

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



Dik Lun Lee
Department of Computer Science and Engineering
Hong Kong University of Science and Technology
July 24, 2008

Objectives of this Talk

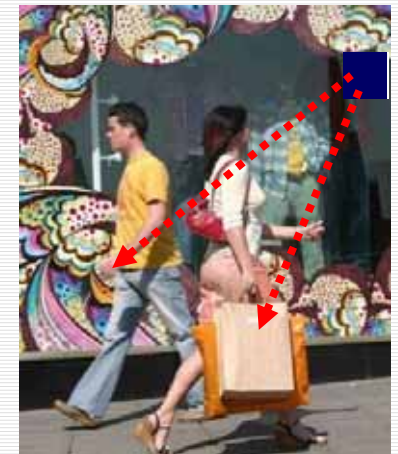
- ❑ Traditional IR vs. mobile IR
- ❑ Information Push as the default information access model
- ❑ User profiling and wireless data broadcast

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?
- ❑

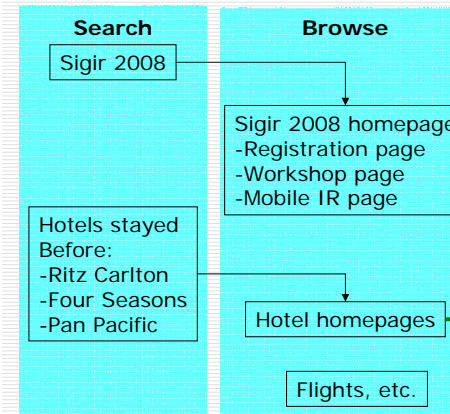


User Profiling

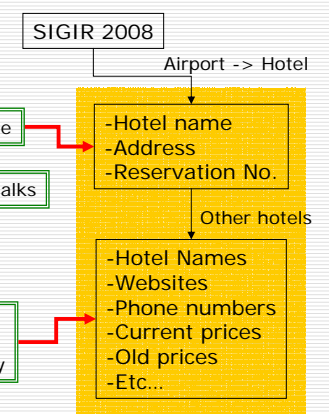
- ❑ Click-stream analysis for website personalization
 - Server-side driven and applies to one website
- ❑ Comprehensive profiling
 - Online and offline tracking
 - Online: search and web browsing
 - ❑ Predictive of future events and needs
 - Offline: movement tracking
 - ❑ Predictive of local interests (both temporal and spatial) and action items
 - ❑ Requires location semantics

User Profiling – An Example

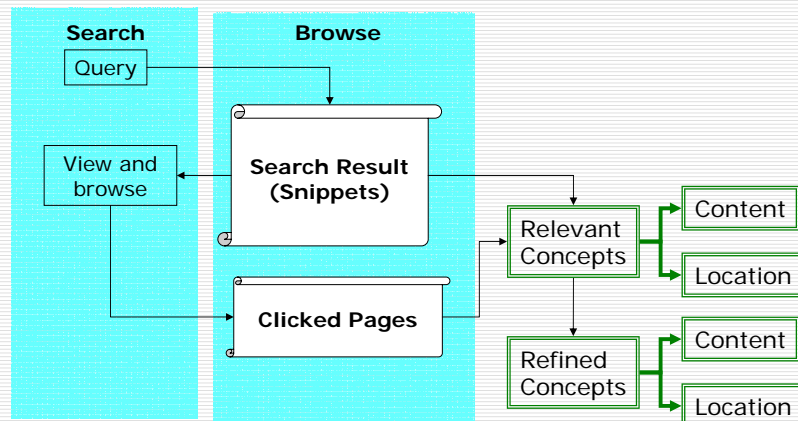
Planning (1 week to 1 month)



Engaging (a few days)



User Profiling – Concept Extraction



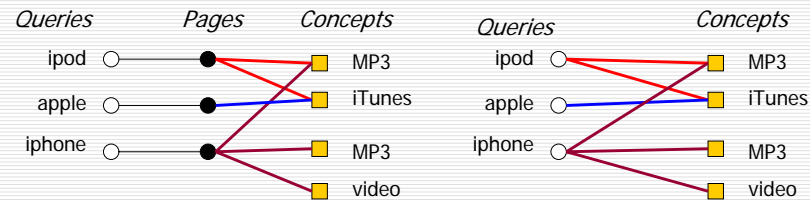
Clickthrough Data

Doc	Clicked	Search results
d_1	✓	Apple Computer
d_2		Apple – Quicktime
d_3		Apple – Fruit
d_4	✓	Apple - Mac
d_5		History of Apple Computer
d_6	✓	Apple Mac News
d_7		Apple tree
d_8	✓	Apple – Support
d_9		AppleInsider

- ❑ Preference mining: Given the clickthrough data, what is the user interested in?

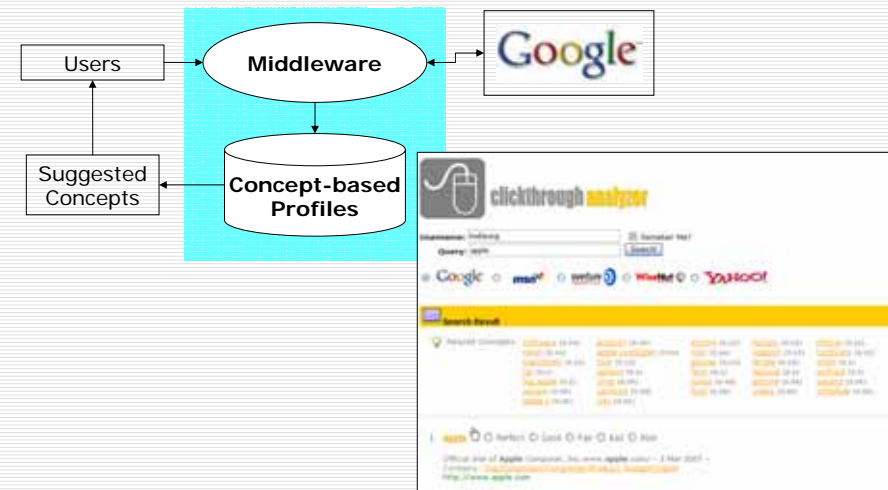
Clickthrough Graph Analysis

- 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

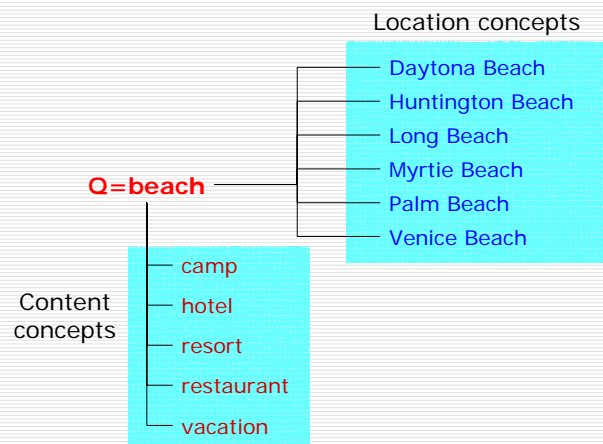


[Extended from Beeferman and Berger KDD 2000]

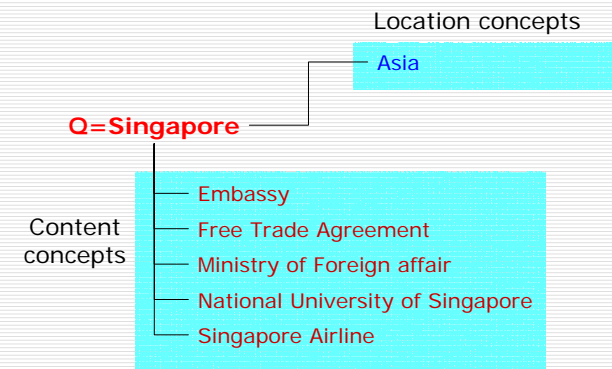
User Profiling – Prototyping



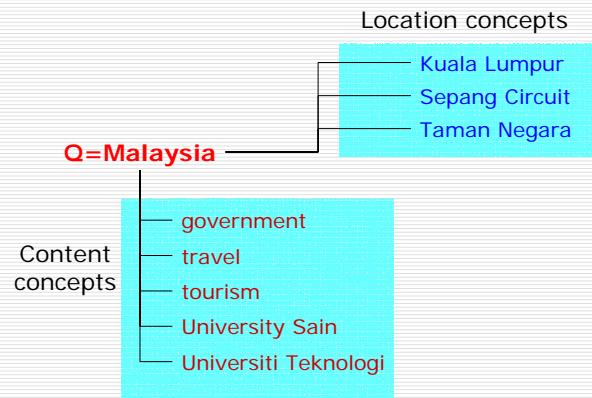
Example: Location Query



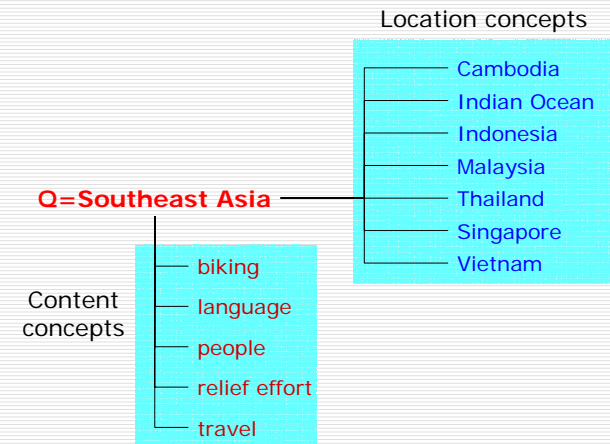
Example: Location Query



Example: Location Query



Example: Location Query



Summary for User Profiling

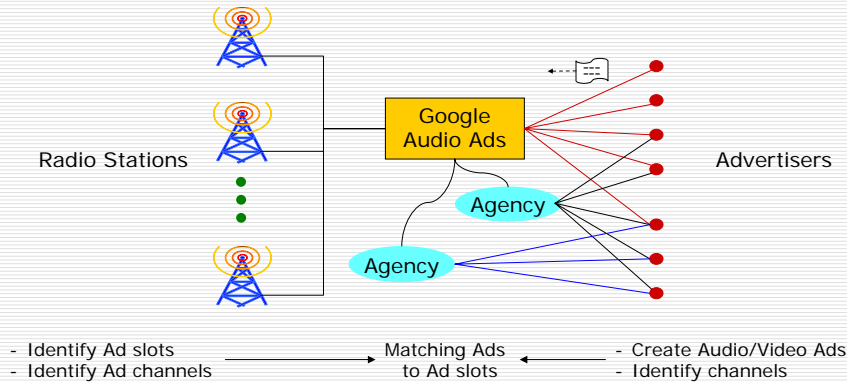
- ❑ Search and browsing activities are useful indicators for users' future activities
- ❑ Locations are useful indicators for user's local/immediate interests
- ❑ Concept-based user profiling
 - Content and location concepts
- ❑ Challenges:
 - Integration of online and offline activities for better profiling 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 ???

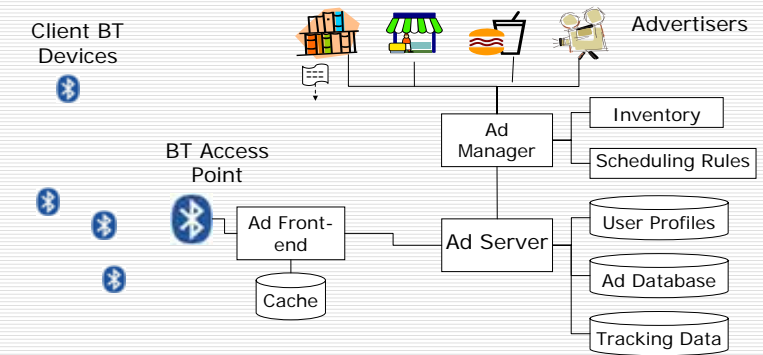
Google Audio/Video

- A commercial example of location-based broadcast
- A 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

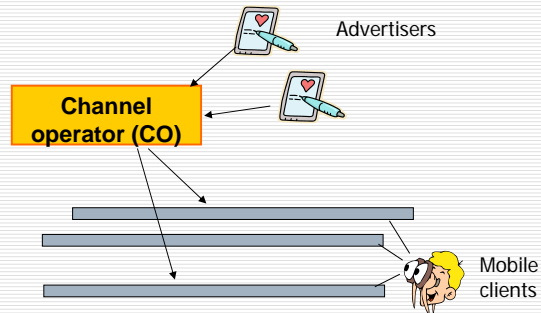
- Google Audio
 - Audience cannot select the ads (passively via the programs they listen to/watch)
 - Local ads (granularity is city); not exactly location based
- BT
 - User discovers (or discovered by) data advertisers
 - User can reject advertisements
 - Small area coverage; clients/customers cannot "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

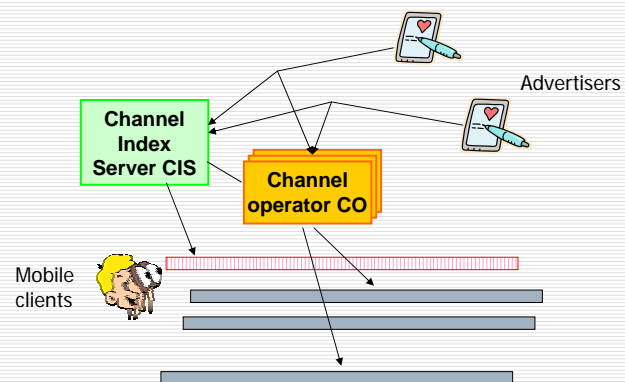
- Wireless channels are controlled by the Channel Operator (CO), e.g., the phone company
 - Advertisers and customers must subject to the rules and limitations of the CO



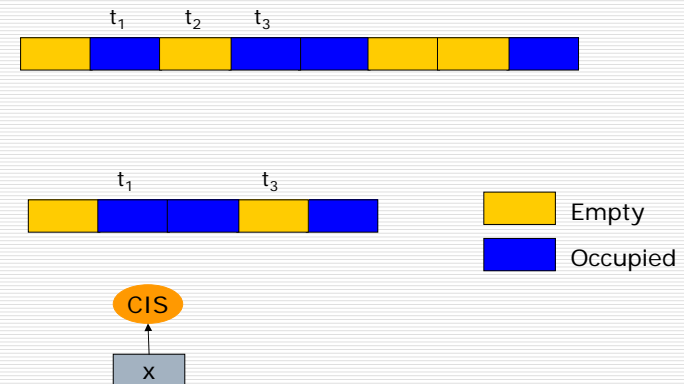
Data Dissemination for the Small (DBFTS)

- 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

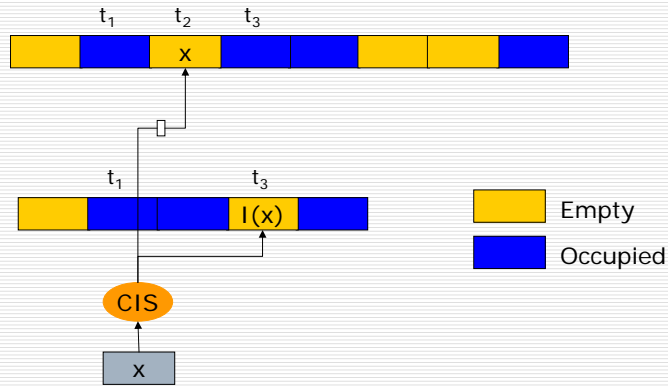
DBFTS Architecture



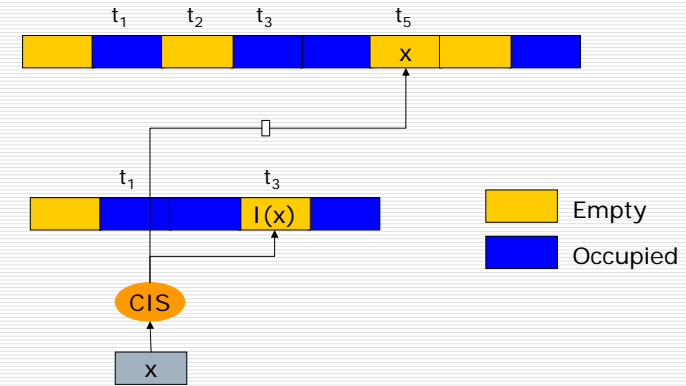
Coordination under Ad Hoc Updates



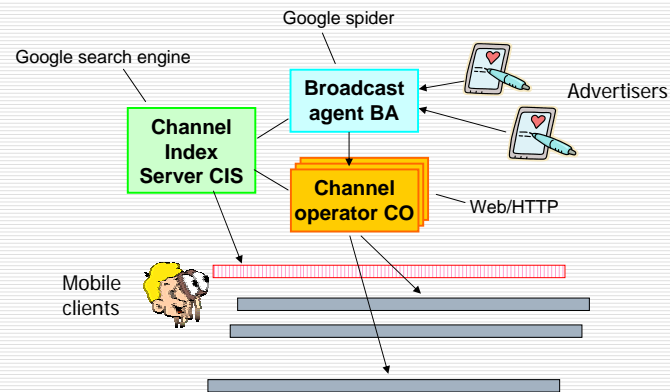
Coordination under Ad Hoc Updates



Coordination under Ad Hoc Updates



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