User Modeling and Personalization
12: Personalization and Privacy

Eelco Herder

L3S Research Center / Leibniz University of Hanover
Hannover, Germany

4 July 2016

Outline I

Outline II

Principles of Fair Information Practices

Privacy-Enhancing Technology

- Pseudonymous users and user models
- Client-Side Personalization
- Distribution, Encrypted Aggregation, Perturbation and Obfuscation

User Model

A User Model is a data structure that characterizes a user U at a certain moment in time.

Which user data can be of relevance

Personal data, demographics

- Name, address, age, birthday, email address, gender, phone number, credit card information, ...
- Education, profession, ...
  Can be used for a rough initial fine-tuning of the interface

Contacts and friends

- Friends' personal data, groups and group membership, chatlogs, ...

Social Media

- User Ids or User Names for social media (e.g. Skype, Twitter, Facebook, LinkedIn, Xing, SchulerVZ)
- Login-Data (direct or via a token) for accessing the contents of the social media profiles
- Privacy controls (which data may be retrieved and used)

Device Information

- System specs, display resolution, network speed and bandwidth, software and tools

Location

- Position, direction, speed, vehicle, ...

Browsing-History & Bookmarks

- Bookmark Folder
- History
- Search history
- Ratings of pages, sites and other objects

Learning actions

- Visited pages
- Test scores
- Number of test attempts
- Time spent learning
- ...

And much more

Personal Data in the Web 2.0

Personal data has always been published and shared in the World Wide Web.

The Web 2.0 has promoted this even more by offering a variety of services where users can publish data without requiring specific technical skills on:

- social networking sites
- photo and document sharing sites
- collaborative work environments
- blogs
- and many other sites.

In addition, many of our activities on our computers and on the Web are logged in some way.

- **Ecommerce sites** register which items we browsed for and which we bought.
- **Social networking sites** keep track of the messages that we broadcast and send to our friends.
- **Browsers** maintain a history of sites that we visited and store cookies.
- **Desktop search engines** index all programs and files that we used.

Who Or What Can Access the Data?

Many of personal data is only accessible by ourselves and are used for **private purposes**

- reflection, archival and refinding.

Many other traces are **shared with others** or broadcast to the world - voluntarily or involuntarily.
These public or semi-public traces define our online presence: the way we are seen by the outside world, based on which others judge who we are.

And even personal data that is not disclosed is often **used for inferring** our interests, in order to provide personalized recommendations.

In most environments it is hard to figure out what exactly is being logged or to inspect and regulate which traces are stored, used by other applications or published to the outside world.

This raises serious privacy issues, of which the average user is often not aware.

The Issue With Linked Data and Mash-Ups

A further trend is the increased interaction between Web applications in **mash-ups**. By clever combinations of existing applications, interesting new applications are developed and user-friendly one-stop portals are created.

Further, many social networking applications create methods for synchronizing - or syncing - contact lists and communication flows.

The increasing monitoring, aggregation and exchange of user data is necessary for better, more integrated services, but brings with it serious threats to informational privacy.

Privacy

Information privacy is the right to decide what information is made available to others.

- Essential to maintain relationships of varying degrees of intimacy.

Online privacy is a more complex construct.

- encompassing technological, legal, and ethical aspects
- therefore an issue of concern for various publics, including consumers, consumer advocacy groups, the media, marketers, and governments.

Information Asymmetry

The fundamental concern to all these groups is the *information asymmetry* between

- website providers as data collectors and
- users as data providers

due to the absence of adequate control mechanisms of how user data are collected and whether they are disseminated.

Privacy Breaches

The combination of online data from various channels has led to practices such as

- cyber-stalking
- cyber-bullying
- online identity theft

It is therefore crucial to equip users of all age groups with a sound understanding of what sharing data online can entail and to help them develop and a stronger sense of responsibility for their own data.

Although users have voiced privacy concerns in privacy surveys, experiments have shown that users do not employ privacy-protection measures.

- they are either not aware that they exist
- or they find them inconvenient to use.

This paradoxical situation calls for privacy-enhancing systems with *high usability* and clear communication about how to use them.

The Externalities of Search 2.0

Search engine providers increasingly use the information from Web 2.0 services to complement both their search results with recent information.

- For example, a Google search for an individual's name routinely returns Facebook and LinkedIn profile pages
- and even the minute and often personal details shared via Twitter.

(http://firstmonday.org/htbin/cgiwrap/bin/ojs/index.php/fm/article/view/2136/1944)

Googling' someone has become common practice. People use search engines to learn about prospective blind dates.

- almost one in four Web users have searched online for information about co-workers or business contacts
- employers are Googling prospective employees before making hiring decisions.

The End of Privacy via Obscurity

Through the powerful reach of search engines, obscure pieces of personal information - such as email messages sent a decade ago to niche forums or newsgroups - are increasingly retrievable by a simple keyword search.

As a result, any 'privacy via obscurity' that generally kept such information from public view has been diminished.

Privacy versus Functionality

Search 2.0 promises breadth, depth, efficiency, and relevancy, but enables the widespread collection of personal and intellectual information in the name of its perfect recall.

O'Really, 2006

If history is any guide, the democratization promised by Web 2.0 will eventually be succeeded by new monopolies, just as the democratization promised by the personal computer led to an industry dominated by only a few companies. Those companies will have enormous power over our lives - and may use it for good or ill.

Facebook Privacy Settings: Who Cares?

(http://firstmonday.org/htbin/cgiwrap/bin/ojs/index.php/fm/article/view/3086/2589)

Mark Zuckerberg, 8 January 2010:

"People have really gotten comfortable not only sharing more information and different kinds, but more openly and with more people. That social norm is just something that has evolved over time."

The above comment came after Facebook's move in December 2009 that prompted users to reconsider their privacy settings.

The *default option* was to make user content publicly accessible to

- all Facebook users
- programs that access the data using the tools that Facebook made available to software developers.

This change outraged many privacy advocates and regulators.

Facebooks's Censorship Problem

(http://michaelzimmer.org/2011/04/21/facebooks-censorship-problem/)

In 2010, Facebook removed a photo of two men kissing from a user's Wall due to an apparent violation of the site's terms of service.

The two men are actors, as the photos is a promotional image from a popular British soap opera.

Hello

Content that you shared on Facebook has been removed because it violated Facebook's Statement of Rights and Responsibilities. Shares that contain nudity, or any kind of graphic or sexually suggestive content, are not permitted on Facebook.

This message serves as a warning. Additional violations may result in the termination of your account. Please read the Statement of Rights and Responsibilities carefully and refrain from posting abusive material in the future. Thanks in advance for your understanding and cooperation.

The Facebook Team

Why was this warning unjust?

1. Nowhere in the Rights statement does it prohibit, or suggest a prohibition, on 'sexually suggestive' content. It merely restricts pornography and nudity.

2. Let's assume for a moment that the Statement does include mention of 'sexually suggestive' content as mentioned in the warning to the user.
   - Does the photo in question fit that description?

Why is this warning disturbing?

It appears that a human took a look at that photo, and then removed it.

Comment from a blog

I think I'd almost rather it had been an algorithm, as it is quite troubling that a Facebook admin, wielding such power, would arrive at this conclusion.

Comment from Facebook on the above blog post: The photo in question does not violate our Statement of Rights and Responsibilities and was removed in error. We apologize for the inconvenience.

Online Advertising, Behavioral Targeting, and Privacy

Data on the online behavior of consumers has allowed companies to deliver online advertising in an extraordinarily precise fashion.

Such behavioral targeting has obvious benefits to advertisers because fewer ad impressions are wasted.

For consumers, however, ads that are behaviorally targeted can appear unauthorized and even creepy.

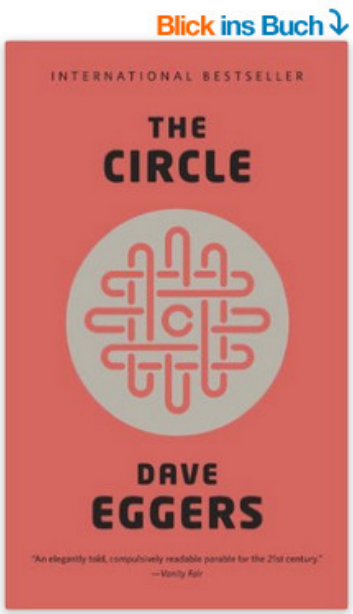
(Communications of the ACM 5 (54), 2011)

Researchers found that in Europe, after privacy protection was enacted, the difference in stated purchase intent between those who were exposed to ads and those who were not dropped by approximately 65%.

There was no such change for countries outside Europe.

In the long run, this may change the kind of Web sites and firms that prosper on the advertising-supported Internet, perhaps leading to fewer free (ad-supported) general-interest Web sites.

Book Recommendation



Dieses Bild anzeigen

The Circle (Englisch) Taschenbuch – 22. April 2014

von Dave Eggers (Autor)

171 Kundenrezensionen

Bestseller Nr. 1 in Technothriller

Alle 16 Formate und Ausgaben anzeigen

Kindle Edition	Gebundene Ausgabe	Taschenbuch	Hörbuch-Download	Audio-CD
EUR 6,99	EUR 14,95	EUR 5,50	EUR 27,08 oder gratis	EUR 35,49

Lesen Sie mit unserer Kostenfreien App

oder **EUR 0,00** im Probeabo von Audible.de

Lieferung bis Samstag, 11. Juli: Bestellen Sie innerhalb **20 Stunden und 39 Minuten** per **Premiumversand**. Siehe Details.

71 neu ab EUR 3,87 | 6 gebraucht ab EUR 4,67

***The Circle* is the exhilarating new novel from Dave Eggers, best-selling author of *A Hologram for the King*, a finalist for the National Book Award.**

When Mae Holland is hired to work for the Circle, the world's most powerful internet company, she feels she's been given the opportunity of a lifetime. The Circle, run out of a sprawling California

Mehr lesen

Factors Fostering Disclosure of Personal Information

The value which Internet users assign to personalization is a very important factor that can override privacy concerns.

- 75% of internet users find it useful if a site remembers basic information *(name, address)*
- 50% find it useful if a site remembers information *(preferred colors, music, delivery options etc.)*
- 62% are bothered if a web site asks for information one has already provided *(e.g., mailing address)*

This suggests that while one should not ignore privacy concerns, it is still possible to persuade users to provide their personal data by improving the quality of personalized services.

Knowledge of and Control over the Use of Personal Information

Extensive work in this direction has been carried out by Judy Kay and her team under the notion of **'scrutability'**.

According to Kay, this means that the user can scrutinise

- the processes used to collect data about the user
- the processes that made inferences based on that data
- the model to see what information the system holds about them

Trust in a Website

Trust in a website is a very important motivational factor for the disclosure of personal information.

Trust-inducing factors include:

Positive experiences in the past
Users are more willing to provide personal details if they obtained positive experiences with the same site or comparable sites in the past.

The design of a website

- the absence of errors, such as wrong information or incorrect processing of inputs and orders
- the (professional) design of a site
- the usability of a site
- the presence of contact information
- links from a believable website
- quick responses to customer service questions

The reputation of the website operator

Users' information disclosure at sites of well-reputed companies is likely to be higher than at sites with lower reputation.

Personalization is therefore likely to be more successful at sites with higher reputation.

The presence of a privacy seal

Privacy seals are logos of certifying agencies such as consumer organizations, data commissioner's offices or private companies.

These agencies assert to web visitors that websites that display their seals respect privacy to some extent.

However, several studies came to the conclusion that websites that decide to 'pay up' for certain privacy seals seem to have more questionable privacy practices than ones that don't.

The presence of a privacy statement (but not necessarily its content)
Most countries that have privacy laws enacted require that users be informed about the data being collected and the purposes for which they are used.

Not too many people seem to view and read privacy policies. Some studies report that only 3% read the statement 'most of the time, carefully', about 14% read it 'frequently', and about 60% 'sometimes'.

What's Wrong with Online Privacy Policies

(Excerpt from Irene Pollach's article in Communications of the ACM 50 (9), 2007)

Irene Pollach analyzed the privacy policies of 50 Web sites covering a broad spectrum of business models (retailers, internet service providers, news sites and travel agents).

The analysis of the vocabulary revealed that companies *sugar-coat data handling practices* by foregrounding positive aspects and backgrounding privacy invasions.

- companies will send email messages to registered users that are of *'interest to them'*
- or they share information with *'carefully selected'* partners
- companies use phrases such as *'not without your permission'*, without making clear whether this permission is *opting-in* or *opting-out*
- model verbs and adverbs make sentences vague (may, might, perhaps, sometimes, occasionally, . . .)

Principles of Fair Information Practices

Over the past three decades, several collections of basic principles have been defined for ensuring privacy when dealing with personal information.

Minimization principles

- Collect and use only the personal information that is *strictly required*..
- Store information for *only as long as it is needed*.
- Implement systematic mechanisms to *evaluate, reduce, and destroy* unneeded and stale personal information on a regular basis.
- *Before deployment* of new activities and technologies that might impact personal privacy, carefully evaluate them for their necessity, effectiveness, and proportionality.

Consent principles

- Require each individual's explicit *informed consent* to collect or share his or her personal information (opt-in);
- Or clearly provide a mechanism for individuals to easily *opt-out*
- When appropriate, offer the possibility to delete user information

Openness principles

- Explicitly state the *precise purpose* for the collection of user data
- and all the things that the information might be used for
- Explicitly state *how long* this information will be stored and used
- Make privacy policy statements *clear, concise, and conspicuous*

Access principles

- Establish and support an individual's right to *inspect and make corrections* to her or his stored personal information, unless legally exempted from doing so.

Accuracy principles

- Ensure that personal information is sufficiently *accurate and up-to-date*..
- Ensure that all corrections are propagated in a timely manner to all parties.

Security principles

- Use appropriate physical, administrative, and technical measures to maintain all personal information securely and protect it against unauthorized and inappropriate access or modification.

Privacy-Enhancing Technology

Pseudonymous users and user models

To ensure their linkability, users would need to employ a 'pseudonym' in all their transactions

- a unique and persistent identifier that differentiates them from all other users.

Client-Side Personalization

Advantages:

- Very few, if any, personal data will be stored on the server.
- Users may be more inclined to disclose ersonal data if personalization is performed locally.

Challenges:

- Methods that rely on data from the whole user population *(such as collaborative filtering and stereotype learning)* cannot be applied any more or will have to be radically redesigned
- Personalization processes will have to operate at the client side.
  - requires processing power
  - confidential algorithms may be reverse-engineered

Distribution, Encrypted Aggregation, Perturbation and Obfuscation

- *Perturbation:* (intentionally) putting the data into disorder (by adding random data or by altering existing data)
- *Obfuscation:* making data confusing, intentionally ambiguous, and more difficult to interpret (for example, by partially encrypting or transforming the data)

Data distribution or encrypted aggregation make it harder, but not impossible, to relate specific information or actions to a specific users.

Perturbation and obfuscation - if done properly - make it impossible to relate data to a specific user (or at least make it not 100% probable)

Disadvantages:

- introduction of uncertainty and ambiguity
- more complicated collaborative filtering techniques needed
- computational power needed increases
- impact on the scrutability of user models