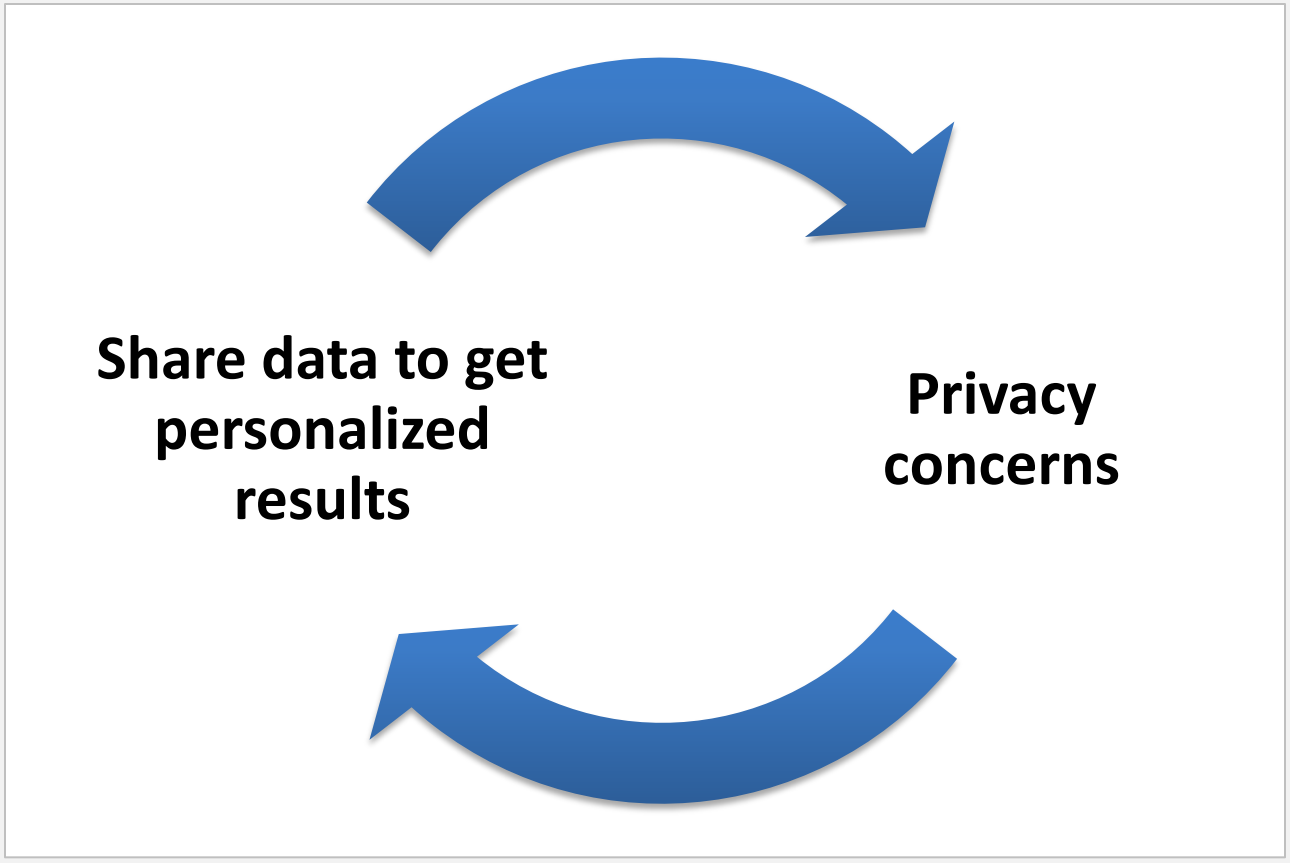
RePriv: Re-Envisioning In-Browser Privacy

Matt Fredrikson
University of Wisconsin

Ben Livshits
Microsoft Research

Microsoft Research


Share data to get personalized results
Privacy concerns

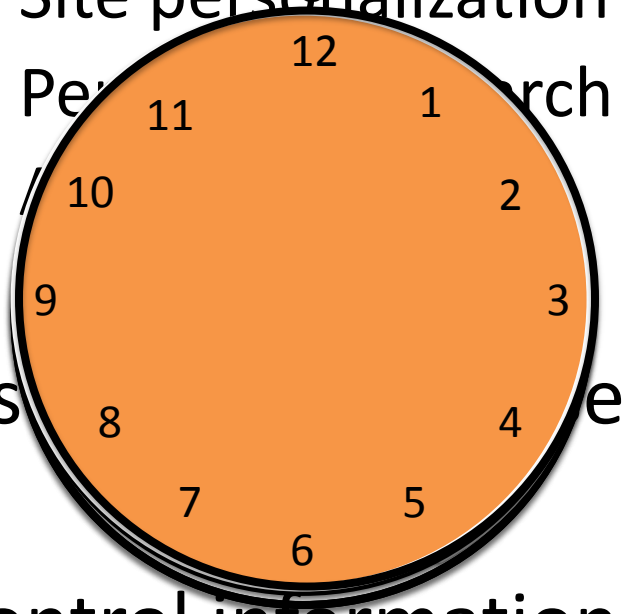
Browser: Personalization & Privacy

Browsing history

Distill

User interest profile

- Broad applications:
  - Site personalization
  - Pe… …rch
  - … …
- Us… …er
- Control information release



bn.com would like to learn your top interests. We will let them know you are interested in:

- Science
- Technology
- Outdoors

Accept Decline

RePriv Protocol

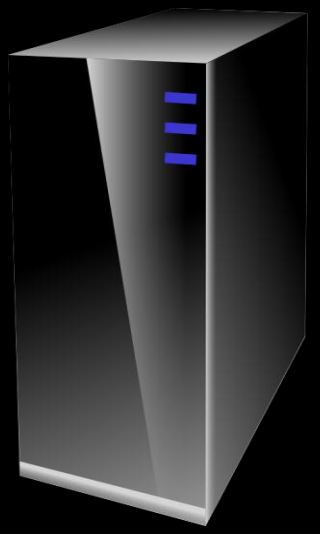
GET /index.html HTTP 1.1
Host: www.example.com
Accept: repriv …

HTTP/1.1 **300 Multiple Choices**
index.html
index.html?top-n&level=m

POST /index.html HTTP 1.1
Host: www.example.com
Content-Length: *x*
category1=c1&…

HTTP/1.1 200 OK

Personalized page content



Would you like to install an extension called “Bing Personalizer” that will:

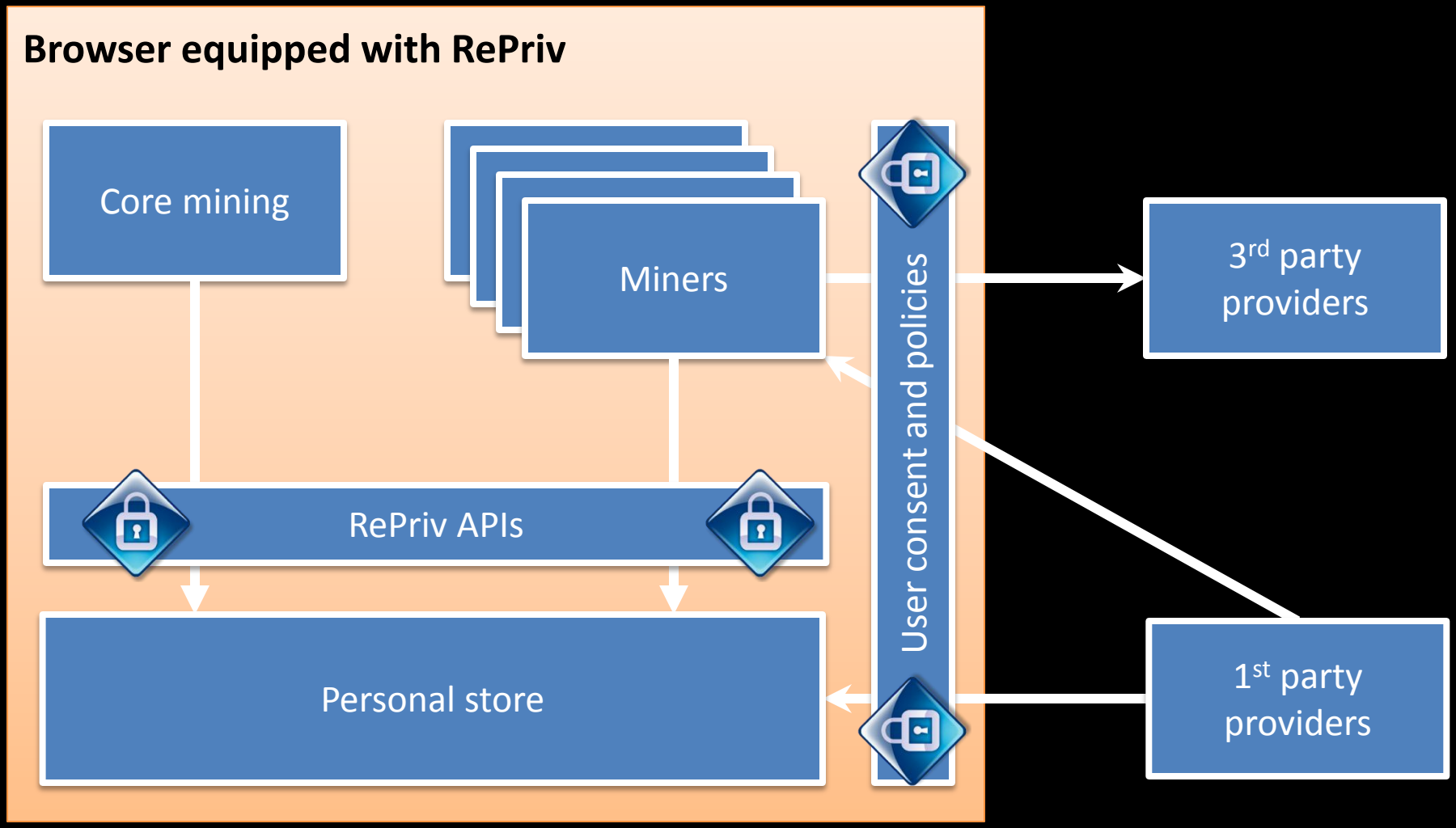
- Watch mouse clicks on bing.com
- Modify appearance of bing.com
- Store personal data in browser

Accept Decline

Contributions of RePriv

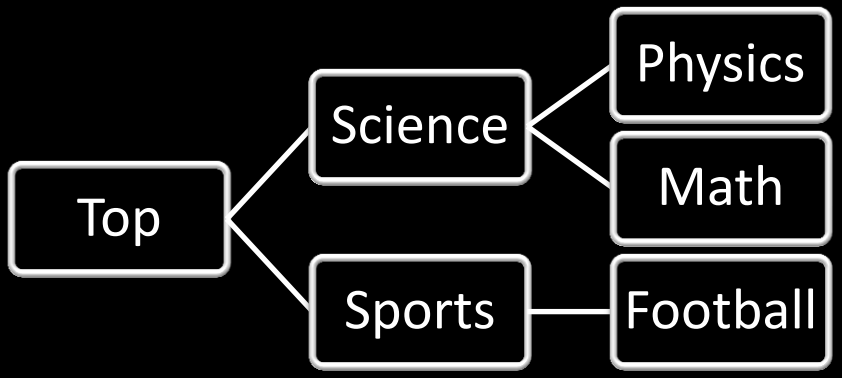
RePriv	• An in-browser framework for collecting & managing personal data to facilitate personalization.
Core Behavior Mining	• Efficient in-browser behavior mining & controlled dissemination of personal data.
RePriv miners	• A framework for integrating verified third-party code into the behavior mining & dissemination of RePriv.
Real-world Evaluation	• Evaluation of above mechanisms on real browsing histories & two in-depth case studies.

RePriv Architecture



Core Mining

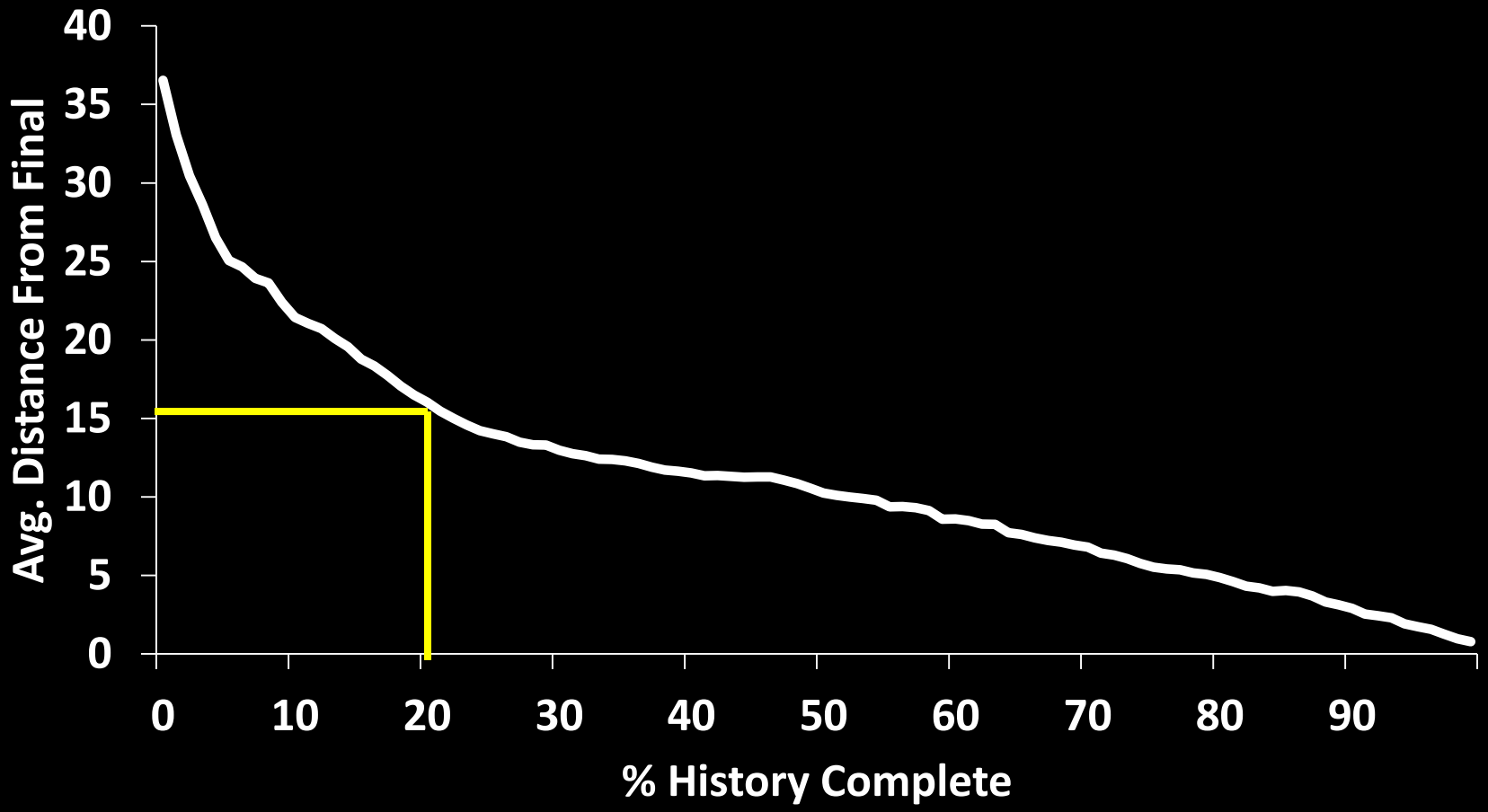
- Taxonomy from first two levels of ODP taxonomy
  - ~450 categories total
  - 20 top-level categories
  - Overlap exists



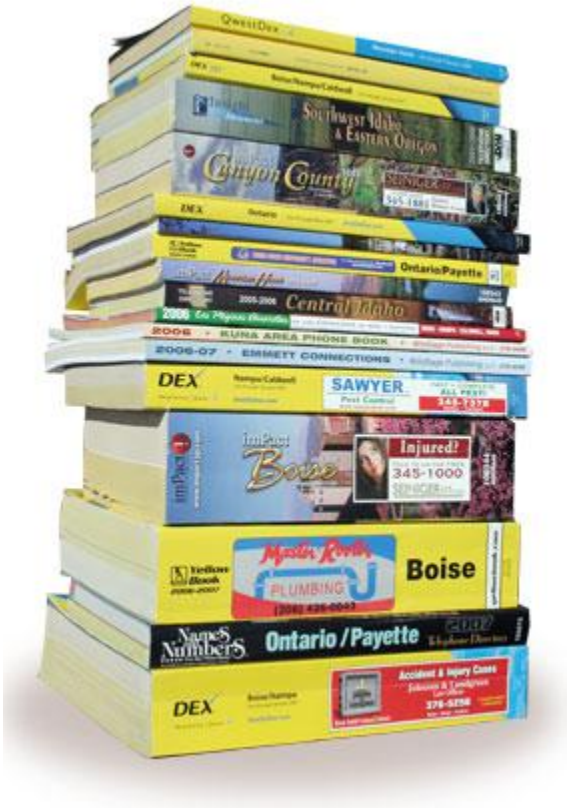
- Naïve Bayes
  - All categories equally likely
  - Training: min(3000, # pages) sites per category
  - Attribute words occur in at least 15% of docs for ≥1 category

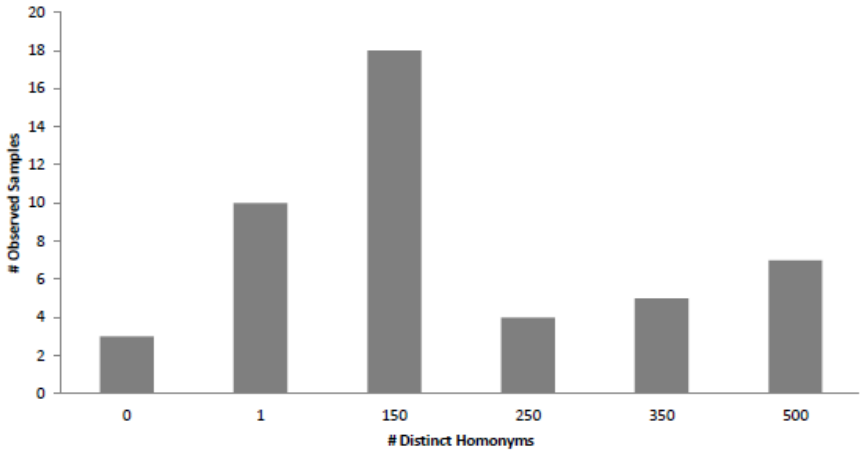
- Classification is fast enough: O(c•n)
  - n is # words in document
  - c is # document categories

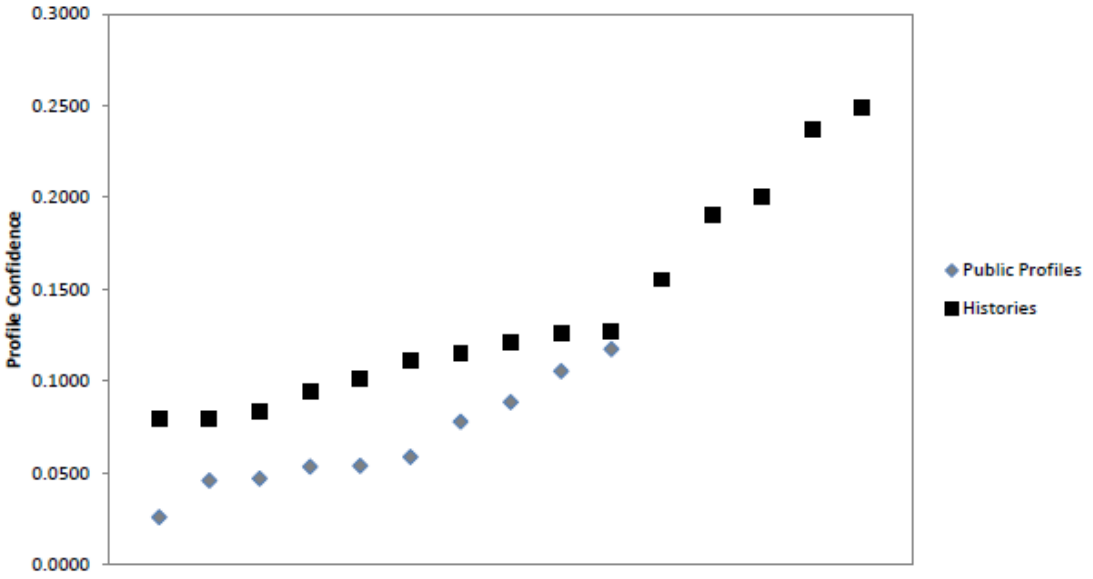
Global Mining Convergence



RePriv vs. the White Pages







Source: WebMii

RePriv

- An in-browser framework for collecting & managing personal data to facilitate personalization.

Core Behavior Mining

- Efficient in-browser behavior mining & controlled dissemination of personal data.

RePriv miners

- A framework for integrating verified third-party code into the behavior mining & dissemination of RePriv.

Real-world Evaluation

- Evaluation of above mechanisms on real browsing histories & two in-depth case studies.

Miner Name	C# LoC	Fine LoC	Verif. Time
TwitterMiner	89	36	6.4
BingMiner	78	35	6.8
NetflixMiner	112	110	7.7
GlueMiner	213	101	9.5

```
assume ExtensionId "twitterminer"
assume CanCommunicateXHR "twitter.com“ Nil
assume  CanUpdateStore("twitter.com“ “twitterminer”)
```

Netflix Example

- [illegible]

```
let doGetMovies genre cdom =
  …
  let flixEnts = GetStoreEntriesByTopic
                 myprov "movie" in
  let genreFlix = bind myprov flixEnts
                  (filterByGenre genre) in
  ExtensionReturn cdom myprov genreFlix
```

.14 lines of Fine code

```
ass
ass
ass
ass
ass
assu
assum
assume CanReadDOMClass "netflix.com" "rv5"
assume CanCaptureEvents "onclick" (P "netflix.com" "netflixminer")
assume CanServeInformation "fandango.com" (P "netflix.com" "netflixminer")
assume CanServeInformation "amazon.com" (P "netflix.com" "netflixminer")
assume CanServeInformation "metacritic.com" (P "netflix.com" "netflixminer")
assume CanHandleSites "netflix.com"
assume CanReadStore (P "netflix.com" "netflixminer")
assume CanReadLocalFile "moviegenres.txt"
```

RePriv	• An in-browser framework for collecting & managing personal data to facilitate personalization.
Core Behavior Mining	• Efficient in-browser behavior mining & controlled dissemination of personal data.
RePriv miners	• A framework for integrating verified third-party code into the behavior mining & dissemination of RePriv.
Real-world Evaluation	• Evaluation of above mechanisms on real browsing histories & two in-depth case studies.

Privacy-Aware News Personalization

Map RePriv interest taxonomy to del.icio.us topics

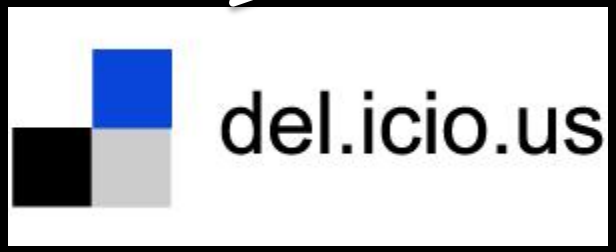
Query personal store for top interests

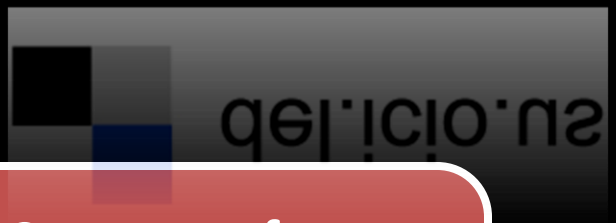
Ask del.icio.us API for “hot” stories in appropriate topic areas from nytimes.com

Replace nytimes.com front page with del.icio.us stories

Privacy Policy

Change "href" attribute of anchor elements on nytimes.com





Change TextContent of selected anchor and div elements on nytimes.com

Technolog
Technolog
Science/C
Science/P

- 2,2
- Ov
- Ty
  -
  -
  -

User profile:

- Games/Card_Games
- Games/Conventions
- Games/Video_Games

Do Video Games Equal Less Crime?

That's one theory for the continuing fall in crime, despite the recession.

Gamers Finally Get Their Wheaties Box ...sort of

Dr Pepper is featuring the Halo 3 player Tom Taylor, who goes by Tsquared, on the labels, which will appear on about 175 million 20-ounce bottles from January to April.

Sony To Shut its SF Metreon PlayStation store

Sony is closing down its one-and-only U.S. PlayStation store at the Metreon mall in San Francisco The recession is clearly to blame, but it's happening at time when Microsoft - which opened and shut its own Microsoft store at the Metreon - is going to open a chain of its own stores.

Microsoft Takes on Cable With Xbox Streaming Video

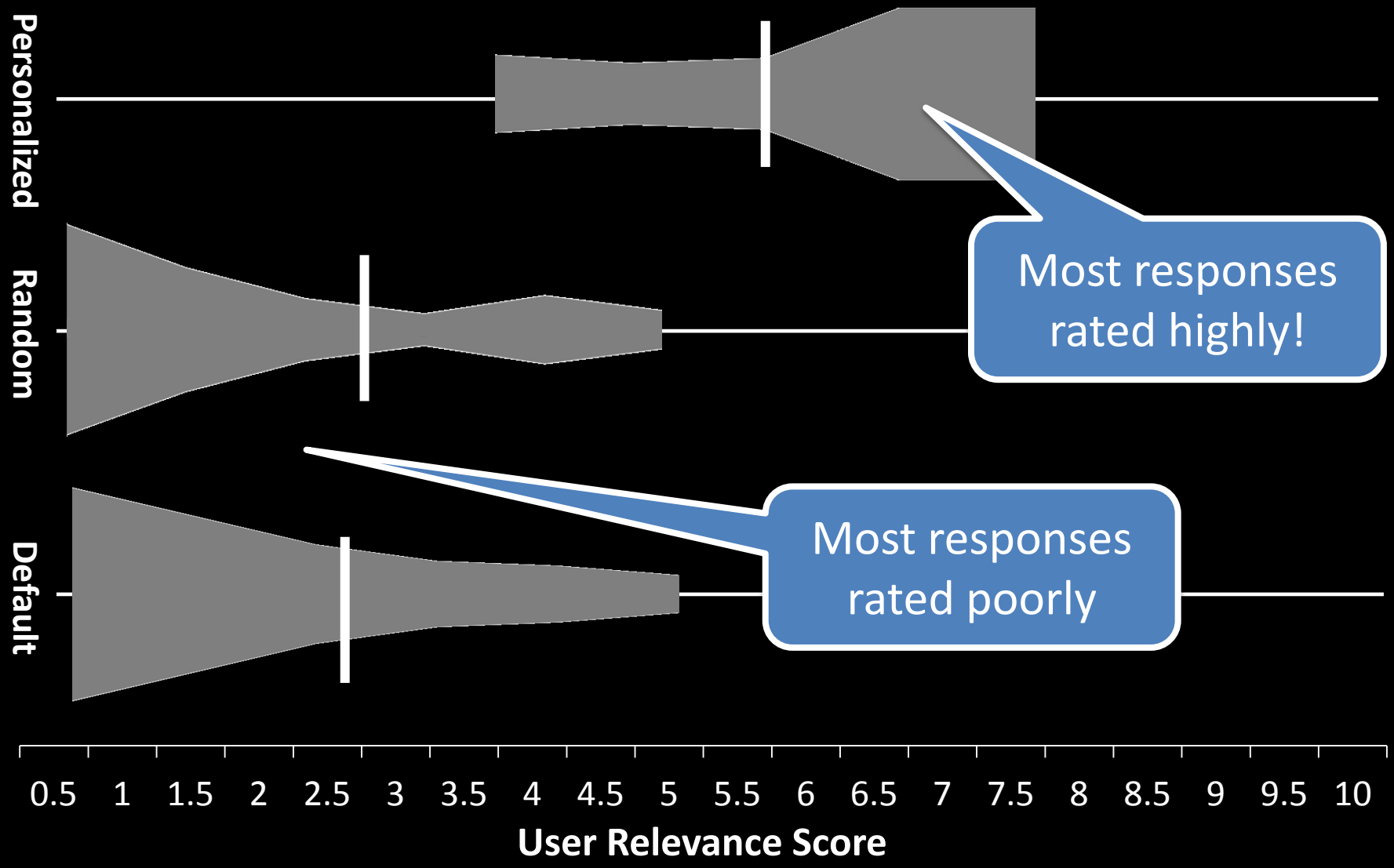
If talks with Disney work out, the game console could stream ESPN content, making it that much easier to watch TV without cable.

Some Video Gamers Leery of Obama's Views

Gamers are worried that the president-elect's positions on video games may signal new regulations or restrictions on the industry.

Relevance: (required)

News Personalization: Effectiveness



RePriv Summary

- Existing solutions require privacy sacrifice
- RePriv is a browser-based solution
  - User retains control of personal information
  - High-quality information mined from browser use
  - General-purpose mining useful & performant
  - Flexibility with rigorous guarantees of privacy
- Personalized content & privacy can coexist