Web Personalisation and Recommender Systems

Shlomo Berkovsky and Jill Freyne

DIGITAL PRODUCTIVITY FLAGSHIP
www.csiro.au

Outline

Part 1: Information Overload and User Modelling

Part 2: Web Personalisation and Recommender Systems

Part 1:
Information Overload and User Modelling

Information Overload

Information Overload

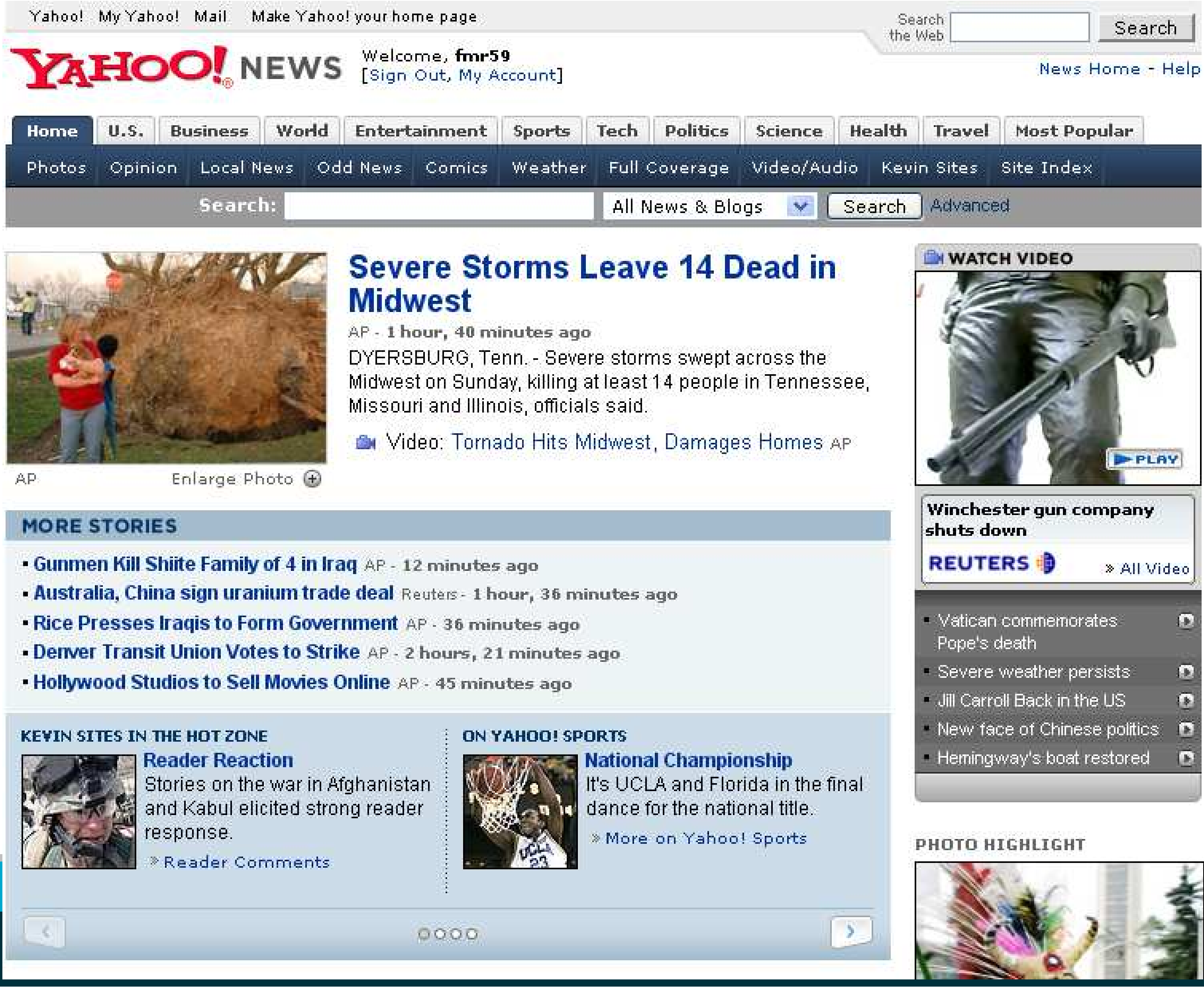
- Information presented at a rate too fast for a person to process

- The state of having too much information to make a decision or remain informed about a topic

Online Information Overload

- Every time we go online, we are overwhelmed by the available options
  - Web Search….which search result is most relevant to my needs?
  - Entertainment….which movie should I download? which restaurant should I eat at?
  - E-commerce….which product is best for me? what's on special now? which holiday will I enjoy most?
  - News….which news stories are most interesting to me? what happened in US last night?
  - Health….which food is healthy for me? which types of exercise should I try? what doctor can I trust?

What news should I read?


Yahoo! My Yahoo! Mail Make Yahoo! your home page

Search the Web | Search

YAHOO! NEWS Welcome, fmr59 [Sign Out, My Account]

News Home - Help

Home | U.S. | Business | World | Entertainment | Sports | Tech | Politics | Science | Health | Travel | Most Popular

Photos | Opinion | Local News | Odd News | Comics | Weather | Full Coverage | Video/Audio | Kevin Sites | Site Index

Search: | All News & Blogs | Search | Advanced

AP Enlarge Photo

Severe Storms Leave 14 Dead in Midwest

AP - 1 hour, 40 minutes ago

DYERSBURG, Tenn. - Severe storms swept across the Midwest on Sunday, killing at least 14 people in Tennessee, Missouri and Illinois, officials said.

Video: Tornado Hits Midwest, Damages Homes AP

WATCH VIDEO

PLAY

Winchester gun company shuts down

REUTERS » All Video

- Vatican commemorates Pope's death
- Severe weather persists
- Jill Carroll Back in the US
- New face of Chinese politics
- Hemingway's boat restored

MORE STORIES

- Gunmen Kill Shiite Family of 4 in Iraq AP - 12 minutes ago
- Australia, China sign uranium trade deal Reuters - 1 hour, 36 minutes ago
- Rice Presses Iraqis to Form Government AP - 36 minutes ago
- Denver Transit Union Votes to Strike AP - 2 hours, 21 minutes ago
- Hollywood Studios to Sell Movies Online AP - 45 minutes ago

KEVIN SITES IN THE HOT ZONE

Reader Reaction

Stories on the war in Afghanistan and Kabul elicited strong reader response.

» Reader Comments

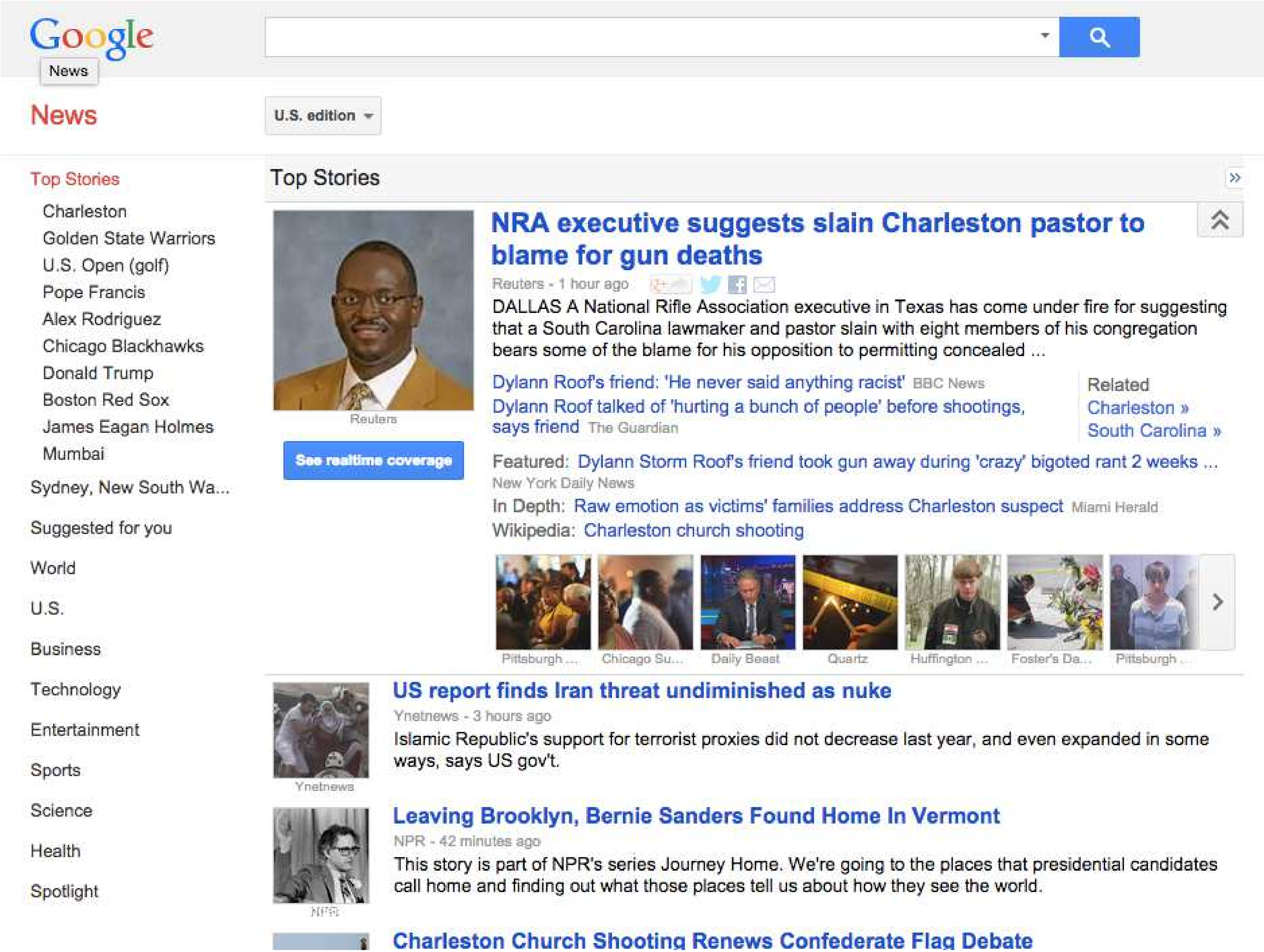
ON YAHOO! SPORTS

National Championship

It's UCLA and Florida in the final dance for the national title.

» More on Yahoo! Sports

PHOTO HIGHLIGHT



Google
News

News

U.S. edition

Top Stories
- Charleston
- Golden State Warriors
- U.S. Open (golf)
- Pope Francis
- Alex Rodriguez
- Chicago Blackhawks
- Donald Trump
- Boston Red Sox
- James Eagan Holmes
- Mumbai

Sydney, New South Wa...

Suggested for you

World

U.S.

Business

Technology

Entertainment

Sports

Science

Health

Spotlight

Top Stories

Reuters

See realtime coverage

NRA executive suggests slain Charleston pastor to blame for gun deaths

Reuters - 1 hour ago

DALLAS A National Rifle Association executive in Texas has come under fire for suggesting that a South Carolina lawmaker and pastor slain with eight members of his congregation bears some of the blame for his opposition to permitting concealed ...

Dylann Roof's friend: 'He never said anything racist' BBC News

Dylann Roof talked of 'hurting a bunch of people' before shootings, says friend The Guardian

Related
Charleston »
South Carolina »

Featured: Dylann Storm Roof's friend took gun away during 'crazy' bigoted rant 2 weeks ... New York Daily News

In Depth: Raw emotion as victims' families address Charleston suspect Miami Herald

Wikipedia: Charleston church shooting

Pittsburgh ... Chicago Su... Daily Beast Quartz Huffington ... Foster's Da... Pittsburgh ...

US report finds Iran threat undiminished as nuke

Ynetnews

Ynetnews - 3 hours ago

Islamic Republic's support for terrorist proxies did not decrease last year, and even expanded in some ways, says US gov't.

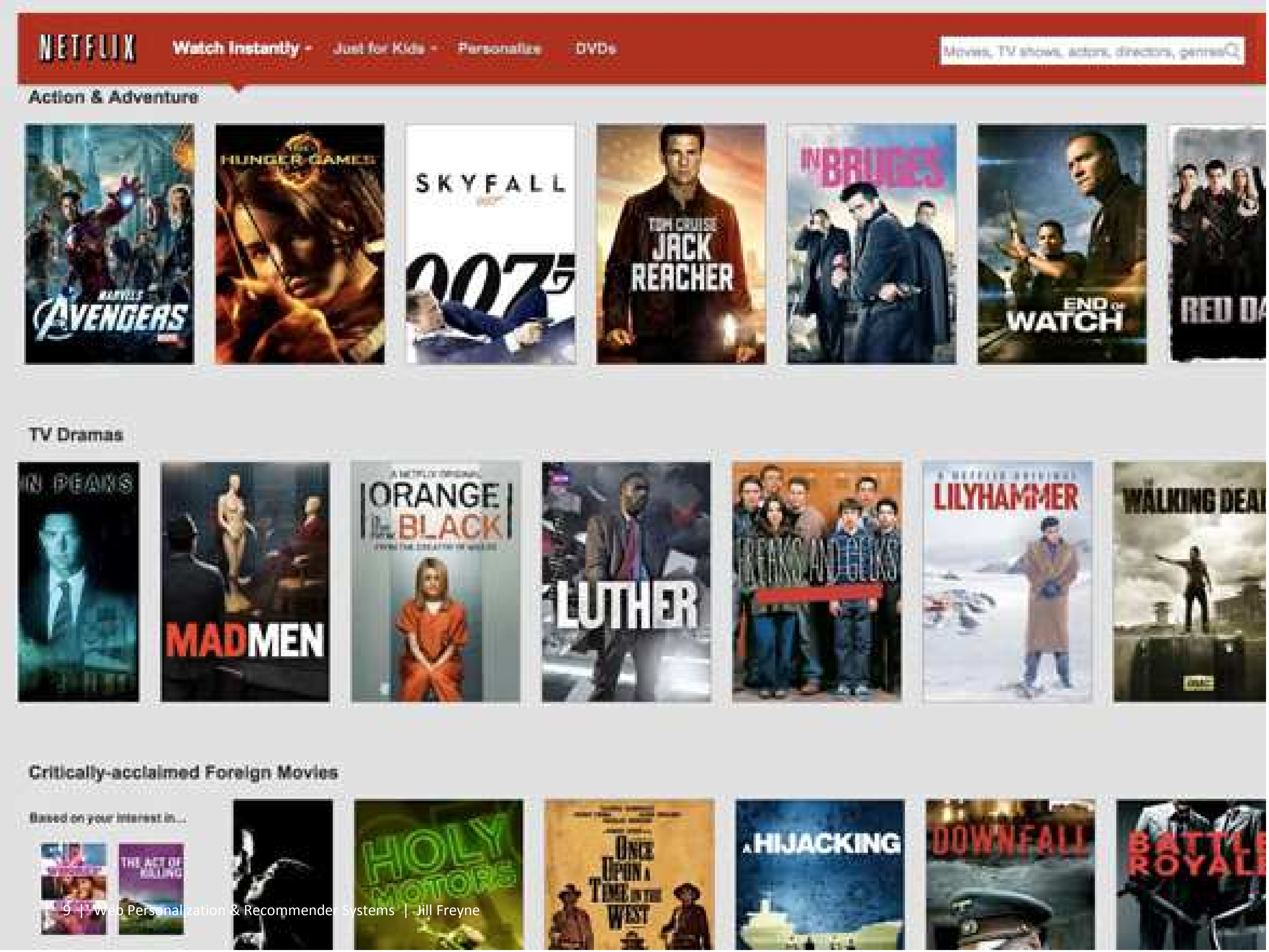
Leaving Brooklyn, Bernie Sanders Found Home In Vermont

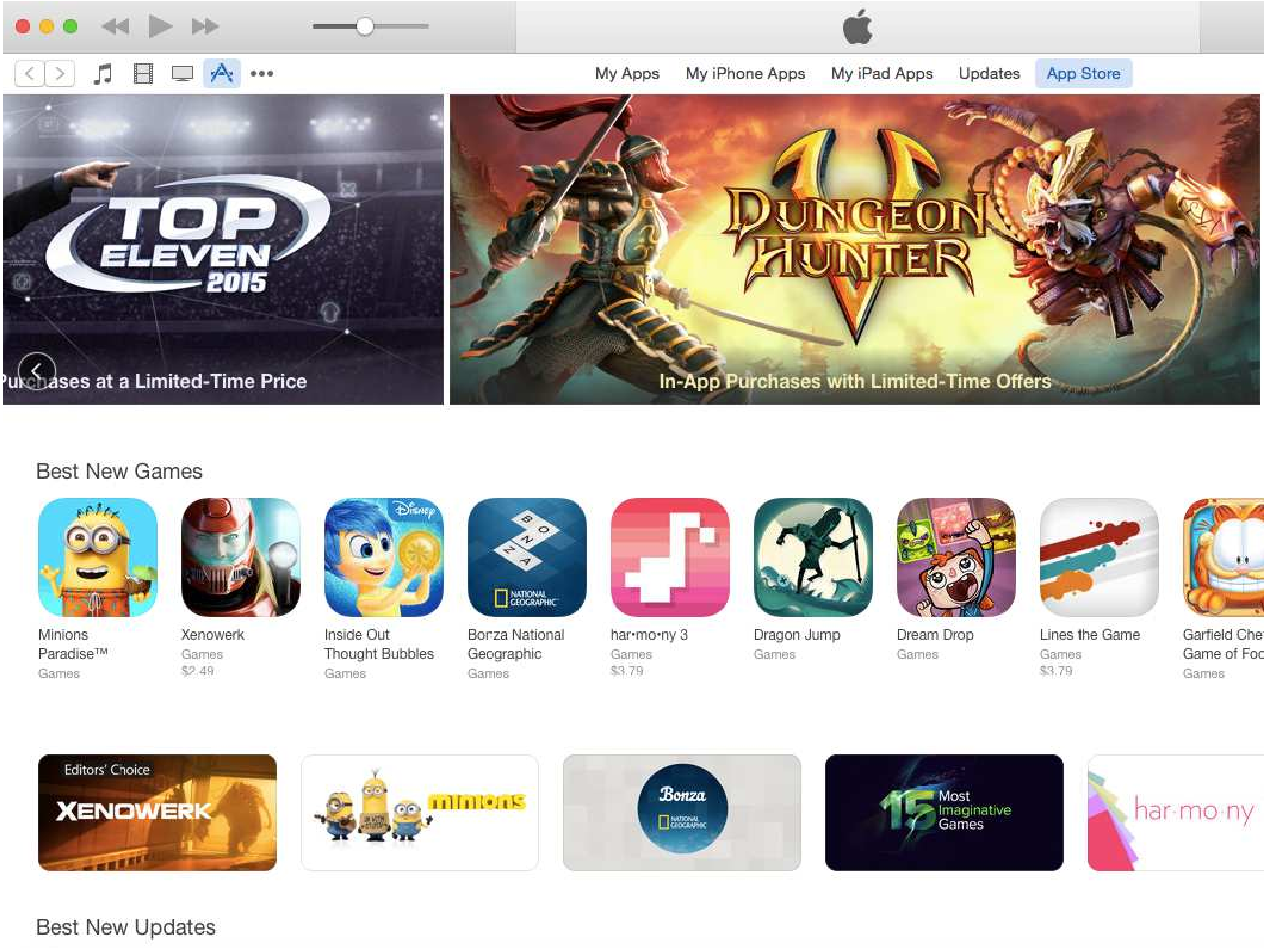
NPR

NPR - 42 minutes ago

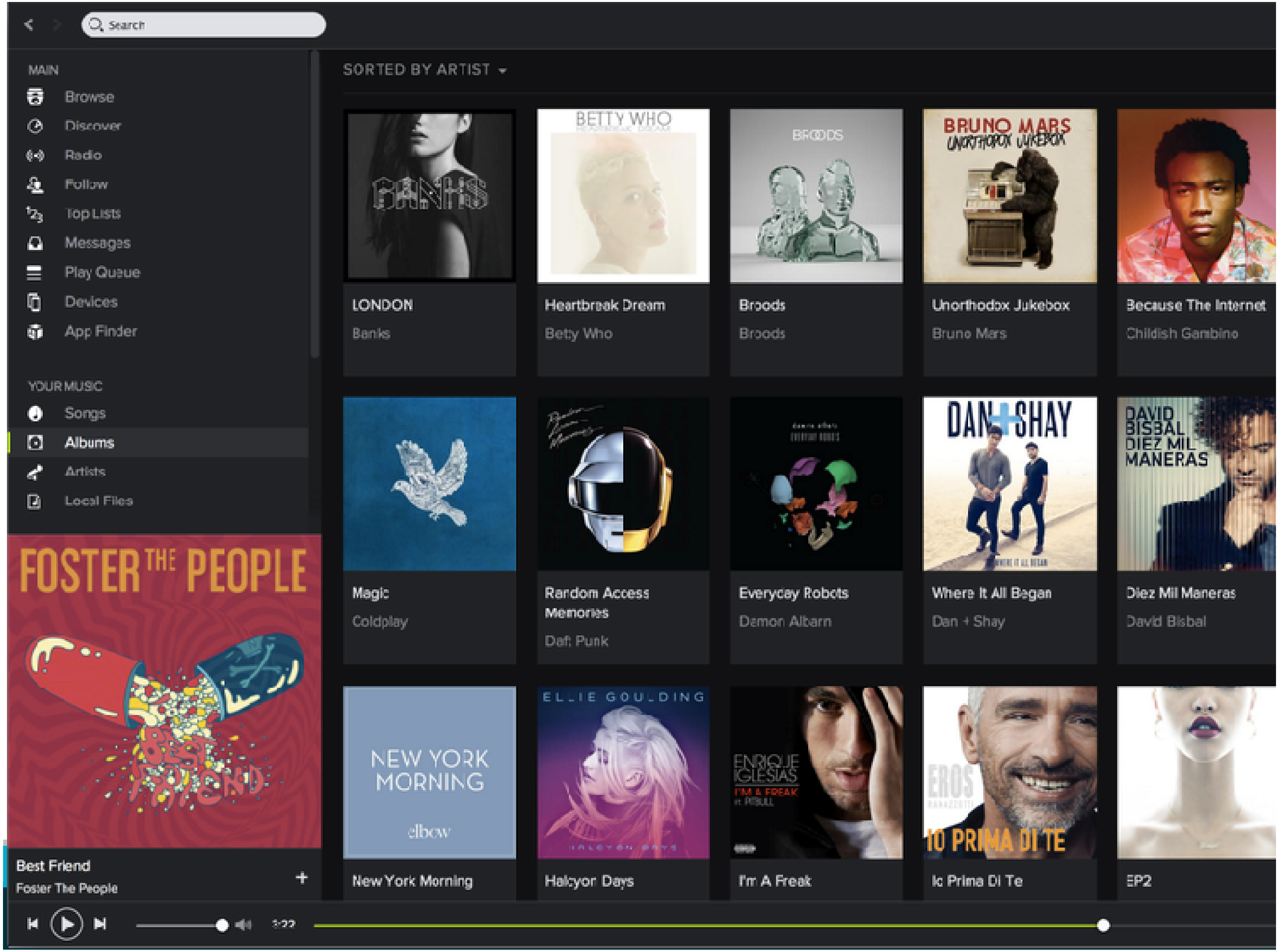
This story is part of NPR's series Journey Home. We're going to the places that presidential candidates call home and finding out what those places tell us about how they see the world.

Charleston Church Shooting Renews Confederate Flag Debate

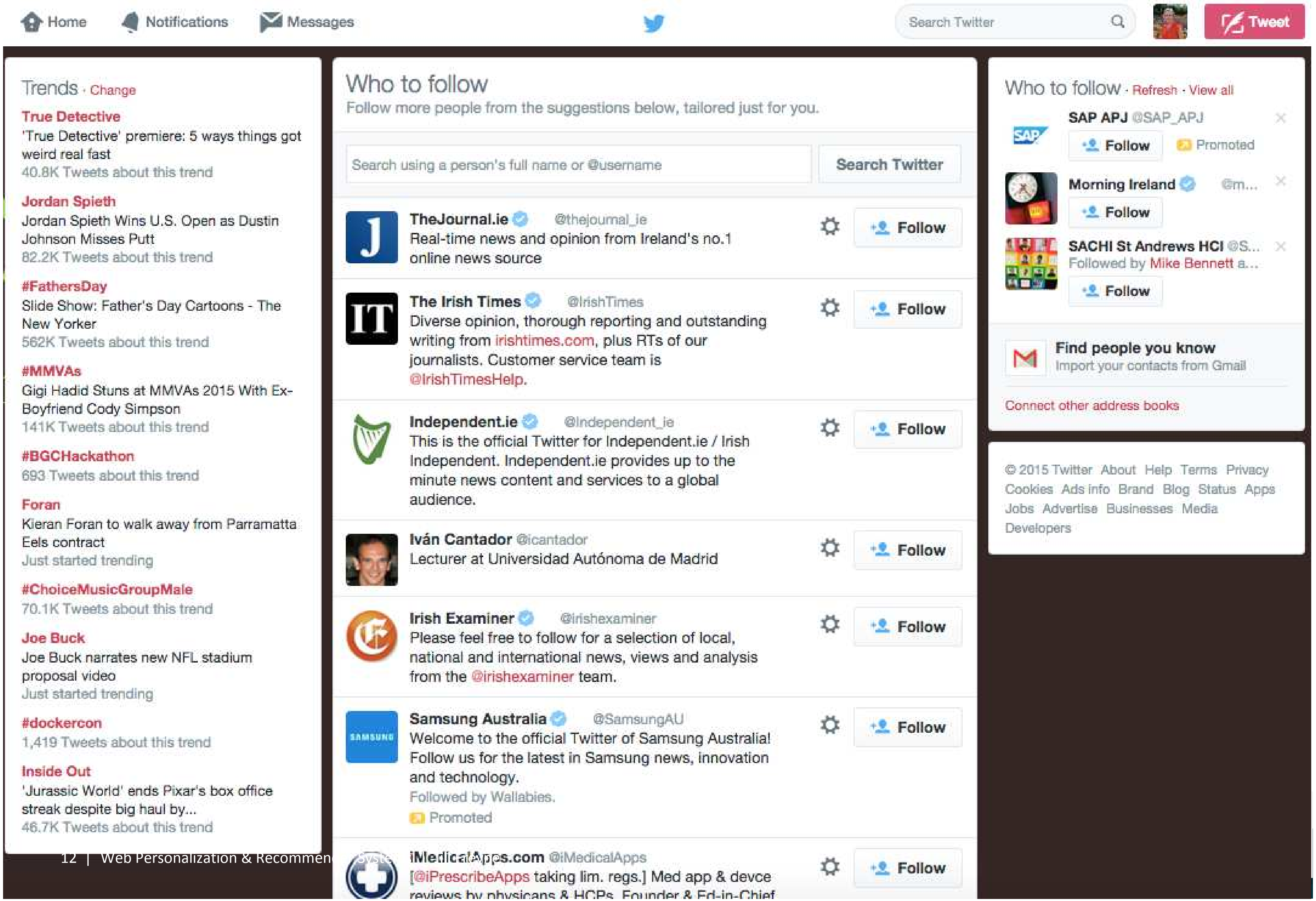

NETFLIX
Watch Instantly
Just for Kids
Personalize
DVDs
Movies, TV shows, actors, directors, genres
Action & Adventure
AVENGERS
SKYFALL
JACK REACHER
END OF WATCH
TV Dramas
MADMEN
ORANGE is the new BLACK
LUTHER
LILYHAMMER
Critically-acclaimed Foreign Movies
Based on your interest in...
HOLY MOTORS
A HIJACKING
DOWNFALL

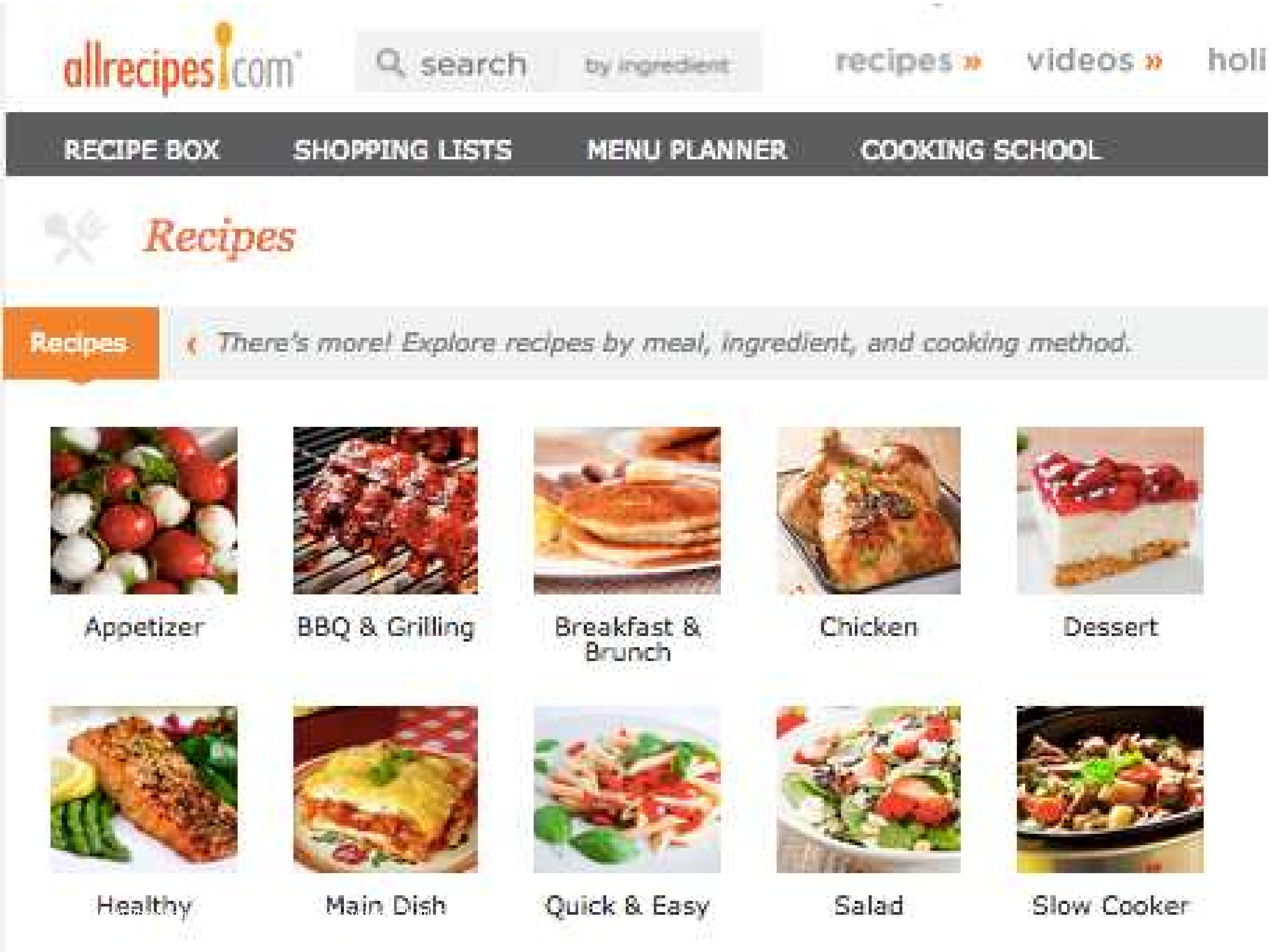


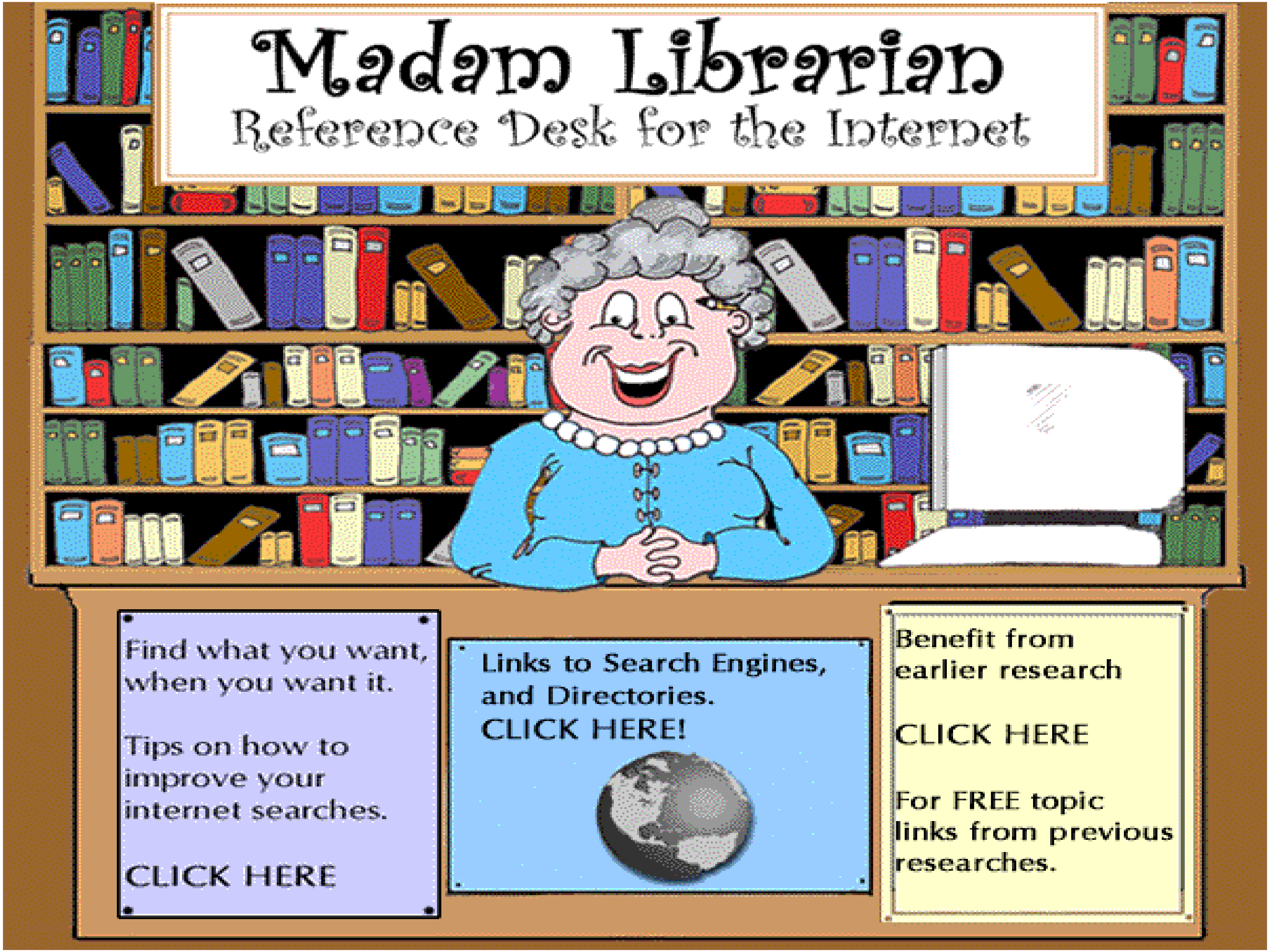
My Apps
My iPhone Apps
My iPad Apps
Updates
App Store
TOP ELEVEN 2015
Purchases at a Limited-Time Price
DUNGEON HUNTER V
In-App Purchases with Limited-Time Offers
Best New Games
Minions Paradise™
Games
Xenowerk
Games
$2.49
Inside Out Thought Bubbles
Games
Bonza National Geographic
Games
har•mo•ny 3
Games
$3.79
Dragon Jump
Games
Dream Drop
Games
Lines the Game
Games
$3.79
Garfield Che
Game of Foc
Games
Editors' Choice
XENOWERK
minions
Bonza
15 Most Imaginative Games
har·mo·ny
Best New Updates



Search
MAIN
Browse
Discover
Radio
Follow
Top Lists
Messages
Play Queue
Devices
App Finder
YOUR MUSIC
Songs
Albums
Artists
Local Files
FOSTER THE PEOPLE
Best Friend
Foster The People
SORTED BY ARTIST
LONDON
Banks
Heartbreak Dream
Betty Who
Broods
Broods
Unorthodox Jukebox
Bruno Mars
Because The Internet
Childish Gambino
Magic
Coldplay
Random Access Memories
Daft Punk
Everyday Robots
Damon Albarn
Where It All Began
Dan + Shay
Diez Mil Maneras
David Bisbal
New York Morning
Halcyon Days
I'm A Freak
Io Prima Di Te
EP2


Home
Notifications
Messages
Search Twitter
Tweet
Trends · Change
True Detective
'True Detective' premiere: 5 ways things got weird real fast
40.8K Tweets about this trend
Jordan Spieth
Jordan Spieth Wins U.S. Open as Dustin Johnson Misses Putt
82.2K Tweets about this trend
#FathersDay
Slide Show: Father's Day Cartoons - The New Yorker
562K Tweets about this trend
#MMVAs
Gigi Hadid Stuns at MMVAs 2015 With Ex-Boyfriend Cody Simpson
141K Tweets about this trend
#BGCHackathon
693 Tweets about this trend
Foran
Kieran Foran to walk away from Parramatta Eels contract
Just started trending
#ChoiceMusicGroupMale
70.1K Tweets about this trend
Joe Buck
Joe Buck narrates new NFL stadium proposal video
Just started trending
#dockercon
1,419 Tweets about this trend
Inside Out
'Jurassic World' ends Pixar's box office streak despite big haul by...
46.7K Tweets about this trend
Who to follow
Follow more people from the suggestions below, tailored just for you.
Search using a person's full name or @username
Search Twitter
TheJournal.ie @thejournal_ie
Real-time news and opinion from Ireland's no.1 online news source
Follow
The Irish Times @IrishTimes
Diverse opinion, thorough reporting and outstanding writing from irishtimes.com, plus RTs of our journalists. Customer service team is @IrishTimesHelp.
Follow
Independent.ie @Independent_ie
This is the official Twitter for Independent.ie / Irish Independent. Independent.ie provides up to the minute news content and services to a global audience.
Follow
Iván Cantador @icantador
Lecturer at Universidad Autónoma de Madrid
Follow
Irish Examiner @irishexaminer
Please feel free to follow for a selection of local, national and international news, views and analysis from the @irishexaminer team.
Follow
Samsung Australia @SamsungAU
Welcome to the official Twitter of Samsung Australia! Follow us for the latest in Samsung news, innovation and technology.
Followed by Wallabies.
Promoted
Follow
iMedicalApps.com @iMedicalApps
[@iPrescribeApps taking lim. regs.] Med app & devce
Follow
Who to follow · Refresh · View all
SAP APJ @SAP_APJ
Follow
Promoted
Morning Ireland @m...
Follow
SACHI St Andrews HCI @S...
Followed by Mike Bennett a...
Follow
Find people you know
Import your contacts from Gmail
Connect other address books
© 2015 Twitter About Help Terms Privacy
Cookies Ads info Brand Blog Status Apps
Jobs Advertise Businesses Media
Developers


allrecipes.com
search
by ingredient
recipes »
videos »
holi
RECIPE BOX
SHOPPING LISTS
MENU PLANNER
COOKING SCHOOL
Recipes
Recipes
There's more! Explore recipes by meal, ingredient, and cooking method.
Appetizer
BBQ & Grilling
Breakfast & Brunch
Chicken
Dessert
Healthy
Main Dish
Quick & Easy
Salad
Slow Cooker


Madam Librarian
Reference Desk for the Internet
Find what you want, when you want it.
Tips on how to improve your internet searches.
CLICK HERE
Links to Search Engines, and Directories.
CLICK HERE!
Benefit from earlier research
CLICK HERE
For FREE topic links from previous researches.

Personalisation

Personalisation is…

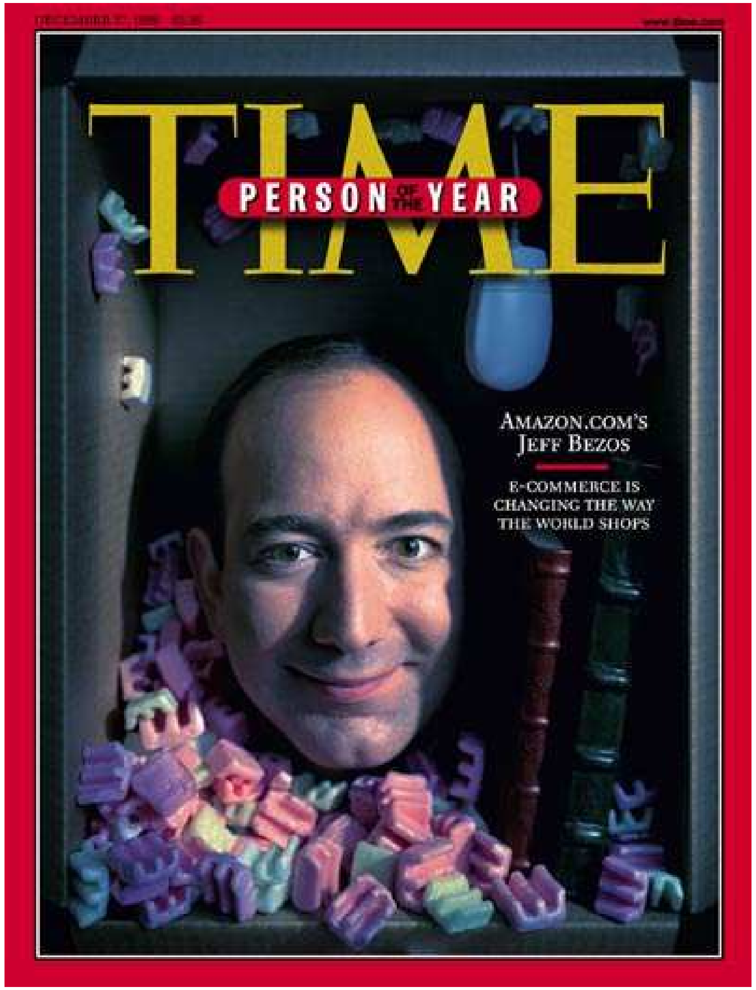
- "… the ability to provide content and *services tailored to individuals* based on knowledge about their preferences and behavior" (tools and information)

- "… the capability to *customize customer communication* based on preferences and behaviors at the time of interaction [with the customer]" (communication)

- "… about *building customer loyalty and meaningful one-to-one relationship*; by understanding the needs of each individual and helping satisfy a goal that efficiently and knowledgeably addresses the individual's need in a given context" (customer relationships)

Amazon and Personalisation

- Jeff Bezos, Amazon CEO
  - Credited with changing the way the world shops
  - Among the first to deploy large-scale personalisation online
- *"If I have 3 million customers on the Web, I should have 3 million stores on the Web"*



For Example…

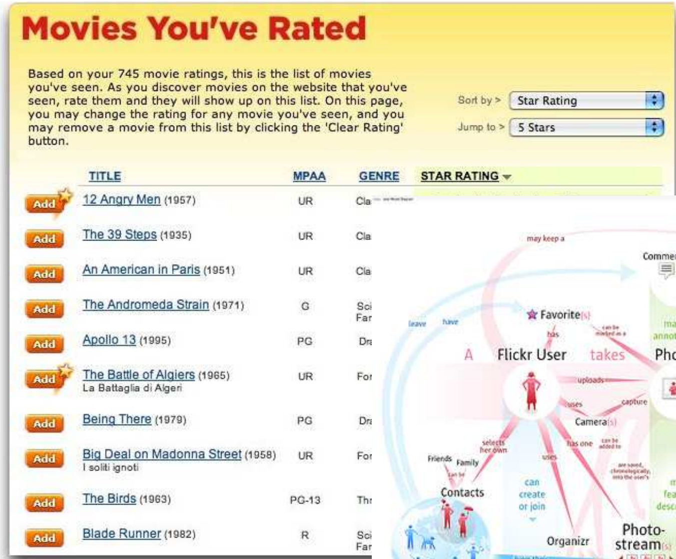
- Amazon maintains shopper profiles
  - Based on products and past interactions
    - Purchased products, feedback, wish list, items browsed, …
- Amazon provides personalised recommendations for items to purchase
  - Instead of showing random or popular or discounted items

How is Personalisation Achieved?

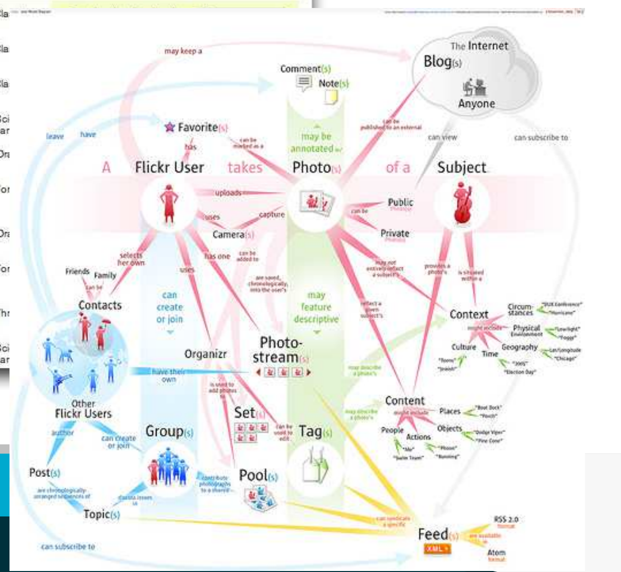
1. Gathering information about the users
   - Explicitly – through direct user input
   - Implicitly – through monitoring user interactions

2. Exploiting this information to create the *user model*
   - Dynamic vs. Static
   - Short term vs. Long term

3. Use the model to adapt some aspects of the system to reflect user needs, interests, or preferences



	TITLE	MPAA	GENRE	STAR RATING
Add	12 Angry Men (1957)	UR	Cla	
Add	The 39 Steps (1935)	UR	Cla	
Add	An American in Paris (1951)	UR	Cla	
Add	The Andromeda Strain (1971)	G	Sci Far	
Add	Apollo 13 (1995)	PG	Dr	
Add	The Battle of Algiers (1965) La Battaglia di Algeri	UR	For	
Add	Being There (1979)	PG	Dr	
Add	Big Deal on Madonna Street (1958) I soliti ignoti	UR	For	
Add	The Birds (1963)	PG-13	Thr	
Add	Blade Runner (1982)	R	Sci Far	



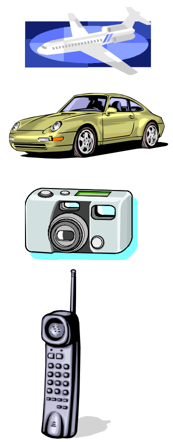
Framework for Personalisation

Mixed-Initiative Systems
Mass Customization
Adaptive Hypermedia
Recommender Systems

Interface
Interaction
Functions
Content

1
2
3
4
5
6

Buy
View
Search
Store
Compare
Select
…



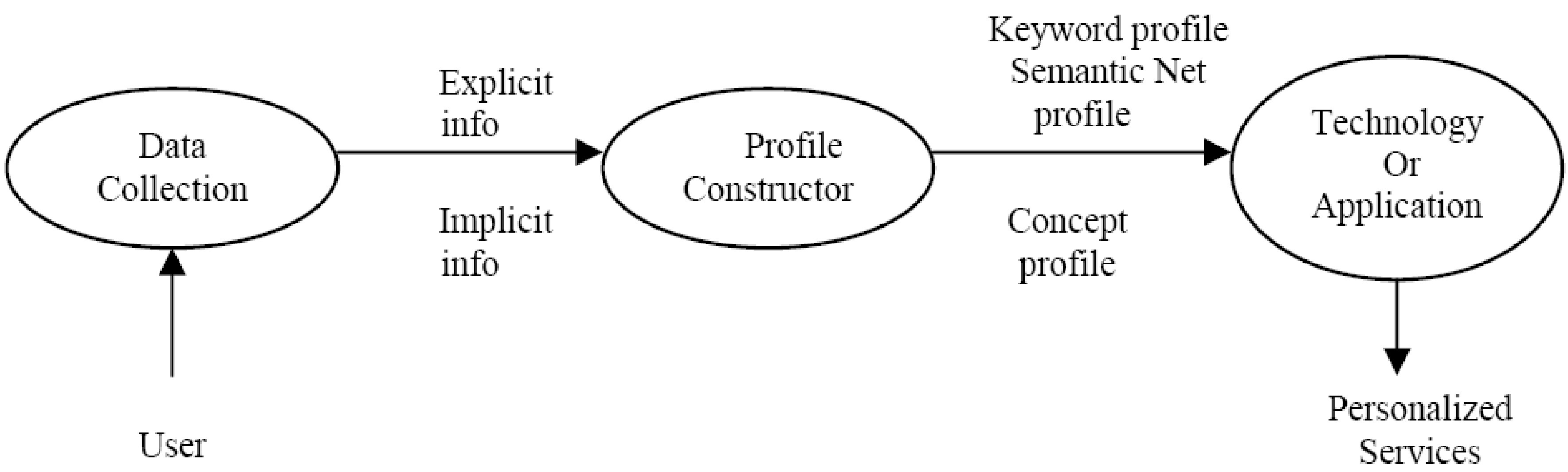
User Models

User Modelling and Personalisation

- People leave traces on the internet...
  - What pages do they visit? How long do they visit for?
  - What search queries are they using?
  - What products do they buy?
  - What movies do they download?
  - Who are their online friends?
- User modelling is about making sense of this data
  - to gain an understanding of the characteristics, preferences, and needs of an individual user
- Personalisation exploits user models
  - to filter information and provide personalised services
    - – that match the user's needs

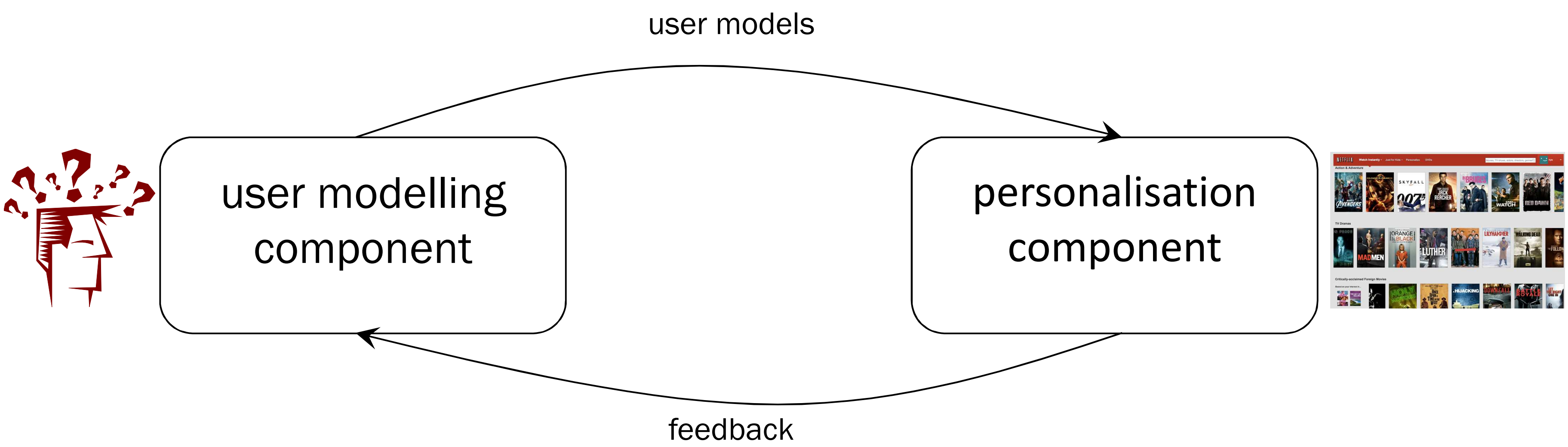
User Model Based Personalisation

- 3 stages
  - User information collection
  - User profile construction
  - Exploitation of profile for personalisation



User Model Based Personalisation

- Two stages
  - User model construction
  - Service personalisation
- But they are linked and inform each other



User Modelling

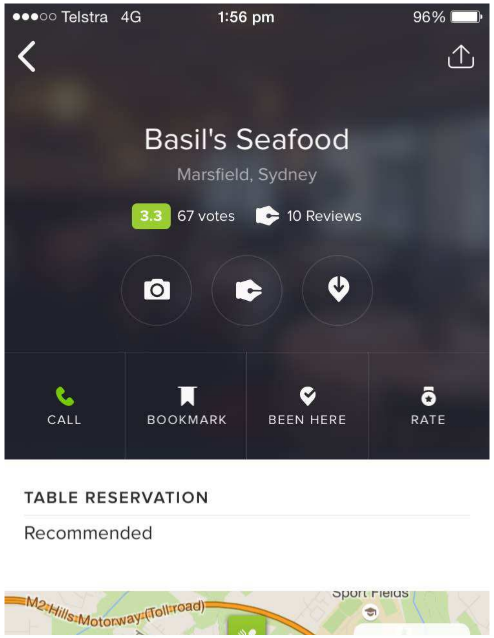
- Different systems require different models
  - Sometimes you model the user in terms of preferences and interests
    - Marketing a product to a user, returning search results, recommending tourist activities
  - Sometimes you model user’s knowledge and goals
    - Adaptive educational systems, online tutorials, video lectures
  - Sometimes model fitness, health or medical conditions
- No single generic user model structure

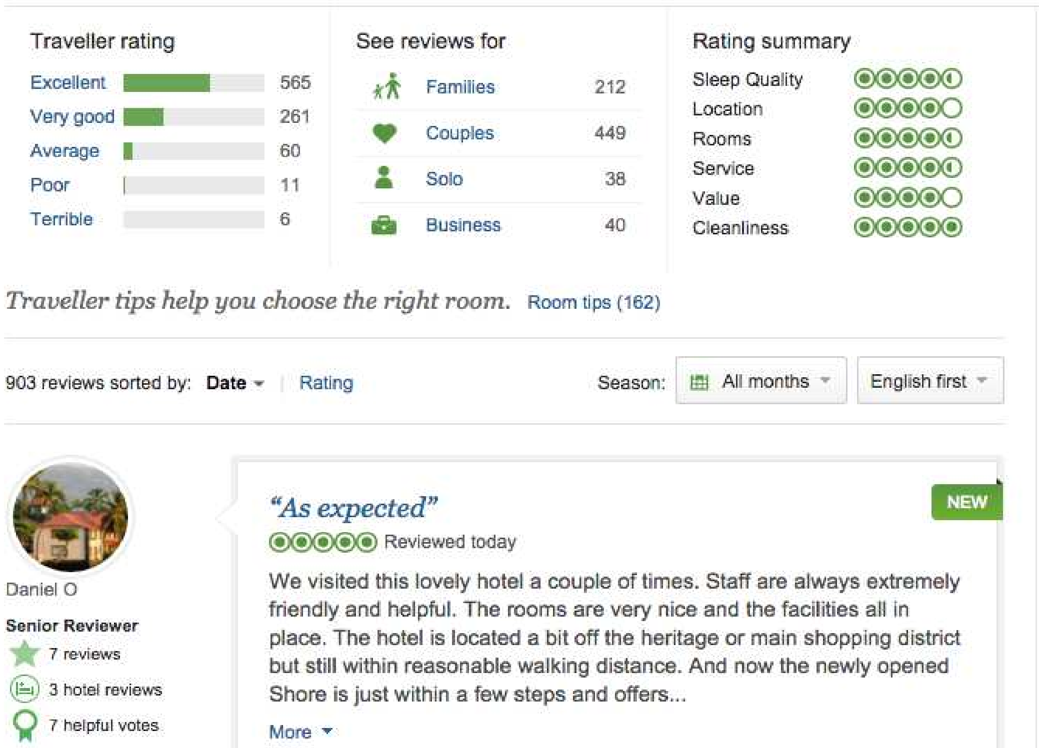

What can be modeled?

- User as an individual
  - Knowledge
  - Interests
  - Preferences
  - Goals and motivation
  - Personality and traits
  - Interactions with system
  - Constraints/limitations
  - …
- External/situational factors
  - Social environment
  - Network conditions
  - End user device
  - …

Explicit User Data Collection

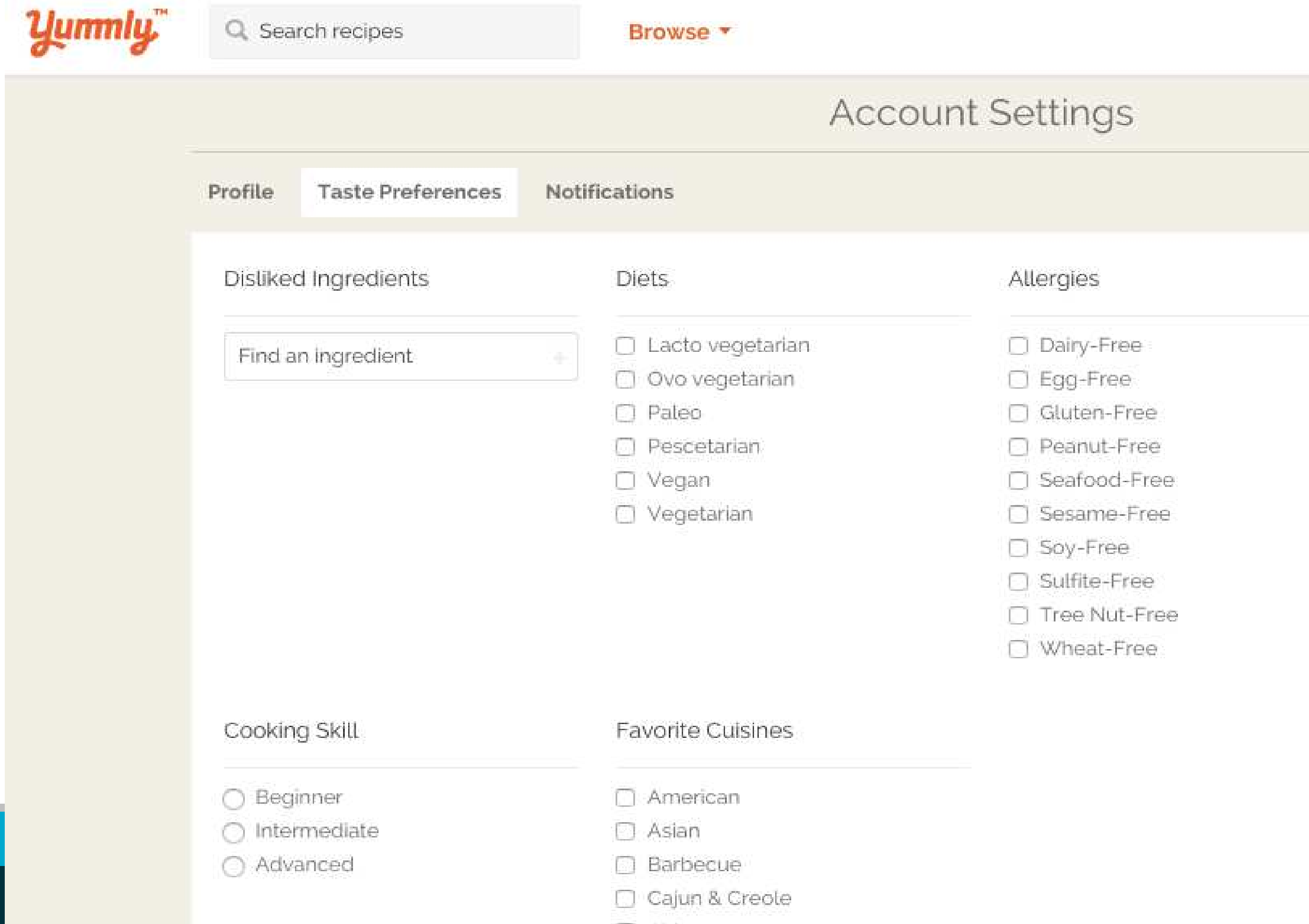
- Relies on information provided by the user
  - Amazon asks for ratings on items purchased
  - TripAdvisor asks for hotel reviews and ratings
- Often contains demographic information
  - Birthday, location, interests, marital status, job …
- Typically accurate, but require time and effort





Explicit User Data Collection

- Often a one-off activity at sign-up



Implicit User Data Collection

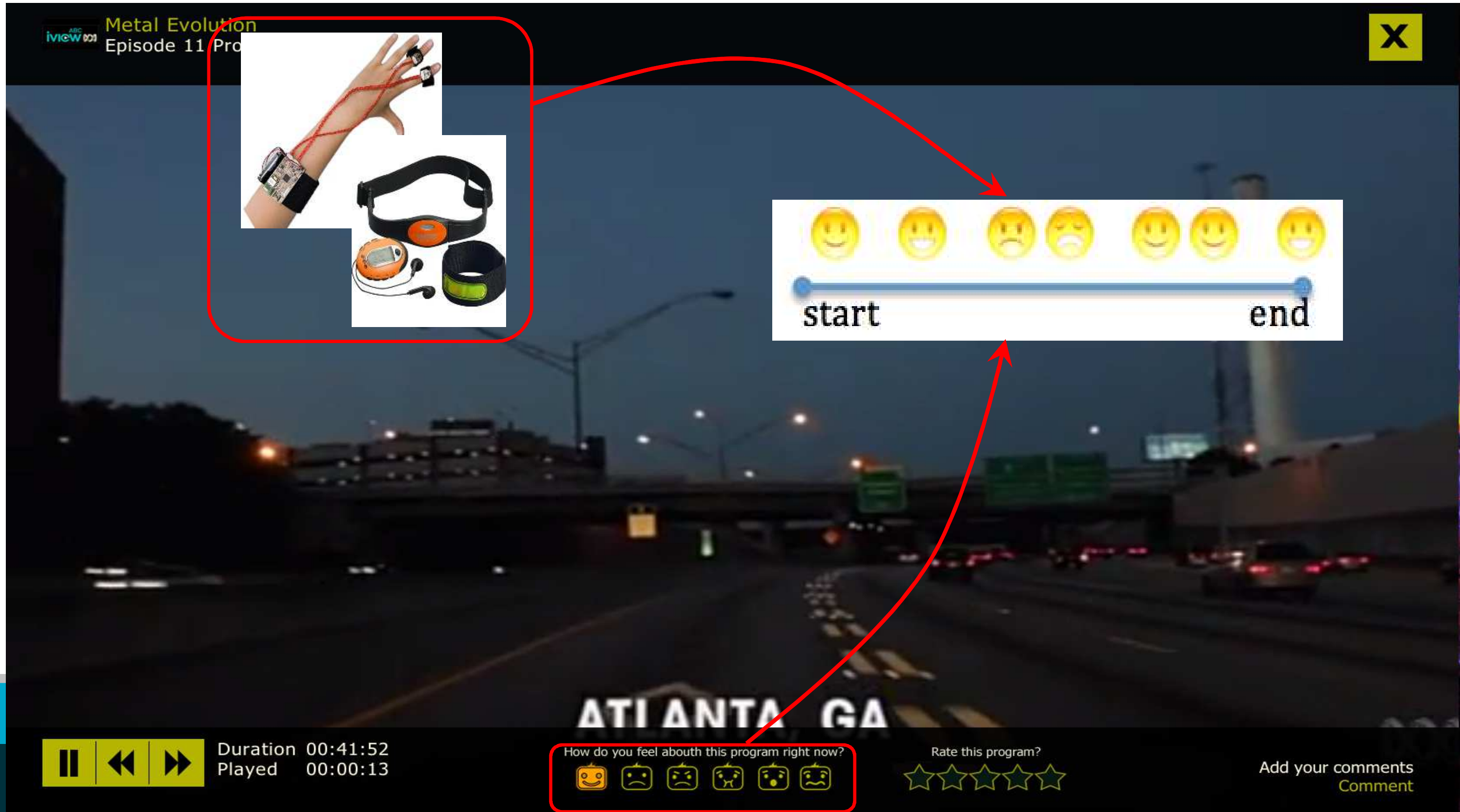
- Derives user modelling data from observable user behavior
  - Monitor users interactions
    - with the system
    - with other users
  - Learn/mine the required user data
- Examples
  - Browser cache, proxy servers, search logs, purchased items, examined products, bookmarked pages, links sent to friends, preferred brands, …
- Typically less accurate than explicit data but
  - more abundant and readily available
  - does not require extra-effort from users

Hybrid Data Collection

- Combines explicit and implicit methods
  - to leverage the benefits of both methods
- Typically achieves the highest accuracy
  - Many things are learned implicitly
  - User feedback is sought for uncertain/important data
- Used by many commercial systems

Emotion Based Modelling

- Relatively new direction in user modelling
- Experienced emotions reflect liked/disliked items
  - Explicit (sentiment analysis) and implicit (sensors)
  - Potentially very fine granularity



Contextualised User Models

- What can be considered as context?
  - Location of the user, presence of other users, time of day, day of week, weather, temperature, mood, …
- Does context matter?
  - Cooking: alone vs. with kids
  - Music: happy vs. sad
  - Movie: home vs. theater
  - Vacation: summer vs. winter
- User preferences are not steady but rather context-dependent
- Only feedback-in-context is meaningful
  - Non-contextualized feedback assumes a default context
    - Default context = most likely context
    - Sometimes true, but often false

Part 2:

Web Personalisation and Recommender Systems

Personalised Search

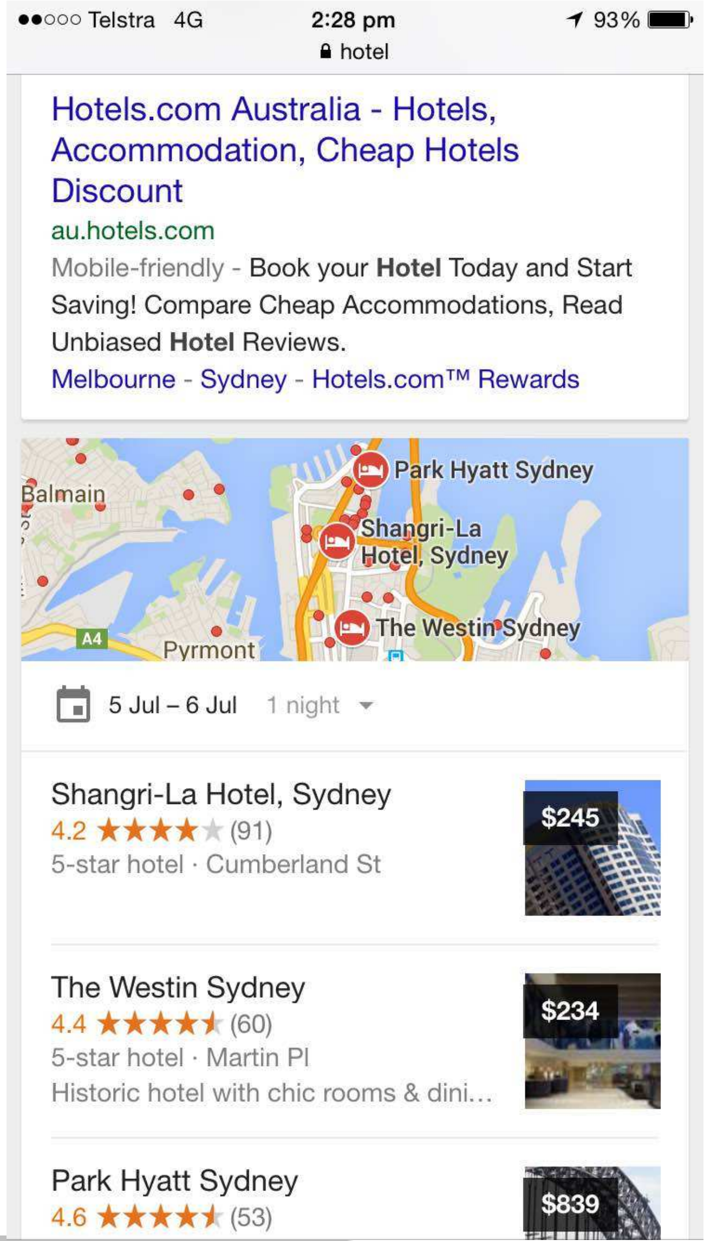
- Search engines can tailor the results to the user

Contextual Search

- Personalisation determined by past searches
- Users are authenticated by accounts or cookies
  - No dedicated user modeling component
- If users enter short queries the profile could indicate the desired meaning
  - If a user has been entering queries about flights, accommodation, or vaccines, they are probably looking for a travel visa

Location Based Search

- Results are tailored to user's geographical location
  - Even though this is not part of the query
- Done automatically through redirection across engines
  - Often switches the language
- Important for mobile search
- Results automatically invoke Maps

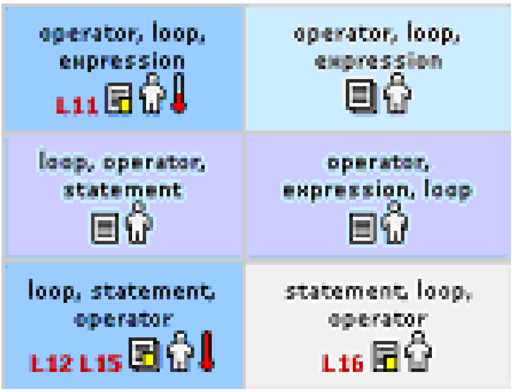


Personalised Navigation Support

- Showing users the way when they browse
- Helping users lost in the Web
  - Direct guidance
  - Sorting lists and links
  - Adding/changing/removing links
  - Adding textual annotations
  - Hiding or highlighting text
  - Increasing font size
  - Adapting images and maps
  - Many more…

Annotations and Signposts

- Annotations
  - Number showing how many times a link have been followed
- Signposts: user feedback regarding past interaction history
- Users may comment on pages or on paths in the social navigation display



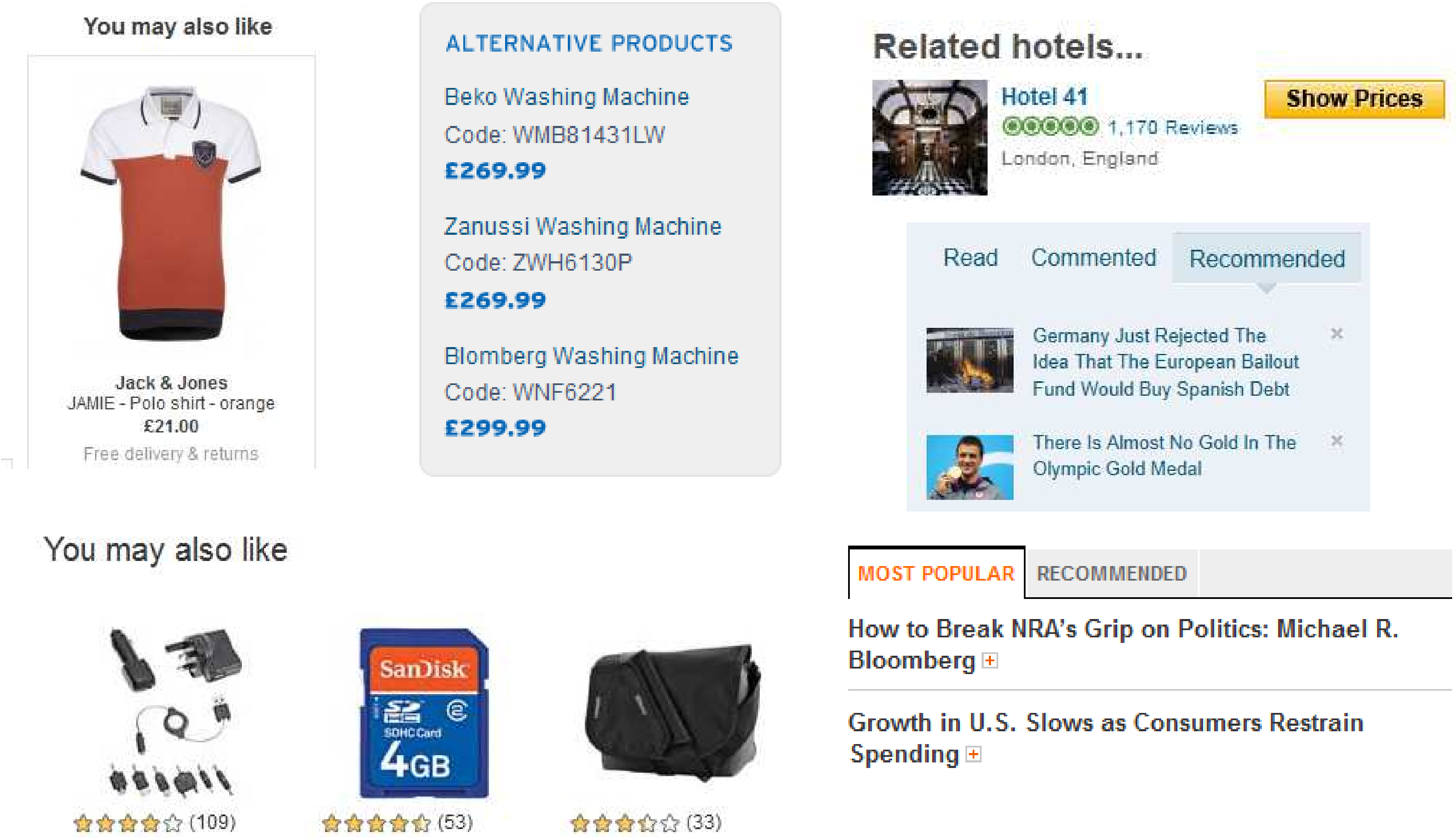
Social Web Personalization

- Unprecedented volume of information
  - Huge contributor to the information overload
  - But non-negligible consumption medium as well
- Personalization use cases
  - News feed filtering and reordering
  - Preselection of tweets/posts
  - Recommendations of friends/followees
  - Recommendations of events/communities
  - Content ranking on behalf of users
  - Content tagging and bookmarking
  - Job/company suggestions
  - Many more…

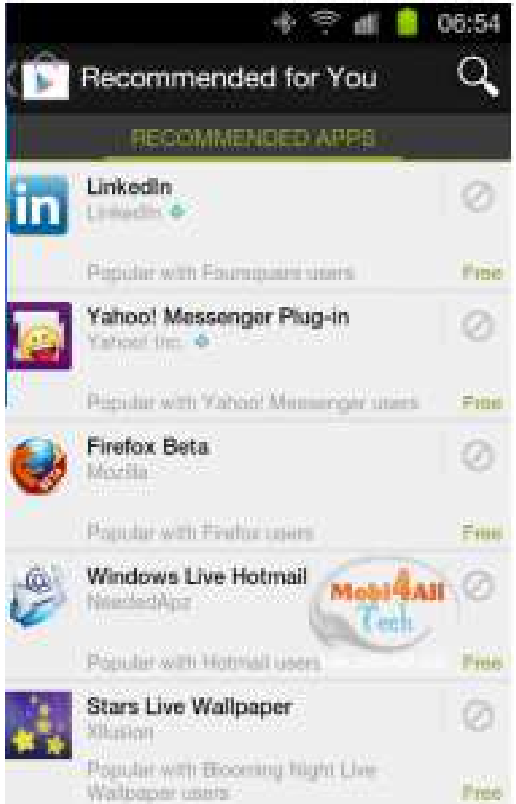
Recommender Systems

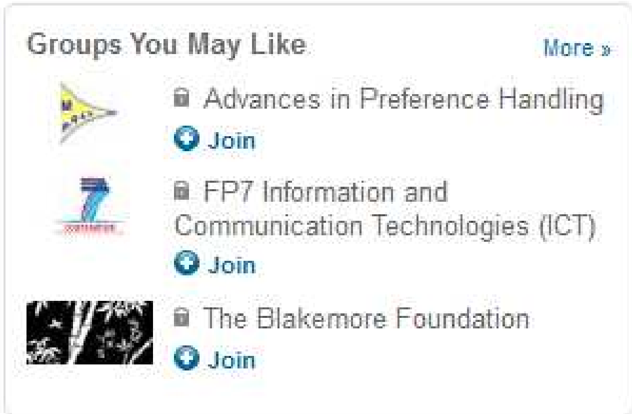
- Recommender systems help to make choices without sufficient personal experience of the alternatives
  - suggest information items to the users
  - help to decide which product to purchase
- “Convert visitors into customers”

Originated in eCommerce



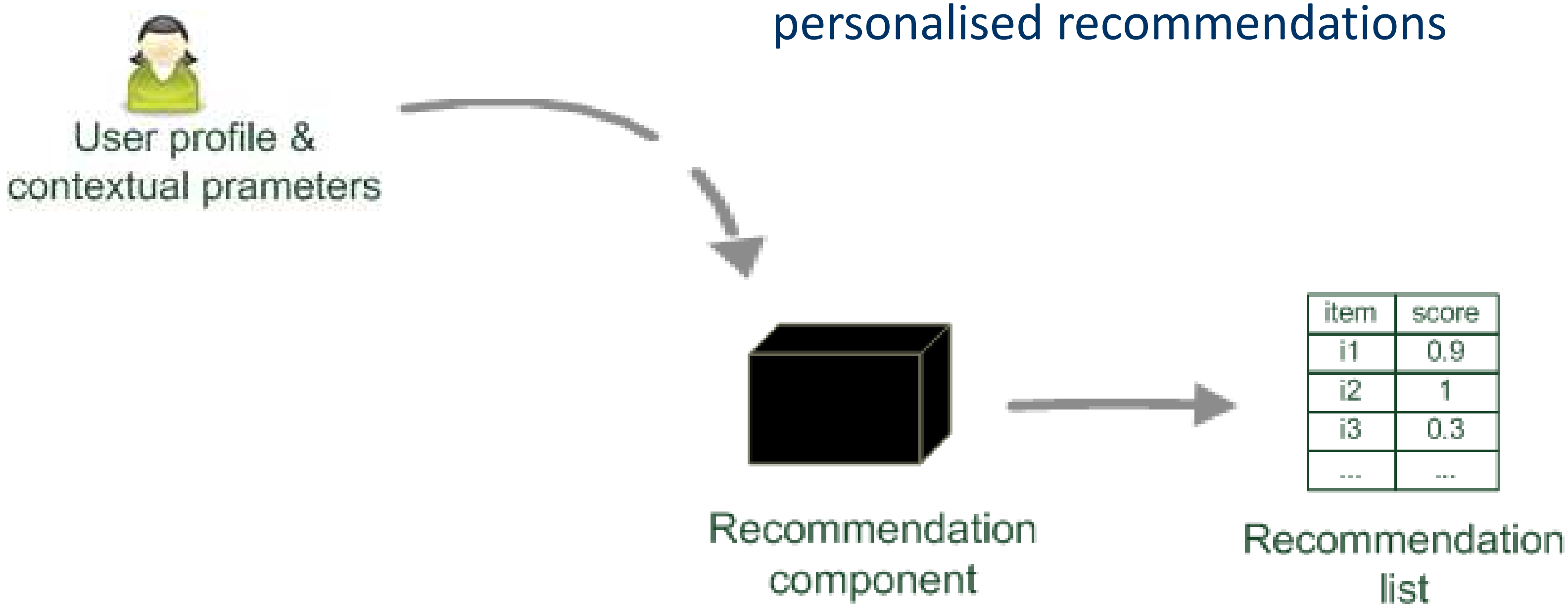
Not only in eCommerce





Paradigms of Recommender Systems

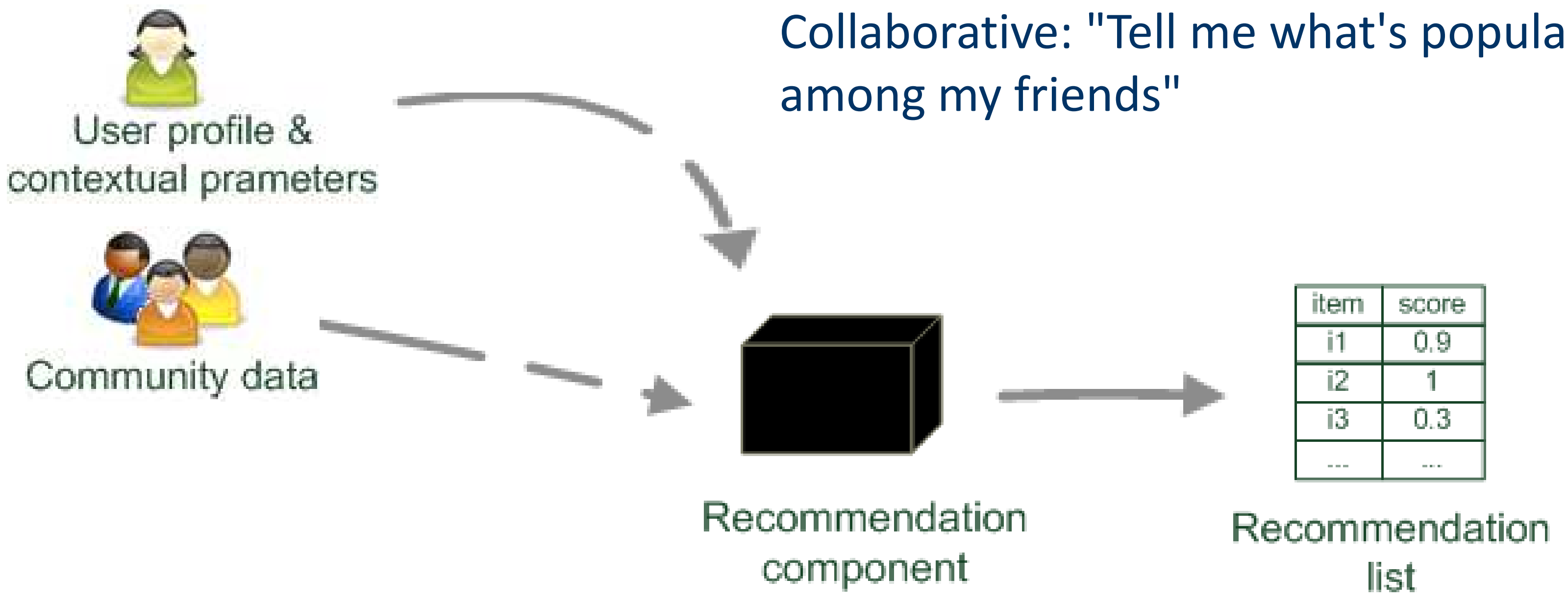
personalised recommendations



item	score
i1	0.9
i2	1
i3	0.3
...	...

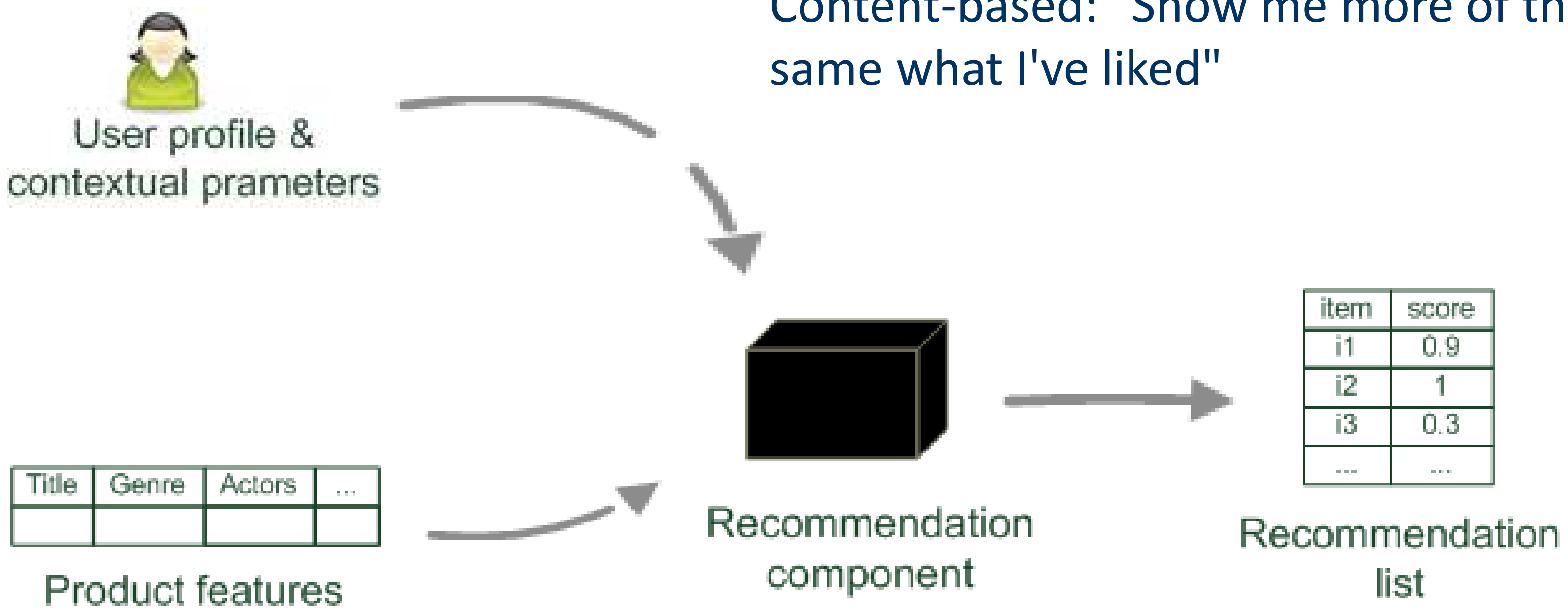
Paradigms of Recommender Systems

Collaborative: "Tell me what's popular among my friends"



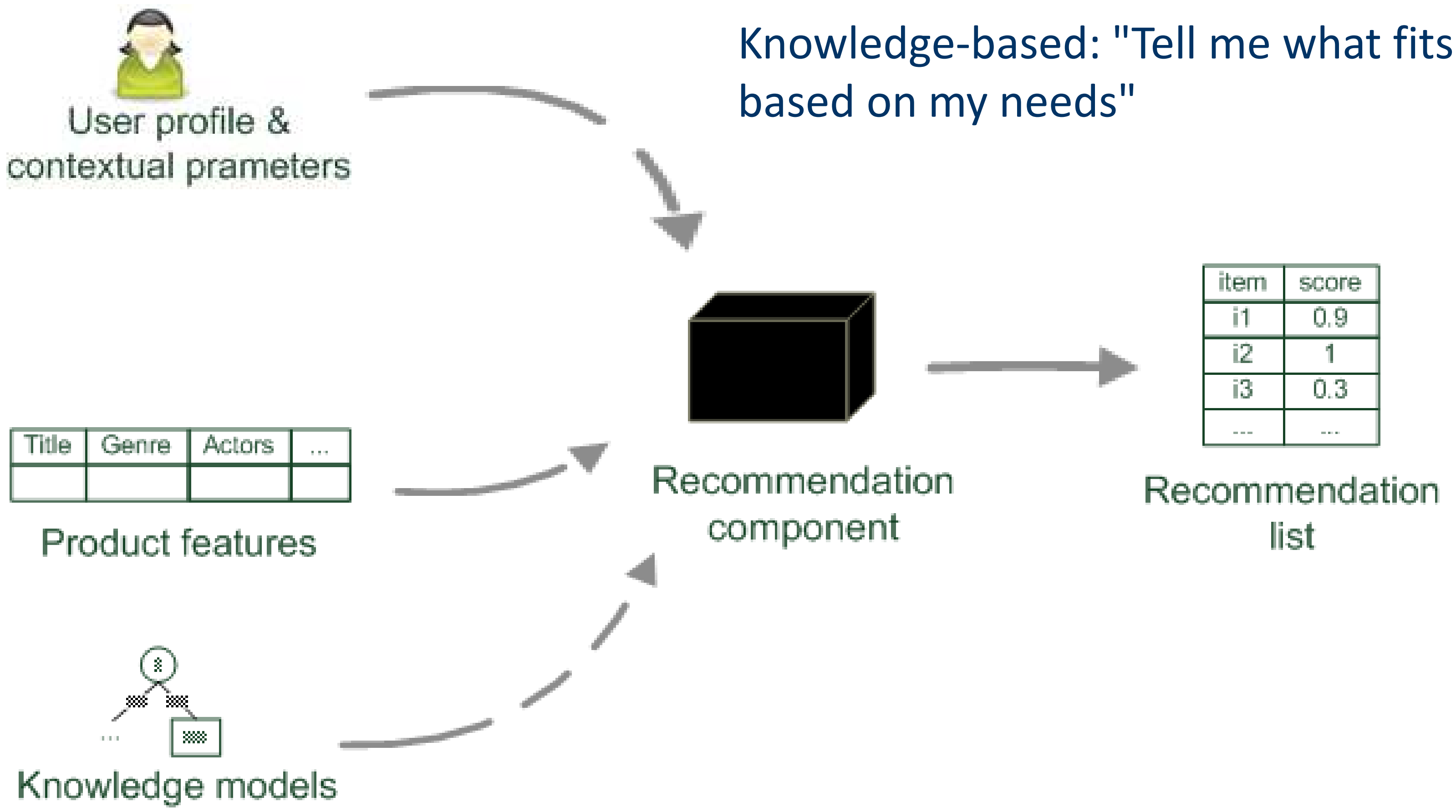
Paradigms of Recommender Systems

Content-based: "Show me more of the same what I've liked"



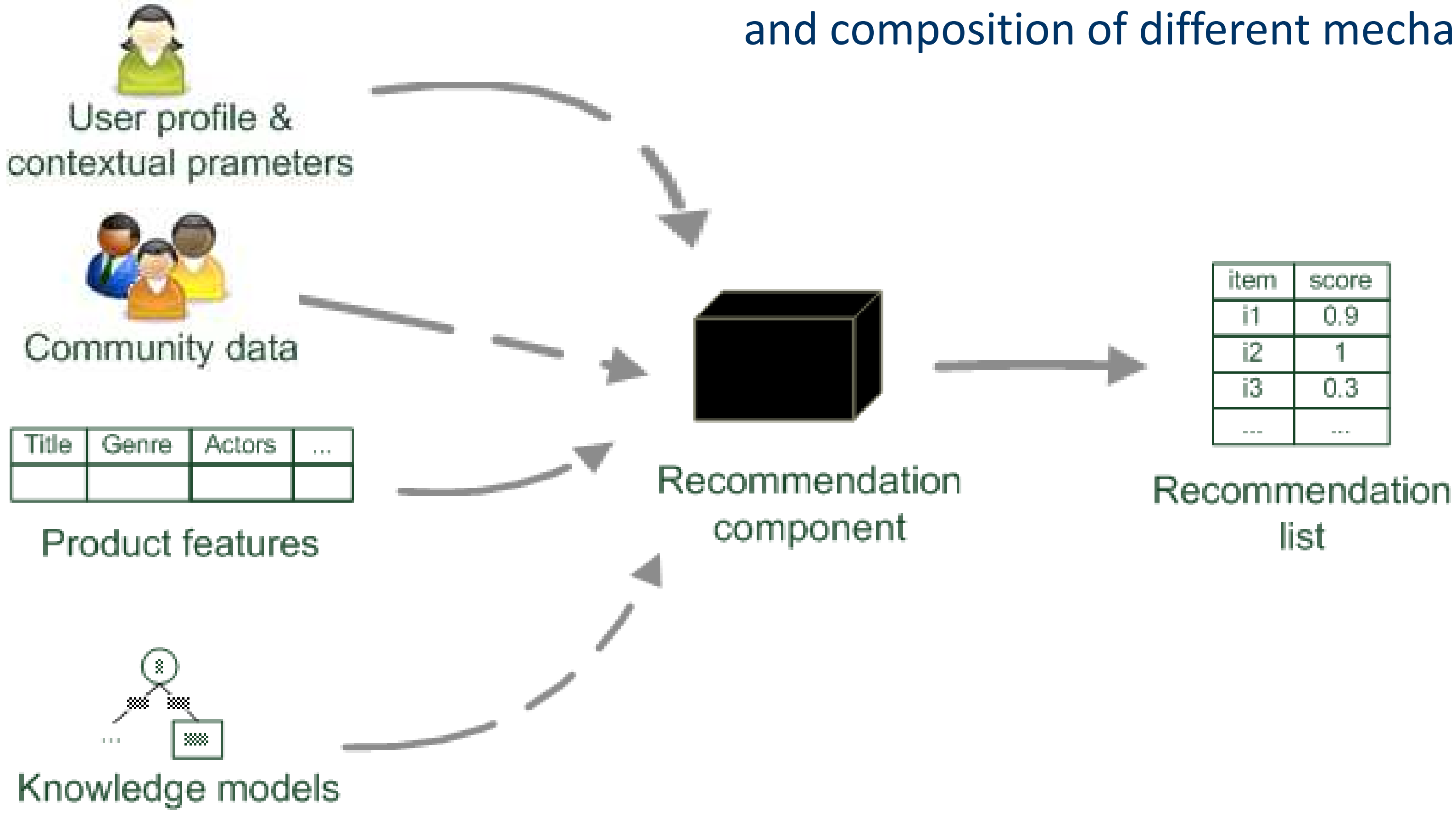
Paradigms of recommender systems

Knowledge-based: "Tell me what fits based on my needs"



Paradigms of Recommender Systems

Hybrid: combinations of various inputs
and composition of different mechanisms



“Core” Recommendation Techniques

Technique	Background	Input	Process
Collaborative	Ratings from **U** of items in **I**.	Ratings from **u** of items in **I**.	Identify users in **U** similar to **u**, and extrapolate from their ratings of **i**.
Content-based	Features of items in **I**	**u**'s ratings of items in **I**	Generate a classifier that fits **u**'s rating behavior and use it on **i**.
Demographic	Demographic information about **U** and their ratings of items in **I**.	Demographic information about **u**.	Identify users that are demographically similar to **u**, and extrapolate from their ratings of **i**.
Utility-based	Features of items in **I**.	A utility function over items in **I** that describes **u**'s preferences.	Apply the function to the items and determine **i**'s rank.
Knowledge-based	Features of items in **I**. Knowledge of how these items meet a user's needs.	A description of **u**'s needs or interests.	Infer a match between **i** and **u**'s need.

Welcome to the new MovieLens!

Existing MovieLens users: We'd like to welcome you back to MovieLens, and let you know we have a new MovieLens FAQ you might want to read. We hope you like what you will see!

Take me to MovieLens!

New MovieLens users: Thank you for joining MovieLens! In order to generate personalized movie recommendations, we need to know a little about what movies you have already seen. MovieLens will now display several lists of movies. If you have seen any of the listed movies, please rate them using the rating scale shown below.

Ratings are on a scale of 1 to 5:

★★★★★ = Must See
★★★★☆ = Will Enjoy
★★★☆☆ = It's OK
★★☆☆☆ = Fairly Bad
★☆☆☆☆ = Awful

Remember: the more movies you rate, the more accurate MovieLens' predictions will be.

To rate a movie, just click on the pulldown next to the title of a movie you have seen. Blue stars will appear to indicate that your rating has been received.

★½☆☆☆	1.5 stars ▾	**Dude, Where's My Car? (2000)** DVD, VHS, info | imdb Comedy	☐

This image shows that the movie 'Dude, Where's My Car?' was rated 1.5 stars.

So far you have rated **0** movies.
MovieLens needs at least **15** ratings from you to generate predictions for you.
Please rate as many movies as you can from the list below.

next >

	Your Rating	Movie Information
★★★	3.0 stars	**Austin Powers: International Man of Mystery (1997)** Action, Adventure, Comedy
★★★★	4.0 stars	**Contact (1997)** Drama, Sci-Fi
???	Not seen	**Crouching Tiger, Hidden Dragon (Wu Hu Zang Long) (2000)** Action, Adventure, Drama, Fantasy, Romance
???	Not seen	**Demolition Man (1993)** Action, Comedy, Sci-Fi
???	Not seen	**Eraser (1996)** Action, Drama, Thriller
???	Not seen	**Maverick (1994)** Action, Comedy, Western
★★★★½	4.5 stars	**Philadelphia (1993)** Drama
★★★½	3.5 stars	**Piano, The (1993)** Drama, Romance
???	Not seen	**Toy Story 2 (1999)** Adventure, Animation, Children, Comedy, Fantasy
★★★½	3.5 stars	**X-Men (2000)** Action, Adventure, Sci-Fi

next >

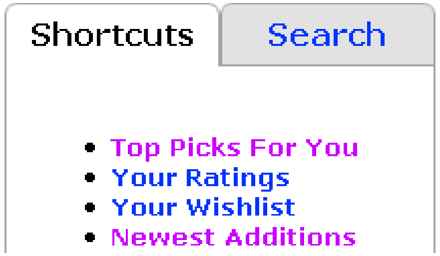
Congratulations!

MovieLens can now generate personalized movie recommendations for you.

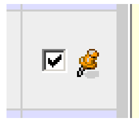
Start Using MovieLens

Remember, you can always keep rating movies you have seen. The more movies you rate, the better your predictions will be. We'd also like to tell you about some other features of MovieLens you might be interested in:

- **Getting recommendations.** MovieLens has shortcuts like Top Picks For You that provide you with quick access to common searches. You can use the Search tab to perform more advanced searches that filter by genre, date, and more, and save your favorite searches as personal shortcuts.



- **Your Wishlist.** Here you can keep track of movies you haven't yet seen. You can even print this list out and take it with you to your video store.



- **Movie buddies.** It can be a pain trying to decide what movie a group of people should see. Let MovieLens choose the right movie for you! You can add MovieLens users to be your buddies and be able to generate group movie recommendations

We will keep adding more great features as time goes on, so look for them!

Start Using MovieLens

You've searched for **all titles**.
Found **8220** movies, sorted by **Prediction**
Genres: **All** | Exclude Genres: **None**
Dates: **All** | Domain: **All** | Format: **All** | Languages: **All**

Show Printer-Friendly Page | **Download Results** | **Suggest a Title**

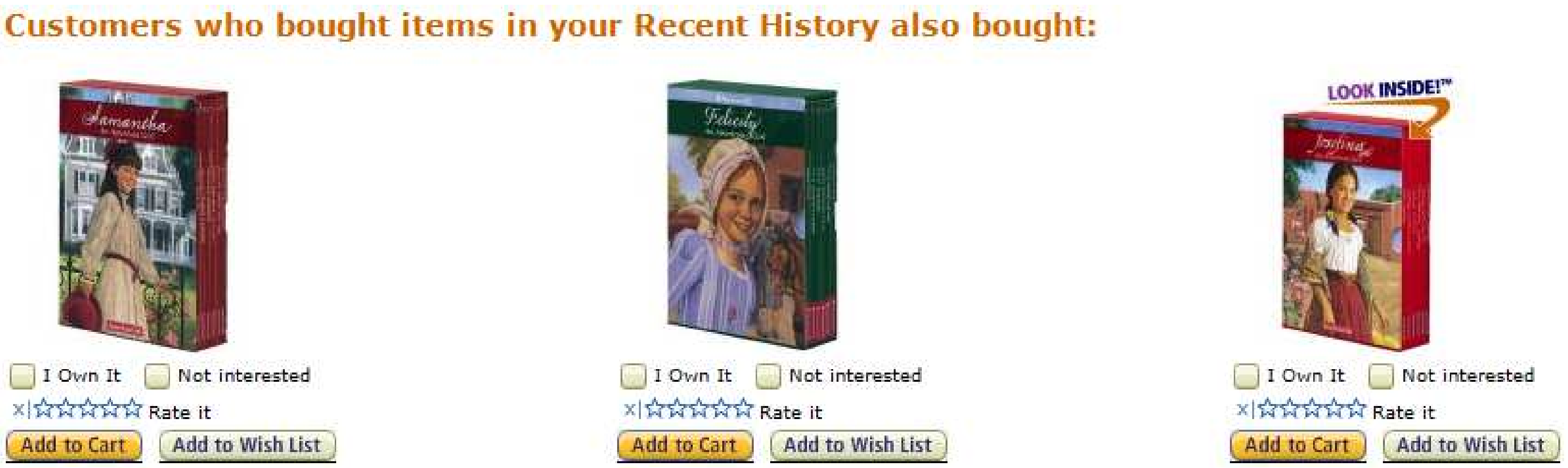
Tags Related to Your Search: **In Netflix queue (178)**, **Futuristmovies.com (134)**, **My DVDs (123)**, **Oscar (Best Cinematography) (90)**, **Oscar (Best Picture) (85)**, **(about tags)**

Page **1** of **548** | Go to page: **1...109...218...327...436...545...last** **page 2>**

(hide) Predictions for you	Your Ratings	Movie Information	Wish List
★★★★★	Not seen	Cat Returns, The (Neko no ongaeshi) (2002) DVD info | imdb Adventure, Animation, Children, Fantasy - Japanese	☐
	[add tag] Popular tags:	anime, cats, In Netflix queue	
★★★★★	Not seen	Immigrant, The (1917) DVD VHS info | imdb | add tag Comedy - Silent	☐
★★★★★	Not seen	Experiment, The (Das Experiment) (2001) DVD VHS info | imdb | add tag Drama, Thriller - German	☐
★★★★★	Not seen	Thesis (Tesis) (1996) DVD info | imdb | add tag Drama, Horror, Thriller - Spanish	☐
★★★★★	Not seen	Howl's Moving Castle (Hauru no ugoku shiro) (2004) DVD info | imdb Adventure, Animation, Children, Fantasy, Romance - Japanese	☐
	[add tag] Popular tags:	06 Oscar Nominated Best Movie - Animation, In Netflix queue	
★★★★★	Not seen	Why We Fight (2005) info | imdb Documentary	☐
	[add tag] Popular tags:	Military, In Netflix queue, controversial	

User-Based Collaborative Filtering

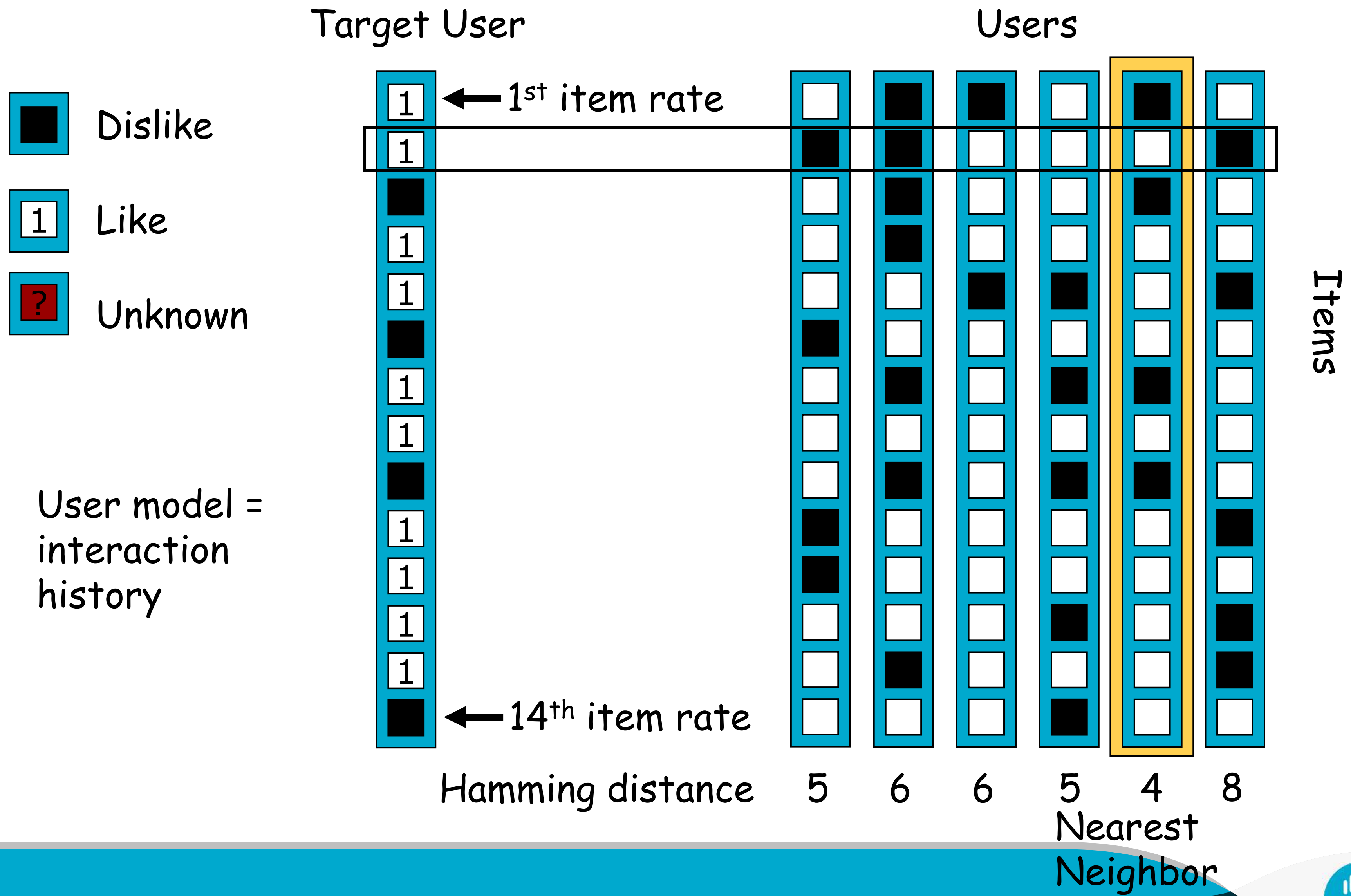
- Idea: users who agreed in the past are likely to agree in the future
- To predict a user's opinion for an item, use the opinions of like-minded users
  - Precisely, a (small) set of very similar users
- User similarity is decided by the overlap in their past opinions
  - High overlap = strong evidence of similarity = high weight



User-Based Collaborative Filtering

1. For a target user (to whom a recommendation is produced) the set of his ratings is identified
2. The users similar to the target user (according to a similarity function) are identified
   - Cosine similarity, Pearson's correlation, Mean Squared Difference, or other similarity metrics
3. Items rated by similar users but not by the target user are identified
4. For each item a predicted rating is computed
   - Weighted according to users' similarity
5. Based on this predicted ratings a set of items is recommended

Collaborative Filtering



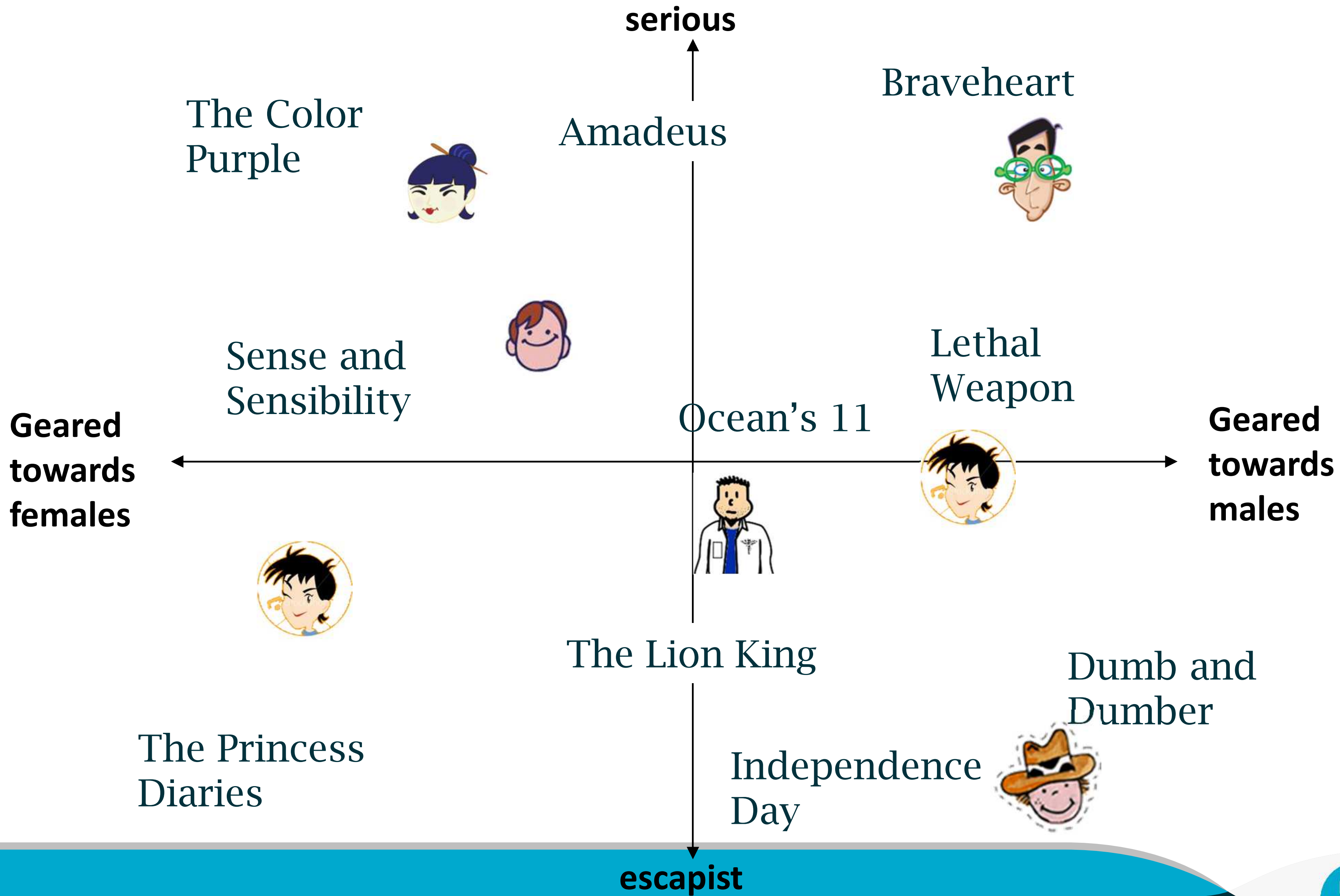
Limitations of Collaborative Filtering

- Sparsity: large product sets and few user ratings
  - Requires many explicit ratings to bootstrap
    - New user and new item problem
  - Sparsity of real-life datasets: 98.69% and 99.94%
  - Amazon: millions of books and a user may have read hundreds
- Drift: popular items are recommended
  - The usefulness of recommending popular items is questionable
    - Recommending top items is obvious for users
  - Recommending unpopular items
    - Is risky, but could be valuable for users
- Scalability – will is scale up to Web size?
  - Quadratic computational time
  - Web recommender will struggle with real-time recommendations

Matrix Factorisation

- Netflix Prize Competition
  - Training data
    - 6 years of data: 2000-2005
    - 100M ratings of 480K users for 18K movies
  - Test data
    - Evaluation criterion: root mean squared error (RMSE)
  - Competition
    - 2700+ teams
    - $1M prize for 10% improvement on baseline
  - Won by the Bellkor-Gravity team
    - Ensemble of more than 100 recommenders
    - Many of them based on Matrix Factorisation
- Boosted Matrix Factorisation for recommender systems

Latent Factor Model



Latent Factor Model

users

items

1		3			5			5		4	
		5	4	?		4			2	1	3
2	4		1	2		3		4	3	5	
	2	4		5			4			2	
		4	3	4	2					2	5
1		3		3			2			4	

~

~

items

.1	-.4	.2
-.5	.6	.5
-.2	.3	.5
1.1	2.1	.3
-.7	2.1	-2
-1	.7	.3

●

users

1.1	-.2	.3	.5	-2	-.5	.8	-.4	.3	1.4	2.4	-.9
-.8	.7	.5	1.4	.3	-1	1.4	2.9	-.7	1.2	-.1	1.3
2.1	-.4	.6	1.7	2.4	.9	-.3	.4	.8	.7	-.6	.1

Estimate unknown ratings as an inner product of latent user and item factors

Latent Factor Model

users

items

1		3			5			5		4	
		5	4	?		4			2	1	3
2	4		1	2		3		4	3	5	
	2	4		5			4			2	
		4	3	4	2					2	5
1		3		3			2			4	

~

~

items

.1	-.4	.2
-.5	.6	.5
-.2	.3	.5
1.1	2.1	.3
-.7	2.1	-2
-1	.7	.3

●

users

1.1	-.2	.3	.5	-2	-.5	.8	-.4	.3	1.4	2.4	-.9
-.8	.7	.5	1.4	.3	-1	1.4	2.9	-.7	1.2	-.1	1.3
2.1	-.4	.6	1.7	2.4	.9	-.3	.4	.8	.7	-.6	.1

Estimate unknown ratings as an inner product of latent user and item factors

Latent Factor Model

users

items

1		3			5			5		4	
		5	4	4		4			2	1	3
2	4		1	2		3		4	3	5	
	2	4		5			4			2	
		4	3	4	2					2	5
1		3		3			2			4	

~

~

items

.1	-.4	.2
-.5	.6	.5
-.2	.3	.5
1.1	2.1	.3
-.7	2.1	-2
-1	.7	.3

•

users

1.1	-.2	.3	.5	-2	-.5	.8	-.4	.3	1.4	2.4	-.9
-.8	.7	.5	1.4	.3	-1	1.4	2.9	-.7	1.2	-.1	1.3
2.1	-.4	.6	1.7	2.4	.9	-.3	.4	.8	.7	-.6	.1

Estimate unknown ratings as an inner product of latent user and item factors

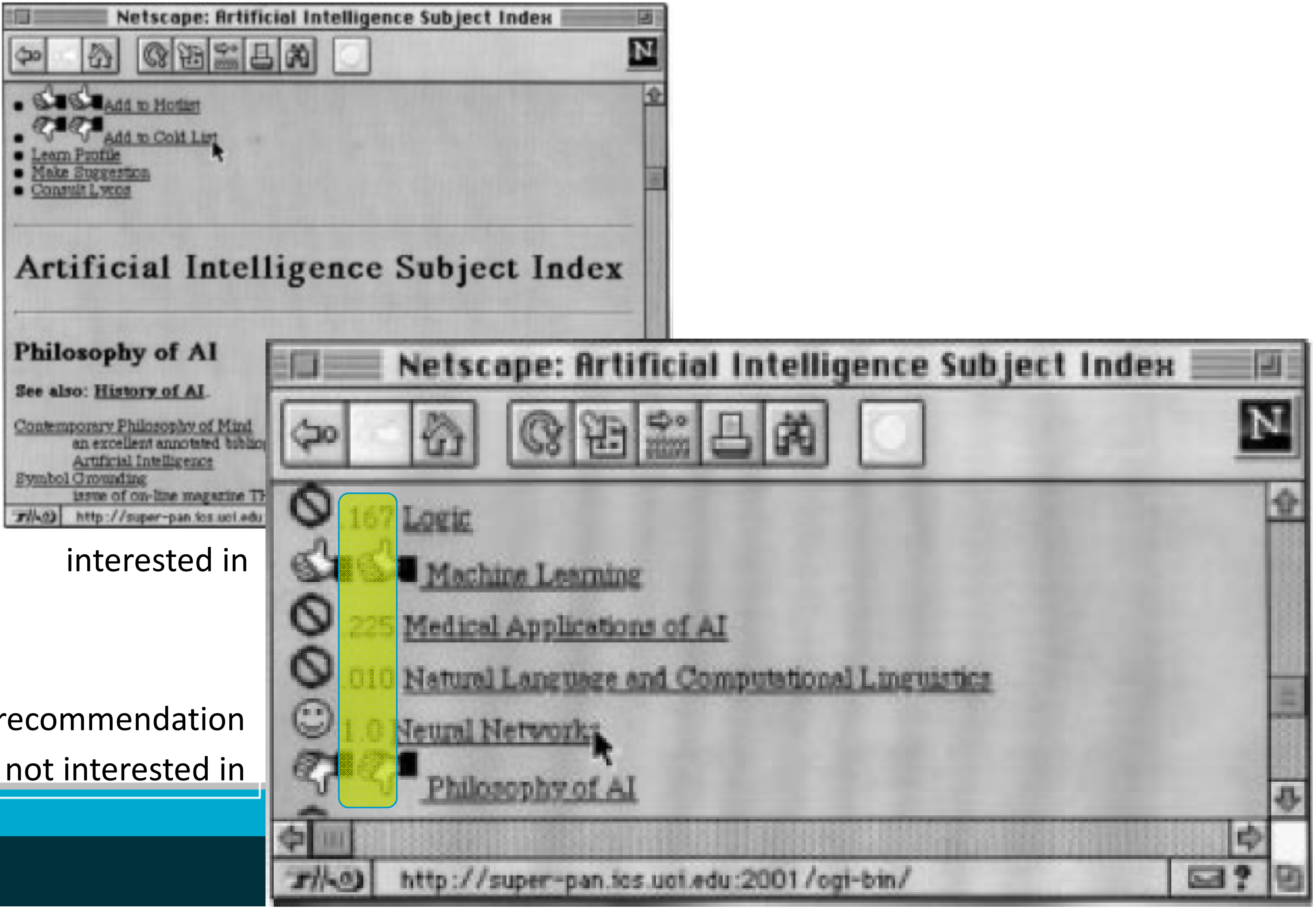
Matrix Factorisation

- Pros
  - Well evaluated in data mining
  - Very strong and accurate model
  - Can scale to Web-size datasets
  - Can incorporate contextual dependency
  - Many variants and open implementations
- Cons
  - Can easily overfit
  - Requires optimisation of parameters
  - Requires regularisation
  - Meaningless latent factors

“Core” Recommendation Techniques

Technique	Background	Input	Process
Collaborative	Ratings from **U** of items in **I**.	Ratings from **u** of items in **I**.	Identify users in **U** similar to **u**, and extrapolate from their ratings of **i**.
Content-based	Features of items in **I**	**u**’s ratings of items in **I**	Generate a classifier that fits **u**’s rating behavior and use it on **i**.
Demographic	Demographic information about **U** and their ratings of items in **I**.	Demographic information about **u**.	Identify users that are demographically similar to **u**, and extrapolate from their ratings of **i**.
Utility-based	Features of items in **I**.	A utility function over items in **I** that describes **u**’s preferences.	Apply the function to the items and determine **i**’s rank.
Knowledge-based	Features of items in **I**. Knowledge of how these items meet a user’s needs.	A description of **u**’s needs or interests.	Infer a match between **i** and **u**’s need.

Syskill & Webert User Interface



interested in

recommendation

not interested in

What is Content?

- Mostly applied to recommending text documents
  - Web pages, emails, or newsgroup messages
- Items are represented using their features
  - With description of their basic characteristics
  - Structured: items are described by a set of attributes

Title	Genre	Author	Type	Price	Keywords
The Night of the Gun	Memoir	David Carr	Paperback	29.90	Press and journalism, drug addiction, personal memoirs, New York
The Lace Reader	Fiction, Mystery	Brunonia Barry	Hardcover	49.90	American contemporary fiction, detective, historical
Into the Fire	Romance, Suspense	Suzanne Brockmann	Hardcover	45.90	American fiction, murder, neo-Nazism

  - Unstructured: free-text
    - – NLP processing and extraction
    - – TF-IDF weighing

Content-Based Recommendations

- The system recommends items similar to those the user liked
  - Similarity is based on the content of items which that the user has evaluated
    - Very different from collaborative filtering
- Originated in Information Retrieval
  - Was used to retrieve similar textual documents
    - Documents are described by textual content
    - The user profile is structured in a similar way
    - Documents are retrieved based on a comparison between their content and a user model
- Recommender implemented as a classifier
  - e.g., Neural Networks, Naive Bayes, C4.5, …

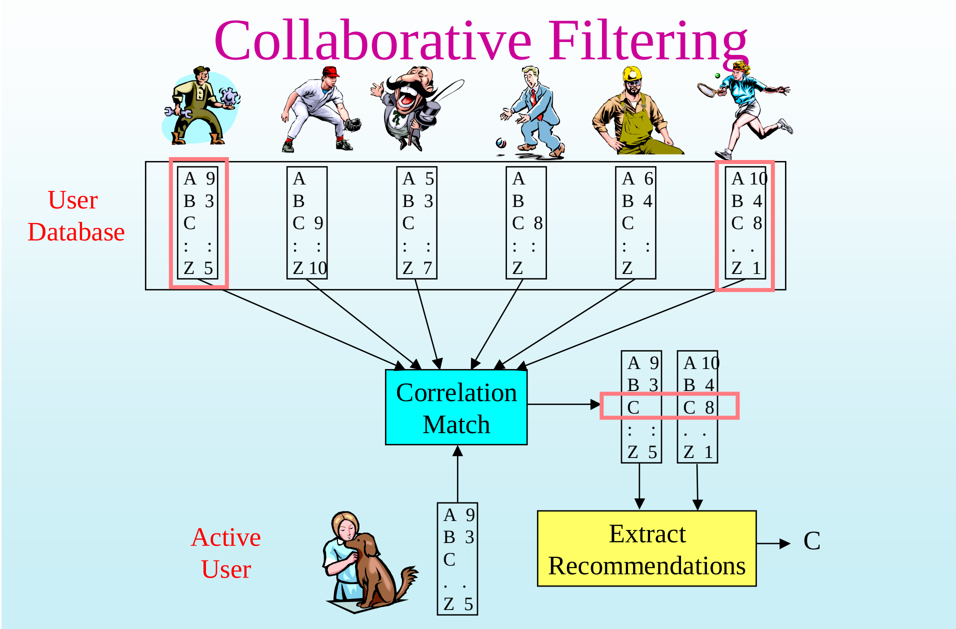
Content-Based Recommendations

- Assist users in finding items that satisfy their information needs
  - User profile describes long-term preferences
- Long-and short-term preferences can be combined
  - Aggregate the level of interest as represented in the long-term and short-term profiles
- Long- and short-term recommendations can be combined
  - Items satisfying short-term preferences can be sorted according to long-term preferences

Limitations of Content-Based Recommendations

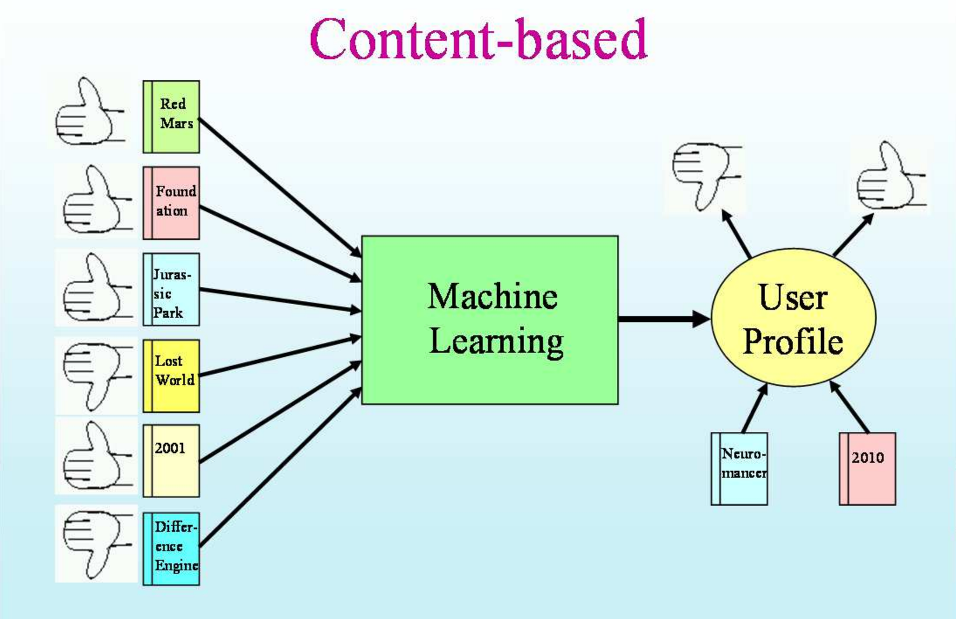
- Only a shallow content analysis is performed
  - Images, video, music, …
- Certain textual features cannot be extracted
  - Quality, writing style, agreement, sentiments, …
    - If a page is rated positively, it could not necessarily be related to the presence of certain words
- Requires considerable domain knowledge
- Even less serendipity
  - Recommends only similar items
  - Trustful but not very useful recommendations

Content-based
vs.
Collaborative



Needs descriptions of items…

Needs only ratings from other users…



“Core” Recommendation Techniques

Technique	Background	Input	Process
Collaborative	Ratings from **U** of items in **I**.	Ratings from **u** of items in **I**.	Identify users in **U** similar to **u**, and extrapolate from their ratings of **i**.
Content-based	Features of items in **I**	**u**’s ratings of items in **I**	Generate a classifier that fits **u**’s rating behavior and use it on **i**.
Demographic	Demographic information about **U** and their ratings of items in **I**.	Demographic information about **u**.	Identify users that are demographically similar to **u**, and extrapolate from their ratings of **i**.
Utility-based	Features of items in **I**.	A utility function over items in **I** that describes **u**’s preferences.	Apply the function to the items and determine **i**’s rank.
Knowledge-based	Features of items in **I**. Knowledge of how these items meet a user’s needs.	A description of **u**’s needs or interests.	Infer a match between **i** and **u**’s need.

Demographic recommendations

- Collects demographic information about users
- Aggregates users into clusters
  - Using a similarity measure and data correlation
- Classifies each user to a cluster that contains the most similar users
- Generates cluster-based recommendation
  - Similar to CF but exploits demographic similarity

	gender	age	area code	education	employed	Dolce
Karen	F	15	714	HS	F	+
Lynn	F	17	714	HS	F	–
Chris	M	35	714	C	T	+
Mike	F	40	714	C	T	–
Jill	F	10	714	E	F	?

“Core” Recommendation Techniques

Technique	Background	Input	Process
Collaborative	Ratings from **U** of items in **I**.	Ratings from **u** of items in **I**.	Identify users in **U** similar to **u**, and extrapolate from their ratings of **i**.
Content-based	Features of items in **I**	**u**’s ratings of items in **I**	Generate a classifier that fits **u**’s rating behavior and use it on **i**.
Demographic	Demographic information about **U** and their ratings of items in **I**.	Demographic information about **u**.	Identify users that are demographically similar to **u**, and extrapolate from their ratings of **i**.
Utility-based	Features of items in **I**.	A utility function over items in **I** that describes **u**’s preferences.	Apply the function to the items and determine **i**’s rank.
Knowledge-based	Features of items in **I**. Knowledge of how these items meet a user’s needs.	A description of **u**’s needs or interests.	Infer a match between **i** and **u**’s need.

Digital Cameras

Get personalized, accurate recommendations with this powerful tool.

Select the features that are important to you. reset **recommend >>**

- [x] **Price Options** what does this mean

  at least $250 at most $605

  ...compared to other features, Price is very important

- [ ] **Brand** what does this mean
- [x] **Effective Pixels** what does this mean - help me decide

  5 megapixels at least

  ...compared to other features, Effective Pixels is extremely important

- [ ] **Optical Zoom** what does this mean - help me decide
- [ ] **Image Capacity (at hi-res)** what does this mean - help me decide
- [x] **Delay Between Shots** what does this mean - help me decide

  0.008 sec at most

  ...compared to other features, Delay Between Shots is extremely important

- [ ] **Camera Size** what does this mean - help me decide
- [ ] **Ease of Download** what does this mean

Utility related information

“Core” Recommendation Techniques

Technique	Background	Input	Process
Collaborative	Ratings from **U** of items in **I**.	Ratings from **u** of items in **I**.	Identify users in **U** similar to **u**, and extrapolate from their ratings of **i**.
Content-based	Features of items in **I**	**u**’s ratings of items in **I**	Generate a classifier that fits **u**’s rating behavior and use it on **i**.
Demographic	Demographic information about **U** and their ratings of items in **I**.	Demographic information about **u**.	Identify users that are demographically similar to **u**, and extrapolate from their ratings of **i**.
Utility-based	Features of items in **I**.	A utility function over items in **I** that describes **u**’s preferences.	Apply the function to the items and determine **i**’s rank.
Knowledge-based	Features of items in **I**. Knowledge of how these items meet a user’s needs.	A description of **u**’s needs or interests.	Infer a match between **i** and **u**’s need.

Knowledge-based recommenders

Hybrid Recommendations

- Each core method has its own pros and cons
- Combine core methods for recommendations
  - Leverage the advantages and hide shortcoming
  - Recall the Netflix winning ensemble!
- Lots of hybrid methods – no standard

Hybridization method	Description
Weighted	The scores (or votes) of several recommendation techniques are combined together to produce a single recommendation.
Switching	The system switches between recommendation techniques depending on the current situation.
Mixed	Recommendations from several different recommenders are presented at the same time
Feature combination	Features from different recommendation data sources are thrown together into a single recommendation algorithm.
Cascade	One recommender refines the recommendations given by another.
Feature augmentation	Output from one technique is used as an input feature to another.
Meta-level	The model learned by one recommender is used as input to another.

Hybrid Recommendations

- Hybrid methods are the state-of-the-art
  - Most powerful and most popular
  - Leverage the advantages of individual methods
  - Generate recommendations superior to individual methods
- Plenty of unexplored options for hybridisation
  - The most simple and widely used methods are weighted, switching, and mixed hybridisations
  - Several focused studies of cascade and feature augmentation hybridisations
  - Very few studies on feature combination and meta-level hybridisations

Evaluating Recommender Systems

- Algorithmic evaluation
  - Offline datasets, statistic evaluations
    1. Measure how good is the system in predicting the exact rating value (value comparison)
    2. Measure how well the system can predict whether the item is relevant or not (relevant vs. not relevant)
    3. Measure how close the predicted ranking of items is to the user's true ranking (ordering comparison).
- User studies
  - Let users play with the system
  - Collect and analyze feedback
  - Compare with non-personalised system

Challenges: Data Sparsity

- Personalised systems succeed only if sufficient information about users is available
  - No user model = No personalisation
- How to gather enough user modelling data in unobtrusive manner?
- If the required data is not available
  - Web of trust to identify "similar users"
  - Use external data sources
    - Web mining
  - The output is always an approximation
- Similarly: new item problem

Challenges: Contextualisation

- Systems should adapt to user context
  - Some methods cannot cope with this
- Largely depends on the definition of context but in practice this includes
  - Short term preferences (“tomorrow I want …”)
  - Information related to the specific space-time position of the user (“less than 5 mins walking)
  - Motivations of search (“present to my wife”)
  - Circumstances (“some time to spend here”)
  - Emotions and mood (“I feel adventurous”)
  - …

Challenges: Privacy

- Personalisation is based on personal data
  - Privacy vs. personalisation tradeoff
    - More user information = more accurate personalisation
    - More user information = less user privacy
- Laws that impose stringent restrictions on the usage and distribution of personal data
  - Systems must cope with these legislation
    - e.g., personalisation systems exchanging user profiles could be impossible for legal reasons
- Personalisation systems must be developed in a way that limits the possibility of an attacker learning/accessing personal data

Challenges: Scalability

- Personalisation techniques rely on extensive user/item descriptions
  - Many of them are hardly scalable
- Techniques that can overcome this
  - Feature selection
  - Dimensionality reduction
  - Latent factors analysis
  - Clustering and partitioning
  - Distributed computing
  - P2P architectures
  - Parallel computing
  - …

Other Open Challenges

- Generic user models and personalisation
- Portable and mobile personalisation
- Emotional and value aware personalisation
- User trust and recommendations
- Persuasive personalised technologies
- Group-based personalisation
- Interactive sequential personalisation
- Complex and bundle recommendations
- Robustness of business recommenders systems
- Semantically enhanced personalisation
- Personalisation on the Social Web
- Personalisation in the Internet of Things
- People recommender systems
- Personalisation or information bubble
- … more and more …

Main Resources

- Books
  - The Adaptive Web – Methods and Strategies of Web Personalization
  - Recommender Systems – An Introduction
  - Recommender Systems Handbook
    - – Second edition is coming up
- Online
  - www.um.org
  - recsys.acm.org
  - www.recsyswiki.com
  - www.coursera.org/learn/recommender-systems

Take home message

It's all about you!

Thank you!

Questions?