Whither Search?

Rakesh Agrawal
Technical Fellow
Search Labs, Microsoft Research – Silicon Valley
Outline

• Current state of affair
• Evolving Search
• Search Labs projects
Outline

• Current state of affair
• Evolving Search
• Search Labs projects
Does Search Work?

**Navigational Queries**

- Library of Congress Home
- www.loc.gov/index.html
- The Library of Congress. The Library of Congress is the nation's library, and it serves as the research arm of Congress. It houses over 130 million items and provides access to information in every format, from books, manuscripts, and phonograph records to maps, films, photos, and much more.
- Library Catalogs
- Digital Collections
- Researchers
- Visitors
- Customer service 202-707-5000

**Images of library of congress**

- see more · large images · black & white

**Library of Congress Online Catalog**

Online access to the collections resident at the Library of Congress.

---

**Pseudo-Navigational Queries**

- Daniel Defoe
- www.bing.com/images?q=daniel+defoe

**Images of daniel defoe**

- see more · large images · black & white

**Daniel Defoe - Wikipedia, the free encyclopedia**

Daniel Defoe (c. 1659-1661 — 24 April 1731 [1]), born Daniel Foe, was a journalist, and pamphleteer, who gained enduring fame for his novels, essays, and the Anglo-American political pamphlets. He is best known for his novels Robinson Crusoe and Moll Flanders.

**Daniel Defoe - Biography and Works**

Daniel Defoe. Biography of Daniel Defoe and a search.

**Daniel Defoe**

Provides an in-depth look at the author's life and his impact.
My Recent Searches

• Car GPS around $300
• Four day trip to Bhutan from Delhi to visit important Buddhist places
Car GPS around $300

Amazon.com: Dash Express Two-Way Internet Connected Portable GPS

Product Description--March 24, 2008

1 new from $200.00
1 used from $175.00

Amazon.com: Customer Discussions: GPS under $300

She really just wants to be able to use this in her car, for directions. ... Magellan 10 December 2007. GPS under $300, 3, December 2007 ... www.amazon.com/GPS-under-300/forum/.../1?_encoding=UTF8... - Cached - Similar

Google search for "car gps around $300"

What is the best handheld GPS around $300 for hunting and fishing? May 12, 2010 ... What is the best handheld GPS around $300 for hunting and fishing? Long do People keep their car GPS systems, (Portable Type): www.60cx.net/what-is-the-best-handheld-gps-around-300-for-hunting-and-fishing.html - Amazon.com: GPS Express Two-Way Internet Connected Portable GPS

$25.99
Free shipping

$319.84
Savingslist

$35.67
Sold by: Newegg

Sonocadie V300 Color GPS Unit (May 1, 2008)

Garmin Nuvi 370 360 350 310 300 GPS Car Cradle

Hard Carrying Case Cover Portable/Handheld GPS 400 550 5501

Hello. Sign in to get personalized recommendations. New customer? Start here

Your search "car gps around $300" did not match any products.

Showing results using some of your search terms

"car gps around $300" (See all 449 results)
Bhutan from Delhi in 4 Days

Travel Tour India, Bhutan, Nepal, Tibet & Sri Lanka, Asian ...
7 Days north India. Destinations: Delhi-Agra-Jaipur-Varanasi. 7 Days north India. Bhutan Days: 4 days. Destinations: Paro - Thimpu...
www.tvlwithus.com/ - Cached - Similar

Bhutan PM to visit India for four days | Top News
Jun 29, 2009 ... Bhutan PM to visit India for four days New Delhi, June 29 : Bhutan's Prime Minister, Lyonchen Jigm Y. Thinley, will pay a four-day working ...
www.topnews.in/bhutan-pm-visit-india-four-days-2184201 - Cached

Bhutan, Bhutan Tours, Bhutan Travel, Bhutan Tour Packages
Bhutan Short Tour 7 Days / 6 Nights New Delhi - Paro - Thimphu - Paro - New Delhi. Thimphu Paro Tour 4 Days / 3 Nights Paro - Thimphu - Paro ...
www.indovacations.net/english/bhutan tours.htm - Cached

Car Rental Delhi - Rent a Car in Delhi India, Affordable Budget Car
Itinerary 4 days: Taj Delhi / Agra / Fatehpur Sikri / Delhi ... NEW DELHI;Bhutan An Eye to History, an exhibition of photographs, presented by Ministry of ...
www.carrentaldelhi.com/ - Cached - Similar

Bhutan Package Tours, Package Tour to Bhutan, Bhutan Tour Packages ...
DAYS, ITINERARY: 4-Day BHUTAN Package Tour - 2010. 1. ARRIVE BHUTAN Fly Kathmandu, Delhi or Bangkok to Paro. Transfer to Thimpu. Overnight hotel in Thimpu. ...
www.visitnepal.com/.../bhutan tours/bhutan_package_tours.htm - Cached - Similar
How about travel sites?

Amari Boulevard Bangkok, Bangkok hotels, Thailand
Centrally located on Sukhumvit Road in downtown Bangkok, with Nana Skytrain station only 50 meters away, this hotel offers easy access to major shops and sights.

Hotel Reviews of Macleay Serviced Apartments, Potts Point hotels, NSW, Australia
... be rather pricey. A traveler from Bhutan recommends this hotel. Overall satisfaction Hotel service Hotel ...

We could not find any vacation packages in 'Bhutan (all)' that match your search criteria.
User Expectations

WebSQL

Anyone?

Holiday destination From Delhi - Team-BHP
10 posts - 7 authors - Last post: Dec 15, 2008
1) cannot extend beyond 4 days weekend included. ... The only places I know of where you don't need a visa are Nepal and Bhutan. ... www.team-bhp.com/.../50450-holiday-destination-delhi.html - Cached - Similar

Delhi To Pune Non Stop In My Wagon R & Return Solo - Page 2 ...
6 posts - 3 authors - Last post: Jan 30
1) I don't think in any case you gonna need 4 days to reach Delhi from Mumbai (Panvel/Khopoli) it's only 1400 KM on 4 lane highway ...
www.indiamike.com/india/india-travel/.../delhi/.../2/ - Cached - Similar

Quote:

Originally Posted by Kool_Kid
i wish to take my wife on a small weekend (since we have to postpone our honeymoon due to my new job)

1) cannot extend beyond 4 days weekend included. 
2) place should not require visa (or visa on arrival) 
3) start point and end point should be delhi
4) budget <= 1 lakh

pls suggest

Since you mention visas, I suppose you are looking at a foreign trip? The only places I know of where you don't need a visa are Nepal and Bhutan. I am sure there are flights directly to Kathmandu from Delhi. For Bhutan, you can go via Kolkata. Not sure of there is anything from Delhi.
Observations

• Search queries are not grammatically correct questions; but they are not bags of words either
• Query terms are often more than strings of characters
• Data often has structure or structure can be derived
• Search can span multiple sessions over several days
• Search often provides entry point for browsing and search and browsing are inter-mixed
• Expectations from Search are increasing
Outline

• Current state of affair
• Evolving Search
• Search Labs projects
From Search to Search-Driven applications

- Health
- Education

*Humanity’s greatest advances are not in its discoveries – but in how those discoveries are applied to reduce inequity.*

Bill Gates

Humane Computing

“Is it right? Is it just?
Is it in the interest of mankind?”


Applications to Benefit Individuals & Society
New Challenge: chronic conditions: illnesses and impairments expected to last a year or more, limit what one can do and may require ongoing care.

In 2005, 133 million Americans lived with a chronic condition (up from 118 million in 1995).
Technology Trends

• Tremendous simplification in the technologies for *effortlessly* capturing useful personal information

• Dramatic reduction in the cost and form factor for personal storage

• Cloud Computing
Personal Health Analytics

Calorie Intake

02/27/2005 - 03/05/2005

Weight

02/27/2005 - 03/05/2005

Fitness Assessment on 07/09/2000

Client: Paul Moore

No of Prev Tests: 1
Last Test: 14 Jun 2000

VO2 - Bike Test Without Watts Meter

Load Kg: 3.0
Bpm: 50
Watts: 150

L/min
Aerobic: 3.58
mt/Kg/min
70%: N/A
90%: N/A

VO2: 44
Above Average VO2

Graph Options

@ Aerobic Capacity

Enter the Load in Kgs.

Options
Back
Personal Data Mining

Charts for appropriate demographics?

Optimum level for Asian Indians: 150 mg/dL (much lower than 200 mg/dL for Westerners) Due to elevated levels of lipoprotein(a)*

Distributed computation and selection across millions of nodes

Privacy and security

India’s Education System: 1951-2002

<table>
<thead>
<tr>
<th></th>
<th>1951</th>
<th>1981</th>
<th>2002(*)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Literacy Percentage</strong></td>
<td>18.33%</td>
<td>43.57%</td>
<td>65.38%</td>
</tr>
<tr>
<td><strong>Educational Institutions</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Primary</td>
<td>209671</td>
<td>494503</td>
<td>664041</td>
</tr>
<tr>
<td>Upper Primary</td>
<td>13596</td>
<td>118555</td>
<td>219626</td>
</tr>
<tr>
<td>High/Hr. Secondary &amp; Inter &amp; Pre Junior College</td>
<td>7416</td>
<td>51573</td>
<td>133492</td>
</tr>
<tr>
<td><strong>Enrollements (in millions)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Primary</td>
<td>19.2</td>
<td>73.8</td>
<td>113.9</td>
</tr>
<tr>
<td>Upper Primary</td>
<td>3.1</td>
<td>20.7</td>
<td>44.8</td>
</tr>
<tr>
<td>High/Hr. Secondary &amp; Inter &amp; Pre Junior College</td>
<td>1.5</td>
<td>11</td>
<td>30.5</td>
</tr>
<tr>
<td><strong>Dropout Rates (%)</strong></td>
<td>NA</td>
<td>82.5</td>
<td>66</td>
</tr>
<tr>
<td><strong>Teachers (in '000)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Primary</td>
<td>538</td>
<td>1363</td>
<td>1928</td>
</tr>
<tr>
<td>Upper Primary</td>
<td>86</td>
<td>669</td>
<td>1157</td>
</tr>
<tr>
<td>High/Hr. Secondary &amp; Inter &amp; Pre Junior College</td>
<td>127</td>
<td>926</td>
<td>1777</td>
</tr>
<tr>
<td><strong>Pupil Teacher Ratio</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Primary</td>
<td>20</td>
<td>38</td>
<td>43</td>
</tr>
<tr>
<td>Upper Primary</td>
<td>20</td>
<td>33</td>
<td>34</td>
</tr>
<tr>
<td>High/Hr. Secondary &amp; Inter &amp; Pre Junior College</td>
<td>21</td>
<td>27</td>
<td>34</td>
</tr>
<tr>
<td><strong>Public Expenditure (% of GDP)</strong></td>
<td>0.64%</td>
<td>2.92%</td>
<td>4.02%</td>
</tr>
</tbody>
</table>

Significant achievements, but problems remain …
Problems of School Level Education in India

**Poor performance**
- 39% dropouts in primary, additional 15.6% in secondary, additional 11.7% in higher secondary
- Pass out ratio is 50% at Class X and majority of them pass in 3rd division
- Less than 8% finish all schooling to qualify for a college education

**Poorly trained teachers**
- 51% of primary teachers are higher secondary or below
- Only 44% have received in-service training
- Absence of learning material for teachers to update their knowledge

**Poor teacher-student ratios**
- Ratio in primary is 1:43, secondary and Higher secondary is 1:34. About 9% of primary schools have a teacher-student ratio > 1:100.
- 1.4% of primary schools have no teachers, 19% have only one teacher for all classes.

**Poor quality of material**
- Poor quality of textbooks, out-dated curriculum

Source: IBM Report on Improving India’s Education System through Information Technology, 2005
Attacking Complex Problems*

**Framework**

1. Define goal
2. Find the highest-leverage approach
3. Discover the ideal technology for that approach
4. In the meantime, make the smartest application of the technology on-hand

**Application to Education**

1. Quality education to all
2. New pedagogy
3. Individualized learning with teacher as a discussant
4. Internet-based mass collaboration to help teachers teach better and improve the educational infrastructure

Education Web

• Participation of experts, teachers, parents and students in the development and revisions of curricula

• Sharing and collaborative development of lectures, assignments, tests, etc.

  • Tools for capturing feedback on textbooks (errors, better explanations, supplementary readings)

  • Collaborative translation and localization of educational material
Technical Challenges

• Tools for creating and updating content (Wikipedia++)
• Trust and authoritativeness of content
• Personalization of search to find the material suitable for one’s own style of teaching
• Bootstrapping and incentives
Outline

• Current state of affair
• Evolving Search
• Search Labs projects
Search Labs

Invent next in Internet search and applications

computational economics
machine learning
game theory
information retrieval
web mining
algorithms
query processing
privacy
data management
parallel mining
link analysis
NLP
ranking
inconsistent data

NLP
Invent next in Internet search
and applications

parallel mining
link analysis

inconsistent data

computational economics
machine learning
game theory
information retrieval
web mining
algorithms
query processing
privacy
data management
Search & Data: Virtuous Cycle

Better Search Results ➤
More Data ➤ Greater Insights ➤
Better Search Results

Web Pages Feeds
Classification Using Click Graph

Algorithm: Random walk with absorbing states
Automatic Generation of Training Data for Ranking

... query/document pairs labeled for relevance (P, G, B, etc.)

- Search results scanned from top to bottom
- On average, read everything above click, 1.4 below click [Cutrell, Guan]

Joachims et al
Assign nodes to k ordered classes s.t. weight of edges that agree with the classes minus weight of edges that do not is maximized
Result Diversification

• Large fraction of search queries are *ambiguous*
  – How to select a list of results to maximize the probability that the average user will find a relevant document?

• Key idea: when deciding whether to include a result, evaluate its *marginal value*, in the context of the results that have already been selected

• Analyze click logs for classifying queries and docs
Lincoln: Experimental Commerce Search Engine

Best car GPS around $300

- Structure and Semantics in Data and Queries
- Insights on user behavior from massive data mining
- From ranking to decision making
- Task-orientation

Category = “Auto GPS”
Price = approx(300)
Order By ReviewRank
**Symphony**

Enable non-developers to create and monetize custom search applications that combine their data and knowledge with Search services.

**Symphony Runtime**

1. Collect initial results
2. Add additional data
3. Generate HTML

**Semi-structured Query Processing**

- ShoeQueen Data
- Customer Data

**Query/Click Logger**

- Query, Click, App-ID

**Rev Sharing Component**

- Advertising
- Forum/review pages
- Images
- Ads

**Data Services**

- Web
- Image
- Advertising
- News
- Video
- Proprietary
- 3rd Party
Concluding Remarks

• Search is becoming an essential “utility”
• Need for innovation to take search to next level
• Data researchers can (and must) play a leading role
Thank you!

Questions

Search Labs’ mission is to invent next in Internet search and applications
Best Car GPS around $300

What to Know When Buying a GPS
May 19, 2005... While some units cost around $100 and may cost as much as $1000 or... I believe it is the best GPS, cabin rental, etc... And the best part, the...

Hello, Rakesh Agrawal. We have recommendations for you. (Not Rakesh?)

Your search "best car gps around $300" did not match any products.

"best car gps around $300" (See all 367 results)
Bhutan from Delhi in 4 Days

Google search results for "Bhutan from delhi in four days"

- Bhutan PM to visit India for four days
- Swine flu
- Michael Jackson's
- June 29th
- News cricket
- Alliance finance
- Hotel Taj Mahal
- Constitutional monarchy
- Dr. Manmohan Singh

Seen April 2009