Personalisation of Web Search: Exploring Search Query Parameters and Information Privacy Implications

Anisha Fernando, Dr. Tina Du, Assoc. Prof. Helen Ashman
School of Information Technology & Mathematical Sciences,
University of South Australia

SIGIR 2014 Workshop on Privacy-Preserving Information Retrieval
Privacy in the Personalised Search Context

Outline

• Web Search Personalisation & Privacy- A Conflict of Interests

• Search Personalisation: Types and Techniques

• Information Privacy and User Privacy Concerns

• A first look at investigating Search Query Parameters and its impact on Personalisation & Privacy

• Conclusions and Future Research Directions
Personalisation aims to provide relevant search results to users based on their information needs*

Personalisation uses personal data

User information needs and preferences are captured to build user profiles

Search engines are profit-oriented: produces more relevant profiles (audiences) for targeted advertising

Conflicts arise:

Targeted Ads: Corporate benefit Vs. Individual Benefit: Is extensive personal information really required for personalisation?

Search Personalisation

- Content & Collaborative based Techniques

- Personalisation Types:
  - Location-based,
  - Social-based,
  - Behavioural Profiling and Data Aggregation

- Benefits of Personalisation:
  - Possible difficulties in expressing Information Needs
  - Information Overload
  - Relevant Results

- Disadvantages of Personalisation:
  - Filter Bubble / lack of it called ‘serendipity’
  - Extensive use of personal data – Is it really required?
Privacy is a social construct with different people having different attitudes towards privacy.

What is Information Privacy? concerned with personal information of users, which can subsequently be used to identify them **

Types of Personal Data:
- Public Personal data
- Non-public Personal data

In the Australian Context:
- Australian Privacy Principles

User Information Privacy Concerns

- Linkability of Personal Data
- Excessive Data Collection
- Behavioural Profiling
- Inferences from ‘Anonymised’ Data
- Targeted Advertising
- Low User Awareness
- Poor Control over Personal Data
- Privacy Policies & Consent
Mechanisms of Personal Data Collection

User Profiles

Implicit Personal Data Collection:
* Cookies / Flash cookies
* HTTP Headers, GET/POST Parameters

Explicit Personal Data Collection:
* Web Search History
* Search Queries
* Click-through Preferences
A First Look at An Experiment

- **Purpose:** to determine what impacts personalisation more: Account Sign-In; Location; ‘pws’ personalisation parameter

- **Method:**
  - Define a scenario
  - Capture parameters
  - Analyse its impact on personalisation and privacy

- **Findings – Early Stages:**
  - ‘pws’ designed to switch personalisation on/off has less impact than location
  - account sign-in has negligible impact in comparison to location
Some Examples...

Example of a search result with account sign-in and the personalisation parameter on without location spoofing

Example of a search result with location spoofing using Tor and the personalisation parameter switched on and account sign-in off
The Big Picture: Future Research Directions

- Investigate how personalisation is performed by search engines
- Identify the impact of personalisation on privacy
- Develop an understanding of the mechanisms of personal data collection used for personalisation and its impact on user privacy in relation to the Australian Privacy Principles
Thank You!
Any Questions?