

Data-Driven Transparency About Online Tracking

Euirim Choi, Claire Dolin, Aaron Goldman, Chang Min Hahn, Shawn Shan, Ben Weinshel,
Michelle Mazurek*, Blase Ur
*University of Chicago, *University of Maryland*

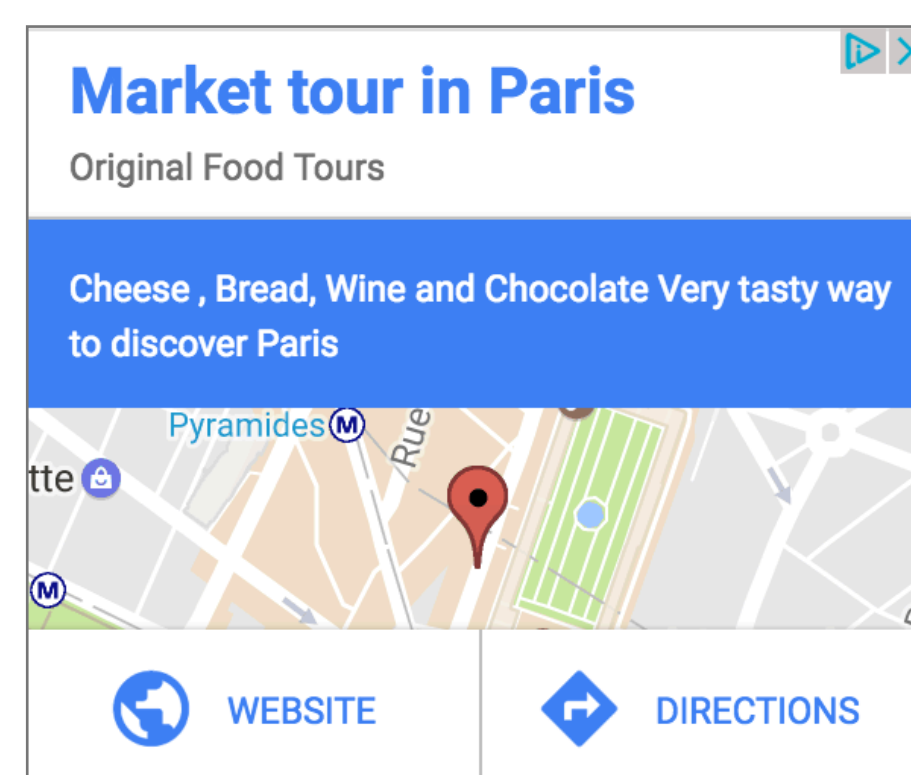
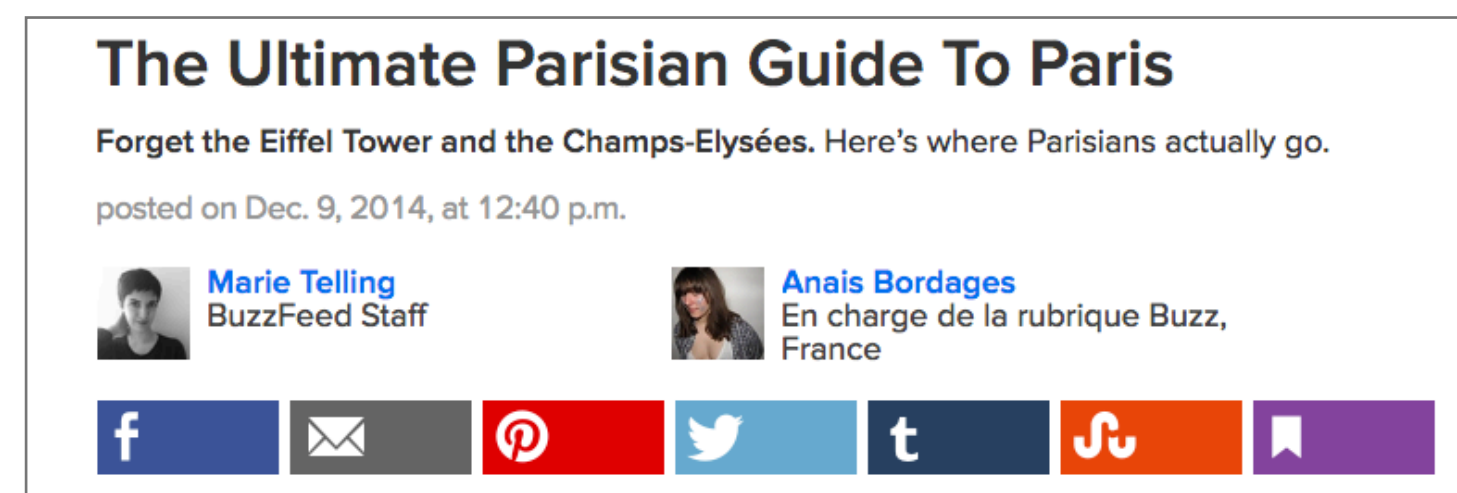
Background & Goals

Online Behavioral Advertising (OBA)

- Browsing → inferences → targeted ads
- Users find OBA useful...
- ... but also creepy

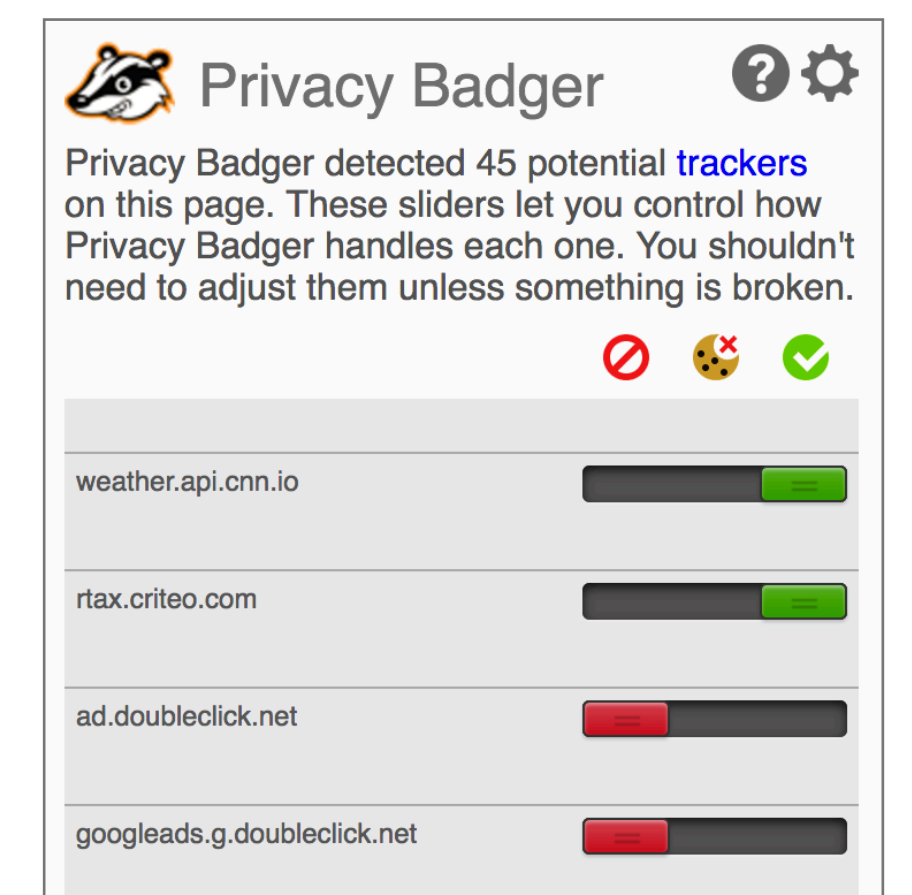
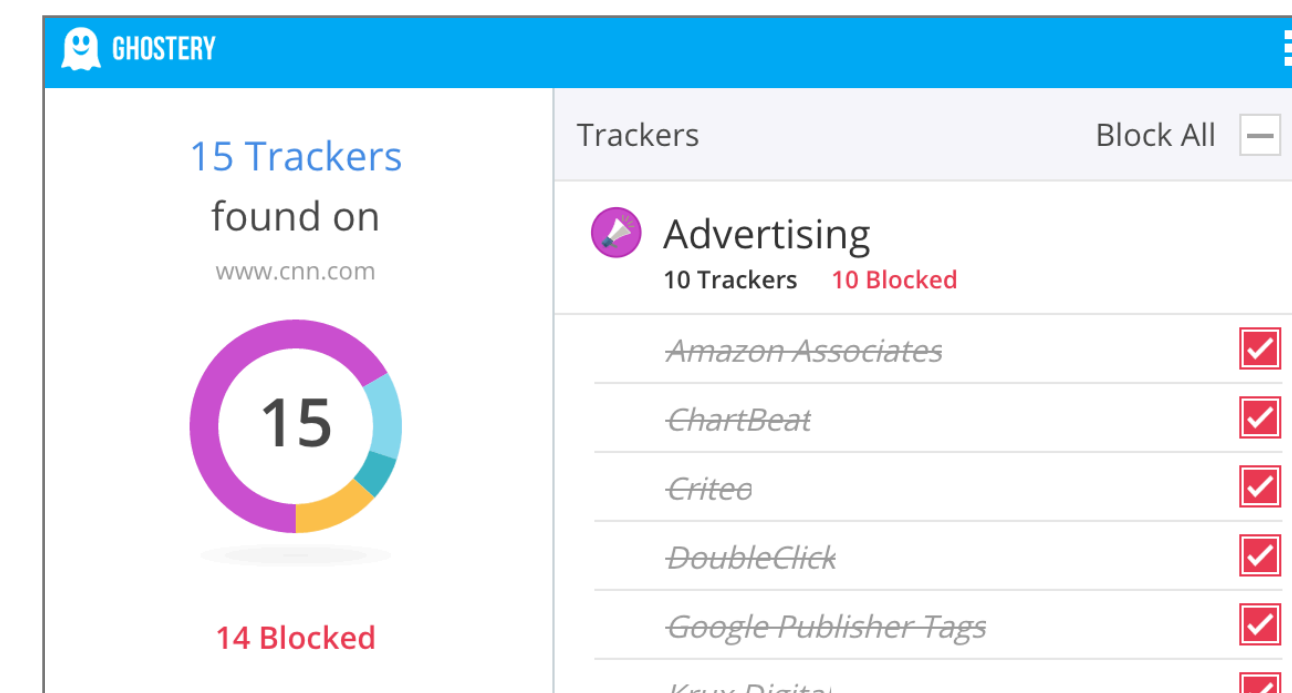
Goals

- Provide data-driven transparency
- “Track the trackers” in a browser extension
- Longitudinal history data → specific examples
- Understand risks; support privacy decisions
- Understand factors influencing OBA perceptions



Current Tools

- Ghostery, Disconnect, Privacy Badger, Mozilla Lightbeam
- Focus on tracking happening now
- Lists of trackers, but little additional information



User Studies: Perceptions of Tracking & Inferencing

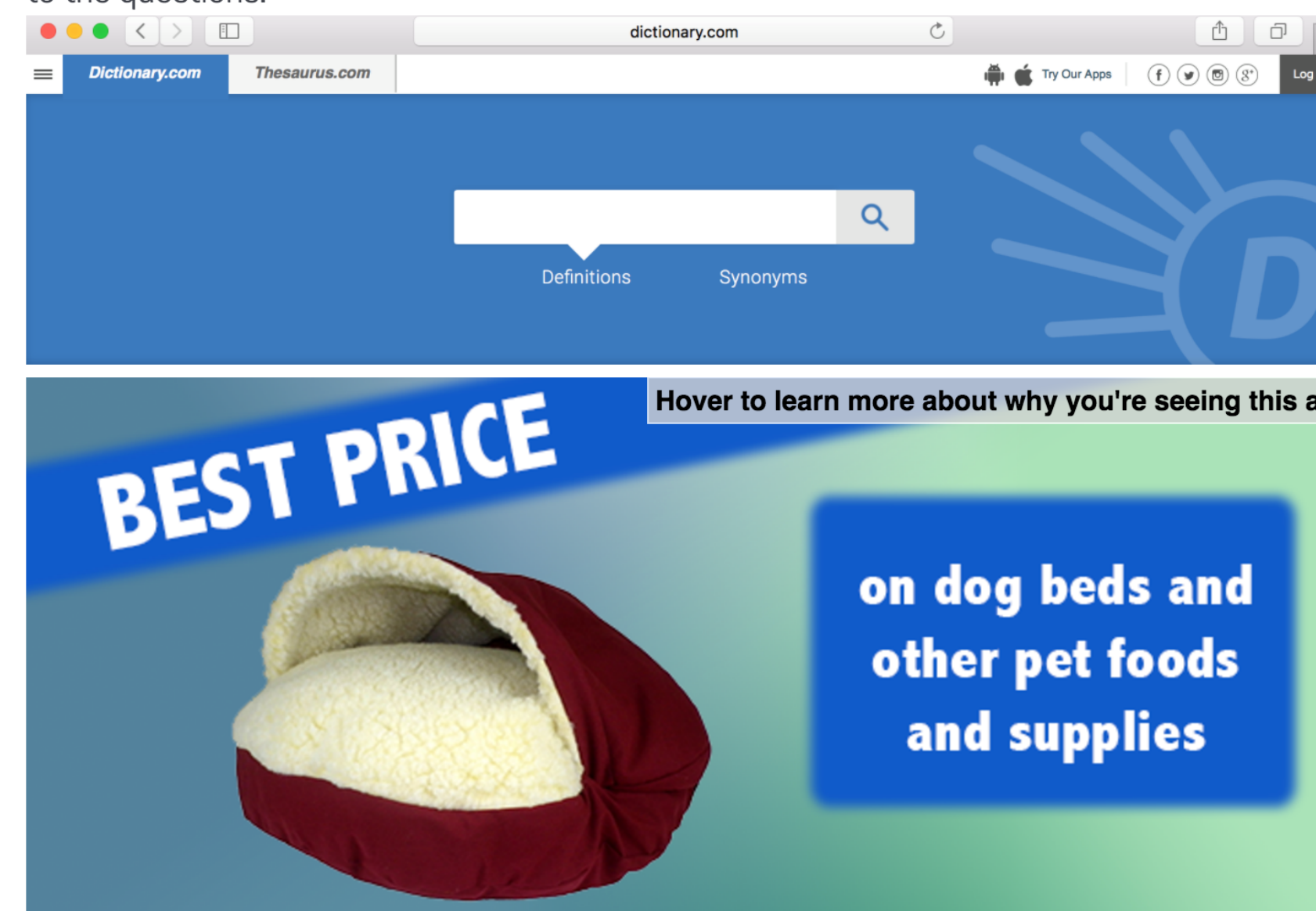
Perceptions of Targeting Mechanisms

- Simulate browsing experience – explain *why* an ad is being targeted
- Compare relative reactions to 11 different ways of making inferences
- Assess user interest and perceived usability of learning about advertising inferences

Inferencing Sensitivity

- Classify of sensitivity of ad categories
- See how classifications vary between users
- Determine factors influencing sensitivity
- Follow up on preliminary study

Hover over the advertisement to learn more. Read the explanation closely before moving on to the questions.



Example ad for targeting-perceptions study

You are seeing this ad because...

the advertiser has inferred based on your online activity that you are likely interested in buying **pet food & supplies**

and this is a related product.

–or–

You are seeing this ad because...

the advertiser has inferred that visitors to this site are often **men between the ages of 25 and 34**

and has inferred that **men between the ages of 25 and 34** are more likely to buy this product.

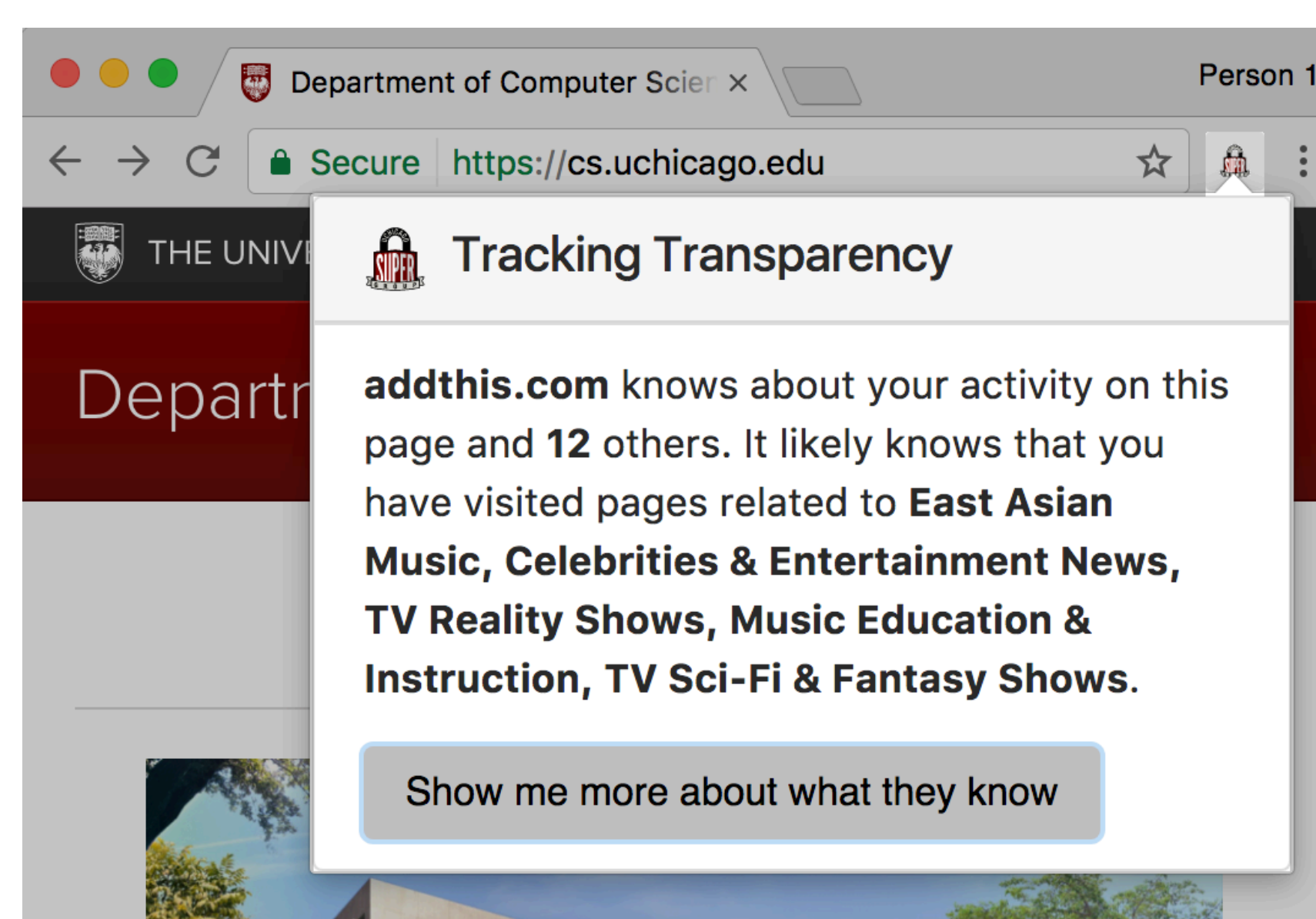
Browser Extension

Browser Extension

- Browser extension in development (Chrome/Firefox)
- Stores page visits, trackers, inferred topic
- Will provide detailed interface for exploration and querying
- Parse page text, compare with ad-interest category keywords

Browser Extension Studies

- Investigate popup taglines
- Test designs for exploration interface
- Planned 2-week field study



Current extension prototype

Inferring Ad-Interest Categories

Topic Keyword Generation

- Find keywords for ad-interest categories
- Categorize Wikipedia articles for training
- Exploring multiple algorithms to extract keywords (TextRank, latent Dirichlet allocation, hierarchical Dirichlet process)
- Associate extracted key unigrams with ad-interest categories

Proposed “hooks” for extension

Exhaustive: "DoubleClick knows you visited 82 pages across 17 sites in the past 3 days"

Recent visits: Yahoo knows that you visited “Things I Wish I’d Known When My Dog Died” (nytimes.com) last night



Security, Usability, & Privacy
Education & Research



THE UNIVERSITY OF
CHICAGO



DATA
TRANSPARENCY
LAB