page chosen nonuniformly. In doing so, we are able to derive PageRank values tailored to particular interests
  - <https://dmoz-odp.org/docs/en/about.html>

- Hubs and Authorities
  - There are authoritative sources of information on the topic; We will call such pages authorities; in the computation we are about to describe, they are the pages that will emerge with high authority scores
  - On the other hand, there are many pages on the Web that are hand-compiled lists of links to authoritative web pages on a specific topic. These hub pages are not in themselves authoritative sources of topic-specific information, but rather compilations that someone with an interest in the topic has spent time putting together
  - A good hub page is one that points to many good authorities; a good authority page is one that is pointed to by many good hub pages

1. Given a query (say leukemia), use a text index to get all pages containing leukemia. Call this the *root set* of pages.
2. Build the *base set* of pages, to include the root set as well as any page that either links to a page in the root set, or is linked to by a page in the root set.

We then use the base set for computing hub and authority scores. The base set is constructed in this manner for three reasons:

1. A good authority page may not contain the query text (such as computer hardware).
2. If the text query manages to capture a good hub page  $v_h$  in the root set, then the inclusion of all pages linked to by any page in the root set will capture all the good authorities linked to by  $v_h$  in the base set.
3. Conversely, if the text query manages to capture a good authority page  $v_a$  in the root set, then the inclusion of pages which point to  $v_a$  will bring other good hubs into the base set. In other words, the “expansion” of the root set into the base set enriches the common pool of good hubs and authorities.

○

## Google targets low-income US women with ads for anti-abortion pregnancy centers, study shows

- Low-income women in some cities are more likely than their wealthier counterparts to be targeted by Google ads promoting anti-abortion crisis pregnancy centers when they search for abortion care

- *The research builds on previous findings detailing how Google directs users searching for abortion services to so-called crisis centers – organizations that have been known to pose as abortion clinics in an attempt to steer women away from accessing abortion care*
- “By pointing low-income women to [crisis pregnancy centers] more frequently than higher-income women in states with restrictive laws, Google may delay these women from finding an actual abortion clinic to get a legal and safe abortion,” says Katie Paul, the director of the Tech Transparency Project.
- The results were not the same in all cities. In Miami, researchers saw the inverse result: high-income women were more likely to get ads from crisis centers than lower-income women
- “Google has a large share of influence, particularly in the United States when people are trying to search for authoritative information. And people generally tend to consider Google’s search engine as an equaliser. They think the results they get are the results that everyone’s going to get. But that’s just not the case,” Paul says.

## Google Location Data

- When you search on Google, like with Maps, Search or Google Assistant, your current location is used to give you more helpful results
- Your location helps to show you nearby results, even if you didn't include a location in your search
- If you're searching for something nearby and not finding local search results, try these fixes:
  - Add your current location to your search, like coffee shops in Chelsea
  - Check that your device sends location to Google when you search. Follow the steps below to manage your device location settings
- **In some cases, areas that you have searched for in the past may be used to estimate your current location. For example, if you search for coffee shops in Chelsea and then nail salon, Google might show nail salons in Chelsea.**

## Consumer Eye Movement Patterns on Yellow Pages Advertising

- *Process tracing data help researchers understand how yellow pages advertisement characteristics influence consumer information processing behavior. In a laboratory experiment eye movement data were collected while consumers chose businesses from telephone directories. Consumers scan listings in alphabetic order. Their scan is not exhaustive. As a result, some ads are never seen. Consumers noticed more than 93% of the quarter-page display ads but only 26% of the plain listings. They perceived color ads before ads without color, noticed more color ads than noncolor ads, and viewed color ads 21% longer than equivalent ads without color. Also, they viewed 42% more bold listings than plain listings. Consumers spent 54% more time viewing ads for businesses they ended up choosing, which demonstrates the importance of attention for subsequent choice behavior.*

# Meeting 2/6

## READINGS

- Sample Yellow Pages
  - Page 71 - [2015 Central Community Directory & Yellow Pages](#)
  - [Pennsylvania - White Pages - Philadelphia - August 1975 | Library of Congress](#)
- Literature on Yellow Pages (via [Google Scholar](#))
  - <https://ieeexplore.ieee.org/abstract/document/1401768>
  - [A yellow-pages service for a local-area network | Proceedings of the ACM workshop on Frontiers in computer communications technology](#)
- Business Permit Search
  - [OEO Registry of Certified Minority/Women/Disable Owned Business Enterprises](#)
  - [Local Businesses](#)
  - [L&I Commercial Activity Licenses \(CSV\) - OpenDataPhilly](#)
  - [\*OpenDataPhilly\*](#)
    - [liquor license directory philadelphia/permit](#)
    - [Register a business | Services | City of Philadelphia](#)
    - [LMS Online - Business License Search By Business Name](#)
    - [Search for property history and business license information | Services | City of Philadelphia](#)

# Week 4 - SERP Assignment

## SERP Assignment (10%)

Collect and analyze the search engine results pages from a set of products from a range of search engines. Write a report summarizing findings.

### INCOGNITO

#### Google

- Shopping
- Maps
- Search

#### DuckDuckGo

#### Bing

#### Yahoo

#### Qwant

#### Neeva

#### StartPage

#### The Real Yellow Pages

#### Valentin.app

#### Brave Search

#### Alternative Engines (cut):

- Ask.com
- Mojeek

1) Running shoes → Philadelphia Runner

2) Book binding and printing → Campus Copy Center

3) tequila store → Fine Wine and Spirits

4) Comic book store → Brave New World

5) See local business referenced:

<https://www.reddit.com/r/philadelphia/wiki/shopping/>

6) See Business in

<https://www.opendataphilly.org/dataset/city-registered-local-businesses/resource/b2e64fb3-36c6-4df6-8286-f030a7054bdf>

7) Apple Computer → Penn Bookstore



Here are some popular and easy-to-use web scraping tools:



1. Beautiful Soup: Beautiful Soup is a Python library that makes it easy to scrape information from web pages. It's commonly used with the Requests library to extract data from HTML and XML files.
2. Scrapy: Scrapy is another popular Python web scraping framework that allows you to write spiders to crawl



continue



Sure, here are some more popular web scraping tools that allow you to easily collect online information:



7. BeautifulSoup: BeautifulSoup is a popular Python library that makes it easy to parse HTML and XML data. With BeautifulSoup, you can extract data from a website and extract specific elements, such as text or links.
8. Scrapy: Scrapy is another popular Python library that is specifically designed for web scraping. It offers a high-level API for making requests to websites, parsing the response, and extracting the data you need
9. Octoparse: Octoparse is a web Regenerate response used without any coding

## Meeting 2/13

How to best utilize these:

- Business Permit Search
  - [OEO Registry of Certified Minority/Women/Disable Owned Business Enterprises](#)
  - [Local Businesses](#)
  - [L&I Commercial Activity Licenses \(CSV\) - OpenDataPhilly](#)

+ registered\_local\_businesses (1)

+ oeo\_registry (1)

- [OpenDataPhilly](#)
  - *liquor license directory philadelphia/permit*
  - [Register a business | Services | City of Philadelphia](#)

- [LMS Online - Business License Search By Business Name](#)
- [Search for property history and business license information | Services | City of Philadelphia](#)
- These can come into play later

## Week 5

Finish SERP

- Check mobile SERPs

Coding

- Commjhuh.asc.upenn.edu
- Association rule learning - Wikipedia
- Python Based scraping:
  - Python-based: requests + BeautifulSoup, Selenium and Playwright

## Old Meeting

SERP Notes:

- Dramatically different results, Yellow pages the worst
- Freaks out for product vs. brand vs store type
  - Need ChatGPT to code products into what type of store would carry them
  - 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
- Scrape the entirety of the engines

SampleData Notes:

-  SampleData
- ChatGPT result changes each time
- Code for replacechars
- Ranking for a sorted dataframe?
  - Is there a way to sort a list of dictionaries - I'm just converting between list of dictionaries and dataframe
- Image display not working

## Reimagining product search

- More filters (?)
- Start with a brand; color/size

Within the product sector of search - Directory

[APIs | eBay Developers Program](#)

## Flow Chart

Product

- Browse (Information)
- Specific (Transactional)

Store

- Locator (Navigational)

All the bike stores

All the specialized bike stores

Stump Jumper

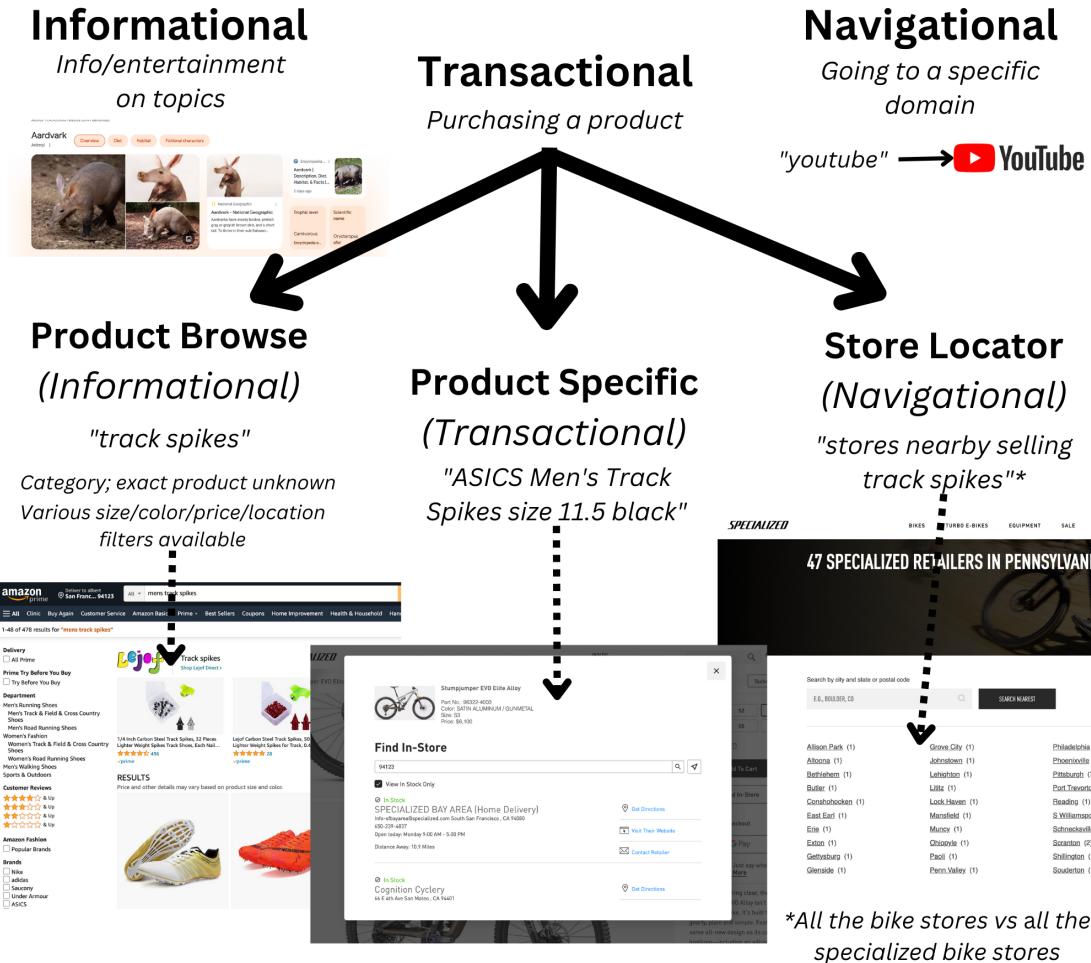
- <https://www.specialized.com/us/en/store-finder?q=94123&country=us&qp=94123&l=en>

(Pivot) Meeting Notes 2/22

# Made infographic

- Don't wanna just rerank Google, but I do want to provide users with a more unbiased and comprehensive search experience centered on a localized directory of businesses and products.

# Search Engine Theory



- Why? Store Locator original path, but seeing Yahoo-Style Name Tag bias / just reranking Google (can come back to this idea down the road) - made me think to start with a store locator for a specific product as the next step:
    - [Store Finder | Specialized.com](#)
  - Focusing on **Product Specific** as the next step

- <https://pabstblueribbon.com/pbr-finder/?gLocation=19104%2C%20Philadelphia%2C%20PA&qBrandNM=Pabst%20Blue%20Ribbon>
- [Stumpjumper EVO Elite Alloy | Specialized.com](http://Stumpjumper%20EVO%20Elite%20Alloy%20-%20Specialized.com)

Tangent:

*code for scraping a website database, request over and over to a database to get another*

- [Ulla Johnson Store Locator](#)
  - Loop Scrape <https://stockist.co/Shopify.dev>

THUS - NEED HELP WITH SAMPLE CODE FOR THIS ("Loop scraping"?)

#### MINI-ASSIGNMENT/TASK:

Develop \*universal/\* code to Scrape all these sites that utilize - [Stockist](#)

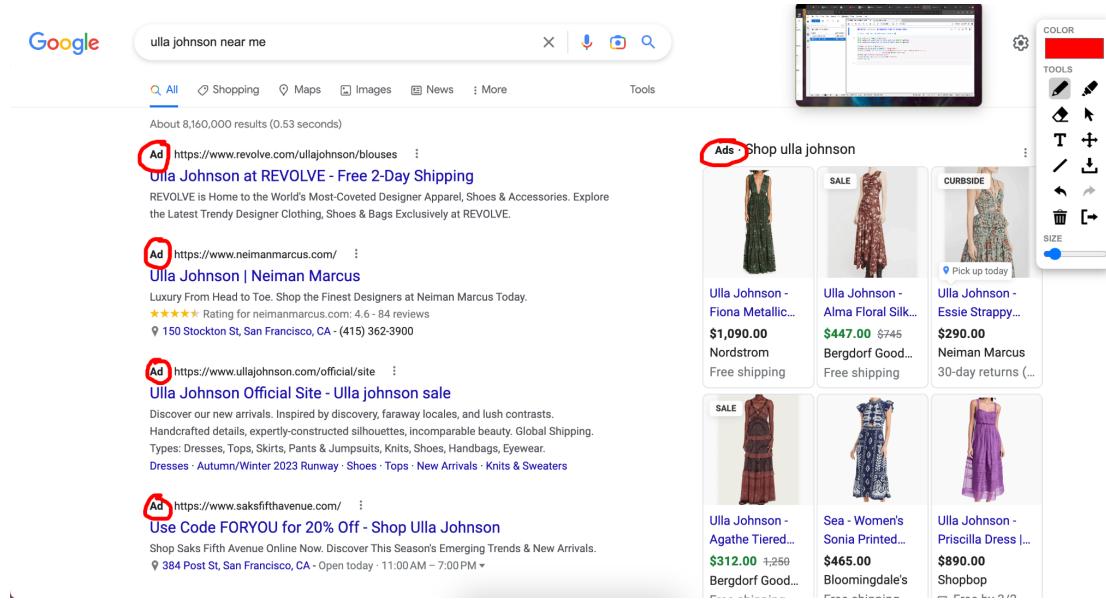
- <https://obeyclothing.com/pages/stockists>
- <https://www.hurley.com/pages/store-locator>
- <https://www.veda.com/locations>
- [https://www.vitamix.com/us/en\\_us/where-to-buy](https://www.vitamix.com/us/en_us/where-to-buy)
- <https://ullajohnson.com/pages/store-locator>
- <https://www.jlaudio.com/pages/dealer-locator>
- <https://www.markcross.com/pages/location>
- <https://www.rocksicare.com/pages/store-locator>
- <https://www.dewalt.com/find-retailer>
- <https://stetson.com/pages/store-locator>
- <https://www.morphe.com/pages/store-locator>
- <https://lespecs.com/pages/stockists>
- <https://pabstblueribbon.com/pbr-finder/>
- <https://locator.casio.com/en-US/store/wat/map/G-SHOCK>

## 2/27 Meeting

### Stockist & Grappos!

- Stockist
  - [Ulla Johnson Store Locator](#)
  - [Stockists](#)
  - [Store Locator | Morphe](#)
  - [Demo - Stockist Store Locator App](#)
  - [Store Locator | Hurley](#)
  - [JL Audio Authorized Dealer / Distributor Locator](#)
  - [Location – Mark Cross](#)
  - [Store Locator – RoC® Skincare](#)
  - [Store Locator | Stetson](#)
  - [Stockists – Le Specs](#)
  - → Find more: <https://stockist.co/>
  -
- Grappos
  - [PBR Finder - Pabst Blue Ribbon](#)
- Other?
  - [Find Locally - Find a Retailer or Distributor Near You | Vitamix](#)
  - [Find Aveda Stores & Hours | Store & Salon Locator](#)
  - [Find a Retailer | DEWALT](#)
  - [CASIO G-SHOCK Store Locator - USA](#)
- <https://destini.co/welcome/>
  - Peet's

[License Lookup | Alcoholic Beverage Control](#)



## SPA - Single Page Application

- Update the page without the URL changing
- Need a different way to get at scraping

Selenium → Headless mode (so you don't see what's happening live)

- ZenRows tutorial?

Bring in idea of metasearch

google displays results a tiny fraction of what is available

- [Why does Google only return 50 pages of 10 results when it claims that there are 560,000 results? - Quora](#)
- <https://blog.xkcd.com/2011/02/04/trochee-chart/>
- Search engines do not index sites equally, may not index new pages for months, and no engine indexes more than about 16% of the Web. As the Web becomes a major communications medium, the data on it must be made more accessible.

# Spring Break Reading

## Death of the Yellow Page Directories

- As the telephone served as the only technology that can be used to contact anyone in the world in real-time, the telephone number listing and organization of all the services in the world were only accessible through the yellow pages
- **If you had a business, you just had to list yourself in the yellow pages, it was the single and most basic and important way to advertise yourself. If you missed out in being printed in the annual yellow page directory, it was as if you didn't exist for the entire year.**
- *R.H. Donnelley focused on buying more yellow page directories, expanding physically, but neglected investing in its digital presence or creating new offerings before collapsing.*
  - *Simply put, the death of R.H. Donnelley and all the other yellow page directories is due to one simple thing: the internet and not adequately adapting to it quick enough.*

## The Story

→ **Thesis:** Google is not trying to sell products; it's trying to sell the big-budget advertising brands' stuff → we need a clean, clear, and complete way to find things

- 1) Proving the problem exists and a stockist solution with statistical analysis/Z-scores

### a) **Store location & Brand**

- b) Aim: Qualtiviatley shows more stores and (assumption) more accurate, *cleaning up what is out there in a clear good way*
- c) Example: PBR
- d) Assumptions: Assuming they are updated recently, assume all companies are on stockist, *Brand may not show product sku (stock keeping unit) vs UPC*
- e) How2:
  - i) **Scraping Stockist, Grappos, Other** → Companies with store locators, City Liquor license,

- 2) Color & Size

### a) **Store location & Item**

- b) Example: PBR: Bottles vs cans for pool party or *Girl: sees all the stores that carry it - but I have a specific item in mind I need in my size - this isn't helpful!*
- c) HOW2:
  - i) Cooperative brands - Email a couple of sample cooperative brands to get their entire inventory list
  - ii) Crowdsourcing Stores - Store collectives to send/share their Shopify
  - iii) Consumer Crowdsourcing - Amazon-style app for people to scan and upload UPC

- 3) Exact color/size - Point of Sale

### a) **Store location & item in stock**

- b) HOW2:
  - i) Microsoft/Shoptify/Visa/... gives out a free POS inventory system that connects to a large cloud base (retailer CAN NOT opt-out) that makes auction style portal to show all available (in-stock) locations across the US of that exact color and size item
  - ii) **Template: Amazon "Other Sellers on Amazon" + Portal example as an idea + Amazon API?**

1. Get a template using (\_\_\_\_ site) for how to scrape a Shoptify Site → with a focus on finding 'barcodes'
  - a. Product name/color/size by number not words -
    - i. [LoveShackFancy Moe Skirt in Pre-Dawn Sky – SINGER22](#)
    - ii. <https://www.barcodelookup.com/840098856185>
  - b. <https://devs.upcitemdb.com/>

→ Hack into the inventory system through API searching UPC

→[\*Retail Barcode Label App Api Access - Shopify Community\*](#)

# First Meeting back from Break

## Selenium Tutorials

- [Selenium with Python Tutorial: How to run Automated Tests | BrowserStack](#)
- [Selenium Python Tutorial - GeeksforGeeks](#)
- [2. Getting Started — Selenium Python Bindings 2 documentation](#)

## Steps in terminal

- 1) cd seleniumTesting
- 2) source venv/bin/activate
- 3) jupyter lab

# Ulla Scraped (3/21)

→ [Zip Codes by Population 2023](#)

- 1) Can we do all 42,000 zip codes?
  - a) **No need for for-loop** → **parallelize looping**
  - b) **Stick to ~500 for now**
  - c) **Have them compare**
- 2) How to scrape Peet's site?
  - a) **Classes within classes vs tags**
  - b) **find\_element(By.CLASS\_NAME,**
    - i) **Classes are usually more unique** → **want something unique to identify the content**
    - ii) **Class names come from CSS**
- 3) How to scrape a Shopify site?
  - a) Example: [LoveShackFancy](#)
  - b) Scrape Shopify sites for "barcodes" -  
<https://www.barcodelookup.com/840098856185>
  - c) [Shopify API reference docs](#) → Does Shopify have an API that lets me access other stores' barcode/internal information
  - d) **Go to the enclosing element (the one above), then go for child elements**

**e) *find.element* (give object not in a list) vs *find.elements* (list of all the matching objects)**

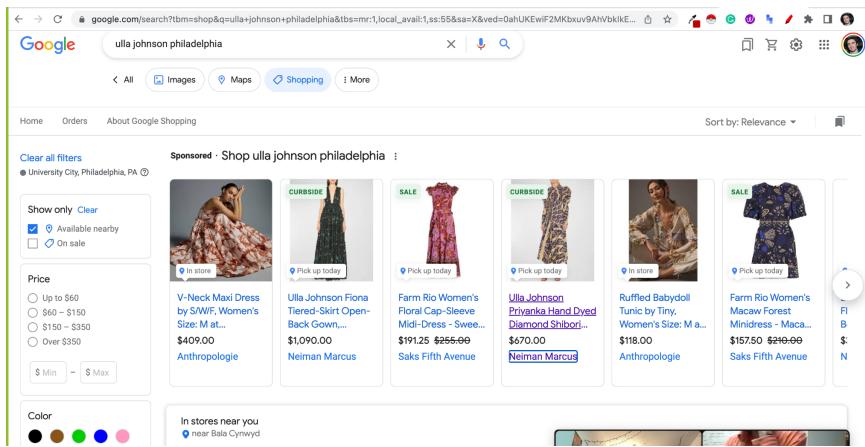
**f) Same selenium structure, find the classes/tags and call them into a dictionary → list of dictionaries to then use as the dataset**

4) Python Web App:

<https://realpython.com/python-web-applications/#build-a-basic-python-web-application>

**a) Flash for when we get there**

**b) Would take code out of notebook**



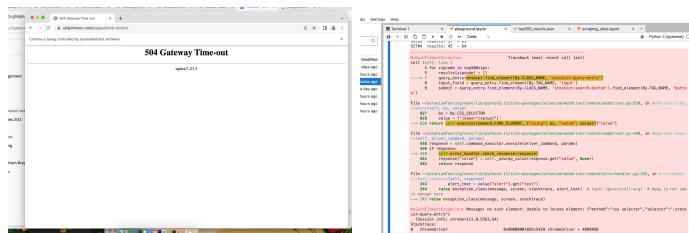
→ Read tutorials

→ Scrape all stockist sites

- <https://www.peets.com/pages/product-locator>
- Le Specs

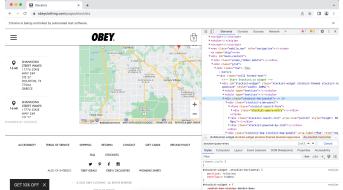
## Solo Work (3/22)

- 1) Trying to scrape all 500 zip codes
  - a) Getting time out & breaking the site? → Fixed with a "try & except"



i)

- 2) Scraping all other stockist sites

- a) 
  - b) Issue with some sites, adding to the dictionary freaked out the system, switched to a mega list
  - c) See 'notes' in 'scraping\_sites'
- 3) Scraping non-Stockist sites
    - a) **Need help locating elements as a list, see 3/22 email;** still need to experiment with the following sites:
      - i) <https://www.aveda.com/locations>
      - ii) <https://www.dewalt.com/find-retailer>
      - iii) <https://locator.casio.com/en-US/store/wat/map/G-SHOCK>
      - iv) <https://www.peets.com/pages/product-locator>
      - v) <https://pabstblueribbon.com/pbr-finder/>
      - vi) Vitamix
        - (1) Find all for the retailer**
        - (2) Use find\_elements for a list, find\_element for object → see comments in notebooks → want to get big object and then dive in**

### 4) Scraping Google + Google Shopping for comparison analysis

- a) Option 1: APIs? Probably not good for free plan
  - b) Option 2: Selenium → 'yuRUbF' as the class
- VitaMix Works → try with the final other ones
  - Selenium scrolling & XPath

## Google Scraping (3/25)

- 1) Vitamix not finding the submit button?
- 2) Google detecting scraping, giving reCaptcha → trying hiding it by going to ESPN.com/CNN.com → have to quit the browser, takes longer but works.
- 3) Import cartopy/geopandas to localhost or another graphing package for zipcodes**
  - a) ChatGPT → <https://nominatim.org/>

websiteType brand results zip

### To-Do

1. Statistics - ONE PAGER
2. Custom Code for ALL SITES
  - a. <https://locations.chipotle.com/>
  - b. <https://www.williams-sonoma.com/stores/>
  - c. <https://valentin.app/>
  - d. SEE SYLLABUS FOR X&Y
3. [Valetin.app](#)

### From the syllabus

I'm working on a start-up that is going to reinvent localized transactional search and make Google worthless.

I'm creating a "one-page" Jupyter notebook with python code and markdown cells to demonstrate the concept of how Google is biased and actually hiding stores in people's areas that have the brands they are searching for. My sample is with Ulla Johnson women's clothing and three zipcodes.

I was able to successfully scrape the Ulla Johnson store locator (giving 17 total "true" results among the three zipcodes) and both Google Search and Google Shopping (giving 68 and 81 results respectively, somewhat serving as our sample space). I have all the results in three data frames

I now need you to come up with a statistical proof via python code and explainer text that can "prove" or show how Google is biased. Would it be some sort of t-test using a sample space of 149 and true answers of 17? What statistical models can I use?

**So I have a sample, and I want to test to see how likely it is to belong in a population. List potential statistical tests/models that are possible in python that can output that and give examples with code**  
 Give me an example comparing results from a search engine to see if the sample search engines results are a part of the original search engine

<https://stats.stackexchange.com/questions/83163/statistical-test-to-tell-whether-two-samples-are-pulled-from-the-same-population>

## March 27th Meeting

### Coding Notes/General Notes

*Coding:*

- !pip install → for importing in the notebook
- pip install → terminal
- Flask → Terminal & app.py, function-based

*Overall:*

- Store list is the goal (like Yellow Pages)
- Ex: "From the store locator, these should be stores that come up, but people aren't not going to the locator, they're going to google, and they're not showing."

*Ranking & Recall:*

- Proving (set size, statistically?)
- **Precision & Recall** metrics
  - To put it in simple terms, precision is about how accurate the search engine or algorithm is at returning relevant results, while recall is about how complete the search engine or algorithm is at finding all the relevant results that exist
  - Precision refers to the proportion of relevant results returned by a search engine or web scraping algorithm among all the results returned. In other words, precision measures how accurate the search engine or algorithm is at returning results that match the user's query or desired criteria. A high precision means that most of the results returned are relevant to the user's query or criteria
  - On the other hand, recall refers to the proportion of relevant results that are returned by a search engine or web scraping algorithm among all the relevant results that exist. In other words, recall measures how complete the search engine or algorithm is at finding all the relevant results that match the user's query or criteria. A high recall means that most of the

relevant results are found and returned by the search engine or algorithm.

- For example, let's say a user is searching for "red shoes" on an e-commerce website. A search engine with high precision would return mostly red shoes that match the user's query, while a search engine with high recall would find all the red shoes available on the website, even if they may not all perfectly match the user's query. Similarly, in web scraping, a high-precision algorithm would only return the data that meets specific criteria, while a high-recall algorithm would return all data that meets the criteria, even if there is some irrelevant data mixed in.

## Meta Search Engine

- 1) Scrape other engines
- 2) Check BingGPT

Hello, I'm here to find information for you. Start by asking me a question.

Where is the best place to buy ulla johnson in philadelphia

✓ Searching for: best place to buy ulla johnson in philadelphia  
 ✓ Generating answers for you...

I couldn't find a specific store that sells Ulla Johnson in Philadelphia. However, you can buy Ulla Johnson clothing online from stores like [Nordstrom](#)<sup>1</sup> and [FARFETCH](#)<sup>2</sup>. Is there anything else you would like to know?

Learn more: 1. [nordstrom.com](#) 2. [farfetch.com](#) 3. [bing.com](#) 4. [ullajohnson.com](#)

1 of 20



a)

Microsoft Bing

Where can I buy ulla johnson in Philadelphia 19104

SEARCH CHAT SHOPPING IMAGES VIDEOS MAPS MORE

About 42,100,000 results Any time Open links in new tab

**Ulla Johnson Tops on Sale | Shop Ulla Johnson Tops**  
<https://www.shopstyle.com/ulla-johnson/tops>

Ad Explore New Arrivals of Ulla Johnson. Browse the World's Largest Catalog. New Members to ShopStyle Get \$25 with First Purchase. Join & Shop Members-Only Offers! shopstyle.com has been visited by 100K+ users in the past month

**Ulla Johnson Women's Tops**  
 Longsleeve, Shortsleeve & Tanks.  
 Shop Latest Ulla Johnson Styles.

**Women's Clothing**  
 Dresses, Tops, Sweaters & More.  
 Shop Latest Ulla Johnson Styles.

**Women's Shoes**  
 Flats, Boots, Sandals & More.  
 Shop the Latest from Ulla Johnson.

**Men's Spring Trend Edit**  
 Shop Curated Spring Trends for Men to Invest in for the New Season.

**Ulla Johnson Handbags**  
 Shop the Latest from Ulla Johnson Shoulder Bags, Totes & Satchels.

b)

Microsoft Bing

Where can I buy ulla johnson in Philadelphia 19104

SEARCH CHAT SHOPPING IMAGES VIDEOS MAPS MORE

About 42,100,000 results Any time Open links in new tab

**Ulla Johnson Tops on Sale | Shop Ulla Johnson Tops**  
<https://www.shopstyle.com/ulla-johnson/tops>

Ad Explore New Arrivals of Ulla Johnson. Browse the World's Largest Catalog. New Members to ShopStyle Get \$25 with First Purchase. Join & Shop Members-Only Offers! shopstyle.com has been visited by 100K+ users in the past month

**Ulla Johnson Women's Tops**  
 Longsleeve, Shortsleeve & Tanks.  
 Shop Latest Ulla Johnson Styles.

**Women's Clothing**  
 Dresses, Tops, Sweaters & More.  
 Shop Latest Ulla Johnson Styles.

**Women's Shoes**  
 Flats, Boots, Sandals & More.  
 Shop the Latest from Ulla Johnson.

**Men's Spring Trend Edit**  
 Shop Curated Spring Trends for Men to Invest in for the New Season.

**Ulla Johnson Handbags**  
 Shop the Latest from Ulla Johnson Shoulder Bags, Totes & Satchels.

**8 Dress Trends for Spring**  
 Refresh Your Spring Wardrobe with Boxy Minis, Sheer Dresses & More.

c)

## \*\*\*\*\*Other Sites to Scrape & Product Location

- 1) Scrape [Product Locator | Peet's Coffee](#)
- 2) Research [ProductSearch.App](#)

<https://kinsta.com/blog/alternative-search-engines/>

- 3) <https://www.veda.com/locations>
- 4) <https://www.dewalt.com/find-retailer>
- 5) <https://locator.casio.com/en-US/store/wat/map/G-SHOCK>
- 6) <https://www.peets.com/pages/product-locator>
- 7) <https://pabstblueribbon.com/pbr-finder/>
- 8) <https://www.pepsico.com/our-brands/for-shoppers/product-locator>
- 9) Vitamix
- 10) <https://locations.chipotle.com/>
- 11) <https://www.williams-sonoma.com/stores/>
- 12) [Find a Whole Foods Market Store Near You](#)

## Flask

- 1) Have ChatGPT experiment with [Flask](#)

# April 3rd Meeting

## Meta Search Results

- 1) DIFFICULT ENGINES - Couldn't scrape **Google Maps, Qwant, Neeva, Valentin**
- 2) MORE PAGES - Issue loading more pages (see Yahoo)

## BingGPT

### 1) See PDF's

- a) *There are several ways to find products available nearby you. One way is to use Google Maps1. You can search for a product you're interested in, and see a list of nearby stores selling it in-store2. Another way is to use Bing shopping2. While you search for products in the Shopping tab, you can simply apply the "Available Nearby" filter to check the products you can buy near you.*
- b) *I hope this helps! Let me know if there's anything else I can help with.*
- c) *Learn more:*
- d) *1. local.google.com*
- e) *2. blogs.bing.com*
- f) *3. producthunt.com*
- g) *4. mywifequitherjob.com*
- h) *5. ecwid.com*
- i) *6. yelp.com*
- j) *7. shopify.com*
- k) *8. lifehacker.com*
- l) *9. Hebobsessed.com*

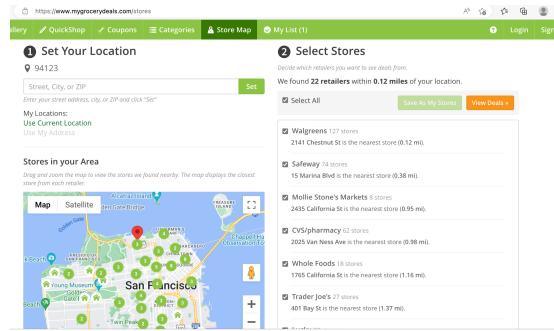
### 2) USEFUL FINDS:

#### a) Product Finder:

- i) [The Locally Product Locator can be used by both brands and retailer to surface local inventory. — Locally](#)

#### b) Brand Locators:

- i) [44 Awesome Store Locator Examples For Design Inspiration \(storemapper.com\)](#)
- ii) [Walgreens Products](#)
- iii) [Product Locator \(pepsico.com\)](#)
- iv) [Find a Whole Foods Market Store Near You](#)
- v) [My Grocery deals map:](#)



vi)

- c) [Bing Shopping Benefits](#)
- d) [What's Near Me? - Find Locations Near You Now \(nearmefy.com\)](#)

To research:

- xpath for selenium
- <https://docs.python.org/3/library/asyncio.html>
- FuzzyWuzzy Documentation → best match out of the list

To add to code:

- try/except for no results / results = []
- **Fuzzywuzzy → how similar two strings are**
- Lambda → in line function
  - Don't write a loop with a DataFrame
- "Metric" of a hit → combine with the title?
- FIGURE OUT METRIC**

BingGPT

- Ask about a specific product

→ progress for transactional blocked by business model

<https://selenium-python.readthedocs.io/waits.html>

<https://offerup.com/>

# April 7th

## Code + FuzzyWuzzy

### 1) Best match out of [the list](#)

Token Sort Ratio

irrespective of position

```
>>> fuzz.ratio("fuzzy wuzzy was a bear", "wuzzy fuzzy was a bear")
91
>>> fuzz.token_sort_ratio("fuzzy wuzzy was a bear", "wuzzy fuzzy was a bear")
100
```

Token Set Ratio

considers duplicate words as a single word.

```
>>> fuzz.token_set_ratio("fuzzy was a bear", "fuzzy fuzzy was a bear")
84
>>> fuzz.token_set_ratio("fuzzy was a bear", "fuzzy fuzzy was a bear")
100
```

Process

```
>>> choices = ["Atlanta Falcons", "New York Jets", "New York Giants", "Dallas Cowboys"]
>>> process.extract("new york jets", choices, limit=2)
[('New York Jets', 100), ('New York Giants', 78)]
>>> process.extractOne("cowboys", choices)
("Dallas Cowboys", 90)
```

a)

There is also one more ratio which is used often called **WRatio**, sometimes its better to use WRatio instead of simple ratio as WRatio handles lower and upper cases and some other parameters too.

```
 fuzz.WRatio('geeks for geeks', 'Geeks For Geeks')
100
 fuzz.WRatio('geeks for geeks!!!!','geeks for geeks')
100
# whereas simple ratio will give for above case
fuzz.ratio('geeks for geeks!!!!','geeks for geeks')
91
```

b)

### 2) Metric of a 'hit' - for how similar 2 items are

a) Combination of WRatio and TokenWTOKENRatio

### 3) Lambda + Don't Loop through data frame

## Selenium Research

### 1) [Waits](#)

a) An explicit wait makes WebDriver wait for a certain condition to occur before proceeding further with execution. An implicit wait makes WebDriver poll the DOM for a certain amount of time when trying to locate an element

b) Also see: [asyncio](#)

### 2) Locating Elements XPath

a) One of the main reasons for using XPath is when you don't have a suitable id or name attribute for the element you wish to locate; Absolute XPaths contain the location of all elements from

*the root (html) and as a result are likely to fail with only the slightest adjustment to the application; By finding a nearby element with an id or name attribute (ideally a parent element) you can locate your target element based on the relationship*

```
<html>
<body>
<form id="loginForm">
<input name="username" type="text" />
<input name="password" type="password" />
<input name="continue" type="submit" value="Login" />
<input name="continue" type="button" value="Clear" />
</form>
</body>
</html>
```

The form elements can be located like this:

```
login_form = driver.find_element(By.XPATH, "//html/body/form[1]")
login_form = driver.find_element(By.XPATH, "//form[1]")
login_form = driver.find_element(By.XPATH, "//form[@id='loginForm']")
```

b)

c)

The "Clear" button element can be located like this:

```
:.find_element(By.XPATH, "//input[@name='continue'][@type='button']")
:.find_element(By.XPATH, "//form[@id='loginForm']/input[4]")
```

d)

e) More: [XPath Nodes](#)

The username element can be located like this:

```
: = driver.find_element(By.XPATH, "//form[input/@name='username']")
: = driver.find_element(By.XPATH, "//form[@id='loginForm']/input[1]")
: = driver.find_element(By.XPATH, "//input[@name='username']")
```

1. First form element with an input child element with name set to *username*
2. First input child element of the form element with attribute *id* set to *loginForm*
3. First input element with attribute *name* set to *username*

## BingGPT

1) Ask about a specific product

a) **See uploaded PDFs - definitely an opportunity to find products**

## \*Current Scraping Capabilities\*

### Brands

- 1) Mark Cross (Stockist)
- 2) JL Audio (Stockist)
- 3) Obey Clothing (Stockist)
- 4) ROC Skincare (Stockist)
- 5) Hurley (Stockist)
- 6) Le Specs (Stockist)
- 7) Ulla Johnson (Stockist)

### Search Engines

- 1) X pages of Google
  - 2) 2 pages of DuckDuckGo
  - 3) First page Mojeek (not always results for all zip codes)
  - 4) First page Yahoo
  - 5) First page of Bing
  - 6) First page of Yellow Pages
  - 7) First page of Brave Search (**RECAPTCHA ISSUES when many zipcodes pasted**)
  - 8) First page of StartPage
- 
- More brand sites/engines → Google Maps
  - Flask → Proof of concept

# Meeting Notes for 4/17

## Flask

- 1) Regular link not always working



a)

- 2) How to HTML in JupyterLab

a) Flask tutorial for absolute beginner | How to create web app...

**b) Template gives you structure**

**c) Text file → HTML**

- 3) Online hosting: Building it with it in mind?

a) [flask-ngrok · PyPI](#)

b) <https://www.linode.com/>

c) <https://replit.com/pricing>

**d) MAKE SURE NOT TO HARD CODE IT, no domain**

- 4) How to connect data?

## Presentation

- Length?
  - **15 minute presentation + FAQ**

**Work of CSV (for now) → prototyping what it would look like**

**Final product mock up visuals → get ot as close as possible via code**

Get vs post

## MOCK UP

- Do heavy organize in python, simple data pass through template

**What the meta search should look like so people can see what you what it to be**

### Meeting Summary:

- Use Safari for running the link
- Build HTML templates for structure; save text files as .html
- Don't hard code anything, so it can be put on a public server

**What the meta search should look like so people can see what you what it to be**

- 1) ~~Ideal Website visual Canva mockup~~ What ~~data do you want to be displayed?~~
- 2) Learn HTML/Flask
  - a) [Quickstart](#), GPT, LinkedIn Learning
  - b) Work from CSV dataset (do heavy organize already)
- 3) ~~Prepare for 15 minute (+FAQ) presentation~~  
~~NEXT FRIDAY~~

# \*Final Presentation

The screenshot shows the Costco Customer Service website. At the top, there's a navigation bar with links for Membership, Costco.com, Warehouse, Returns & Exchanges, Products & Services, Payment & Tax Info, Technical Help & Warranties, and Business Delivery. A search bar is also present. The main content area features a large blue header box with the text "How can I find if an item is available at my local Costco?". Below this, there's a list of ways to check item availability, followed by a note about contacting customer service. On the left side, there's a sidebar with icons and links for My Orders, Membership Auto-Renewal, Return/Replace Order, Price Adjustment, Cancel Order, Verify Membership, and a Print link.

## How can I find if an item is available at my local Costco?

Brought to you by SeatGeek: [FuzzyWuzzy: Fuzzy String Matching in Python - ChairNerd](#)

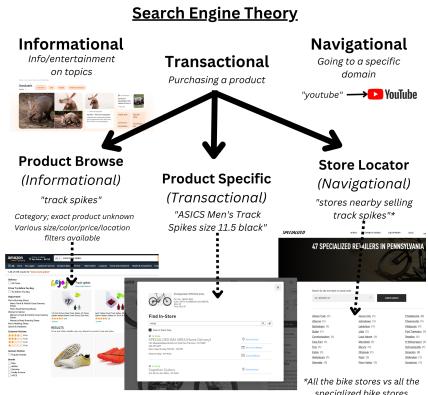
<https://www.fastcompany.com/90563920/this-map-tells-you-if-you-live-in-a-15-minute-city>

Hardware store, radio kismet

### →ANTI SEO

**RECORD A VIDEO SKETCH & TED TALK STYLE SPEECH**

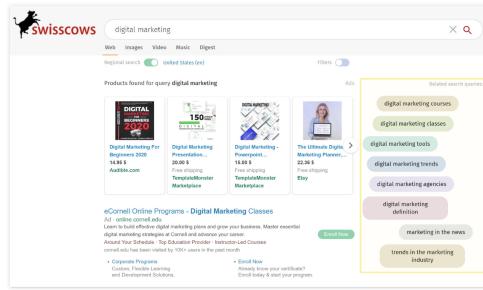
"Oh no! I need to find PBR quickly tonight!"



Solution: Cloud-based point of sale for processing and inventory

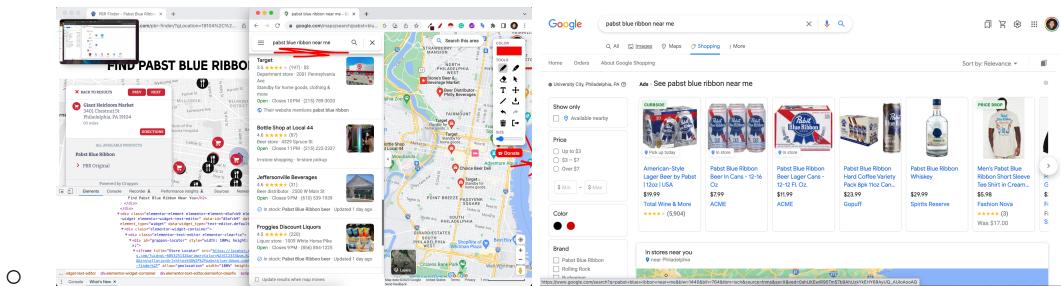
- One issue: Different results on SERPs at different times
- New HBO Max Ads Look Like Scenes From 'Sex Lives'
- Tagging Bias - How to cross-list stores like Target and Walgreens with several types of products

#### Feature spotlight



○ Swisscows has “semantic maps” to help you refine your searches.

- Contact [@RebLieuw](#) for PR
- Direct Contrast:



# Week of April 19th

<https://ieeexplore.ieee.org/abstract/document/1401768>

- trust and fairness that characterize the printed Yellow Pages
- They are both trustworthy and inclusive—they contain at least minimal information on all businesses

## Cut Sections

- *Proving the problem exists and a stockist solution with statistical analysis/Z-scores*
  - a) **Store location & Brand**
  - b) *Aim: Qualtiviatley shows more stores and (assumption) more accurate, cleaning up what is out there in a clear good way*
  - c) *Example: PBR*
  - d) *Assumptions: Assuming they are updated recently, assume all companies are on stockist, Brand may not show product sku (stock keeping unit) vs UPC*
  - e) *How2:*
    - i) **Scraping Stockist, Grappos, Other** → Companies with store locators, City Liquor license,
- [https://en.wikipedia.org/wiki/Telephone\\_directory](https://en.wikipedia.org/wiki/Telephone_directory)

## 1) Color & Size

- a) **Store location & Item**
  - b) *Example: PBR: Bottles vs cans for pool party or Girl: sees all the stores that carry it - but I have a specific item in mind I need in my size - this isn't helpful!*
  - c) *HOW2:*
    - i) *Cooperative brands - Email a couple of sample cooperative brands to get their entire inventory list*
    - ii) *Crowdsourcing Stores - Store collectives to send/share their Shopify*
    - iii) *Consumer Crowdsourcing - Amazon-style app for people to scan and upload UPC*
1. *Product name/color/size by number not words -*
    - a. [LoveShackFancy Moe Skirt in Pre-Dawn Sky - SINGER22](#)
    - b. <https://www.barcodelookup.com/840098856185>
    - c. <https://devs.upcitemdb.com/>

d. [Retail Barcode Label App Api Access - Shopify Community](#)

PA What's the difference between yellow pages and Business White Pages

 Yellow Pages and Business White Pages are directories that serve different purposes in terms of listing and searching for businesses and organizations.

1. Yellow Pages: This directory focuses on categorizing businesses and services based on their industry or type. Listings in the Yellow Pages often include business names, addresses, phone numbers, and sometimes additional information such as websites or email addresses. The Yellow Pages are typically used by consumers to find and contact local businesses or services that fit their needs. The term "Yellow Pages" originates from the fact that these directories were traditionally printed on yellow paper to distinguish them from the White Pages.

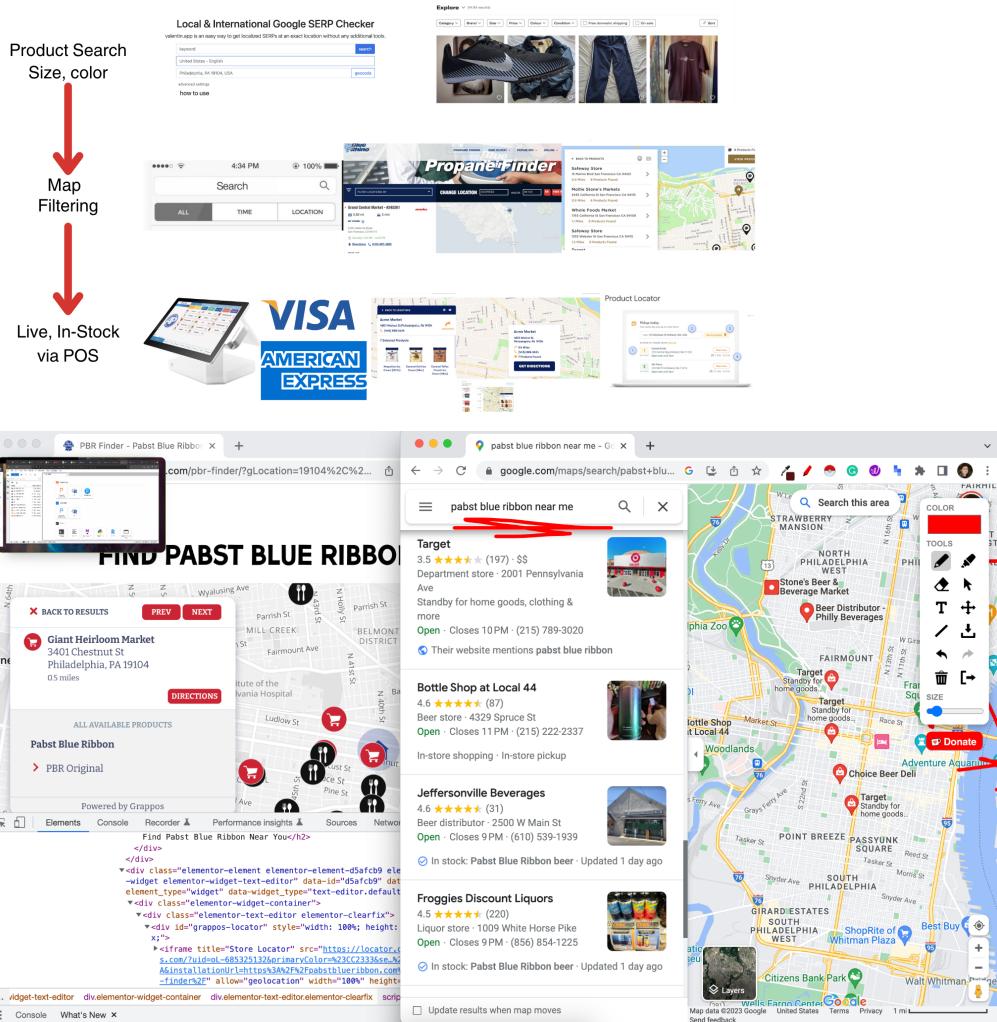
2. Business White Pages: This directory primarily lists businesses, organizations, and government agencies in alphabetical order by name, regardless of their industry or type. Business White Pages usually provide basic contact information such as business names, addresses, and phone numbers. The main purpose of the Business White Pages is to offer a comprehensive list of contact information for businesses and organizations in a specific area, making it easier for users to find contact details for a specific business or organization they already have in mind.

- 
- <https://www.loc.gov/resource/usteledirec.usteledirec08212/?sp=482&r=0.214,0.189,0.364,0.196,0>

<https://www.blog.google/products/search/making-search-results-more-local-and-relevant/>

<https://wheretobuy.io/>

<https://centralcitynews.us/wp-content/uploads/2015/03/2015-CORRECTED-Central-Directory1.pdf>



# April 24th

**Finish Code:** Demo with HTML & Fuzzy matching

**Present:** FRIDAY IS THE BIG DAY

- 12:15 - Annenberg 225

====

## Submitting on GitHub

- Annotated Biblio → May 5th
- Push notebooks to GitHub since all local
- Live Log Book → Send as PDF

====

### 1) Watch practice presentation

#### 1. Talk some Tech

- a. Demo Slide
  - i. FuzzyWuzzy
- b. Semester Learnings Slide
  - i. Selenium WebScraping

#### 2. Make clear two distinct ideas: **What I did vs Big Picture**

- a. 1) Store Searcher (based on product locator, and then some)
  - i. Modern day yellow pages → based on product locator since database weak
- b. 2) Google Maps crossed with inventory: Started out as a search engine, but actually very different from a search engine (ended up with more of an eBay auction/inventory system that you can search through)
  - i. User Interface sample UPDATE SLIDE

#### c. **EVOLUTION**

#### d. ***List listing vs Live Bidding***

- Critical Mass, would Nordstroms be on it?

- Is it in stock? Implementation → stolen staffer
- Less like Google, more like Destini
  - eBay + Google Maps with a search component

What I did vs. Big picture concept → Evolved as the presentation went on

- NearBuy PAUSE Philadelphia Zoo

Live Listing vs Live Bidding

Modern day yellow pages → based on product locator since database weak

BIGGER IDEA:

- POS, Live inventory system

VC:

UI for POS

<https://squareup.com/us/en?v=all>

Tagging and indexing physically things so everything can be found

- Scan a physical object → SEE PINNED TWEET

<https://flask.palletsprojects.com/en/2.2.x/quickstart/>

<https://www.linkedin.com/learning/flask-essential-training/flask-s-development-environment?autoplay=true&resume=false&u=2262250>