

CSE 303: Concepts and Tools for Software Development

Dan Grossman

Winter 2006

Lecture 11— 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-site data was one that was all over the news *this week*.

10% of your grade: a short paper on something we discuss.

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.

This has worked pretty well for 30 years.

The Web has $1e9$ users instead of $1e3$.

The Web is complicated

My homepage is basically just a world-readable file.

- Subject to government regulations

But many sites are more complicated:

- Google makes an automatic copy of other sites, which may have illegal content.
- KodakEasyShare and Facebook post strangers' content.
- Wikipedia is a free-for-all information repository.

And sites are interactive, not just static content:

- Amazon tracks much of what you do for business purposes.
- Google stores searches and is fighting a subpoena.

The Web is big

And there's a lot of data out there:

- King County Assessor's Office
- Massachusetts health records and voter lists

And you know know how to get it at “computer speed”

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. Is “hard to get” public data more permissible than “easy to get” public data?
5. For what information is “Web-scale” voting a good idea?
6. Do technical solutions (passwords, log-deletion, ...) help or is this “just” a legal/ethical issue?
7. ...