

CrowdLogger as a Community Platform for Searcher Behavior Experiments

Henry Feild

Center for Intelligent Information Retrieval
University of Massachusetts Amherst

November 28, 2012



Things we like to do in IR

- Observe and model user behavior

Modeling and Measuring the Impact of Short and Long-Term Behavior on Search Personalization

Personalization of Search Results Using Interaction Behaviors in Search Sessions

Improving Searcher Models Using Mouse Cursor Activity

Search, Interrupted: Understanding and Predicting Search Task Continuation

User Evaluation of Query Quality

- Compare search algorithms / interfaces

- which do users prefer?

- time to completion

- which result in more/fewer clicks, etc.

Absence time and user engagement:
Evaluating Ranking Functions

Optimized Interleaving for Online
Retrieval Evaluation

What's currently done

- software:
 - make a toolbar from scratch
 - modify the Lemur Search Log Toolbar
- study:
 - recruit some users and conduct controlled lab study
 - install on campus computers, observe users in situ
 - well, in situ specifically in a library setting
- This is slow, expensive, and generally a lot of effort

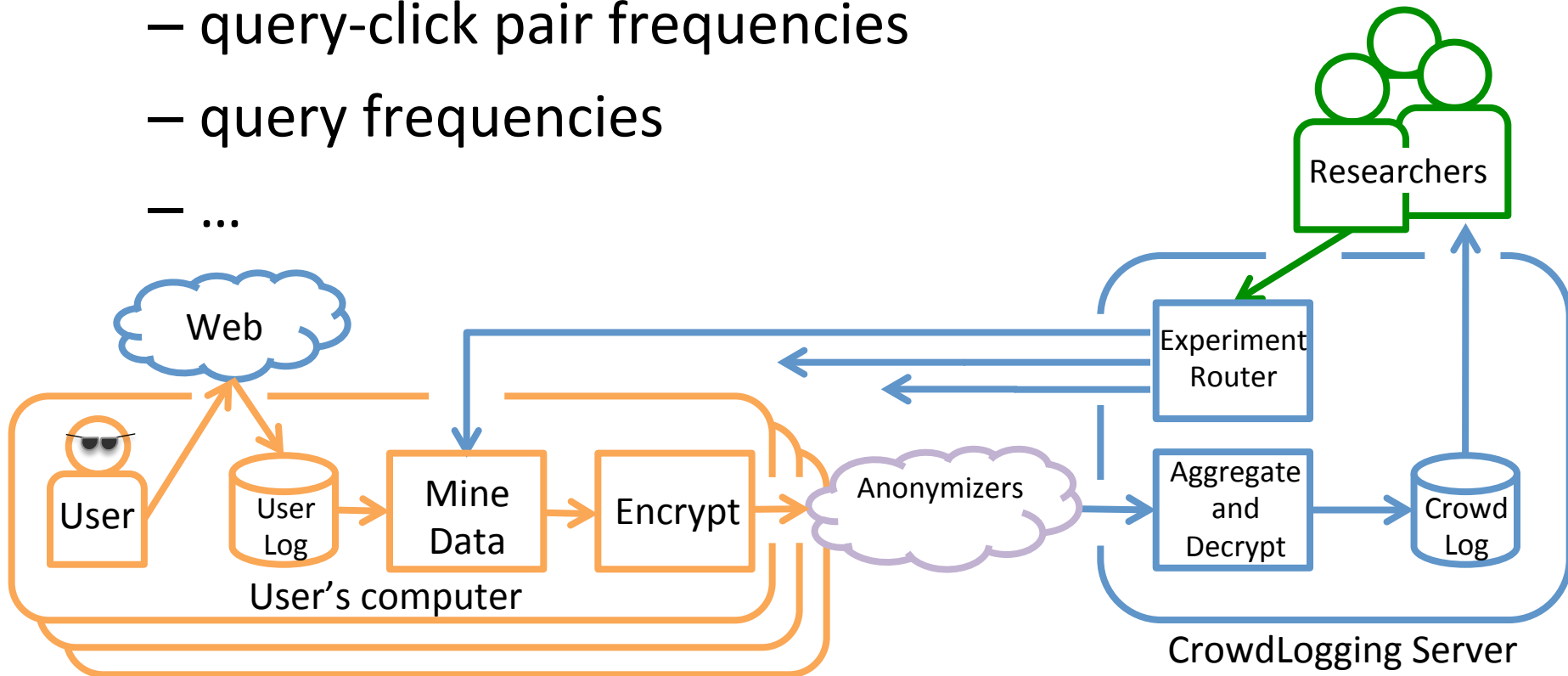


What we want

- a common, open source platform that deals with the basics
 - interaction data collection
 - data storage
 - privacy
- a common user base
 - can recruit some new users, but already have a significant pool of participants
- an interface for implementing novel studies



- Current a private data aggregation system
 - query reformulation pair frequencies
 - query-click pair frequencies
 - query frequencies
 - ...



Interactive experiments

System comparisons



Explicit feedback

Page Satisfaction Feedback

Page Evaluation

Please evaluate how well the page you just viewed addresses the current task, even if it is not the best page for the task.

http://en.wikipedia.org/wiki/Nixes_Mate

Pirates

Perfect -- the page completely satisfied my information needs
 Excellent
 Good
 Fair
 Bad -- the page did not satisfy my information needs
 This page does not exist, is not viewable, or is not working

Hint: Use the up and down arrows to cycle through the options.

BING IT ON WHICH GAVE YOU THE BEST RESULT FOR: "snow farmers almanac"

RESULTS ON LEFT DRAW RESULTS ON RIGHT

1

Snow | Farmers' Almanac
www.farmersalmanac.com/tag/snow/
The 2013 Farmers' Almanac. Watch Video · Farmers' ... Farmers' Almanac snow ...
Maori legends say snow was born from the marriage of the rain and the wind.

Snow | Farmers' Almanac
www.farmersalmanac.com/tag/snow
If you notice a hole in the upper left-hand corner of your Farmers' Almanac, don't return it to the store! That hole isn't a defect; it's a part of history.

chrome-extension://bepdacjnhiacjhbhbhhihmbpdpqmba/html/search_task_assistant.html

Task "portland airport shuttle" [edit] [delete task]

Saturday, August, 11, 2012

- portland airport to portland state university google @ 1:06 PM
- portland state university google @ 1:05 PM

Friday, August, 10, 2012

- portland restaurants google @ 4:13 PM
- portland restaurants - Google Search @ 4:13 PM
- portland airport to portland state university google @ 1:06 PM
- portland state university to Portland Marriott Downtown... google @ 4:12 PM
- Portland State University Press to Portland Marriott Downtown... @ 4:12 PM
- Portland Oregon Hotel | Portland Marriott Downtown... @ 4:12 PM
- Portland Oregon Hotel | Portland Marriott Downtown... @ 4:12 PM
- portland airport to portland state university google @ 4:11 PM
- portland airport shuttle google @ 4:10 PM
- portland airport shuttle - Google Search @ 4:10 PM
- Blue Star Transportation, Portland's largest local... @ 4:10 PM
- PDX Portland International Airport Shuttles @ 4:10 PM

2012 | Farmers' Almanac
8/29/2012-us-winter-forecast
Cause now it does not look like we will get the farmers almanac every ...

Forecast - Old Farmers' Almanac
winter-weather-forecast
their forecasts and snow conditions report for the Almanac.

Almanac 2012 Winter Weather
farmers-almanac-2012-winter-weather-forecast
s. Winter temperature and snow fall data. The Farmer's Almanac has been in ...

recasts gardening moon ...

ers' Market Directory; ... Your Daily Almanac way! 'Tis pie season.

2012 winter outlook: calls for a ...
ic-releases-2011-2012-winter-forecast
is released their outlook for this upcoming winter across the eastern half of the ...

Labeling

Challenges

- data management across multiple experiments
- an API that allows researchers sufficient control over accessing user data and implementing experiments
- controlling what data is shared with researchers
- incentivizing users to download the extension and participate in experiments

Data management

What to log

- browser interactions
 - add/remove/move tab
 - back/forward buttons
 - favorites/home/minimize/exit ...
- web page interactions
 - page un/loads
 - page focus
 - clicks
 - scrolls
 - mouse movements
- SERP interactions
 - query
 - top 10 results
 - urls, summaries, etc.
- complex interactions
 - opening links in new tabs
 - search tasks
 - study data

How to log it

Lemur Toolbar format

```
...  
SEARCH 1354022000 google wikipedia  
CLICK 1354023824 wikipedia.org google.com/search?  
q=wikipedia  
LOAD 1354023900 wikipedia.org  
...
```

A tab-delimited text file

CrowdLogger format

```
...  
{event: search, time: 1354022000, query: wikipedia, se: google}  
{event: click, time: 1354023824, destUrl: wikipedia.org,  
srcUrl: google.com/search?q=wikipedia}  
{event: load, time: 1354023900, url: wikipedia.org}  
...
```

JSON stored in IndexedDB

Data management

Benefits of JSON: easily extensible

```
{event: search, time: 1354022000, query: wikipedia, se: google}
```

```
{event: search, time: 1354022000, query: wikipedia, se: google, results: [{rank: 1, title: Wikipedia, url: wikipedia.org, snippet: "Wikipedia, the free encyclopedia that anyone can edit."}, {rank: 2,...}]}
```

Benefits of IndexedDB

- versioning built in
- entries can be updated in place
 - no need to re-write entire log file
- can build multiple indexes over data store
- HTML5 standard

API Categories

User Data



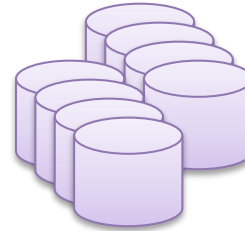
Historical data

- *get all clicks*
- *get all searches*

Real time data

- *on new search, do ...*

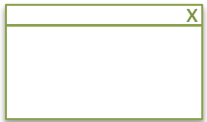
Aggregate User Data



Already collected data

- *get all query rewrites*
- *get all query-click pairs*

User Interface



Add to CrowdLogger interface

- *add widget to tools page*

Stand-alone windows/pages

- *present dialog when user searches*
- *modify search page ranking*

Uploading/Privacy



Encryption

Anonymization & aggregation

- *upload via anonymizers*
- *privately aggregate data*

Client-server communication



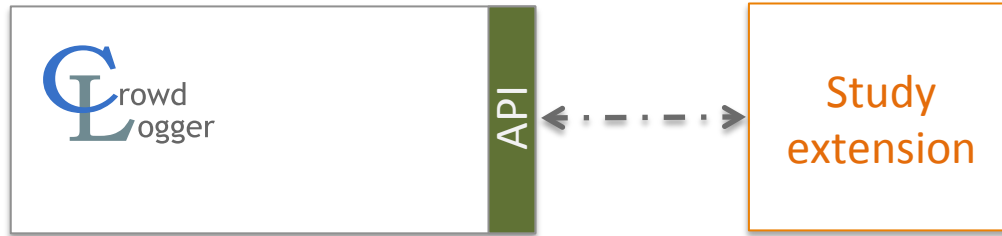
Request server-side computation

- *run retrieval algorithm for query*

Access server-side data

- *send me synonyms for ...*

API Layer Options



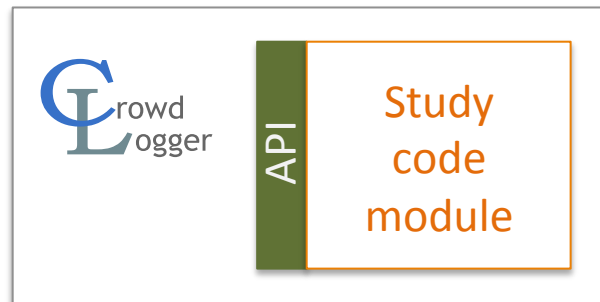
Communicate via inter-extension event calls

Pros:

- flexible study extension

Cons:

- limited communication
- no control over extension
- user has to download separate extension



CrowdLogger executes study code module

Pros:

- single module
- browser independent

Cons:

- remote JavaScript execution
- requires code approval
- potentially complex study code formulation
- less study code flexibility

Privacy controls

- what data get's shared with researchers?
- under what conditions?

What data is being collected
and how it will be used

What is minimally
useful to researchers

What users are
comfortable with

Query rewrites for public
release

Whatever users are
comfortable with

User 1: only if shared by
9+ other users ($k=10$)
User 2: $k=1$ rewrites

Feedback on retrieval system
preference for researcher use
only

$k=1$ anonymized
feedback from users

~~**User 1:** $k=5$ feedback~~
User 2: $k=1$ feedback

Privacy controls

What will be collected:

All search reformulations. For example, if you search for “blueberry pie” and then “blueberry pie recipes”, the pair:
“blueberry pie”, “blueberry pie recipes”
will be collected.

How the collected data will be used:

Reformulations will be anonymized and made **publically accessible** and used to, for example, generate search suggestions for you and other users.

Privacy settings:

For each search reformulation collected from you, select the anonymization level: the number of other users that must also share the same reformulation for it to be included in the final data set:

I have read the [consent](#) form and agree to participate in this study.

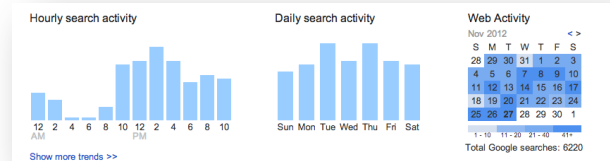
Cancel

Continue

Incentivization

Provide a service

- research prototypes
- visualizations
- re-finding tools
- citizen scientist



Google Search History

The screenshot shows a 'Current Tasks' list with items like 'portland airport to portland s...', 'portland state university', and 'House Centipede Information'. Below it are 'Related Tasks' such as 'clkm 2012 hotel' and 'flights from bos to maui'. At the bottom, there are 'Related Searches' like 'portland airport to portland s...' and 'portland state university'.

Search Task Assistant



Financial incentives

- gift cards
- virtual currency to 'buy' research apps

Gamification

- study-specific
- could also be a service



The screenshot shows a 'Google-a-day' game interface. It features a Google search bar at the top, a question about a 'Friend of J. K. Rowling' below it, and a 'View instructions' button. The game is part of an 'ESP Game' by Concentrix, with a score of 51M and a timer of 2:09. The game is titled 'EPS game' and has a 'Got it! Let's Play!' button.

Google-a-day

Service: Search Task Assistant

The image shows a screenshot of a Google search page. On the left side, there is a sidebar with three sections: 'Current Tasks', 'Related Tasks', and 'Related Searches'. The 'Current Tasks' section lists several search queries, including 'portland airport to portland s...', 'portland state university', 'House Centipede Information - ...', 'massachusetts spider identific...', and 'Common Brown House Spiders: Ho...'. The 'Related Tasks' section lists 'cikm 2012 hotel', 'flights from bos to maui', and 'NIKE, Inc.— Inspiration and In...'. The 'Related Searches' section lists 'portland airport to portland s...', 'portland state university', 'portland airport shuttle', 'portland restaurants', 'Portland Oregon Hotel | Portla...', and 'sigir events portland'. The main search area shows a search query 'portland airport to portland state university' with a search button and a search bar. Below the search bar, there are search results for 'Portland International Airport - Intensive English Language Program ...', 'Portland State University Computer Science Department | Home', and 'AAPT Summer Meeting 2010 - Portland State University'. Two red arrows point from the text annotations to the 'Current Tasks' and 'Related Tasks' sections of the sidebar.

As you search, your most recent searches are organized by task

Related tasks and searches are updated for each new search

Service: Search Task Assistant

The image shows a browser window with a Google search page. The search query is "portland airport to portland state university". The search results show about 2,470,000 results. The first result is a PDF titled "Transportation to Portland State University" from oia.pdx.edu. Other results include "Portland International Airport - Intensive English Language Program...", "Portland State University Computer Science Department | Home", "Amherst, MA Change location", "AAPT Summer Meeting 2010 - Portland State University", "Visiting Portland, Oregon - Portland State University", and "Homestay - Intensive English Language Program - Portland State University".

Search your history

portland

Matched tasks

- portland airport to portland s... (google)
- portland state university (google)
- portland state university (google)
- cikm 2012 hotel (google)
- cikm 2012 (google)
- sigir 2012 keynote (google)

Matched Searches

- portland airport to portland s... (google)
- portland state university (google)
- portland state university (google)
- sigir 2012 keynote (google)
- sigir events portland (google)
- portland restaurants (google)
- from: portland state universit... (maps.google)
- portland state university to P... (google)
- Portland Oregon Hotel | Portla... (google)
- portland airport to portland s... (google)
- portland airport shuttle (google)

Search About 2,470,000 results (0.22 seconds)

Web [\[PDF\] Transportation to Portland State University](#)
oia.pdx.edu/images/iss/Transport.pdf
File Format: PDF/Adobe Acrobat - [Quick View](#)
Transportation to **Portland State University**. **Airport** Arrival. Free transportation to **Portland State University**: Friends of Overseas Citizens and University. Students ...

Images

Maps

Videos

News

Shopping

More

Amherst, MA
Change location

Show search tools

Portland State University Computer Science Department | Home
[www.cs.pdx.edu/resources/directions](#)
The MAX light rail serves Eastern **Portland** (with stops at the convention center) ... The MAX light rail goes between the **airport** and downtown (2 zones) and runs ...

AAPT Summer Meeting 2010 - Portland State University
[aapt.physics.pdx.edu/](#)
Portland is hosting the American Association of Physics Teachers AAPT Summer ... Transportation and Parking Services · **Portland Airport** Directions · Find the ...

Visiting Portland, Oregon - Portland State University
[pdx.edu/family/visiting-portland-oregon](#)
Visiting **Portland**, Oregon ◇ Visting **Portland State University**. How to get here. **Portland International Airport** - Flying into the **Portland International Airport** is an ...

Homestay - Intensive English Language Program - Portland State University
[www.ielp.pdx.edu/housing_host.php](#)
Airport pickup at **Portland International Airport** (PDX). Support from ANDEO staff

Service: Search Task Assistant

chrome-extension://bepdacjnhiacjjhbahbhimbpdogpmba/html/search_task_assistant.html

Current Tasks

- portland airport to portland s... (google)
- portland state university (google)
- portland state university (google)
- House Centipede Information - ... American House Spider Informat...
- massachusetts spider identific... (google)
- massachusetts house spider (google)
- Common Brown House Spiders: Ho...
- Common House Spiders: How to

Related Tasks

- portland airport to portland state university (google) @ 1:06 PM
- portland state university (google) @ 1:05 PM
- Common Brown House Spiders: Ho...
- Common House Spiders: How to
- NIKE, Inc. - Inspiration and In...

Related Searches

- portland airport to portland s... (google)
- portland state university (google)
- portland state university to P... (google)
- portland airport shuttle (google)
- portland restaurants (google)
- Portland Oregon Hotel | Portla...
- sigir events portland (google)

Task "portland airport shuttle" [edit] [delete task]

Saturday, August 11, 2012

- portland airport to portland state university google @ 1:06 PM
- portland state university google @ 1:05 PM

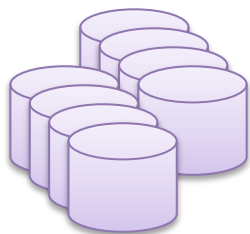
Friday, August 10, 2012

- portland restaurants google @ 4:13 PM
- portland restaurants - Google Search @ 4:13 PM
- portland state university to: Portland Marri... maps.google @ 4:12 PM
- portland state university to Portland Marriott Dow... google @ 4:12 PM
- Portland State University Press to Portland Marrio... @ 4:12 PM
- Portland Oregon Hotel | Portland Marriott Downtown... @ 4:12 PM
- Portland Oregon Hotel | Portland Marriott Downtown... @ 4:12 PM
- portland airport to portland state university google @ 4:11 PM
- portland airport shuttle google @ 4:10 PM
- portland airport shuttle - Google Search @ 4:10 PM
- Blue Star Transportation, Portland's largest local... @ 4:10 PM
- PDX Portland International Airport Shuttles @ 4:10 PM

Use case example

- evaluate a query recommendation algorithm
 - APIs:

Aggregate User Data



Extract query reformulations from users

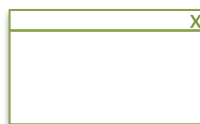
User Data



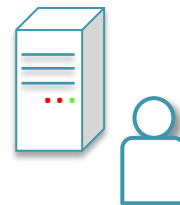
- When a user enters a new query, modify the search page to include a list of query suggestions

- Ask user to rate the suggestions

User Interface



Client-server communication



On each new query, contact a server to compute the recommendations

Uploading/Privacy



- k=5 for query reformulations

- k=1 for feedback

Use case example: Privacy consent

What will be collected:

Your feedback about the quality of search suggestions generated by our algorithm. In addition, you will have the option of providing the search you entered along with the feedback.

How the collected data will be used:

The feedback will be used only by the researchers in charge of this study (listed in the consent form).

Privacy settings:

For each set of feedback collected from you, select the anonymization level: the number of other users that must share the same feedback for it to be included in the final data set (max to participate: 1):

1



I have read the [consent](#) form and agree to participate in this study.

Cancel

Continue

Use case example: code snippet

```
api.userData.onQuery( function(data){
    var suggestions = getSuggestions(data.query);
    api.ui.addOverlay(queryData.page, showSuggestions);
});
```

```
function getSuggestions(query){
    return api.clientServer.callServer(
        QUERY_SUGG_SERVER_URL, query);
}
```

```
function showSuggestions(page, suggestions){
    var overlay = page.jquery("<div>");
    ...
    button.click(onFeedbackSubmitted);
}
```

```
function onFeedbackSubmitted(data){
    api.upload.private(data.toString, {k: 1});
}
```

Use case example: interface

The image shows a screenshot of a web browser displaying a Bing search for "laptops". The search results page includes a search bar with the query "laptops", a search button, and a results count of 777,000,000. Below the search bar, there are several search results, including a shopping section with product images and prices, and several text-based search results from Best Buy, Overstock.com, Amazon.com, and ZDNet. A "Suggestions" dialog box is overlaid on the right side of the page, featuring a "Useful?" poll with "Yes" and "No" radio buttons, a list of suggestions (MacBook Air, MacBook Pro, Dell laptops, Windows 8 Laptops), an "Include query ('laptops')" checkbox, and a "Continue" button.

laptops - Bing






www.bing.com/search?q=laptops&go=&qsn=&form=QBLH&pq=laptops&sc=8-7&sp=-1&sk=

WEB IMAGES VIDEOS MAPS NEWS SHOPPING MORE

bing laptops

777,000,000 RESULTS

[Shop for laptops](#)
bing.com/shopping

				
HP ProBook VM814AV 15.5 inch Laptop \$116.62	Apple iPad 2 Wi-Fi + 3G \$317.99	Toshiba Satellite P7700 \$229.95	Dell Passion Purple 10.1 inch Laptop \$149.95	Sony VAIO AW Series 15.5 inch Laptop \$127.94

[Laptops & Netbook Computers: PC, Apple MacBook - Best Buy](#)
www.bestbuy.com/site/Computers-PCs/Laptop-Computers/abcat0502000.c?...
All Laptops · PC Laptops · HP · Refurbished Laptops · Toshiba
Free shipping on Apple's, PC laptops and MacBook computers at BestBuy.com. See the latest deals, compare specs & reviews, and find the one that's right for you.

[Laptops | Overstock.com: Notebook Computers and Netbooks](#)
www.overstock.com/Electronics/Laptops/133/subcat.html
When you need your cheap laptop fast, Overstock.com is the place to shop; our expedited shipping is more affordable than ever.

[Amazon.com Laptops, Netbooks, & Accessories: Laptop ...](#)
www.amazon.com/Notebooks-Laptop-Computers/b?ie=UTF8&node=565108
Amazon.com offers you the largest selection of laptops from you favorite brands. Shop Amazon.com for laptops, ultrabooks, chromebooks & more

[Best Buy Cyber Monday 2012 deals on tablets, laptops, desktop ...](#)
ZDNet · 2 days ago
Nov 26, 2012 · Best Buy Cyber Monday 2012 deals on tablets, laptops, desktop computers. Summary: Didn't buy a new PC on Black Friday? You can get sales on a ...

[Laptops | HP® Official Store](#)

rowd ogger Suggestions X

Useful?
Yes No

MacBook Air

MacBook Pro

Dell laptops

Windows 8 Laptops

Include query ("laptops")

Continue

Other challenges

- cross-device synchronization
- mobile device support
 - requires browsers to allow extensions on mobile platforms
- neutral code review panel