To Find or To Be Found, That is the Question in Mobile Information Retrieval

Objectives of this Talk

- Traditional IR vs. mobile IR
- Information Push as the default information access model
- User profiling and wireless data broadcast

Web Search vs. Mobile Search

- Simple mobile search model
  - Shrink the desktop/web search onto a mobile device
  - Voice I/O, auto-completion (Google Suggest), query suggestion, aiming at reducing the user I/O effort
  - Vertical search services to cater for common mobile search
    - Route, restaurant, directory search
  - Yahoo Go!, Google Mobile
- Proactive model
  - Up-to-date and relevant information are pushed to mobile device, replacing explicit requests by local browsing
  - Make possible by large local storage and high bandwidth
  - Require profiling user interests and context awareness
  - Best-effort suggestions

Example: While you are shopping...

- Do you want your mobile devices to be loaded with useful coupons?
- What about store information, sales items?
- What about a bookstore selling a book that you browsed on Amazon yesterday?
- What about the time for the next bus that you appear to take every day?
- ... ...
User Profiling

- Click-stream analysis for website personalization
  - Server-side driven and applies to one website

- Comprehensive profiling
  - Online and offline tracking
  - Online: search and web browsing
  - Predictive of future events and needs
  - Offline: movement tracking
  - Predictive of local interests (both temporal and spatial) and action items
  - Requires location semantics

User Profiling – An Example

Planning (1 week to 1 month)

Engaging (a few days)

Search

- Sigir 2008

Browse

- Sigir 2008 homepage
- Registration page
- Workshop page
- Mobile IR page

Hotels stayed

Before:
- Ritz Carlton
- Four Seasons
- Pan Pacific

Search

- AppleInsider
- Apple – Support
- Apple tree
- Apple – Quicktime
- Apple – Fruit
- Apple – Mac
- History of Apple Computer
- Apple Mac News
- Apple – Support
- AppleInsider

Preference mining: Given the clickthrough data, what is the user interested in?
Clickthrough Graph Analysis

- Pages clicked by the users in response to a query are likely to be on the same topic.
- Different queries leading to the same clicked pages are likely to be on the same topic.

Queries | Pages | Concepts
--- | --- | ---
ipod | MP3 | ipod
apple | iTunes | apple
iphone | MP3 | iphone
video | | video

[Extended from Beeferman and Berger KDD 2000]

User Profiling – Prototyping

Queries | Pages | Concepts
--- | --- | ---
apple | MP3 | apple
ipod | iTunes | ipod
iphone | MP3 | iphone
video | | video

Example: Location Query

Q=beach
- Location concepts: Daytona Beach, Huntington Beach, Long Beach, Myrtle Beach, Palm Beach, Venice Beach
- Content concepts: camp, hotel, resort, restaurant, vacation

Q=Singapore
- Location concepts: Asia
- Content concepts: Embassy, Free Trade Agreement, Ministry of Foreign Affairs, National University of Singapore, Singapore Airlines
Example: Location Query

Q=Malaysia
- Location concepts:
  - Kuala Lumpur
  - Sepang Circuit
  - Taman Negara
- Content concepts:
  - government
  - travel
  - tourism
  - University Sain
  - Universiti Teknologi

Example: Location Query

Q=Southeast Asia
- Location concepts:
  - Cambodia
  - Indian Ocean
  - Indonesia
  - Malaysia
  - Thailand
  - Singapore
  - Vietnam
- Content concepts:
  - biking
  - language
  - people
  - relief effort
  - travel

Summary for User Profiling

- Search and browsing activities are useful indicators for users’ future activities
- Locations are useful indicators for user’s local/immediate interests
- Concept-based user profiling
  - Content and location concepts
- Challenges:
  - Integration of online and offline activities for better profiling of user interests
  - Profiling: how deep and how wide
  - Reasoning and planning
  - Community-based concept extraction
  - Live experiments

Information Push

- Suppose user interests are captured
- We need an infrastructure for pushing information to the mobile devices
  - 3G ???
  - Wifi ???
  - Bluetooth ???
Google Audio/Video

- A commercial example of location-based broadcast
- An example where publishers (radio stations) and advertisers take control of the broadcast schedule

Bluetooth-Based Data Broadcast

- Fine location granularity (<100m)
- An example of data broadcast for the small (DBFTS)

Users Control

- Google Audio
  - Audience cannot select the ads (passively via the programs they listen to/watch)
  - Local ads (granularity is city); not exactly location based

- BT
  - User discovers (or discovered by) data advertisers
  - User can reject advertisements
  - Small area coverage; clients/customers cannot "look ahead" to distant advertisers

Challenges

- Provide a directory of advertisers
  - Program guide
  - Fixed download points (via dedicated Access Points)

- How to provide immediate access to directories?

- User modeling via tracking on locations and interests

- Push ads based on user model
Walled-Garden Broadcast Environment

- Wireless channels are controlled by the Channel Operator (CO), e.g., the phone company
  - Advertisers and customers must subject to the rules and limitations of the CO

Data Dissemination for the Small (DBFTS)

- Channels are owned by many COs with different but typically small capacities
  - Coffee shops, convenient stores, bus stops, light pools, you and me, on the roof top, balcony, outside every window, etc.

- Information may be disseminated through different channel operators on their broadcast channels; it is up to the publisher or information provider (i.e., you and me) to decide whom and where.

- How to find the data? Search all channels one by one?
  - An index (directory) channel is needed

DBFTS Architecture

Coordination under Ad Hoc Updates
Coordination under Ad Hoc Updates

I(x)

Empty
Occupied

DBFTS Architecture

Conclusion

- Proactive information pushing
  - Pre-programmed: download the price of a particular stock
  - Situational: learn user interests and deduce actions based on user interests and local context
  - Online activities, physical movement and their integration

- Data broadcast for the small (DBFTS)
  - Current wireless operators are not compatible cross platform, making data collections and profiling from diverse sources impossible
  - Small broadcast channel operators based on open standards for cross-platform compatibility
  - Data publishers can decide when and where to publish
  - Clients can search and download any data from any source