

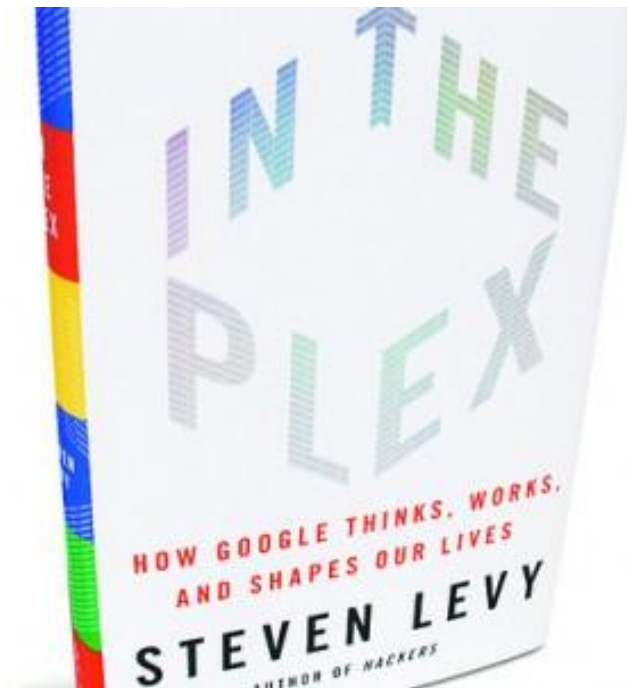
# BOOK: "IN THE PLEX"

AUTHOR: STEVEN LEVY

# “IN THE PLEX”

**Book: How Google Thinks, Works and Shapes our Lives**

**Author: Steven Levy**



# BEFORE SEARCH ENGINES

**When the Internet Started there was no search**

- **Static Indexes at some sites**
- **Web sites were found by someone referencing them in an article or E-Mail**
- **There was a book called “The Yellow Pages of the Internet” with lists of web sites organized by subject**

# EARLY WEB SEARCH SITES

- **Archie (Manual search of Public Sites) - 1990**
- **Gopher Index System-1991**
- **WebCrawler - 1994**
- **Alta Vista, Magellan, Excite, Yahoo -1995**
- **Ask Jeeves – 1996**
- **Google -1998**
- **Ask – 2005**
- **Microsoft Live Search - 2006**
- **Microsoft Bing -2009**
- **Full list at [http://en.wikipedia.org/wiki/Web\\_search\\_engine](http://en.wikipedia.org/wiki/Web_search_engine)**

# GOOGLE FOUNDERS

## **Larry Page**

**Grew up in Lansing, Michigan**

**Undergrad: Computer Science at University of Michigan**

## **Sergey Brin**

**Born in Russia**

**Mathematical Wiz**

**Undergrad: University of Maryland, Mathematics**

**Both were at Stanford graduate school**

**They met when the guide for Larry Page for Stanford introduction tour was Sergey Brin**





# ORIGINAL NEW CONCEPT FOR WEB SEARCH



**Rather than listing all the links found in order of receipt**

## **Rank the links by**

- Enumerating all the links that referenced that web page in the link
  - Rank the pages found by which pages are referenced the most by other sites
  - Similar to how scientific articles are evaluated as important

**Problem: You would have to capture the entire web link structure to determine all the links back to a given page**

**Solution: A web crawler that found and indexed the entire web**

# OTHER GOOGLE KEY IDEAS

- **Keep Hardware costs low**
  - Build your own servers
  - Use redundancy rather than reliability
  - Build servers where electric power and land is cheap
- **Execute Searches Fast**
  - Detailed analysis of code to improve performance
  - Each search is timed
- **Hire only the Brightest and Best**
  - Difficult Entrance tests
  - Many Interviews for each candidate
- **Motto**
  - “Don’t Be Evil” (Hard to do as company became bigger)



# GOOGLE INDEXING FILE SYSTEM

## **In-RAM**

- **As much as possible of the indexing file is kept in memory**

## **Sharding**

- **Split up the web into regions with each region is on its own computer**

## **Computer system redundancy**

- **If one computer goes down another one can take over**

## **Index Updated**

- **The entire web is crawled in parallel every night, to update the index**



# UNDERSTANDING LANGUAGE ARTIFICIAL INTELLIGENCE NEEDED

**New York is 2 words, a city**

**New York Times is 3 words, a newspaper**

**New York Times Square is 4 words, a location**

- The search engine needs to understand not just individual words but strings of words, so AI is needed to understand

## **More Complex Search strings**

- “mike siwek lawyer mi”
  - Lawyer is not the last name but profession attorney
  - Mi is Michigan
- “rock”
  - Is it a rock, a stone, a city Little Rock



**The search engine is constantly learning from everyone's daily searches**

# USING LOCATION



**Google search provides different results depending on where the computer that is doing the search is located**

**Google tries to figure out your location through your IP address**

**If you do not put in location information it tries to rank searches depending on your IP address location**

- “Newspaper” or “Newspaper San Jose, Ca”
- “Movie” or “movie Sunnyvale, ca.”
- “Chinese restaurant” or “Chinese restaurant, palo alto, ca.”

# ADWORDS AND ADSENSE



**Use auctions to sell ads**

**If Joe bids 10 cents a click, Alice bids 6 cents a click and Sue bids 2 cents a click, the Joe wins and pays 1 penny more than the smallest bid, that is 3 cents per click and gets the best placement of their ad and Alice is the runner up also paying 3 cents per click**

**Auctions are automatic and happen when someone types in a word in the search**

**Ads are placed on the right hand side of the page and are unobtrusive to those only wishing to see the results of the search**

**2002 First profitable year for Google**

**Some advertisers tried to game the system and use many search words in their ads. The AdSense had to prevent this and also find the best ads to use using AI techniques**

**Salespeople had to sell the ads to advertisers and convince them the system worked best for them and the consumer**

# GOOGLE UNIVERSAL SEARCH

**How to show images, products, text search, videos all on one page?**

**Google just wanted 1 search engine for everything**

## Google Universal Search Explained

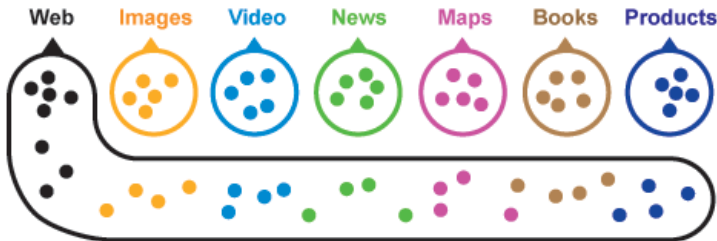
### Old Google Model

Previously, a web search on Google would only return web pages and document contents. Items such as images and video were not included.



### New Google Model: Universal Search

A web search on Google can now include multiple types of information such as images, video, news, maps, books, and products.



# GETTING RID OF SPAM



## Search Engine Optimization (SEO)

- **Some web sites tried to get around Google's blocking system**
  - They would try and game the system so their pages ranked high
  - Google had to fix the SEO every month to adjust to web sites gaming the SEO

# CONTENT PROVIDERS VS GOOGLE

Content farms, which have flourished on the Web in the past 18 months, are massive news sites that use headlines, keywords and other tricks to lure Web-users into looking at ads. The texts are generated by paid writers. Examples:

- Associated Content
- Answerbag
- Demand Media

You can't mess with Google forever. In February 2011, Google concocted what it concocts best: an algorithm, Panda.

- Panda, affects some 12 percent of searches, and it has — slowly and imperfectly — been improving things
- Panda 2.2 came out in June 2011

# GOOGLE CHROME



## Google Chrome

- Fast browser
  - Loads Quickly
  - Use a prediction service to help complete searches and render pages
- Many features hidden
- Cleaner interface
- More Difficult to setup and find the hidden features
- URL bar is also the search input bar
- Has natural language voice search



# IMAGE SEARCH



**Images.google.com**

**Drag image to the search area and it will look for a matching image**

# GOOGLE PRODUCTS

- |                    |                 |                   |
|--------------------|-----------------|-------------------|
| 1. Web search      | 10. Panorama    | 19. Translate     |
| 2. Browser: Chrome | 11. Docs        | 20. SketchUp      |
| 3. Maps            | 12. Gmail       | 21. Desktop       |
| 4. YouTube         | 13. Calendar    | 22. Finance       |
| 5. Books           | 14. Google+     | 23. iGoogle       |
| 6. Images          | 15. Sites       | 24. Blogger       |
| 7. News            | 16. Talk (Chat) | 25. Picasa/Picnik |
| 8. Earth           | 17. Voice       | 26. Blogger       |
| 9. Reader          | 18. Knol        | 27. Groups        |

# GOOGLE PRODUCTS REMOVED

1. ~~Google Labs~~
2. ~~Google Powermeter~~
3. ~~Google Health~~

# CONFLICTS AND PROBLEMS



**As Google grows to be a big company problems and conflicts happen**

- **Gathering router information during street views**
- **Some government restrictions on street views**
  - Switzerland, Italy, Czechoslovakia
- **Censorship in China**
  - Google moved China operations to Hong Kong
- **Larry Page, Google CEO, to testify at Congress on abusing search dominance**

# CONFLICTS AND PROBLEMS

## Google Sued for \$421 Million by Paris Rival for Ad Policies

(Updates with Google comment in fifth paragraph.)

June 28 (Bloomberg) -- Google Inc. was sued in France by a local competitor over claims the world's largest search engine blocks rivals from reaping advertising revenue and gives preference to its own sites in query results.

1PlusV, a Paris-based web publisher, filed the lawsuit to the Paris Commercial Court today, seeking 295 million euros (\$421 million) in damages and an order that Google post details of alleged "anti-competitive behavior" on its French home page for three months, 1PlusV said in an e-mailed statement.

---

STORY TOOLS

# GOOGLE MARKET CAPITALIZATION ON 8/14/11

**IBM Market Cap                      \$200B**

**Microsoft Market Cap           \$210B**

**Apple                                        \$349B**

**Intel Market Cap                    \$109B**

**Google Market Cap                \$182B**

- Revenue \$8B in 2010
- Profit        \$2B in 2010