CSE 303:
Concepts and Tools for Software Development

Dan Grossman

Spring 2007

Lecture 10— Societal Implications: Web-Site Data

Why are we doing this?

An educated computer-scientist should think about the broader implications of what they do.

Conversely, people not trained in computer-science may not be equipped to make ethical / practical decisions about relevant technology.

There are a million topics we could pick; web data seems particularly interesting given "Web 2.0" (and Web 1.0).

10% of your grade: a short paper on topic(s) we discuss; more information later.

The plan

I'll share 5–10 minutes of thoughts, overview, examples, and questions.

We'll divide into groups of 5–8 for 20 minutes.

- Discuss questions of interest
- Bring up new examples (preferred) or discuss existing ones
- Pick a speaker to report back (1 minute)

We'll reconvene for whole-class discussion

- (1-minute presentations will guide us)

A file system

For sake of comparison, consider `attu`.

- You have files and permissions on them. By default, nobody checks for appropriateness.
- You can look at other people's files, but you usually don't.
- The department (i.e., the government) owns the disk.
- `turnin` lets you put whatever you want in a place that "belongs" to the TAs.
- Most actions are not *logged*; system actions are to *diagnose problems and detect intrusions* (simple example: last login)

This worked pretty well for 30 years.

The Web has 1e9 users instead of 1e3.

Simple (?) Questions

- What data should be publicly available?
- What should be stored about where users web-surf?
- Who should be to blame when the wrong data / surf-histories become known?

Data Content

- Is it clear "who controls" a web-site? Should it be?
- My homepage, but on a government site.
- Search engines copy other possibly illegal (in what country?) sites.
- YouTube, Facebook, Flickr, ... post strangers' content.
- What if gmail or Google Calendar "got hacked"?

A universal dilemma: replication increases recovery (backups) and decreases security.

Anachronistic laws? Is "who owns the hard-drive" the point? If not, then what is?

Surfing / Search-query privacy

- Search for "drug rehab", "domestic-violence shelter" or "how to make a bomb", "child pornography"
- Jealous ex-S.O. snooping IM or parent snooping 8-year-old's IM
- Amazon suggesting "similar products" or totalitarian regime finding buyers of "banned books"
- Search-query personalization (jaguar the car) or like a security camera (everything you search for)
- Employers have right to ensure productivity and network-security or employees have right to "close the door".

A universal dilemma: data can improve user-experience, but once collected can be used for other purposes.

Is there a "best answer" for all/most web-sites or networks? How can a user / site protect him/her/itself?

Questions

1. How responsible is the "equipment owner" for content?
2. Does a "big site" have greater obligations than a "small site"?
3. What bounds should there be on "terms of service" for web sites? Should you read them?
4. Do technical solutions (passwords, log-deletion, ...) help or is this "just" a legal/ethical issue?
5. How long should search logs be kept? Who should decide that?
6. How much privacy would you give up for "something really cool"?
7. What web-privacy ethical issues have gotten more complicated in the last 5-ish years?
8. ...