

CS6120: Intelligent Media Systems

Derek Bridge

Teaching and examining

- Lectures: 13.00-15.00, Fridays
- Labs: 15.15-16.15, Fridays
- Handouts: copies of slides & labs on M.Sc. web site
- Written exam:
 - 90 minutes in the Summer
 - 60 marks
- Continuous assessment:
 - PHP programming
 - 40 marks

People

- | | |
|--|-------------------------|
| ▪ Dr. Derek Bridge | Lecturer |
| ▪ Room 304, Kane Building | |
| ▪ d.bridge@cs.ucc.ie | |
| ▪ Adrian O'Riordan | Lecturer |
| ▪ CompSci Prefab | |
| ▪ a.oriordan@cs.ucc.ie | |
| ▪ Prof Barry Smyth | Possible guest lecturer |
| ▪ University College Dublin | |
| ▪ Aidan Waugh | Demonstrator |
| ▪ CompSci Prefab | |
| ▪ a.waugh@cs.ucc.ie | |

Outline

- Lecture 1: Background & motivation
 - Information overload
 - Adaptive systems
- Lecture 2: Personalization infrastructure
 - Types of personalization; user models
 - User identification; privacy
- Lecture 3: Personalized search
 - Search engines and spiders
 - Ranking algorithms
- Lecture 4: Collaborative filtering I
 - Core, user-based techniques
 - Evaluation

Outline

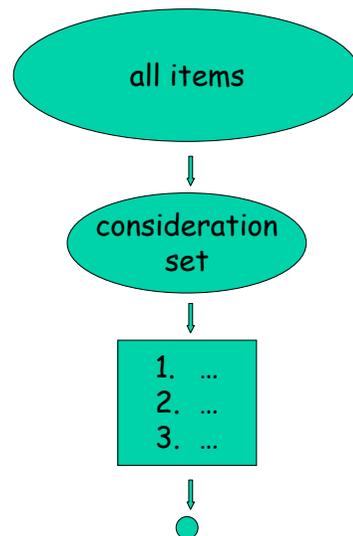
- Lecture 5: Collaborative filtering II
 - Challenges
 - Item-based techniques and hybrid systems
- Lecture 6: Recommender systems for e-commerce
 - Filter-based, similarity-based and utility-based retrieval
 - Diversity-enhanced retrieval; personalization
- Lecture 7: Conversational recommender systems for e-commerce
 - Navigation-by-asking
 - Navigation-by-proposing
- Lecture 8: Music recommender systems
 - Recommending tracks
 - Recommending play-lists

Outline

- Lecture 9: Web 2.0
 - Social navigation & networking
 - Blogs & news, feeds & syndication
- Lecture 10: Recommenders for the mobile web
 - Contextualization
 - Personalized web navigation
- Lecture 11: No lecture
 - Good Friday
- Lecture 12: Conclusions
 - Future prospects
 - Revision

Consumer decision making (idealized)

1. Perception of need
2. Establish criteria
3. Discovery of options
4. Identify consideration set
5. Compare items
6. Choose winner
7. Consume item
8. Evaluate item
9. Learn new needs and criteria



Which movie should I watch?



- The Internet Movie Database (IMDb)
 - Launched 1990; owned by Amazon since 1998
 - Database statistics

Which book should I buy?

Amazon.co.uk Bestsellers: The most popular items in Books. Updated hourly.

More in Books

Bestsellers

Hot Future Releases

Most Gifted

Most Wanted For

Movers & Shakers

Narrow by Category

Books

Art, Architecture & Photography

Audio CDs

Audio Cassettes

Biography

Business, Finance & Law

Calendars, Diaries, Address & More

Children's Books

Comics & Graphic Novels

Computers & Internet

Crime, Thrillers & Mystery

Fiction

Food & Drink

Gay & Lesbian

Health, Family & Lifestyle

History

Home & Garden

Bestsellers in Books

The most popular items in Books. Updated hourly.

- Nigella Express**

by Nigella Lawson (Author)

4.5 (14 customer reviews) | 1 customer question

In stock

RRP: £25.00

Price: **£11.25**

You Save: £13.75 (55%)

27 used & new from £10.41

[Add to Basket](#) [Add to Wish List](#)
- Do Ants Have Arseholes?: And 101 Other Bloody Ridiculous Questions**

by Jon Butler; Bruno Vincent (Author)

4.5 (22 customer reviews) | 4 customer questions

In stock

RRP: £7.99

Price: **£3.94**

You Save: £4.05 (51%)

22 used & new from £3.06

[Add to Basket](#) [Add to Wish List](#)
- My Booky Wook**

by Russell Brand (Author)

4.5 (24 customer reviews) | 2 customer questions

In stock

RRP: £19.99

Price: **£9.49**

[Add to Basket](#) [Add to Wish List](#)

Which papers should I read?

Computer and Information Science Papers CiteSeer Publications ResearchIndex

SPONSORS

NATIONAL SCIENCE FOUNDATION

Microsoft Research

NASA

Interested in sponsoring CiteSeer? [Contact](#)

CiteSeer.IST
Scientific Literature Digital Library

[CiteSeer\(Docs\)](#) [Google\(Docs\)](#) [Citations](#) [Acknowledgments](#)

Documents indexed by CiteSeer.IST

Mirrors of CiteSeer are available at the following locations:
[U. of Kansas](#) [MIT](#) [U. of Zurich](#) [National U. of Singapore](#)

[Submit Documents](#) | [Statistics](#) | [Help](#) | [CiteSeer Metadata](#) | [Announcements](#)
[Copyright NEC and Penn State](#) | [Privacy Policy](#) | [About](#) | [Feedback](#)

Searching 767,558 documents.

Hosted by Penn State's [College of Information Sciences and Technology](#)

Which news articles shall I read?

Google News Ireland - World

Web Images News Groups Books Gmail more

Search News Search the Web

Search and browse 4,500 news sources updated continuously.

News archive search | Advanced news search | Blog search

Top Stories

World

European Union Leaders Sign Bloc's New Treaty Amid Public Doubts
FOX News - 1 hour ago
LISBON, Portugal - European Union leaders signed a new treaty on Thursday that would give the 27-nation bloc a long-term president and streamline its decision-making process.
[Brown files to Lisbon for belated EU treaty signing](#) Guardian Unlimited
[The European treaty signs the apocalypse](#) Economist
[AFP - Bloomberg - Telegraph.co.uk - The Canadian Press](#) all 562 news articles.

Bitter Divisions Exposed at Climate Talks
New York Times - 1 hour ago
By THOMAS FULLER and ELISABETH ROSENTHAL NUSA DUA, Indonesia - Amid growing frustration with the United States in deadlocked negotiations at a United Nations conference on global warming, the European Union threatened Thursday to boycott separate talks...
[AI Qora lays blame for Bali stalemate on US](#) Reuters
[AI Qora lambastes US climate negotiations](#) Times Online
[CNN International - Washington Post - AFP - Canada.com](#) all 1,456 news articles.

Men questioned over Lebanon blast
BBC News - 87 minutes ago
Lebanese police are questioning several men after Wednesday's car bombing in Beirut which killed a general expected to become army chief, officials say.
[Recent attacks in Lebanon](#) Guardian Unlimited
[Lebanese Army investigates Slaying of Top General](#) Voice of America
[Yalowitz - AFP - Aljazeera.net](#) all 1,456 news articles.

Duke of Edinburgh's 'scarier' letters revealed at Diana inquest
Guardian Unlimited - 1 hour ago
Extracts from affectionate "Dearest Pig" letters between Diana, Princess of Wales, and the Duke of Edinburgh in which she thanked him for his marriage guidance skills were revealed at the inquest into her death today.
[Jury sees Duke's letters to Diana](#) BBC News
[Diana inquest shows 'Dearest D' letters](#) Times Online
[Reuters UK - AFP - The Associated Press - huffmagazine.com](#) all 500 news articles.

EU nations poised to go it alone over Kosovo
Guardian Unlimited - 1 hour ago
The EU's biggest four countries are pushing to impose and oversee independence in Kosovo without a fresh UN mandate, risking a showdown with a resurgent Russia and fierce resistance from Serbia.

Where should I take my vacation?

Reviews of holidays, hotels, resorts, holiday and travel packages - TripAdvisor

10,000,000+ traveller reviews & opinions of hotels, holidays & more
get the truth, then go™

18,078,725 Travelers From 191 Countries Planned Trips Here This Week!

Search:

Plan Your Next Trip!

Find Hotels | Flights | Read & Write Reviews

The best hotels based on millions of reviews from travellers like you.

Town / City:

Check-in: 14 Dec | Check-out: 18 Dec | Price level: Adults | Currency: Euro

Browse Destinations

Europe | Middle East | Africa | Asia | United States | Canada | Caribbean | South Pacific | Mexico | Central America | South America

Rants & Raves

The good, the bad and the ugly: real stories from real travellers

[Howard Johnson Express - Redding, California](#) [GrabSkry Kanuhara Kanuhara, Lhanyisi Atoll](#)

"...We have just returned from a fabulous trip to Kanuhara. There is nowhere quite like it in the world. It was nearly sick. The mattress was filthy and the carpets were full of bum holes." [Read More](#)

2 of 10,000,000+ unbiased reviews and opinions you can trust

Show your friends where you've travelled.

Free Newsletter

Subscribe to our free weekly TripWatch newsletter!

Get the latest deals, reviews & articles Tailored for the destinations you choose

Inside Pages Written By Our Members

Planning a trip? Read TripAdvisor Inside™, where real travellers like you have

TripAdvisor.com

- "...features real advice from real travellers, with more than 5 million unbiased reviews and opinions, covering 250,000+ hotels and attractions"

'Shopping'

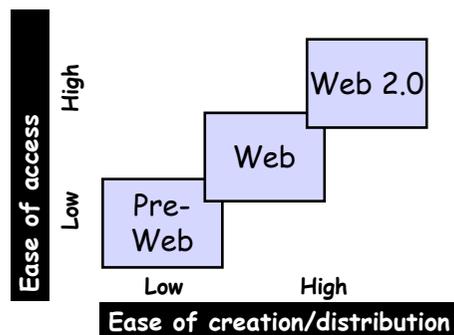
- ...in the material world
 - Physical limitations on range of items available
 - Irreproducibility or limits on reproducibility of items
 - Can experience the item (test drive, try it on, try it out, taste it,...)
 - Guidance from salesperson or other intermediary (librarian, travel agent, tutor,...)
 - ...online
 - Lower physical limitations means wider choice
 - Some items (digital ones) are infinitely reproducible
 - Cannot experience physical items; but may be able to play samples of digital items
 - No human intermediary ("disintermediation")
- Greater reliance on information

Information about items

- Types of information about items
 - Descriptions
 - Technical
 - Lifestyle
 - Reviews
 - Expert
 - End-user
 - Meta-descriptions
 - Reliability
 - Reputation
- Format of information about items
 - Multimedia: text, images, video, audio

Information overload

- Increased volume
 - More items
 - More information about items
 - More information about information
- Increased access
 - Browsing
 - Search engines
 - Alerts, feeds, ...
 - Ubiquitous access



Estimating the scale of the problem

- "How much information?"
 - "...an attempt to estimate how much new information is created each year."
 - Peter Lyman, Hal R. Varian & others at the University of California at Berkeley
 - 1999
 - 2 exabytes (10^{18} bytes)
 - Print is only .003%
 - Perhaps 99.96% created by individuals
 - 2002
 - 5 exabytes
 - Print is only .001%
- <http://www2.sims.berkeley.edu/research/projects/how-much-info-2003/>

Estimating the scale of the problem

- A quick experiment
 - Issue queries to Google and count the results
 - Assumes Google's indexes are representative of the Web as a whole
 - Assumes that all results are relevant

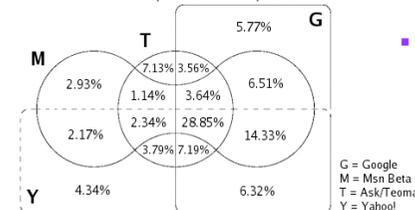
Query	Results
ireland football	1,070,000
lung cancer	6,030,000
irish tourism	2,170,000
search queries	9,780,000
information overload	276,000
ucc	2,410,000
ucd	389,000
train timetables	288,000
george orwell	344,000
united states	694,000,000
china	74,800,000
sun holiday	7,450,000

Estimating the scale of the problem

- Surface web
 - Indexed by web search engines
- Deep web (invisible web, hidden web)
 - especially dynamic web pages
 - typically not indexed by web search engines
 - estimated in year 2000 to be 500 times larger than the surface web
<http://www.brightplanet.com/resources/details/deepweb.html>
 - See also B. He et al (2007): Accessing the deep web, Communications of the ACM, vol.50(5), pp.94-101
 - Try the Turbo10 search engine

Estimating the scale of the problem

- "The Indexable Web is 11.5 Billion Pages"
 - A.Gulli & A.Signorini, 2005
 - Using the method of K. Bharat & A. Broder, 1998
 - Issue a random query to a search engine A
 - Choose a URL, u, from the result list
 - Visit u and construct queries from its content
 - See whether u appears in the result list when you issue these queries to search engine B
 - On the basis of how often this does/does not happen, we get an estimate of the overlap/non-overlap in A's and B's indexes



- Of the 11.5 billion pages
 - Google indexes 76.2%
 - Yahoo indexes 69.3%
 - MSN indexes 61.9%
 - Ask/Teoma indexes 57.6%

Adaptive systems

- In the face of information overload,
 - abandon one-size-fits-all
 - develop *adaptive systems*, which tailor content, presentation and behaviour to their context of use
- E.g.
 - *adaptive search engines*: in result lists, promote items that are more relevant to this user
 - *adaptive hypermedia*: select and highlight content, inc. hyperlinks, that are of more interest to this user
 - *adaptive recommender systems*: recommend items that best satisfy user's goals and preferences

Personalization

- In particular, develop *personalization technologies*, which adapt to the user
- "...personalization tailors certain offerings (such as content, services, product recommendations, communications, and e-commerce interactions) by providers (such as e-commerce Web sites) to consumers (such as customers and visitors) based on knowledge about them, with certain goal(s) in mind."
[Adomavicius & Tuzhilin 2005, p.84]
- G. Adomavicius & A. Tuzhilin (2005): *Personalization Technologies: A Process-Oriented Perspective*, Communications of the ACM, vol.48(10), pp.83-90
- But be aware that personalization is not the only form of adaptation: a system can adapt to time, place, consumer's companions, consumer's device,...

Google News: How it works

- **Ranking**
 - timeliness of the article
 - whether it is an original piece
 - placement by the editors on the source page
 - apparent scope and impact
 - popularity (click rate of the article by users of Google News)
 - authority of the source based on
 - number of non-duplicate stories
 - length of articles
 - breadth of articles
 - number of important/breaking stories
 - click rate of the site by users of Google News
 - average quality of the writing
 - PageRank of the source's website
 - real world data on the news source (e.g. number of employees)
- **Agglomerative hierarchical clustering**

Case study: Google News

- *Google News is a news aggregator*
 - developed by Krishna Bharat in 2001
 - now different versions available in 42+ regions and 17+ languages
 - English-language versions cover the last 30 days' articles from 4,500 sites
 - ranking and clustering of articles is automatic
 - but humans select the source sites
 - accusations of bias, censorship, failure to censor, illegal reuse,...



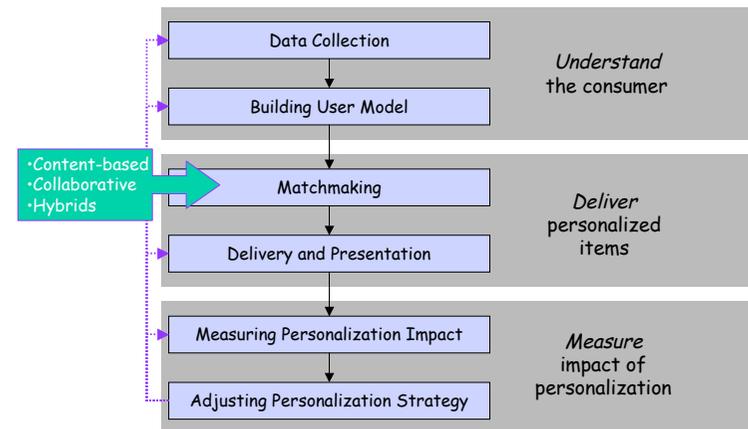
Personalization

- **User can choose**
 - standard sections (even from different versions)
 - new sections (based on keywords)
 - number of items per section
 - layout (JavaScript drag-and-drop)

Personalization

- **Recommended stories**
 - different recommendations for different users
- **It records your selections**
 - hence, when using Google News, you must sign-in to your Google account and you must have "Personalized Search" enabled
 - the more you use Google News in this fashion, the better the recommendations become
- **It recommends stories to you that similar users have selected**
 - a similar user is one who has selected stories similar to the ones you have selected
 - A. Das et al (2007): Google News Personalization: Scalable Online Collaborative Filtering, Procs. Of the 16th International World Wide Web Conference, pp.271-280

Personalization process model



[Adomavicius & Tuzhilin 2005, p.84]

Consumers' objectives

- **Better outcomes**
 - better information about items
 - better items chosen for consumption
- **Better interaction**
 - more efficient
 - more confidence-inspiring
 - more enjoyable experience
- **Prestige**
- **Social interaction**

Providers' objectives

- **Increase conversion rate (visit-to-buy ratio)**
 - measured by volume
 - measured by value
- **Make interactions more efficient**
 - pre-sales, sales and post-sales
- **Increase likelihood of repeat visits/customer loyalty**
- **May depend on their business model**
 - Are consumers charged per use, per transaction, by subscription?
 - Or is revenue generated via some third-party, e.g. from ad click-through?
 - The business model raises questions about bias: how much providers want to influence recommendations