Closing remarks

Where we started

“Google’s mission is to organize the world’s information and make it universally accessible and useful”

World Wide Web invented by Tim Berners-Lee 1989

“A memex is a device in which an 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.

Where we have been: major themes

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

2. Algorithms & data structures
   - indexing
   - graph traversal
   - random walks
   - eigenvector computation
   - clustering
   - sampling

Algorithms meet BIG practice
   - distributed computation

3. Costs: what optimizing?
   - Quality of solutions
     - Almost always approximations 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
4. Evaluation

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

Where are "we" going?

1. new modes of search
   - location-based search
   - real-time search
   - non-text based input
     - spoken requests
     - content-based input
   - concise output
     - text versus non-text
     - summarizing
     - single source
     - multiples sources

Where are "we" going?

2. search everything
   - multi-media
     - obvious: image, music, video,
     - maps
     - other?
   - data
     - Deep Web
   - social networks
     - influence in many ways
   - cloud computing

Where are "we" going?

3. real semantic-based search
   "... It is an enlarged intimate supplement to his memory."  
   Vannevar Bush

   - question answering
   - understanding user intent
     - personalization

Where are "we" going?

1 + 2 + 3 + ?  →  Semantic Web?

A bedtime story by
Berners-Lee, Hendler and Lassila
Scientific America, May 2001

The agent promptly retrieved information about Mom's prescribed treatment from the doctor's agent, looked up several lists of providers, and checked for the ones in-plan for Mom's insurance within a 20-mile radius of her home and with a rating of excellent or very good on trusted rating services. It then began trying to find a match between available appointment times (supplied by the agents of individual providers through their Web sites) and Pete's and Lucy's busy schedules. In a few minutes the agent presented them with a plan.

... Lucy's agent, having complete trust in Pete's agent in the context of the present task, automatically assisted by supplying access certificates and shortcuts to the data it had already sorted through.
Semantic Web

- **structured** data
  - information?
- **interoperability**
  - unrelated sites
- **functionality**
  - active agents vs passive retrievers

W3C Semantic Web Activity Statement
http://www.w3.org/2001/sw/Activity

Excerpt:
"The goal of the Semantic Web initiative is as broad as that of the Web: to create a universal medium for the exchange of data. It is envisaged to smoothly interconnect personal information management, enterprise application integration, and the global sharing of commercial, scientific and cultural data."

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
- open specifications, open source
  - Allow independently written tools interoperate

Frameworks and Methods

- **publishing and linking** data
  - Resource Description Framework (RDF)
- define **structure**
  - Web Ontology Language (OWL)
- **linking** “knowledge organization systems”
  - Simple Knowledge Organization System Reference (SKOL)
- **query** language
  - SPARQL for RDF
- **inference**
  - Rule Interchange Format (RIF)

RDF

- Graph model to represent **resources** and relationships between them
  - Documents and other resources
- Formal semantics
- XML syntax
- URIs for naming –Uniform Resources Identifiers
  - Generalization of URLs
- Most recent standard update Feb 2004

RDF representation

- Represents “Web resources”
  - Documents on Web
  - Generalizes to “objects” identifiable but not directly retrievable, e.g. shopping facility
- Represents **metadata** for resources
  - Title, author, copyright of document
  - Price, shipping date of an item for sale
RDF Graph Model

- Nodes: resources and property values
- Edges: labeled with property identifiers (i.e. attribute names)

Example from http://www.w3.org/TR/rdf-primer/
"there is a Person identified by http://www.w3.org/
People/EM/contact#me,
whose name is Eric Miller, whose email address is em@w3.org, and whose title is Dr." 

OWL

- Advanced support for
  - software agents
    Programs that "understand" and can plan and act
  - knowledge management
    Finding and exploiting complex interactions of information across sources
- Builds on RDF
- Represents ontologies
- OWL 2 standard published Oct. 27, 2009

OWL expressiveness

- Ontology: "representation of terms and interrelationships"
  - very general
  - not just trees
- Has formal semantics
- Can represent relationships between classes

Major concerns going forward

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

Internet Penetration

Wikipedia file Internet Penetration.png; provided by World Internet Users and Population Stats, June 2009