Adaptive Search on a Web Scale

David Harper, Tech Lead Manager, Google
Outline

• **Purpose:** To share some things I have learnt working at Google that might have made my research more relevant and impactful.
• Review state-of-the-art in practice of adaptive web search
• What can be adapted in web search?
• Examine implications of web scale for research on adaptive search
• **Evaluation** of adaptive web search
About Me

- **Tech Lead Manager** at Google
- Previously academic and academic researcher
- Currently leading teams working on various search personalization projects
Selective Review of s-o-t-a

- Query Formulation
- Search ranking adaption: by geography, by personalization
- Search result adaption: specialized snippets, blended search results, host crowding, site maps
- UI adaption: user type (labs), user tasks (vertical), mobile search
Query Formulation

Google

[Google search results for "elephant"

Yahoo!

[Yahoo search results for "elephant" with suggested search terms: elephant bar, elephant man, elephant car insurance, elephant pictures]
Adaption to User?

- **Topic adaption**: diverse results as topic anchors, related searches
- **Task adaption**: task anchors, e.g. Home work on elephants, Find info on new film
Search Results (2)

http://dusty.com/search?v:file=viv_1032@32:yjRmRx&v:frame=list&v:state=(root)&root&v:N355&v:action=list&v:

Did you mean: garland

1. Tarland
   Information on the village of Tarland, Aberdeenshire, for visitors on the fiddle tradition of North East Scotland.

2. Tarland
   Tarland Woodland Trust: Poplar trees planted at Tarland, Aberdeenshire.

3. Paul
   In every rural community there are Pauls and Johnnies

4. Tarland
   Buying a house in Tarland

5. Tarland
   Par Sands

6. The
   Found in Aberdeenshire rural area.

7. Tarland
   Visit Scotland

8. Web
   Hunter

9. Tarland
   Visit Aberdeenshire

Results per page: 10
Number of clusters: 10
Result target: Current Window
Result display: Summary, Link Actions (opener, preview, show in clusters), URL, Cache, Sources
Adult filter: No

Tabs: web, news, images, wikipedia, blogs, jobs, shopping, more
Adaption to Context: Browsing

Google Video BETA

Recommended videos

- Palin: I'm Obama to Biden's McCain
- McCain: I Might Suspend Again
- John McCain Thinks Hugo Chavez is in the
- Südafghan auf Flughafen Zürich Kloten
- SVP'song
- Sparse and large-scale learning with het
Adaption to Context: Chatting

Duncan appears to be offline and will receive your messages after signing in. You can also send a message to Duncan’s mobile device.

Send an SMS Message (Ctrl+T)

You currently appear offline to Duncan.

David H: what games are being played at Murrayfield in the Six Nations?
David H: Ignore message … just doing an experiment
Keyboard shortcuts

Navigate search results quickly and easily, minimizing use of your mouse. Current keyboard shortcuts include:

Key  Action
---  ---
J    Selects the next result.
K    Selects the previous result.
O    Opens the selected result.
<Enter> Opens the selected result.
I    Puts the cursor in the search box.
<Esc> Removes the cursor from the search box.

Try out this query: rattlesnake
Adaption to Device/Task: Mobile Search

http://www.youtube.com/watch?v=JKxzX3p1iRs
Query Formulation

- Sources of query expansion
- Types of “expansion”: spell corrections, left and right extensions, phrases
- Diversity of expansions
- Navigating and selecting expansion suggestions
- When/how to surface expansions
- UI
Result Ranking

• User context
  ▪ Language
  ▪ Location
  ▪ ...

• User history (of interactions)
  ▪ Individual interactions
  ▪ Session history
  ▪ All history

• Histories of “Similar” users: aggregated data

• User histories ("wisdom of the crowds"): aggregated data


San Diego Zoo: Animals from Africa including elephants, lions and leopards ...
[www.sandiegozoo.com/](http://www.sandiegozoo.com/)

...
Result Display

- Blending results from different corpora;
  challenges:
  - Balancing relevance and diversity
  - Regular versus distinguished results
  - UI
- Specialized snippets
  - Query-biased
  - Action-biased
  - Property/vertical-biased
  - Answers in ...

**elephant**

**African elephants** live in **Africa**.
www.wikipedia.com/....

Elephant images

Map search: “Elephant and Castle”
... Getting to, Local travel

Film: Elephant Man ... Purchase ticket
Search UI

- Adaption in the Large
  - User language, e.g. Chinese
  - By search vertical
- Adaption to user type
  - 8-year old primary pupil
  - 20-year old University student
  - 38-year old car mechanic
  - 75-year old retired...
- Will this be the de facto standard search UI for web search?

www.wikipedia.com/....

San Diego Zoo: Animals from Africa including elephants, lions and leopards ...  
www.sandiegozoo.com/...

General: ....  
Geograpahy: ....  
Conservation: ...

....

....
(Search) UI - Adaption

• Level of Content

• Interaction mode: type, mouse, voice, gesture

• Interaction preference: search, browse, ...

• UI Complexity: prefer simplicity over complexity

• Type of adaption
  ▪ User selected and/or determined
  ▪ Adaption by selected action
  ▪ Automatic adaption

• User configurable UIs
Web user: there is no such person as a typical web user. They are distinguishable on many axes:

- Language
- Location
- Age group
- Educational level
- Job/task
- ...

There are huge opportunities for research on adaptive search that meet the needs of specific user groups.
Typical round-trip for query to results in web search is 250 msecs. Much of this is due to networking. Therefore, processing for purposes of adaption needs to be very, very fast.

- “On the fly” adaption needs to be:
  - Intrinsically fast (generally linear processes) and/or
  - Able to be parallelized and/or
  - Applied to small datasets and/or
  - Processed client-side

- Pre-compute slower adaptations and store/serve these fast

- Consideration of constraints on processing adaptive processes can result in (more) applicable research
Implications of Web – Logging user actions (1)

Logging individual data for individual adaption:

- Agreement to store, use, for how long, ...
- Must be protected from unauthorized use, and able to be display/modified by user
- Intrinsically harder to achieve user agreement for this!

Logging of accumulated user data

- Aggregate user interactions (not individual)
- Anonymized and protected from statistical attack
- Needs to be processed, stored and served efficiently
Implications of Web – Logging user actions (2)

• Adaption based on smaller “chunks” of user history
  ▪ Easier to satisfy above requirements re authorization, storage, etc
  ▪ Higher impact on users as more users will benefit
  ▪ Constraints can result in interesting research problems, e.g. Recommending Xs with limited click data (say)

• Adaption based accumulated user data
  ▪ Generic adaption: users who viewed this X ... also viewed these Xs (books, products, articles, videos, ...)
  ▪ Can be used for limited adaption to individual user
Evaluation of Adaptive Search – Challenges (1)

• Access to representative subsets of (web) users
  ▪ Stratified samples of query and/or session logs, e.g. informational, navigational, transactional query sets, by language, etc [very difficult]
  ▪ Access to subsets of actual web search users. e.g. “Open” experimental labs
  ▪ Constrain set of users by type and/or availability to you. Examples:
    • Piggy-back on some existing search service, or specialised service established for research/experimental purposes (e.g. IRF)
    • Client-side search adaption (and logging), but sharing data still difficult
    • Plug UI (adaption mechanism) into “open” search service
  ▪ Handling logs data appropriately is still an issue for researchers!
Tools and Services for Researchers

- Evaluation tools
- Logging tools, including dashboards to read/understand logs
- Services to store and share datasets, including results of experiments
- Standardized mark-up format for all above
- Note: probably all “in hand” but if not IR research community should find a way to support this.

Google has developed "Google Research Datasets", which will enable research datasets to be persistently stored and referenced, and made available across the web. These datasets must be open and public (although can be embargoed while publications go the press). Currently, this service is in closed beta testing.

For more information, please contact research-datasets@google.com.
Take Aways

• Research in adaptive (web) search should be informed by the state-of-the-art in both research and practice

• Adaptive search extends beyond the adaption of result ranking, and such extensions might have higher impact on user effectiveness and efficiency

• Interesting research problems will emerge through addressing the specific requirements of web scale (adaptive) search

• Web search covers a diverse range of user types, search services, kinds of search ... with consequent challenges in adaptive search

• Question: are the resources, tools and techniques used by the research community fit for purpose for research on adaptive search?