



Chapter 6

Four Hot Topics:
Personalization, Web 2.0, Social
Networks and the Semantic Web

1. Personalization

- Much of our information access is alternating between search and browse in supporting our tasks; SEs encourage this;
- Visual media has more browse, and text media has more search;
- Personalization is the sweet spot between these two, where “the system” takes the initiative and pre-determines what you may need by exploiting a number of underlying techniques;
- Most of these exploit information about you, knowing things about you, knowing your preferences, habits, etc., and often also knowing those same things about others, and then exploiting all that;
- Personalization appears as a standalone function i.e. a recommender system, or embedded in something else;
- Sounds great, except there is a tension between personalization and privacy;

Personalization

- The *techniques* for personalization and recommendation are IR-based and IR-derived
 - *Content-based or case-based* - match the user's scenario against a case-base or match an ideal 'product' against a collection based on content; no cold start problem;
 - *Collaborative Filtering* - (MovieLens) store ratings from other users (anonymously) and match this user against other user trends; issues of cold start;
 - *Probabilistic models* - combinations

Personalization & Privacy

- Personalization & Privacy ... a challenge to combine them and to develop techniques that offer a user the balance choice depending on how important the application is.
- A case study

Personalization & Mobile

CA557 - Information Access -

- Personalization is esp. important on mobile devices because of screen, input mode, people in a hurry, usage scenarios
- Case of CW - mobile portals allow users browse icons/text - for example, a user who checks local cinema info at weekends means
 - go to ENTERTAINMENT,
 - then CINEMA,
 - then MOVIE TIMES,
 - then DUBLIN,
 - then OMNIPLEX,
 - then BROWSE MOVIES,
 - then MAKE BOOKING
 - ... 18 clicks !
- Mobile usability -> 30s (10-12 clicks) is our max tolerance (vs. search time?)

Personalization & Mobile

CA 557 - Information Access

- For a sample of 20 European portals, click distance >16 for average info location;
- Leads to users frustrated with navigation time
- So ... improve portal design by personalised navigation ... promote and re-order menu options according to usage patterns of individuals ... move ENTERTAINMENT to the top at weekends;
- Leads to reduced click distance, improved usability, yet users then spend more time online generating more revenue for operators!
- This was the first venture into mobile personalization - nowadays mobile devices (e.g. iPhone) use WLAN so less time pressure, and have touchscreens so more navigable, yet the principal findings still hold true.

Personalization & Web Search

CA557 - Information Access -

- Vague queries, ambiguity in NL terms, generic search tools (G, Y!, etc. dominating), no sense of community of like-minded, lots of repetition in web searching;
- If we knew where a user was coming from, this could disambiguate and personalise:
 - ‘jaguar’ from motoring vs. wildlife site ?
 - Corporate search, directory search

Personalization & Web Search

- Collaborative web search exploits repetition and regularity among searchers “locally”
- Record and reuse past search session histories (Qs, reldocs) to promote previously relevant results
- Example I-Spy - no individual user profiles but community behaviour is stored and leveraged;

2. Web 2.0

- Web 2.0 ? Nobody knows, everybody has an opinion as to what it is;
- Term was popularised by Berners-Lee only in 2005;
- Refers to the second phase of the web in terms of architecture and applications and interactions with people;
- Moving from the browser as a networked interface to remotely stored information .. to .. a place where users meet, collaborate, share, and experience things together;
- Tim O'Reilly's mindmap of the concepts that make up Web 2.0 is

O'Reilly's Web 2.0

CA557 - Information Access

- O'Reilly's principles and features of Web 2.0 are:
 - Web as a platform
 - Harnessing collective intelligence of groups of people
 - Data is crucial
 - No “software release cycles”
 - Lightweight programming - lots of mash-ups
 - Software above the level of a single device
 - Rich user experiences
 - The long tail

Web 2.0 uses Mashups

CA557 - Information Access

- Lightweight programming is achieved through “mashups”;
- Suppose you’re looking for accommodation near DCU and near shops - go to daft.ie to find possibilities limited by price, availability, no rooms - and then look up each one to see how close it is to shops;
- Clever hacker can fetch apartment listings, parse contents, look up address in a gazeteer, get GPS, look up shop locations from GPS, calculate distances, rank apartments, and plot on Google maps;
- This is a *mash-up*, relatively light programming but there are few clever hackers ... so ... there is now *Yahoo! Pipes* which allows us to do this on a GUI.

Pipes: editing 'IWMW pipe' - Mozilla Firefox

File Edit View History Simpy Bookmarks Tools Help del.icio.us

pipes IWMW pipe*

Layout Expand All Collapse All Back to My Pipes New Save a copy

Sources

- Fetch CS
- Fetch Feed
- Fetch Data
- Flickr
- Google Base
- Yahoo! Local
- Yahoo! Search

User inputs

Operators

- Url
- String
- Date
- Location
- Number
- My pipes

Fetch Feed

- URL
- http://del.icio.us/rss/tag/iwmw20

Fetch Feed

- URL
- http://api.flickr.com/s

Truncate

Truncate feed after 10

Truncate

Truncate feed after 10

Regex

Use regular expression patterns here:

- Rules
- In item.title replace (.*?) with [del.icio.

Regex

Use regular expression patterns

- Rules
- In item.title replace item

Union

Sort

- Sort by

Debugger: Regex (10 items)

Time taken: 0.260048s Refresh

Web 1.0 to Web 2.0 (Tim O'Reilly)

CA557 - Information Access -

- DoubleClick --> Google AdSense
- Ofoto --> Flickr
- Akamai --> BitTorrent
- mp3.com --> Napster
- Britannica Online --> Wikipedia
- personal websites --> blogging
- page views --> cost per click
- screen scraping --> web services
- publishing --> participation
- content management systems --> wikis
- directories (taxonomy) --> tagging ("folksonomy")

3. Social Networks

- A lot of Web 2.0 is connected to social networks, an idea that long since pre-dates the web;
- We are social creatures, and when we work together in groups creates value that is greater than the sum of the parts;
- Research coalitions, University, church, girl guides, a band, a team ...
- With +1B people on the internet, social networks = “social software focused on the building and verifying of online social networks for whatever purpose” (Wikipedia).
- Surf from your list of friends to their fiends and their friends and onwards

Social Networks

- Model of information production changes from top-down, to bottom up, with consumers now being producers, enabled by new web apps;
- From information repository, to a tool for connecting people;
- 12 months ago, more than half US citizens 12 to 17 use online social networks, 55% of teens have a personal profile;
- Growth in social networks is through viral marketing - word of mouth, but successful sites \neq profitable sites, though some (Wikipedia) are not-for-profit

Social Networks - stats

- Broadcaster.com - video sharing & webcam chat - 26M
- Classmates.com - school, college groups - 40M
- Facebook.com - upload photos, tag friends - 58M
- Flickr.com - photo sharing - 4M
- Fotolog.com - photo sharing - 12M
- Friendster - search for and connect with friends and classmates - 50M
- MySpace.com - vids, movies, IM, news, blogs, chat - 217M
- Spaces.live.com - Windows live spaces blogging - 40M

Social Networks

- MySpace
 - Formed 2003 as a peer and media-based SN for 16-34 yo, users create miniature websites about themselves, bands create pages, original media is shared, equal M-F, very open in terms of signup, members retain ownership of content (bands are happy).
- Facebook
 - Launched 2004, peer relationship based SN, personal profiles customised with media snippets, many 3rd party tools & applications enabled via the backend turning it into a platform

Social Networks

- Wikipedia
 - Collaborative online encyclopedia - 9.25M articles from 70,000 authors in 250 languages - dynamic and ever-changing, each article's history can be viewed but no restriction to publishing - no adverts so no revenue, funded by donations as a public service - articles navigated via search a/o hyperlinks - susceptible to vandalism and debate about authenticity and trust;
- Youtube
 - Video upload and sharing network - video can be tagged by authors, and comments, ratings, etc. enrich content description.

Whither SNs ?

- Social networks going where ?
 - Getting more social
 - Getting more embedded
 - Getting more functional by getting more application-supportive
 - There is a future because we are social creatures;

- What's next ?

4. Web 3.0 & the Semantic Web

- Web 3.0 == the Semantic Web ?
- Most web info is text and tags - which tells is about the information between the (HTML?) tags, but not what it means.
- Semantic Web or Web 3.0 - "an extension of the current web in which information is given well-defined meaning, better enabling computers and people to work in cooperation"
T. Berners-Lee
- Requires "pages" to have metadata with underlying ontologies to link them;

Semantic Web

- Metadata elements are used to provide structure to describing some resource, at the level of a single resource
- Title, description, keywords, date, author, location, etc.
- An ontology then provides further structure
- Metadata about a book says it always has a title, author, date, etc. Metadata about a module says it always has a lecturer, number code, programme, etc.
- Isn't this an XML DTD ... yes, but in the SW, it is for text, not structured information !
- "Semantic" means "meaning of"

Semantic Web

- The Resource Description Framework (RDF) is an open markup format and is the fundamental data model of the SW, endorsed by W3C since 1999, and OWL for encoding ontologies to allow interoperability;
- RDF describes relationships between paid - A is part of B, C is owned by D, E works for F, etc.;
- These are then represented as triples, and the information content is represented as a graph of linked triples;

Semantic Web

CA 5 5 7 - I n f o r m a t i o n A c c e s s

- Graphically ... so this is a triple



- So building a small closed world in some domain means linking triples so that information retrieval turns into graph exploration and graph traversal;
- How big ? Advantage is the ease for me to publish material in RDF and my ontology in OWL.
- The challenge is in combining my ontology with yours and everyone else's - allowing interoperability by ontology mapping, commonality of expression
- The push for this comes from dissatisfaction with the quality of search results, and the pull for this comes from the search engine vendors.

Semantic Web

- But it may be too late, or may not be.
- Certainly it is a case of a couple of classic chick-and-egg problems:
 - SW vendors and developers reluctant until there is a market, no market without SW !
 - Applications require data to use, including machine-readable vocabularies for building ontologies, data owners won't release this until they see SW technologies payback.
- So it's a case of deadly embrace, but the pace of action, the interest, is growing