

## Closing remarks

1

## Where we started

“Google's mission is to organize the world's information and make it universally accessible and useful”  
[Google's mission statement](#), ~ 1998.

World Wide Web invented by Tim Berners-Lee 1989

“A memex is a device in which an individual stores all his books, records, and communications, and which is mechanized so that it may be consulted with exceeding speed and flexibility. It is an enlarged intimate supplement to his memory.” [Vannevar Bush](#), *As we may think*, *Atlantic Monthly*, July 1945.

2

## Where we have been: [major themes](#)

3

## 1. Mathematical models fundamental

- how model information
  - Capture structure within?
- [vector](#) models + linear algebra
- [graph](#) models
  - links and paths
- [probabilistic](#) models
  - Markov models
- applications: text, Web, other media

4

## 2. Algorithms & data structures

- indexing
- graph analyses
- random walks
- eigenvector computation
- clustering
- sampling
- aggregating
- compressing

### Algorithms meet BIG practice

- distributed computation

5

## 3. Costs: what optimizing?

- **Quality** of solutions
  - Almost always approx' tions or heuristics
  - Humans add information
    - users: characteristics & feedback
    - authors: semi-structured content
- **Performance**
  - Disk I/O
    - Drives times of algorithms
    - Drives main memory needs
      - buffers, caches
  - Network latency
    - highest time and \$ cost

6

## 4. Evaluation

- Must have **quantified**
  - Not just “see how well works”
- May need more than one measure
- Quantifying does not preclude **human perception**

7

What did we miss?

8

# Information Summarization and Visualization

For illustrations of Metro Maps, see [Trains of Thought: Generating Information Maps](#), Dafna Shaha, Carlos Guestrin, and Eric Horvitz, *Intern. World Wide Web Conf. (WWW)*, ACM, 2012, pp. 899-908.

For illustration of social graph of feelings see

[We Feel Fine and Searching the Emotional Web](#), Sepandar D. Kamvar and Jonathan Harris, *Proc. of the Intern. Conf. on Web Search and Data Mining (WSDM)*, ACM, 2011, pp. 117-126.

# Privacy

The screenshot shows a web browser window with the URL 'www.nytimes.com/2010/08/02/technology/02search.html'. The article title is 'A Face Is Exposed for AOL Searcher No. 4417749'. The text of the article reads: 'Buried in a list of 20 million Web search queries collected by AOL and recently released on the Internet is user No. 4417749. The number was assigned by the company to protect the searcher's anonymity, but it was not much of a shield. No. 4417749 conducted hundreds of searches over a three-month period on topics ranging from "mumb fingers" to "No single ones" to "log that satirizes on everything."'. Below the article, there is a section titled "'10 Concerts' Facebook Meme May Reveal More Than Musical Tastes' by CHRISTOPHER HILL and DANIEL VICTOR APRIL 28, 2017'. The image shows a person sitting on a car with a 'facebook' logo.

# Security

The screenshot shows a web browser window with the URL 'thehackernews.com/2017/02/yahoo-hack.html'. The article title is 'Yahoo Hacked Once Again! Quietly Warns Affected Users About New Attack'. The text of the article reads: 'The Hacker News™ Security in a serious way. MANAGED SECURITY FOR YOUR DATA. rockspace. Yahoo Hacked Once Again! Quietly Warns Affected Users About New Attack. Beginner's Guide to Hybrid Cloud Security: From On-Premise to the Cloud. TRENDING STORIES: How Microsoft Warned, Signed With Last Apple OS, Found Spying in Windows 10. Eminent New Android Malware: WPMAL2 Malware Google Play. An in-the-know hacker master is using tactics for privacy invasion, slipping the more intel corporate makes it possible for her to be used when it disclosed insufficiently anonymous information about nearly half-a-million customers as part of its \$1.1 billion contract to improve its accounting system.

## Where are we going

13

## 1. new modes of search

- real-time search
- non-text based input
  - spoken requests
  - content-based input
- concise output
  - text versus non-text
  - summarizing
    - single source
    - multiples sources

handhelds  
drive  
much of

14

## 2. search everything

- multi-media
  - obvious: image, music, video,
  - maps
  - other ?
- social networks
  - Influence in many ways
- cloud computing
- objects: “internet of things”

15

The screenshot shows a web browser window with the URL [www.computerworld.com/article/310341/10341-the-internet-of-messy-things.htm](http://www.computerworld.com/article/310341/10341-the-internet-of-messy-things.htm). The page is from ComputerWorld and features an article titled "The Internet of messy things" by Steven J. Vaughan-Nichols. The article's sub-headline is "How much damage can a 'smart' toaster do? Lots, and not just burning your bread." The page includes a green header with the year "2015" and a social media share bar. Below the article is a large image of a man standing in a colorful, futuristic-looking environment, possibly a data center or a smart home. The year "2017" is visible on the left side of the image.

### 3. real semantic-based search

“ ... It is an enlarged *intimate* supplement to his memory.”

Vannevar Bush

- question answering
- understanding user intent

17

### Where are “we” going?

1 + 2 + 3 + ? → Semantic Web ?

vision of use of information by active agents  
Berners-Lee, Hendler and Lassila  
Scientific America, May 2001

18

### Semantic Web Overview

- Initiative of W3C: WWW Consortium
  - academic, government and industry
  - begun 1994 by Tim Berners-Lee
- common frameworks for data specification
- frameworks allow sophisticated functionality
  - automated understanding and use of information
  - Inference!
- open specifications, open source
  - Allow independently written tools interoperate

19

### Major concerns going forward

- Security!
- Privacy!
- Data explosion?
- Universal access?
  - Resource limitations
    - developing nations
- Others?

20

- For the statistics on world internet usage and percent penetration of the population by region see [Internet Usage Statistics](#), *Internet World Stats*