



# Adaptive Search on a Web Scale

David Harper, Tech Lead Manager, Google

- **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

- **Tech Lead Manager** at Google
- Previously academic and academic researcher
- Currently leading teams working on various search personalization projects

- 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




elephant		<a href="#">Advanced Search</a>
elephant man	4,820,000 results	<a href="#">Preferences</a>
elephant bar	612,000 results	<a href="#">Language Tools</a>
elephant.co.uk	57,300 results	
elephant and castle	758,000 results	
elephantitis	130,000 results	
elephant ears	505,000 results	<a href="#">Switzerland</a>
elephant in the room	13,800,000 results	
elephant walk	821,000 results	
elephant pharmacy	176,000 results	
elephant pictures	15,200,000 results	
	<a href="#">close</a>	

[Advertising F](#)

[Web](#) | [Images](#) | [Video](#) | [Local](#) | [Shopping](#) | [more](#) ▾

elephant



elephant bar	Explore related concepts: elephant "Alex Frost" elephant "African elephants" elephant "Gus Van Sant" Asian Elephant
elephant man	
elephant car insurance	
elephant pictures	
white elephant	

Search In:  the Web  pages from Switzerland

# Search Results (1)

**Google** elephant   [Advanced Search](#) [Preferences](#)


[Web](#) [Images](#) [News](#) [Video](#)

**Elephant** - Wikipedia, the free encyclopedia  
Elephants (family: Elephantidae) are large land mammals of the order Proboscidea . There are three living species: the African Bush Elephant, the African ...  
[en.wikipedia.org/wiki/Elephant](http://en.wikipedia.org/wiki/Elephant) - 175k - [Cached](#) - [Similar pages](#)


**Elephant (film)** - Wikipedia, the free encyclopedia  
Elephant is a 2003 crime-drama film written and directed by Gus Van Sant. It is set on the day of a massive school shooting. The film takes place a short ...  
[en.wikipedia.org/wiki/Elephant\\_\(film\)](http://en.wikipedia.org/wiki/Elephant_(film)) - 50k - [Cached](#) - [Similar pages](#)  
[More results from en.wikipedia.org »](#)

[cheap online car insurance and home insurance quotes - elephant.co.uk](#)  
Get cheap car insurance and home insurance quotes from [elephant.co.uk](#). It's easy , fast and you could save money! Get an online motor insurance quote.  
[www.elephant.co.uk/](http://www.elephant.co.uk/) - 17k - [Cached](#) - [Similar pages](#)

[Image results for elephant](#) - [Report images](#)





[News results for elephant](#)

 [Kenya's elephants send text messages to rangers](#) - 11 Oct 2008  
OL PEJETA, Kenya (AP) — The text message from the elephant flashed across Richard Lesowapi's screen: Kimani was heading for neighboring farms. ...  
[The Associated Press](#) - [314 related articles »](#)

[Metro Canada - Toronto](#) - [Fund managers told 'don't panic' over the elephant in the room](#) - [Financial Times](#) - [2 related articles »](#)

[Video results for elephant](#)

 [YouTube - Elephant - Trailer](#)  
2 min 3 sec  
[www.youtube.com](http://www.youtube.com)

 [Kat DeLuna featuring Elephant Man - Whine Up](#)  
3 min 52 sec  
[www.youtube.com](http://www.youtube.com)

Searches related to: **elephant**

<a href="#">african elephant</a>	<a href="#">elephant habitat</a>	<a href="#">elephant video</a>	<a href="#">elephant restaurant</a>
<a href="#">elephant gifts</a>	<a href="#">elephant jokes</a>	<a href="#">elephant info</a>	<a href="#">grizzly bear</a>

## 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)

Navigation icons: back, forward, refresh, home, star. Address bar: [http://clusty.com/search?v:file=viv\\_1032@32:yjRmRx&v:frame=list&v:state=\(root\)|root&id=N355&action=list&sv](http://clusty.com/search?v:file=viv_1032@32:yjRmRx&v:frame=list&v:state=(root)|root&id=N355&action=list&sv)

Clusty logo

web news images wikipedia blogs jobs more »

Search bar: tarland [Search] [advanced preferences](#)

clusters sources sites remix

All Results (173)

- Golf Club (33)
- Deeside, Royal (14)
- Photos (14)
- School (14)
- Cromar (10)
- Hotels (12)
- Property, Sale (13)
- Tarland And Migvie (7)
- Regional, Europe (7)
- Village in Aberdeenshire, Scotland (3)

[more](#) | [all clusters](#)

find in clusters:  [Find]

Top 173 results of at least 11,077 retrieved for the query **tarland** ([details](#))

Did you mean: [garland](#)

1. <a href="#">Tarland</a>	
Information on the village of Tarland, Aberdeenshire, for website on the fiddle tradition of North East Scotland <a href="#">www.</a>	
2. <a href="#">Tarl</a>	<b>* Results per page</b> Show how many search results? <input type="text" value="10"/> Default: 10
Tarla Popul <a href="#">en.wi</a>	<b>* Number of clusters</b> Start with how many clusters? <input type="text" value="10"/> Default: 10
3. <a href="#">Paul</a>	<b>* Result target</b> Open search results where? <input type="text" value="Current Window"/> Default: Current Window
In eve as we <a href="#">www.</a>	<b>Result display</b> Show what for each search result?
4. <a href="#">Tarl</a>	<input checked="" type="checkbox"/> Link Actions (opener, preview, show in clusters) <input checked="" type="checkbox"/> Summary <input checked="" type="checkbox"/> URL <input checked="" type="checkbox"/> Cache <input checked="" type="checkbox"/> Sources
Buyin <a href="#">www.</a>	
5. <a href="#">Tarl</a>	
Par/S <a href="#">www.</a>	
6. <a href="#">The</a>	<b>* Adult filter</b> Filter adult content? <input type="text" value="No"/> Default: No
Found rural l <a href="#">www.</a>	Note: the adult filter reduces, but may not completely eliminate explicit results (only applies to web tab)
7. <a href="#">Tarl</a>	<b>Tabs</b>
VisitS Scott golf.vi	<input checked="" type="checkbox"/> web <input checked="" type="checkbox"/> news <input checked="" type="checkbox"/> images <input checked="" type="checkbox"/> wikipedia <input checked="" type="checkbox"/> blogs <input checked="" type="checkbox"/> jobs <input type="checkbox"/> shopping <input type="checkbox"/> gov
8. <a href="#">Wel</a>	
Huntin maint <a href="#">www.</a>	
9. <a href="#">Tarl</a>	

Font size:

# Adaption to Context: Browsing



← → ↻ 🏠 ☆ <http://video.google.com/>

[Web](#) [Images](#) [Maps](#) [News](#) [Shopping](#) [Gmail](#) [more](#) ▼ | [New Features!](#) | [Web History](#)

**Google**™ Video BETA   [Advanced Video Search](#)  
[Preferences](#)  
[Moderate SafeSearch is on](#)







Search:  all videos  videos playable on Google

[Most blogged](#) [Most shared](#) [Most viewed](#) [Movers & Shakers](#)

[Uploaded Videos](#)

---

**Recommended videos** Page 1 of 5 ◀ ▶

 <p><a href="#">Palin: I'm Obama to Biden's McCain</a> 1 min - ★★★★★ <a href="#">youtube.com</a></p>	 <p><a href="#">McCain: I Might Suspend Again</a> 3 min - ★★★☆☆ <a href="#">youtube.com</a></p>	 <p><a href="#">John McCain Thinks Hugo Chavez is in the...</a> 34 sec - ★★★★★ <a href="#">youtube.com</a></p>	 <p><a href="#">Südanflug auf Flughafen Zürich Kloten</a> 61 min - ★★★☆☆ <a href="#">video.google.com</a></p>	 <p><a href="#">SVPsong</a> 4 min - ★★★★★ <a href="#">video.google.com</a></p>	 <p><a href="#">Sparse and large-scale learning with het...</a> 55 min - ★★★★★ <a href="#">video.google.com</a></p>
---	--	---	--	---	--

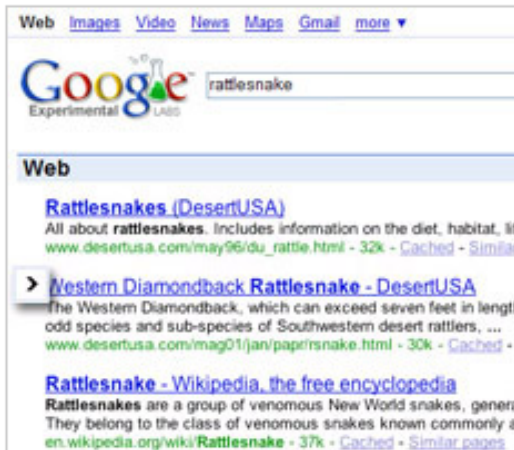


# Adaption to Context: Chatting

The screenshot shows a desktop environment with three overlapping windows:

- Instant Message Window:** Titled "Duncan - Instant Message". It shows a conversation with "Duncan". A status message indicates Duncan is offline. A message from "David H:" asks "what games are being played at Murrayfield in the the Six Nations?". The input field contains "David H: ingore message ... just doing an experiment".
- Web Browser Window:** Titled "Y!Q Search Results(what games are being ... the the six natio". The address bar shows "http://yq.search.yahoo.com/search?context=what+games+are-". The search results are displayed in a sidebar on the right.
- Search Results Window:** Titled "YAHOO! SEARCH". It shows "Web Results (1 - 3 of about 88,400)". The results list includes:
  - 1. [Official RBS 6 Nations Rugby : Venue - ...](#)  
A new-look Murrayfield was opened in 1994 by the Princess Royal, after a ... Team Statistics onlyinclude Five & Six...  
[www.rbs6nations.com/en/1584.php](#)
  - 2. [Official RBS 6 Nations Rugby : Home](#)  
... but Murrayfield next May could not be further away from coach Tony ... Team Statistics onlyinclude Five & Six...  
[www.rbs6nations.com](#)
  - 3. [Rbs six nations tournament Tickets : Find...](#)  
RBS Six Nations Tournament - Scotland VS Wales tickets in Edinburgh, EB at Murrayfield ... inventory, pictures, games and...  
[www.ticketsinventory.com/rugby/rbs-six...](#)Sponsored Results include:
  - [National Brand Cigar Outlet](#)  
Buy National Brand at Famous for 21% - 39% Off Retail Price. Shop Now.  
[www.Famous-Smoke.com/nationalbrand](#)

## 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.
/	Puts the cursor in the search box.
<Esc>	Removes the cursor from the search box.

Try out this query: [rattlesnake](#)

# Adaption – Search Vertical



[Web](#) [Images](#) [Video](#) [News](#) [Maps](#) [more »](#)

cj van rijnsbergen

Search

[Advanced Scholar Search](#)

[Scholar Preferences](#)

[Scholar Help](#)

**Scholar** All articles - [Recent articles](#)

Resu

book Information Retrieval

**CJ Van Rijnsbergen** - 1979 - Butterworth-Heinemann Newton, MA, USA

[Cited by 3804](#) - [Related articles](#) - [Web Search](#) - [Library Search](#) - [Find in RERO](#) - [All 4 versions](#)

book Information retrieval

**CJ Rijnsbergen** - 1975 - Butterworths London, United Kingdom

[Cited by 350](#) - [Related articles](#) - [Web Search](#)

[A non-classical logic for information retrieval](#)

**CJ Van Rijnsbergen** - The Computer Journal, 1986 - Br Computer Soc

A non-classical logic for information retrieval **CJ VAN RIJSBERGEN\*** Department of Computer Science, University College, Dublin 4 Implicit in many information retrieval models is a logic. These logics are hardly ever formalised. This ...

[Cited by 298](#) - [Related articles](#) - [Web Search](#) - [All 7 versions](#)

[A theoretical basis for the use of co-occurrence data in information retrieval](#)

**CJ van Rijnsbergen** - Journal of Documentation, 1977 - emeraldinsight.com

This paper provides a foundation for a practical way of improving the effectiveness of an automatic retrieval system. Its main concern is with the weighting of index terms as a device for increasing retrieval ...

[Cited by 251](#) - [Related articles](#) - [Web Search](#) - [All 2 versions](#)

[The Use of Hierarchic Clustering in Information Retrieval.](#)

**M. Lalmas, C. van Rijnsbergen** - Information Science and Statistics, 1974 - info.sci.umd.edu

[Probabilistic Models of Information Retrieval Based on Measuring the Divergence from Randomness](#)

**G AMATI, CJ VAN RIJSBERGEN** - ACM Transactions on Information Systems, 2002 - portal.acm.org

We introduce and create a framework for deriving probabilistic models of Information Retrieval. The models are nonparametric models of IR obtained in the language model approach. We derive term-weighting models by measuring the ...

[Cited by 147](#) - [Related articles](#) - [Web Search](#) - [BL Direct](#) - [All 7 versions](#)

[An evaluation of feedback in document retrieval using co-occurrence data](#)

**DJ Harper, CJ van Rijnsbergen** - Journal of Documentation, 1978 - emeraldinsight.com

... Title: AN EVALUATION OF FEEDBACK IN DOCUMENT RETRIEVAL USING CO-OCCURRENCE DATA

Author(s): DJ HARPER, CJ VAN RIJSBERGEN Journal: Journal of Documentation ISSN:

0022-0418 Year: 1978 Volume: 34 Issue: 3 Page: 189 - 216 DOI: 10.1108/eb026659 ...

[Cited by 126](#) - [Related articles](#) - [Web Search](#) - [All 4 versions](#)

Key authors: [C van Rijnsbergen...](#) - [C Rijnsbergen](#) - [F Crestani](#) - [N Jardine](#) - [M Lalmas](#)

<http://www.youtube.com/watch?v=JKxzX3p1iRs>

elephant

elephant

indian elephant

african elephant

elephant conservation

elephant man

elephant and castle

pink elephants

- 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

elephant

Wikipedia: African elephants.  
**African elephants** live in **Africa**.  
[www.wikipedia.com/....](#)

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

....

- 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

elephant

Wikipedia: African elephants.  
**African elephants** live in **Africa**.  
[www.wikipedia.com/...](http://www.wikipedia.com/...)

Elephant images

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

Film: Elephant Man ... Purchase  
ticket

- 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

Wikipedia: African elephants.  
**African elephants** live in **Africa**.  
[www.wikipedia.com/....](http://www.wikipedia.com/....)



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

...

 Wikipedia: African elephants.

**General:** ....

**Geographay:** ....

**Conservation:** ...

....

....

- 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?



- 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 msec. 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**

## 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

- 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

- 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](mailto:research-datasets@google.com).

- 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**?
-