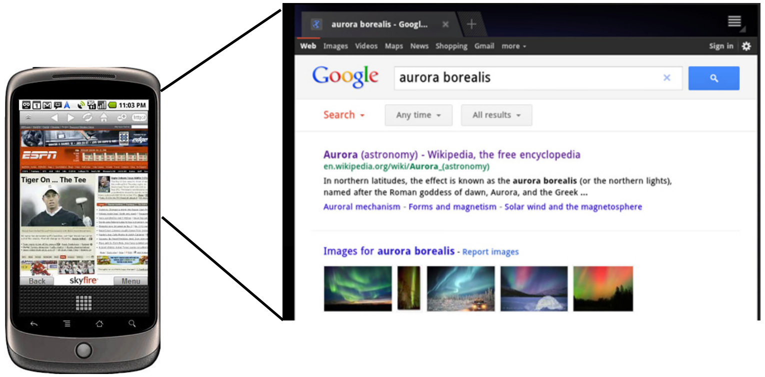
FindAll: A Local Search Engine for Mobile Phones

Aruna Balasubramanian, Niranjan Balasubramanian,
Sam Huston, Donald Metlzer, David Wetherall

Motivation: Mobile web search depends heavily on connectivity



Aurora (astronomy) - Wikipedia, the free encyclopedia
en.wikipedia.org/wiki/Aurora_(astronomy)
In northern latitudes, the effect is known as the **aurora borealis** (or the northern lights), named after the Roman goddess of dawn, Aurora, and the Greek ...
Auroral mechanism - Forms and magnetism - Solar wind and the magnetosphere

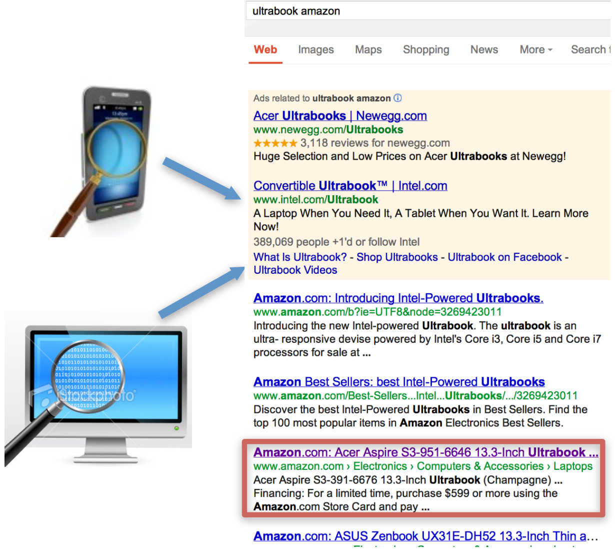
Images for **aurora borealis** - Report images

- But 3G connectivity is often poor (and expensive), or may not be available
  - Search is >10X slower than on desktop

Idea: Improve web search with better re-finding

Re-finding is very common

1) Visit a page on mobile (or desktop)



ultrabook amazon

Ads related to **ultrabook amazon**

Acer **Ultrabooks** | Newegg.com
www.newegg.com/**Ultrabooks**
★★★★★ 3,118 reviews for newegg.com
Huge Selection and Low Prices on Acer **Ultrabooks** at Newegg!

Convertible **Ultrabook**™ | Intel.com
www.intel.com/**Ultrabook**
A Laptop When You Need It, A Tablet When You Want It. Learn More Now!
389,069 people +1'd or follow Intel
What Is Ultrabook? - Shop Ultrabooks - Ultrabook on Facebook - Ultrabook Videos

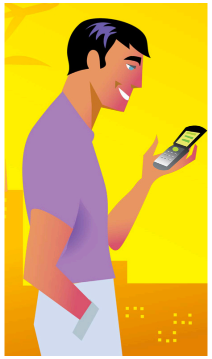
Amazon.com: Introducing Intel-Powered **Ultrabooks**.
www.**amazon**.com/b?ie=UTF8&node=3269423011
Introducing the new Intel-powered **Ultrabook**. The **ultrabook** is an ultra- responsive devise powered by Intel's Core i3, Core i5 and Core i7 processors for sale at ...

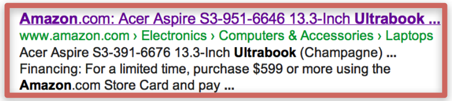
Amazon Best Sellers: best Intel-Powered **Ultrabooks**
www.**amazon**.com/Best-Sellers...Intel...**Ultrabooks**/.../3269423011
Discover the best Intel-Powered **Ultrabooks** in Best Sellers. Find the top 100 most popular items in **Amazon** Electronics Best Sellers.

Amazon.com: Acer Aspire S3-951-6646 13.3-Inch **Ultrabook ...**
www.amazon.com › Electronics › Computers & Accessories › Laptops
Acer Aspire S3-391-6676 13.3-Inch **Ultrabook** (Champagne) **...**
Financing: For a limited time, purchase $599 or more using the **Amazon**.com Store Card and pay **...**

Amazon.com: ASUS Zenbook UX31E-DH52 13.3-Inch Thin a...

2) Later, search for same page on mobile



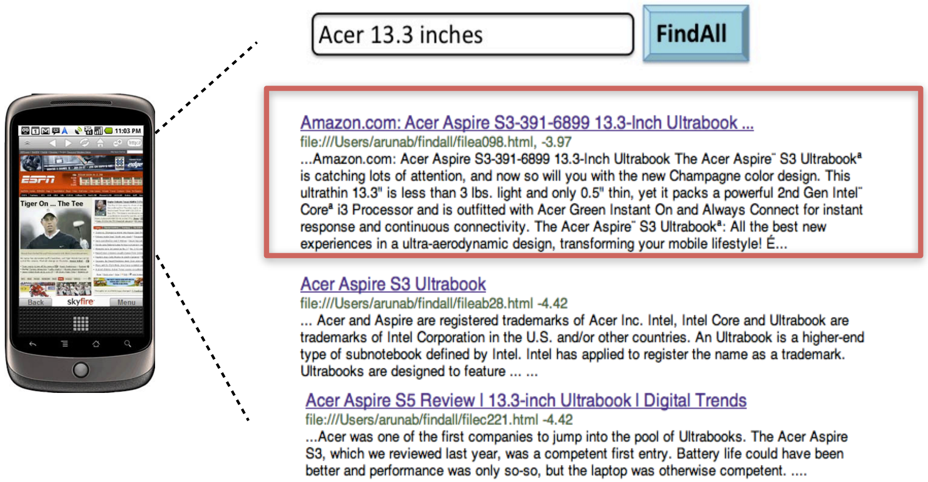


Amazon.com: Acer Aspire S3-951-6646 13.3-Inch **Ultrabook ...**
www.amazon.com › Electronics › Computers & Accessories › Laptops
Acer Aspire S3-391-6676 13.3-Inch **Ultrabook** (Champagne) **...**
Financing: For a limited time, purchase $599 or more using the **Amazon**.com Store Card and pay **...**

- Mobile: 70% of searches for 50% of users
- Non-Mobile: 40% to 60% of all searches

Our Solution: FindAll

- Local web search on mobile for re-finding with a Google/Bing-like search interface



Acer 13.3 inches

FindAll

Amazon.com: Acer Aspire S3-391-6899 13.3-Inch Ultrabook ...
file:///Users/arunab/findall/filea098.html, -3.97
...Amazon.com: Acer Aspire S3-391-6899 13.3-Inch Ultrabook The Acer Aspire¨ S3 Ultrabookª is catching lots of attention, and now so will you with the new Champagne color design. This ultrathin 13.3" is less than 3 lbs. light and only 0.5" thin, yet it packs a powerful 2nd Gen Intel¨ Coreª i3 Processor and is outfitted with Acer Green Instant On and Always Connect for instant response and continuous connectivity. The Acer Aspire¨ S3 Ultrabookª: All the best new experiences in a ultra-aerodynamic design, transforming your mobile lifestyle! É...

Acer Aspire S3 Ultrabook
file:///Users/arunab/findall/fileab28.html -4.42
... Acer and Aspire are registered trademarks of Acer Inc. Intel, Intel Core and Ultrabook are trademarks of Intel Corporation in the U.S. and/or other countries. An Ultrabook is a higher-end type of subnotebook defined by Intel. Intel has applied to register the name as a trademark. Ultrabooks are designed to feature ... ...

Acer Aspire S5 Review | 13.3-inch Ultrabook | Digital Trends
file:///Users/arunab/findall/filec221.html -4.42
...Acer was one of the first companies to jump into the pool of Ultrabooks. The Acer Aspire S3, which we reviewed last year, was a competent first entry. Battery life could have been better and performance was only so-so, but the laptop was otherwise competent. ....

Trades compute/storage for connectivity

- Will show lower search latency and improved availability with reduced energy cost

Why a local search engine?

- Search with indexes (ranking) is very effective compared to alternatives:
  - Browsing history is cumbersome
  - Keyword matching on cached pages is cheap and fast, but not very effective
  - Database of search queries misses query changes and non-searched pages

Challenge: search is memory/energy intensive

Talk & Contributions

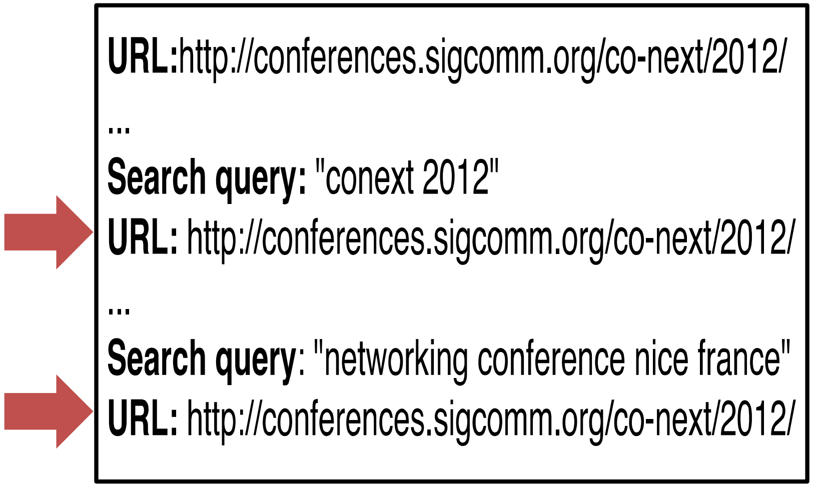
1. User study
   - Identifies re-finding behaviors
2. FindAll
   - Design of search engine for mobile re-finding
3. Evaluation
   - Results of the tradeoffs in practice

User study

- Monitored 23 participants for 1 month
  - Grad and under-grad students
- Collected logs from user’s mobile/desktop
  - Visited URL and search query (anonymized)
- Mark URL re-found if:
  - Page revisited via search query, and unchanged

Examples

Re–finding



URL:http://conferences.sigcomm.org/co-next/2012/

...

Search query: "conext 2012"

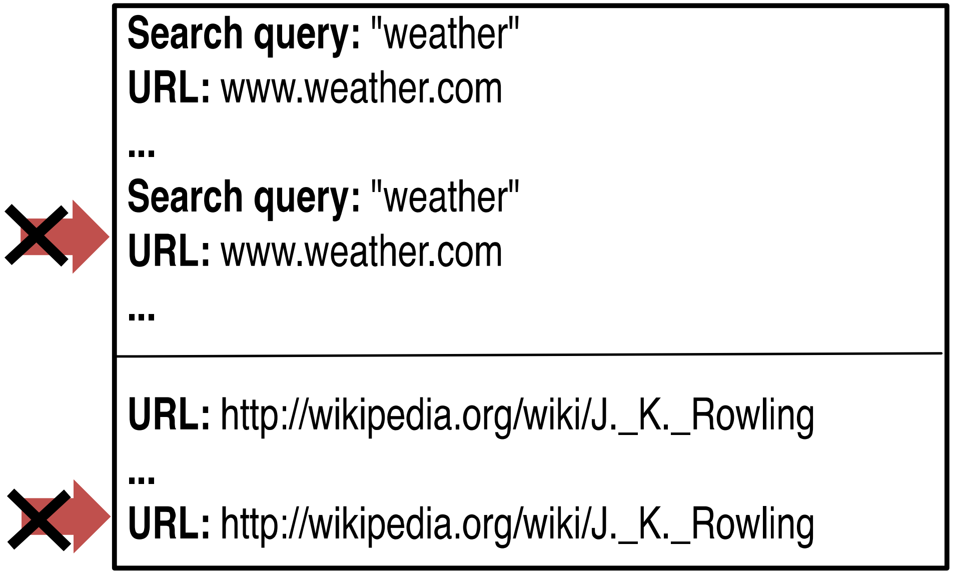
URL: http://conferences.sigcomm.org/co-next/2012/

...

Search query: "networking conference nice france"

URL: http://conferences.sigcomm.org/co-next/2012/

Not Re–finding



Search query: "weather"

URL: www.weather.com

...

Search query: "weather"

URL: www.weather.com

...

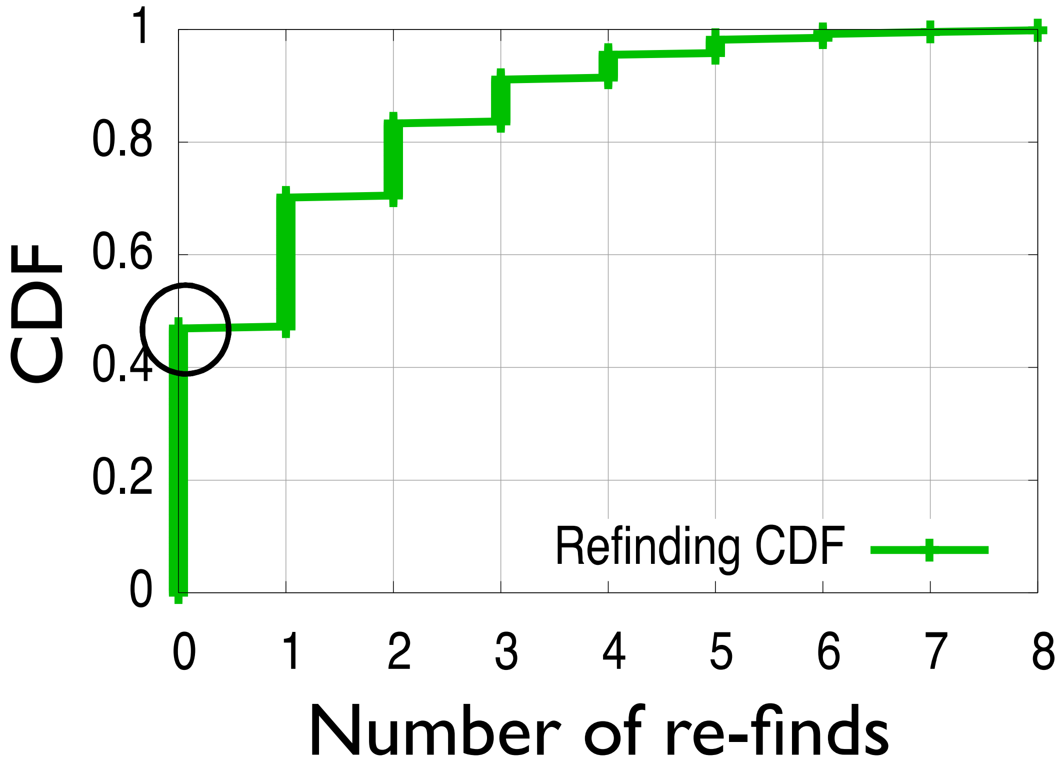
URL: http://wikipedia.org/wiki/J._K._Rowling

...

URL: http://wikipedia.org/wiki/J._K._Rowling

(Conservative rules, under–estimate re–finding)

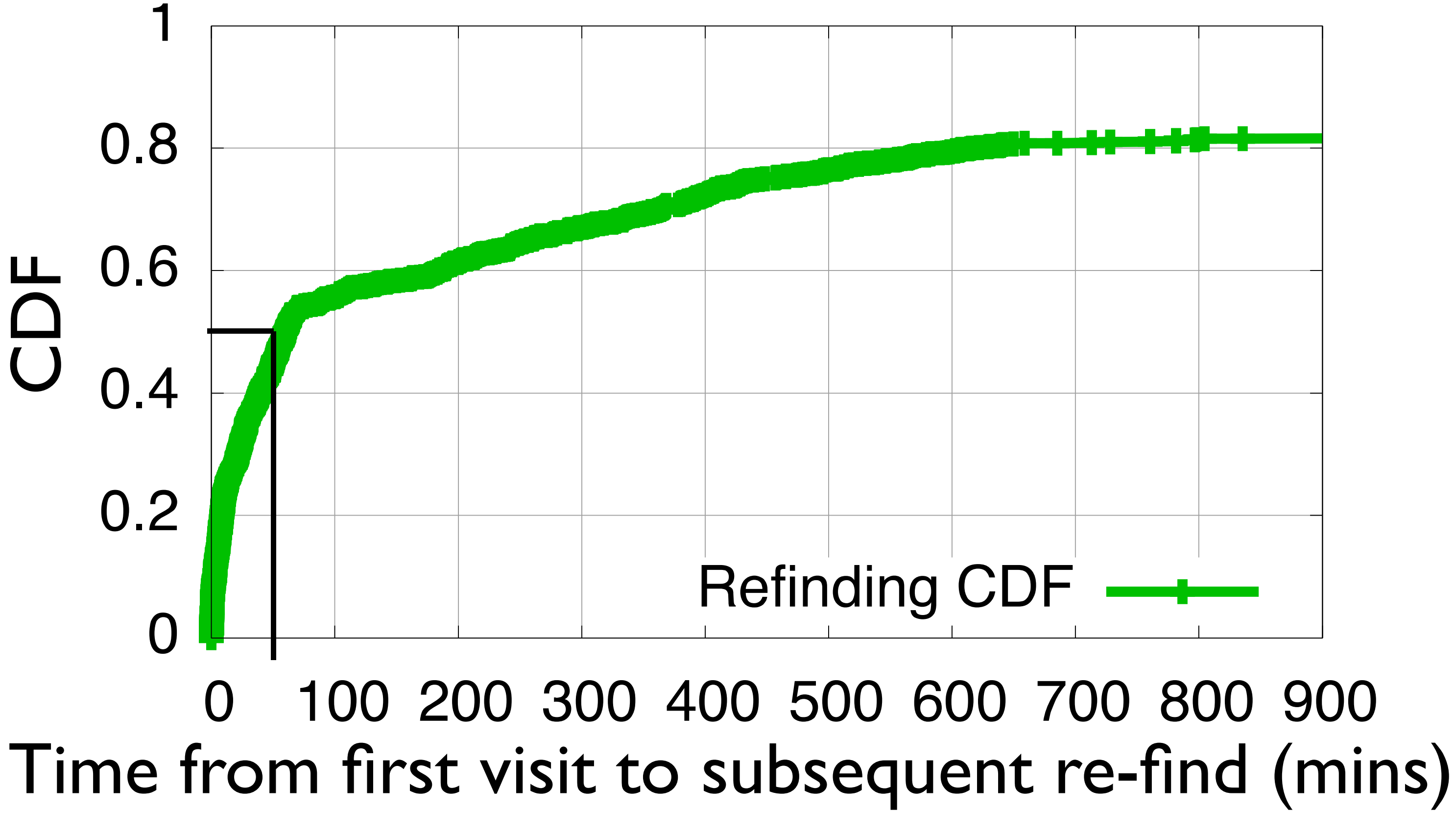
Most search is re-finding



>20% of re–finds have different query

Lots of opportunity for improvement!

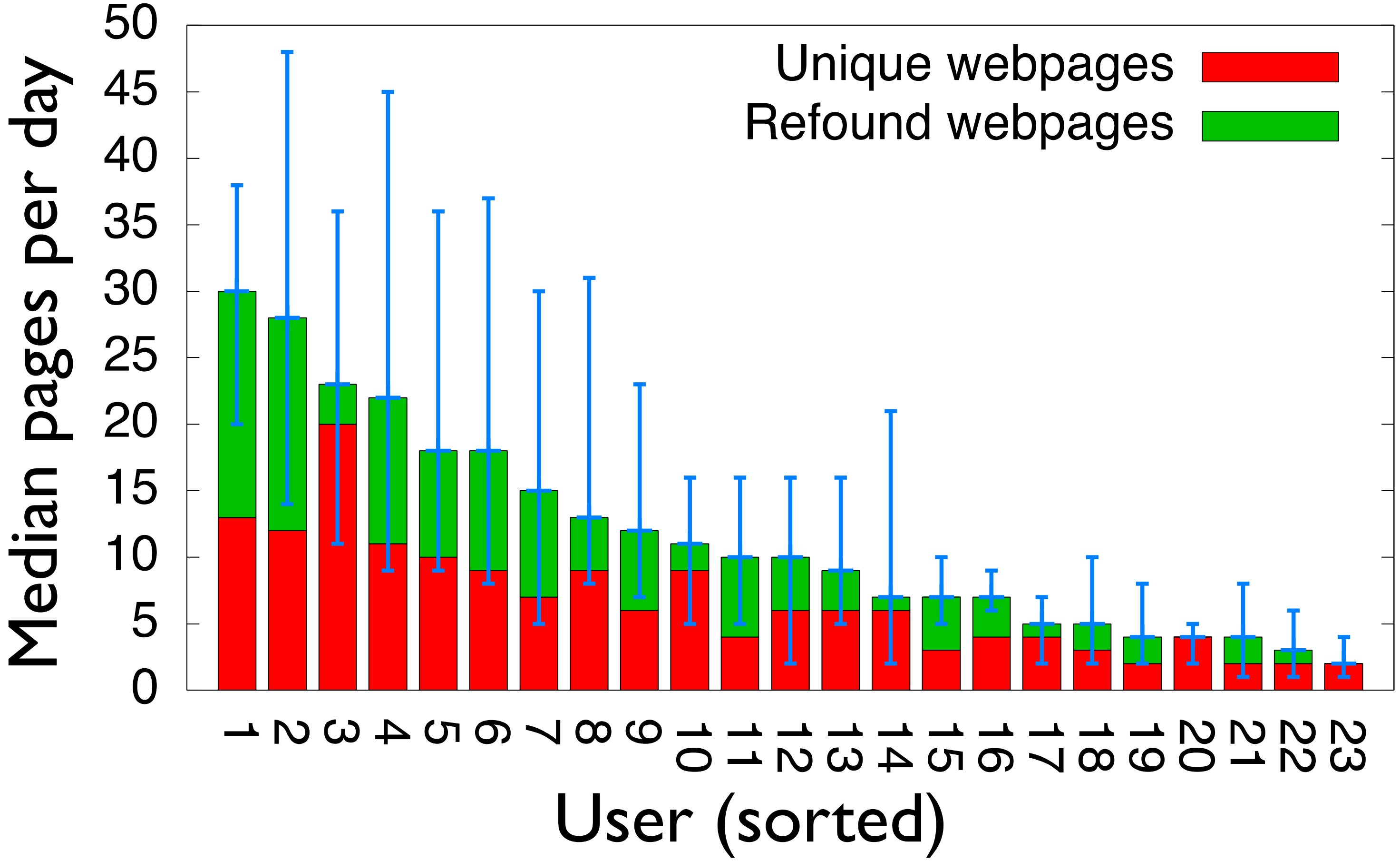
Most re-finds happen quickly



Need to index near when page is first accessed

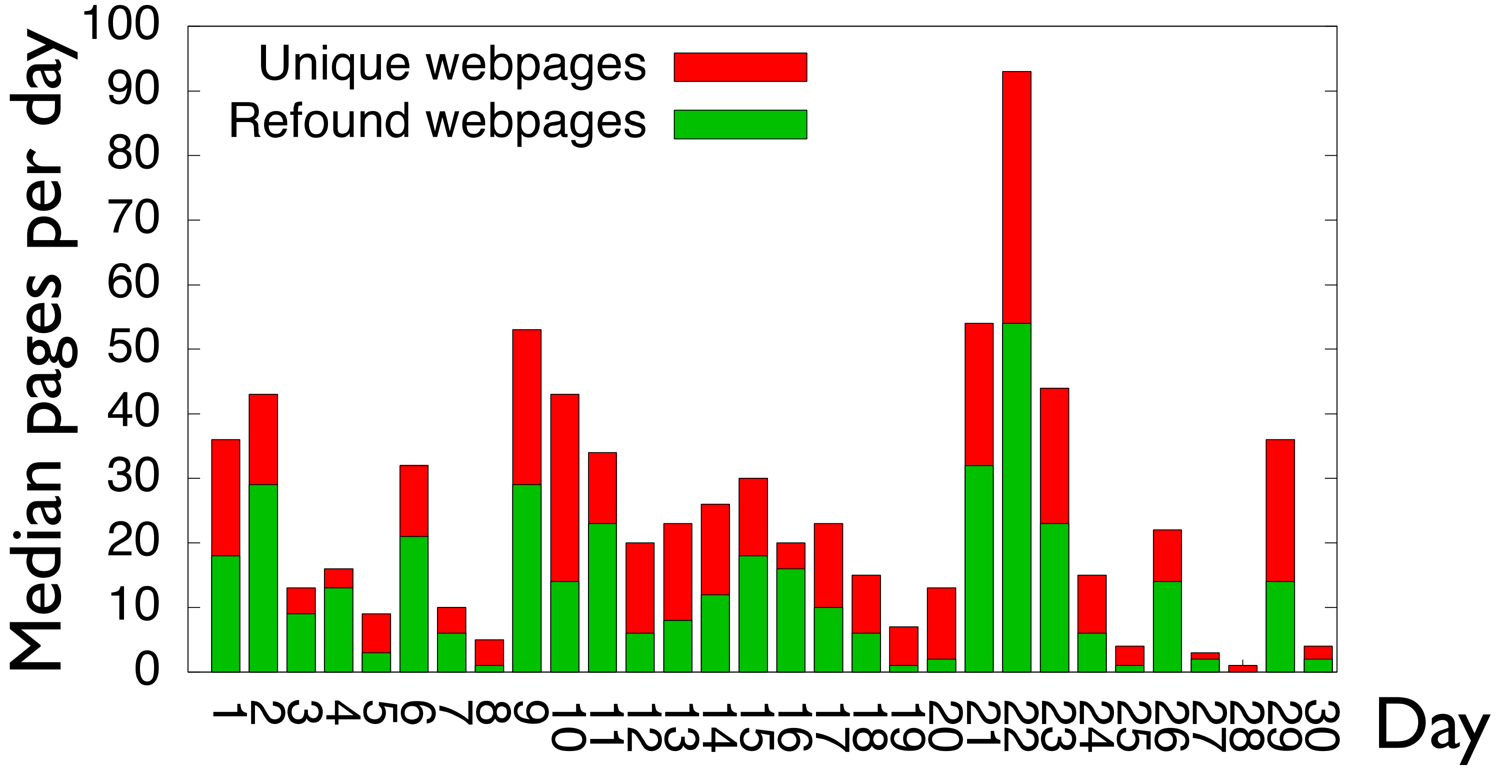
Diverse behavior across users

Varying numbers of pages, fraction of re-finding



Re-finding behavior is fairly consistent for a given user

Sample user with avg. 43% re-finding with 9% std dev.

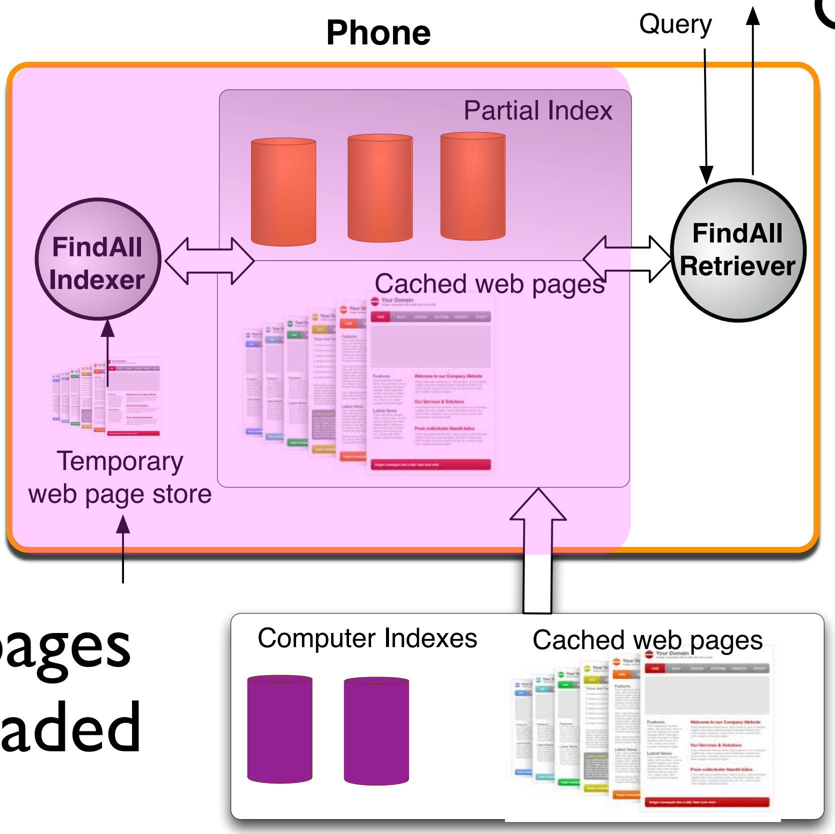


Can adapt to user behavior!

Outline

1. User study
    - Identifies key re-finding behaviors
2. FindAll
    - Design of search engine for mobile re-finding
3. Evaluation
    - Results of the tradeoffs in practice

FindAll architecture



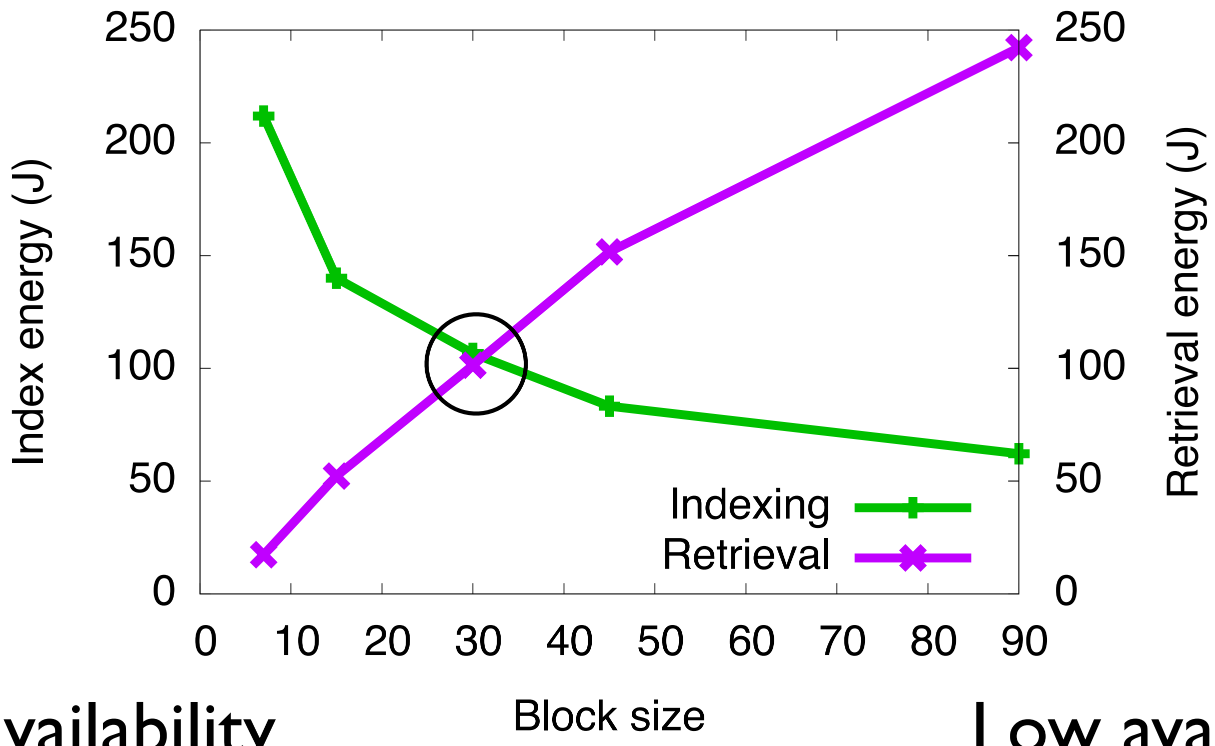
Output: ranked list of pages

Input: pages downloaded

[See paper for cross-device]

- Indexes as pages are browsed for availability
- Merges partial indexes to save energy

Key question: When to index?



High availability
High index energy

Low availability
High retrieval energy

Tradeoff depends on user's behavior

Indexing strategy

- Index to maximize availability with energy consumption no higher than default search
- Index when:

$$\underset{\text{Energy of indexing}}{E[I]} < \underset{\text{Energy of default search}}{E[\neg I]}$$

- Estimates need to be based on user behavior

Energy estimates [see paper]

- Energy of indexing, E[I]:
  - Cost of current block plus future penalty
- Energy of default search, E[¬I]:
  - Sum of page download x re-find prob. of page
- Train a classifier for re-find prob. of page
  - Depends on several user features

Prototype on Android OS

- Adapt Galago search engine for phones
  - Implement partial indexing
- Implement online energy cost estimator
  - Train classifier when mobile is charging
  - Make an indexing decision every 5 mins
- Runs on a Motorola DroidX mobile
  - Has WiFi and Verizon 3G
  - Energy from Monsoon power monitor

Outline

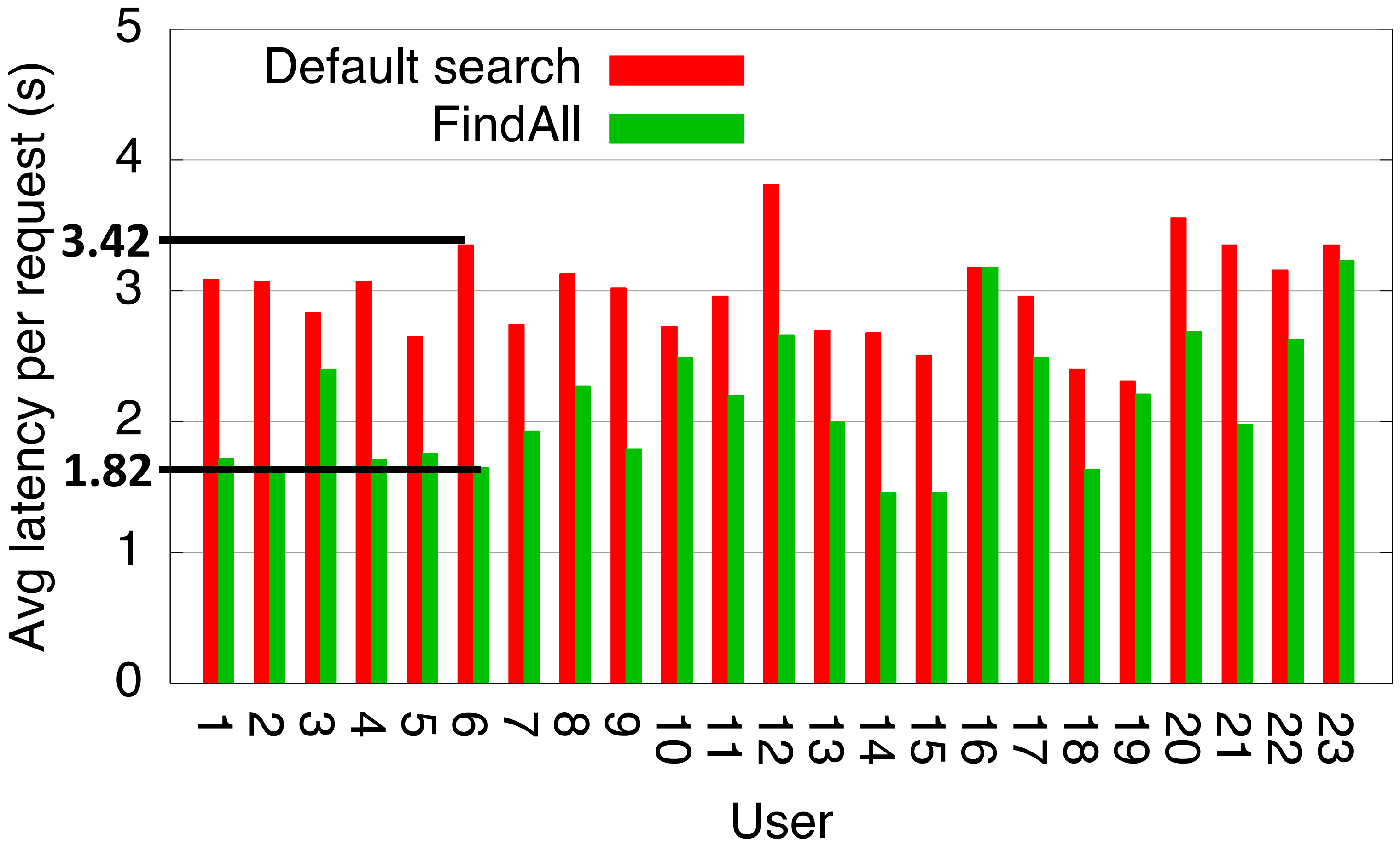
1. User study
   - Identifies key re-finding behaviors
2. FindAll
   - Design of search engine for mobile re-finding
3. Evaluation
   - Results of the tradeoffs in practice

Experiments

- Benefits and Costs
  - Latency, 3G data usage, availability
  - Energy, storage
- Alternate approaches
  - Cache keywords, query database
- Alternate indexing strategies
  - Cloud index, Always index, Fixed index

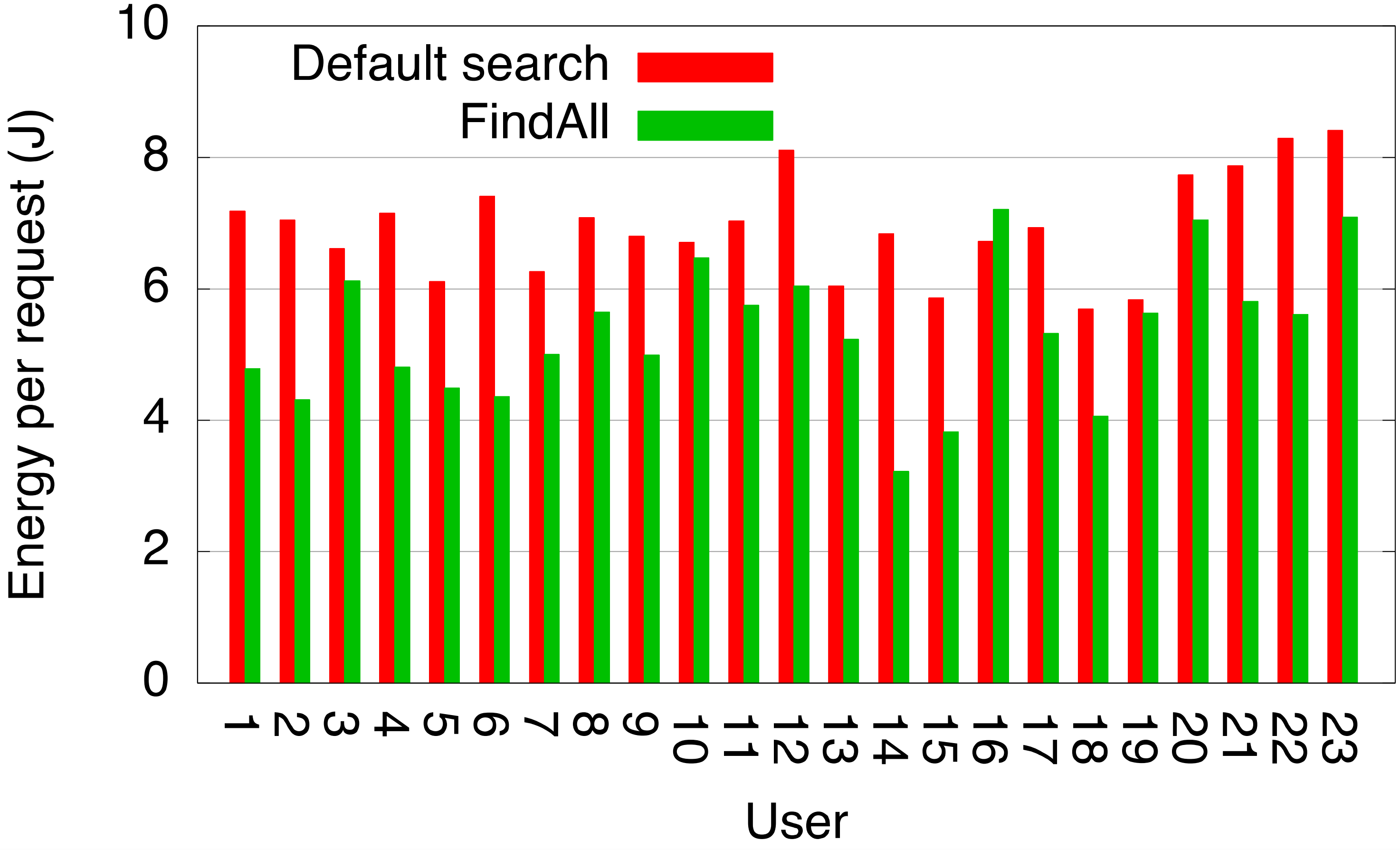
Results based on prototype and user trace

FindAll improves search latency



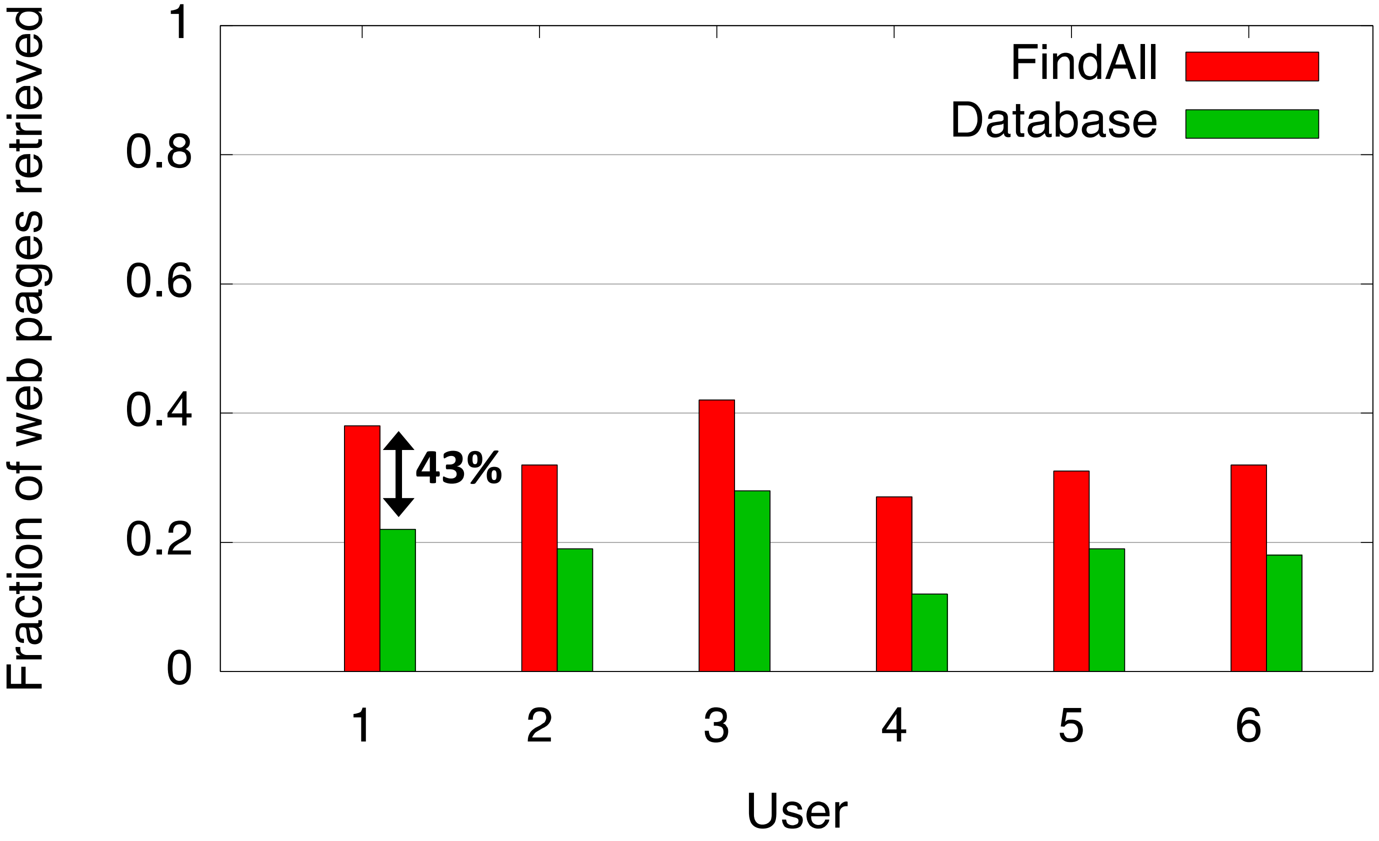
Average up to 2-fold better for high re-finders

FindAll reduces energy costs



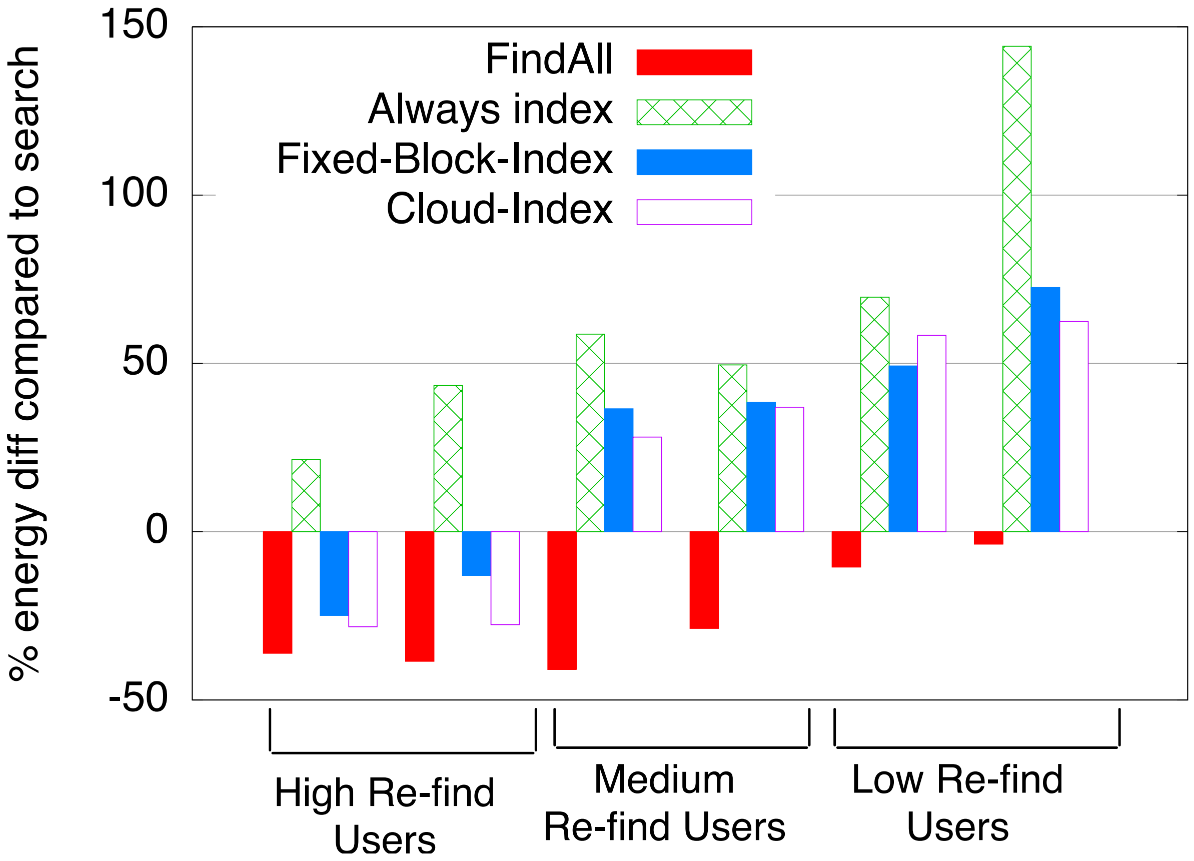
Energy up to 30% lower for high re-finders

FindAll improves availability when disconnected



(Under a random 50% connectivity model)

FindAll indexing strategy is important for energy benefits



Plus better availability than Cloud/Fixed

Conclusions

- FindAll makes a win-win tradeoff for search
  - Can decrease latency and increase availability, all with reduced energy and bandwidth
  - But need to adapt to user behavior
- Future directions
  - Integrating re-finding with mobile apps
  - Prediction across devices and users

Thank you. Questions?