

CmpE 473

Internet Programming

Pınar Yolum
pyolum@cmpe.boun.edu.tr

Department of
Computer Engineering
Boğaziçi University

Chapter 12

Personalization and Privacy

Fixed Web Pages

- Content of a Web page is fixed
 - All users view the same Web page
- Different people may be interested in different parts of the page (or Web site)
 - Browse through too many pages to find the desired page
 - View adds that are not relevant
- How to show people what they are interested in?

Personalization

- How to determine what people are interested in?
- Gather user information to handle customer needs better
 - Information based on previous interactions
 - Information of others that have similar demands
- Customization (my.yahoo.com)
 - User inputs information (e.g., by selecting from menus)
- Personalization (amazon.com)
 - Done by the system

Personalization

- Improve usability
 - Personalize Web pages based on user's interests
 - User closer to information related to her interest
- Improve sales
 - Understand customers' service needs
- Requires customer modeling and learning

Personalization Data

- Data mining
 - For usability (Web usage)
 - Which Web pages are visited more often?
 - Consider frequency and duration of visit
 - Web server access logs
 - For sales (Content data)
 - What types of products are browsed?
 - Sequence in which products are browsed
- Integrate data from multiple sources

Data Analysis

- **Association rules**
 - Capture patterns of cooccurrences in multiple transactions
 - Example: Find groups of pages that are viewed together
- **Sequential patterns**
 - Sequence of items that occur frequently
 - Important to understand causal relations
 - Discovering Web site usability (which paths users take to reach the end Web pages)
- **Clustering approaches**
 - k NN: Cluster users based on the similarity of browsing, buying, and so on.

Spring 2005— Pinar Yolum

7

Usage

- **Association rules**
 - Used for business intelligence
 - Gives clues about sales
 - Example: Diapers and beers
- **Sequential patterns**
 - If a sequence of hyperlinks never used, redesign a system
 - Example: /products =>toys => PlayStation2.html never followed
- **Clustering approaches**
 - Recommend products based on what others in the cluster are buying
 - Predict future sales

Spring 2005— Pinar Yolum

8

Privacy

- How much of the personal data can be used for personalization?
- Web sites announce their privacy regulations through privacy policies
- Privacy policies (targeted for humans)
 - In natural language
 - Exist in different places depending on the Web site
 - Time-consuming to read and track changes
- Design privacy policies targeted for machines

P3P

- Platform for Privacy Preferences Project (P3P)
- Industry standard to manage personal information on the Web
- Standardized set of multiple-choice questions
 - Specifies (in XML) how a Web site controls its users' personal information
 - Note which privacy policies apply to which parts of the Web site in a policy reference file
 - Users' agents (mostly embedded in browsers) can decide whether they want to interact with the Web site
- No legal enforcement
- In principle, the privacy policy for the humans should be compatible with the P3P policy

HTTP with P3P

Before the transaction starts:

- Browser requests policy reference file
- Web server sends policy reference file
- Browser requests P3P file
- Web server sends P3P file

P3P Client

- Checks the privacy policy before every request from the Web server
- If the page contains information from multiple sites, checks all privacy policies
- Example: www.privacybird.com
 - Add-on to IE
 - Reads and displays P3P policies
 - Display warning messages for discrepancies
 - Display summaries
 - Settings from their Web site (next slide)

Privacy Settings

Privacy Preference Settings

These settings control when a warning icon will be displayed at the top of your browser window. You can click on the warning icon for more information.

Select Privacy Level Low Medium High Custom Imported

HEALTH OR MEDICAL INFORMATION

Warn me at web sites that use my health or medical information :

For analysis, marketing, or to make decisions that may affect what content or ads I see, etc.

To share with other companies (other than those helping the web site provide services to me)

FINANCIAL OR PURCHASE INFORMATION

Warn me at web sites that use my financial information or information about my purchases :

For analysis, marketing, or to make decisions that may affect what content or ads I see, etc.

To share with other companies (other than those helping the web site provide services to me)

PERSONALLY IDENTIFIABLE INFORMATION (name, address, phone number, email address, etc.)

Warn me at web sites that may contact me to interest me in other services or products :

Via telephone

Via other means (email, postal mail, etc.)

And do not allow me to remove myself from marketing/mailling lists

Warn me at web sites that use information that personally identifies me :

To determine my habits, interests, or other characteristics

To share with other companies (other than those helping the website provide services to me)

Warn me at web sites that do not allow me to find out what data they have about me

NON-PERSONALLY IDENTIFIABLE INFORMATION (demographics, interests, web sites visited, etc.)

Warn me at web sites that use my non-personally identifiable information :

To determine my habits, interests, or other characteristics

To share with other companies (other than those helping the website provide services to me)

Help Import Settings Export Settings OK Cancel

Spring 2005— Pinar Yolum

13

Example Policy File

- Name and contact information for site
- The kind of access provided
- Mechanisms for resolving privacy disputes
- The kinds of data collected
- How collected data is used, and whether individuals can opt-in or opt-out of any of these uses
- Whether/when data may be shared and whether there is opt-in or opt-out
- Data retention policy

Spring 2005— Pinar Yolum

14

Example Policy File

```
<POLICIES xmlns="http://www.w3.org/2002/01/P3Pv1">
  <POLICY name="forBrowsers"
    discuri="http://www.catalog.example.com/PrivacyBrowsing.html"
    xml:lang="en">
    <ENTITY>
      <DATA-GROUP>
        <DATA ref="#business.name">CatalogExample</DATA>
        <DATA ref="#business.contact-
          info.telecom.telephone.number"> 3926753</DATA>
      </DATA-GROUP>
    </ENTITY>
    <ACCESS><nonident/></ACCESS>
  </POLICY>
</POLICIES>
```

Example Policy File (cont'd)

```
<STATEMENT>
  <PURPOSE><admin/><develop/></PURPOSE>
  <RECIPIENT><ours/></RECIPIENT>
  <RETENTION><stated-purpose/></RETENTION>
  <!-- Note also that the site's human-readable privacy policy MUST
  mention that data is purged every two weeks, or provide a link to this
  information. -->
  <DATA-GROUP>
    <DATA ref="#dynamic.clickstream"/>
    <DATA ref="#dynamic.http"/>
  </DATA-GROUP>
</STATEMENT>
</POLICY>
</POLICIES>
```

Example Policy Reference File

```

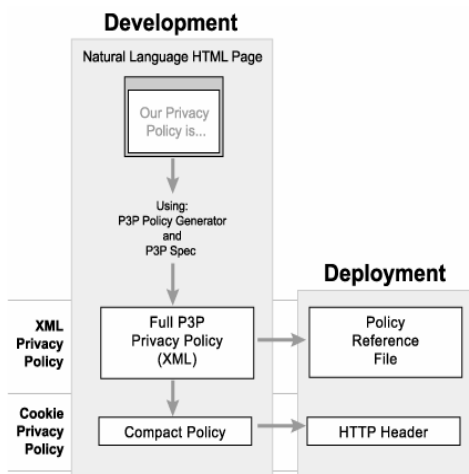
<META xmlns="http://www.w3.org/2002/01/P3Pv1">
  <POLICY-REFERENCES>
    <EXPIRY max-age="172800"/>
    <POLICY-REF about="/P3P/Policies.xml#first">
      <INCLUDE>*/</INCLUDE>
      <EXCLUDE>/catalog/*</EXCLUDE>
      <EXCLUDE>/cgi-bin/*</EXCLUDE>
      <EXCLUDE>/servlet/*</EXCLUDE>
    </POLICY-REF>
    <POLICY-REF about="/P3P/Policies.xml#second">
      <INCLUDE>/catalog/*</INCLUDE>
    </POLICY-REF>
  </POLICY-REFERENCES>
</META>

```

Spring 2005— Pinar Yolum

17

Enabling P3P Policy



Compact policy

- Optional summary of full P3P policy for cookies
- Must be used in addition to full policy
- Must commit to following policy for lifetime of cookies

* *Diagram from microsoft.com*

Spring 2005— Pinar Yolum

18

Example Policy Check

Privacy Policy Check:

- **AT&T's privacy policy *does not match your preferences:***
 - Site may use financial information or information about your purchases for analysis or to make decisions that may affect what content or ads you see, etc.
 - Unless you opt-out, site may use financial information or information about your purchases for marketing

Challenges

- **Balance between personalization and privacy**
- **Better user agents for modeling users' privacy requirements**
 - **Handling context-dependent cases**
 - **Learning whom to trust for preserving privacy**
- **Legal actions for privacy violation**