How Internet Search Works

Van Londen
How Search Works

When asking questions of a search engine, it is able to list search results very quickly because it has already indexed most of the files online around the world, according to The Internet: How Search Works. Google, for example, sends a spider algorithm out to the Internet to crawl through the linked web pages so it could categorize and archive them for later. Hyperlinks from one web page to another help the spider find even more pages to index in the database of web pages. Web pages are then ranked by the page title, key words and phrases, and by how many other pages link to them, resulting in page rank. Spammers try to out-smart the algorithms so their pages will list higher in search results, so Google must constantly improve them using machine learning, whereby the algorithm understands cultural meanings and not just single words.
Using Search Phrases

Not everyone understands the most efficient ways to search Google, according to Hack College’s [Tips & Tricks for Students](https://www.hackcollege.com/tips-for-students) infographic. Use operators such as `site:`, `~` tilde, `””` quotation marks, `-` hyphens/minus, `…` ellipses between keywords, `filetype:`, `intitle:`, `define:`, and `*` asterisk to narrow the results. To search professional journals, use [scholar.google.com](https://scholar.google.com) along with `author:` name as well as keywords and quotation marks to narrow those results to academic studies and papers. Within a single article, search for keywords using ⌘ or `Ctrl-F`. Calculations and conversions can be typed right into the address/search bar. Use library databases such as [JSTOR](https://www.jstor.org) and Wikipedia to locate original sources (which are located in the References section at the bottom of each encyclopedia page). Bibliographies from books and papers are also excellent for mining sources.
How Online Advertising Works

Shopkeepers could not easily discover and track what products you liked looking at in their stores prior to the Internet. As Online Advertising explains, shopkeepers had to guess at where you might be next so they could show those products to you...on billboards, in magazines, on the radio or tv, etc. and rely on general demographic data to make ad-buying decisions. Now, shopkeepers buy online ads which are driven by algorithms focused on niche products and other content that you search for. Cookies in the browser provide the mechanism that tracks your Internet searches so they can serve sponsored ads next to editorial content in the next websites you visit. These ads hyperlink to pages that promote the product, effectively getting you to purchase.
How Online Advertising Works (continued)

These landing pages also keep accurate records (analytics) of how many people saw the product, their ages, interests, income and education levels, countries of origin, internet speed, devices, etc., allowing behavioral targeting. They are allowed to use your personal data, text, and media as they see fit when you inadvertently agree to their terms and conditions.
Truth in Advertising Laws

The USA Federal Trade Commission (FTC) provides requirements and guidelines for businesses in Advertising and Marketing on the Internet: Rules of the Road to “maintain credibility” in the industry. The two overarching guides are that the ads must be truthful. They cannot mislead or claim something that is not provable. Ad agencies, website owners, and catalog marketers are responsible for assuring truthful display. They must provide disclaimers, demonstrations, refunds, and control what is presented to children (COPPA). Some types of advertising, such as MLM, Pyramid schemes, lending and billing, fair credit reporting, warranties, and guarantees have more guidelines. Environmental and jewelry claims must be substantiated. Free products must be explicitly explained. Mail/telephone orders must ship within 30 days.
Truth in Advertising Laws (continued)

Testimonial and endorsement relationships must be disclosed. The country of origin must be provided on textiles and “Made in the USA” products must be made from USA components. **Non-compliance** may result in cease and desist notices, fines, and lawsuits.
Who Profits from Online Advertising

Among global digital advertisers, the USA gained the most in revenue which was about 80 billion dollars in 2016 and in 2020 it will probably be 94 billion. The forecast for ad spending worldwide is $204 billion dollars. Retail, automotive, and financial services spend the most on ads. Search advertising via Google nets the highest revenues ($35 billion in 2017) and Facebook display ads ($17 billion in 2017) comes in second. Microsoft and Amazon are the next highest. Mobile device ads have superseded desktop computer ad sales (2017). Recall rates among consumers are highest (58%) when seeing full-screen ads on their mobile versus on their desktop/laptop devices (38%). Mobile devices are less likely to have ad blocking software, which helps account for the popularity. Sixty-one percent of retailers worldwide think online ads are more effective than print ads.
How Data Mining Works
Advertising companies gather data about people in a variety of ways, such as through **loyalty programs**, microtargeting, geotracking, and cyberpsychology. Data Mining explains how our personal information often falls into the hands of those who should not have it. One unethical example is that the father received advertising about baby items from Target because they learned his daughter was pregnant. The baby shower registry and purchase list helped form a **prediction score**. Companies can predict when you are moving, when someone is graduating, and looking for a new car. Our personal shopping details are sold by **third party advertisers** who then target our other online reading habits with ads related to those future purchases. Many **surveys are ‘bribes’** for market research. When we have location services turned on, that **geolocation data** gets sold as well.
How Artificial Intelligence Works

**Machine Learning & Artificial Intelligence** algorithms allow computers to make decisions and predictions. **Classification** can be done by algorithms which assign features to text/media based on human experts’ labeled data. **ML** can find optimal separations between a sets of overlapping data to create decision boundaries and identify outliers on a confusion matrix. ML can predict, then, whether unlabeled data likely belongs to one set using **Decision Trees. Support Vector Machines** use math to divide up the space on a classification matrix in logical fluid lines rather than straight, arbitrary lines. Adding one more feature to the classification table results in a 3D representation.
How Artificial Intelligence Works (continued)

Real-world statistical problems often have many more dimensions/features, requiring **AI Neural Networks**, which calculate at high speeds in complex layers (**deep learning**) using graphics processing units (**GPUs**). The results are self-driving cars, facial recognition, translation, etc. Specific task intelligence is **Narrow AI** and well-rounded human-like intelligence is called **Strong AI**. **Reinforcement learning** is possible through hundreds of practice sessions.
Problems with AI

In These are three of the biggest problems facing today's AI, James Vincent talks with experts about what is needed for AI to work efficiently and provide helpful results for human problems. Most importantly, only large corporations have access to the large quantities of data needed for machines to make informed decisions or spot trends. Some data is private and not available. A second problem is that deep learning can only take place when neural networks are well trained...which requires the ability of machines to learn multiple tasks at once. One promising method is progressive neural networks where machines share solutions with each other. A third big problem is understanding how AI comes to a conclusion. Helping machines to synthesize thinking with deep symbolic reinforcement learning, like scientists at Imperial are doing, will increase their efficiency.
Solutions Provided by AI

Murati Techlabs discusses **8 problems that can be easily solved by machine learning**.

**Manual data entry** is boring and time consuming, so images and text can be analyzed by machine and appropriately placed in clean datasets using natural language processing. Social media companies use neural network ML rules to detect **spam** email, posts, and ads. Companies like Amazon provide ML-driven product **recommendations** based on previous purchases and searches. Other statistics related to **segmentation**, **churn prediction**, and **lifetime value** help marketers make better product and ad-buying choices. Datasets of anonymized patient data helps machines **diagnose medical conditions** and **make predictions**. Financial institutions take advantage of **algorithmic trading** and **fraud detection**, among other tasks.
Solutions Provided by AI (continued)
The manufacturing industry can easily locate patterns in **factory maintenance** with AI and ML. historical data, flexible analysis, workflow visualization, and feedback loops as well as simulations of sensor failure help keep machines running. Computer vision helps them **recognize images** for data mining.
Bibliography

3. Online Advertising: Crash Course Media Literacy #7 Retrieved 27 June 2018, from https://www.youtube.com/embed/cmRcoJZRXEY