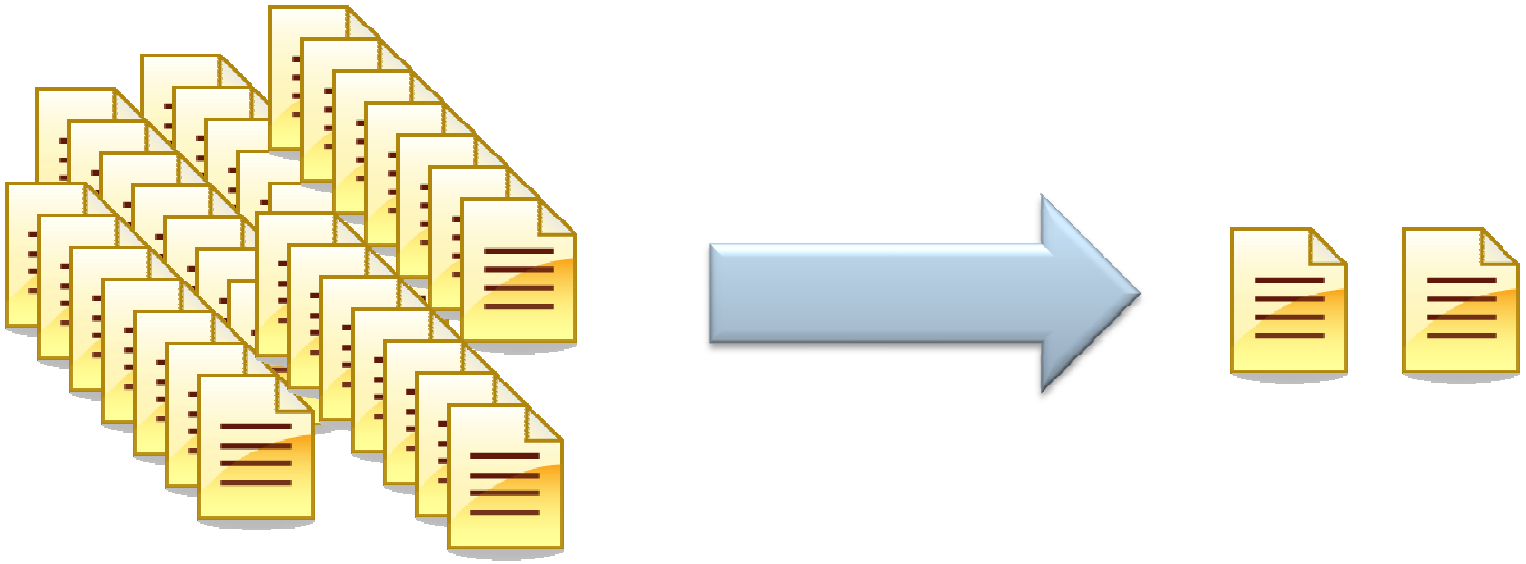
ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Collaborative Filtering

Presentation by Alex Hugger

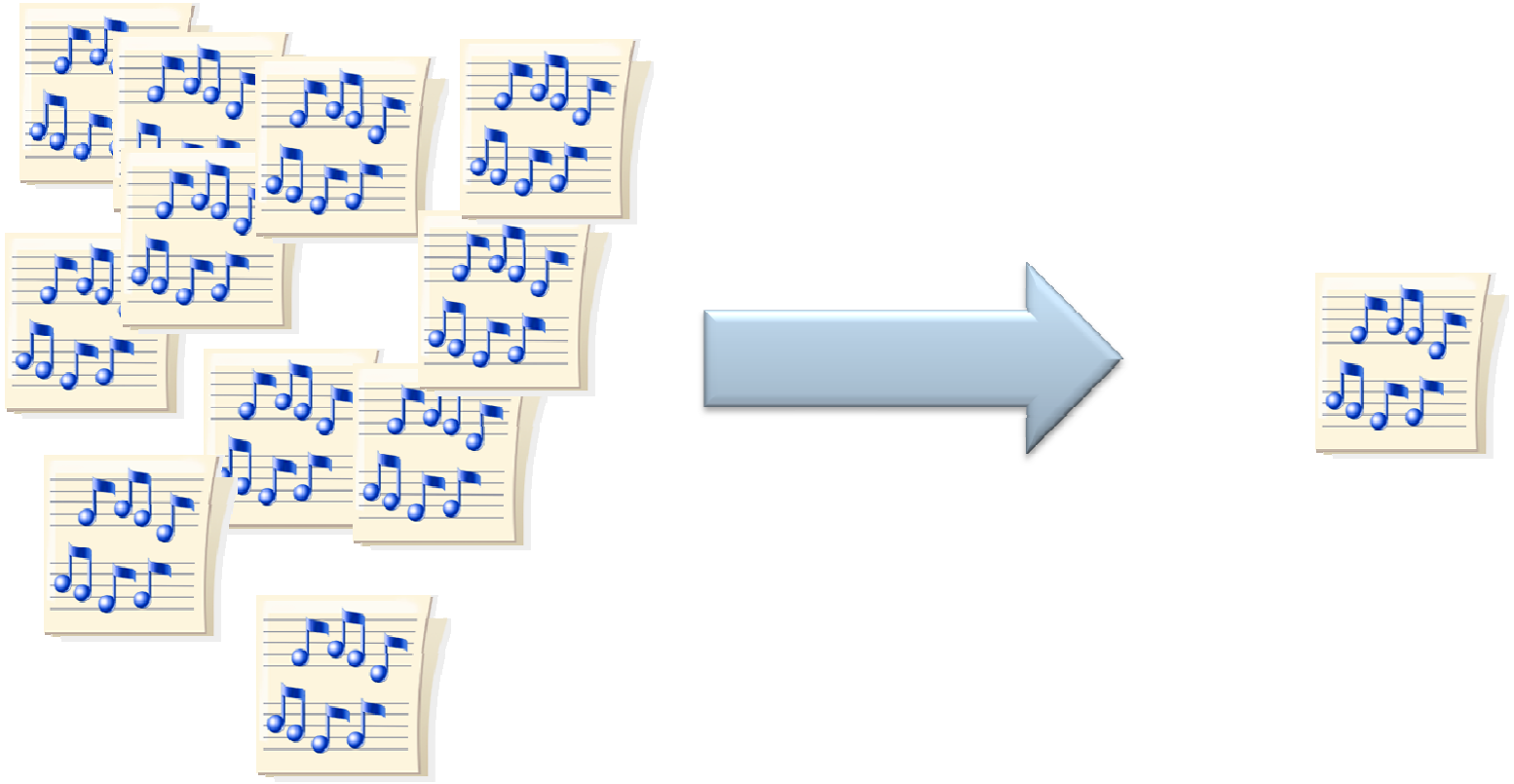
Filtering Documents



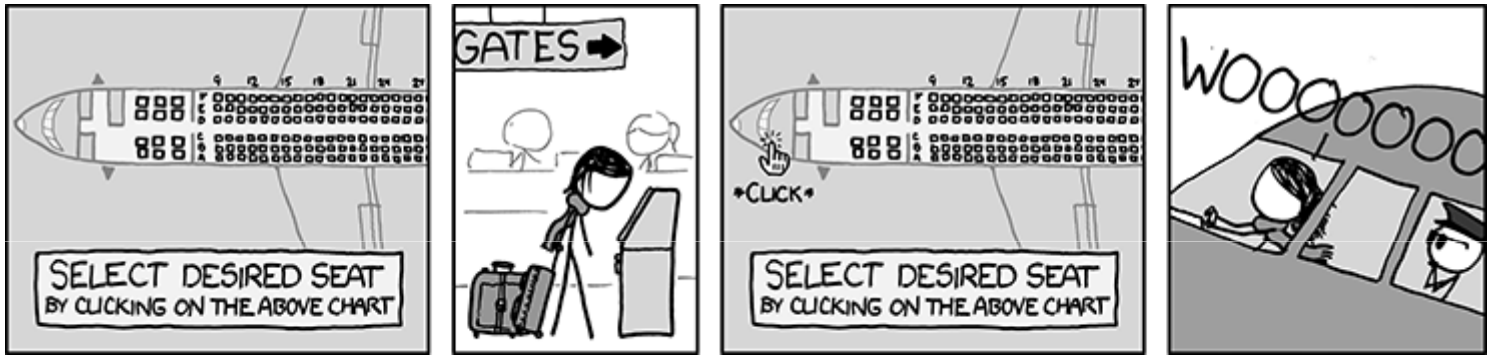
Content-Based Methods

- Find other popular items by the same author or similar keywords
- Recommendation quality is relatively poor

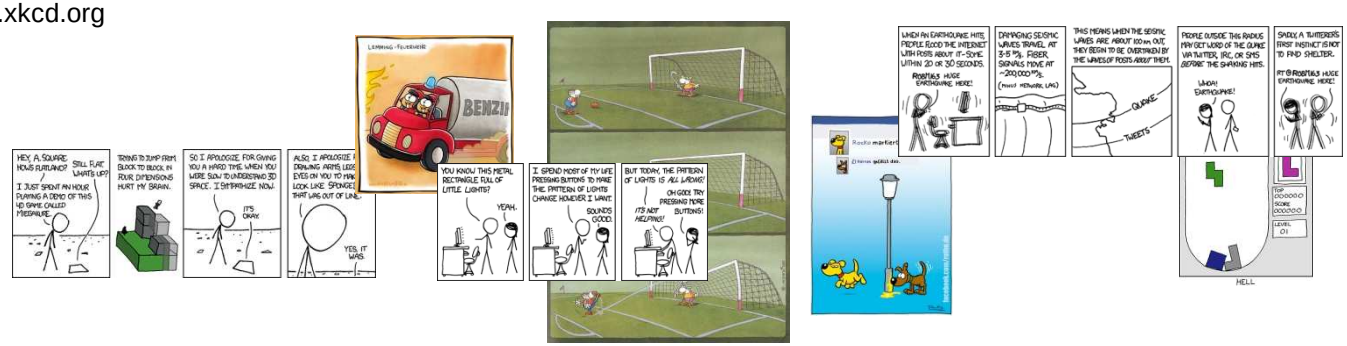
Filtering Music



Filtering Jokes

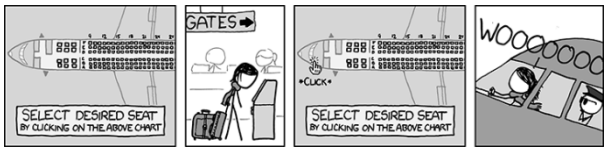


www.xkcd.org



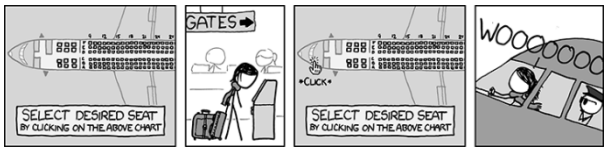
Filtering Jokes

- Let the users rate the jokes
- Sort by average rating

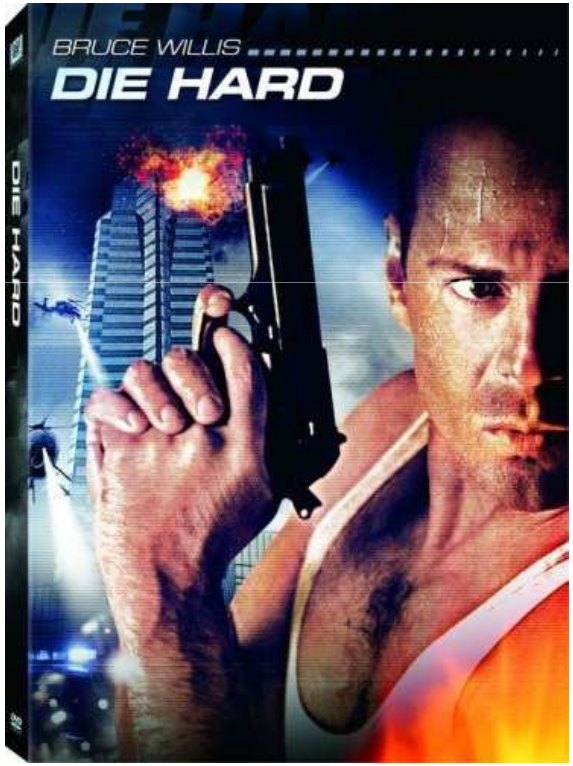


Collaborative Filtering

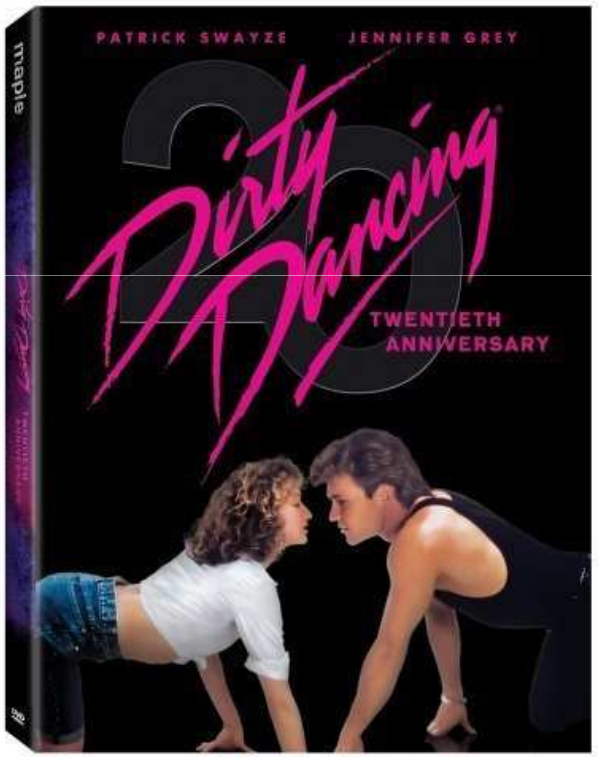
- People who have agreed in the past tend to agree in the future



Good or Bad?



Die Hard (1988)



Dirty Dancing (1987)

Good or Bad?

KEANU REEVES DENNIS HOPPER SANDRA BULLOCK
GET READY FOR RUSH HOUR.
SPEED
JUNE 10

movielens
helping you find the *right* movies

Jester 4.0
Jokes for *Your* Sense of Humor

last·fm™
the social music revolution

amazon.com

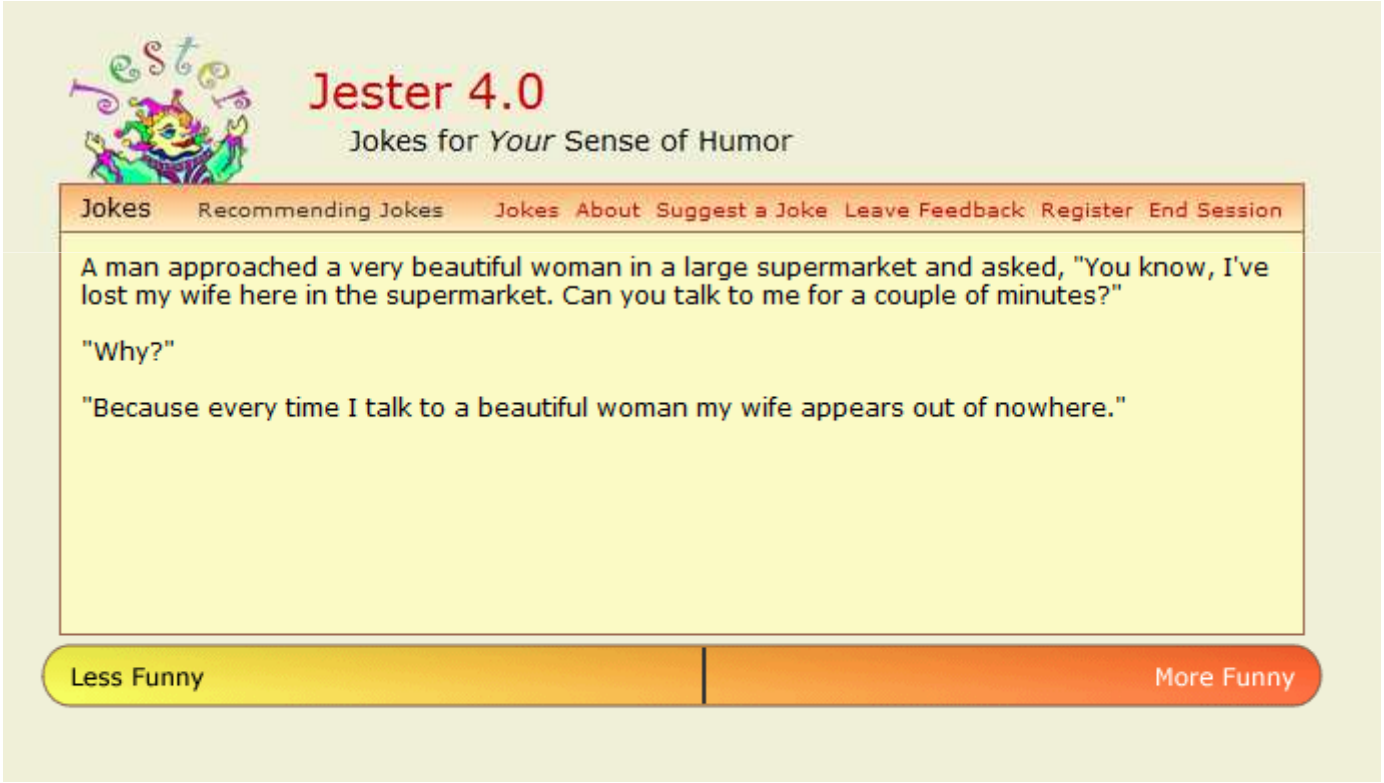
digg

ebaY

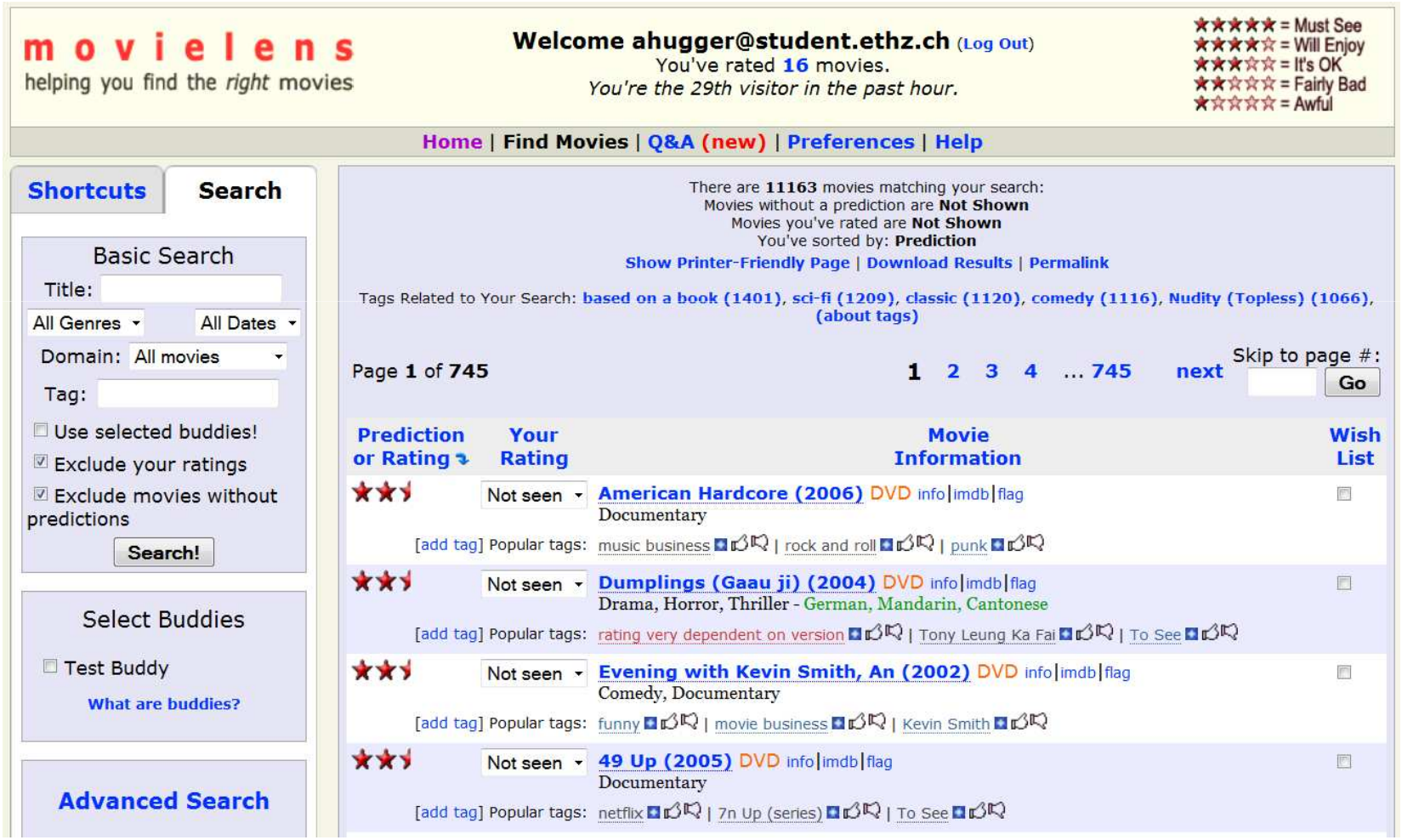
IMDb®

givealink.org
Share your links!

Jester 4.0 (http://eigentaste.berkeley.edu)



MovieLens (http://movielens.org)



Netflix

- www.netflix.org
- DVD/Blue-Ray rental and video streaming
- 1’000’000$ for the first beating the current recommendation algorithm by 10%

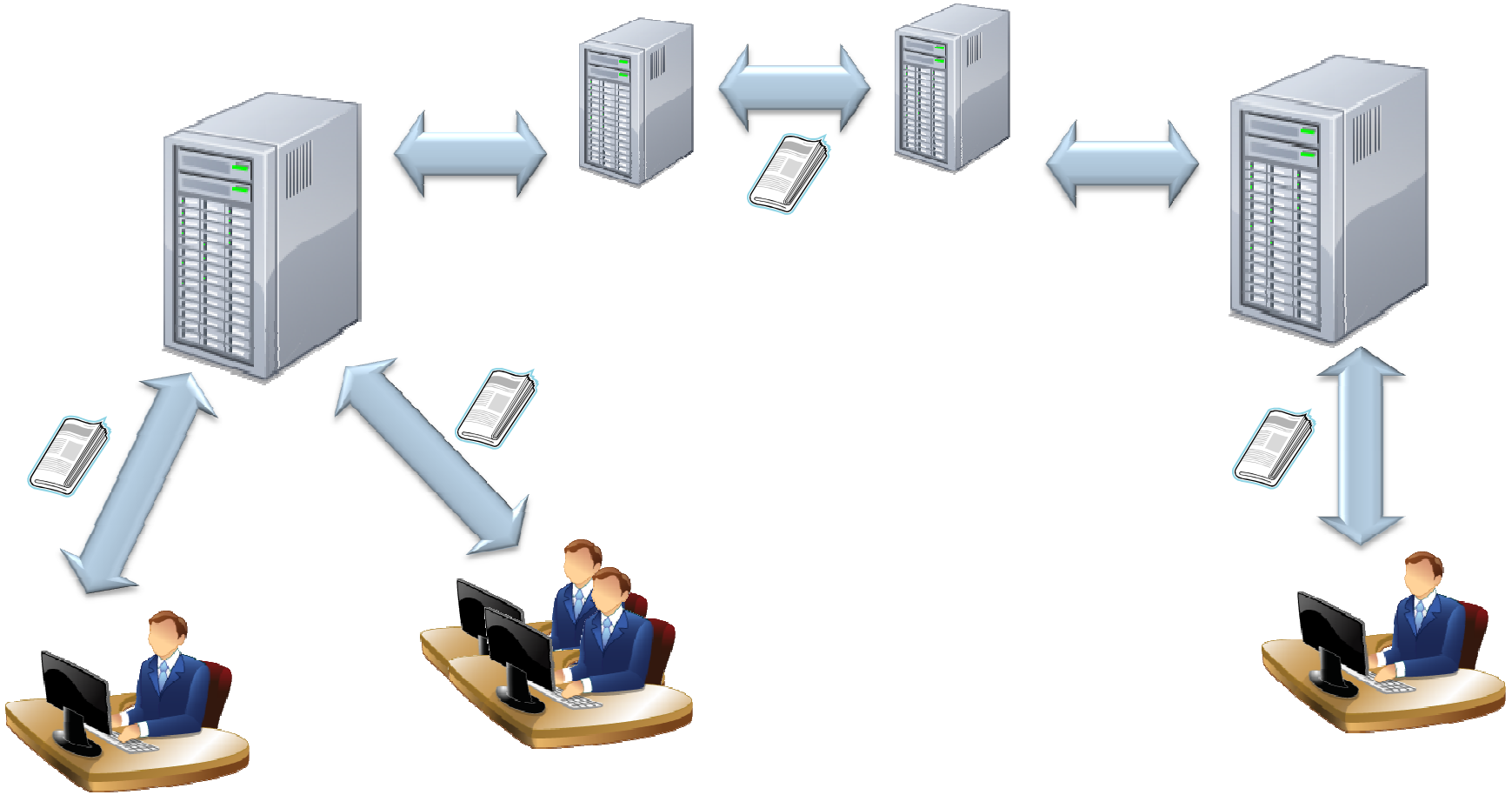
- Competition started in October 2006
- Ended July 2009

GroupLens: An Open Architecture for Collaborative Filtering of NetNews

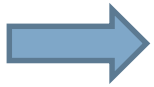
Research paper from 1994 by:

- Paul Resnick, MIT Center for Coordination Science
- Neophytos Iacovou, University of Minnesota
- Mitesh Suchak, MIT Center for Coordination Science
- Peter Bergstrom , University of Minnesota
- John Riedl , University of Minnesota

NetNews



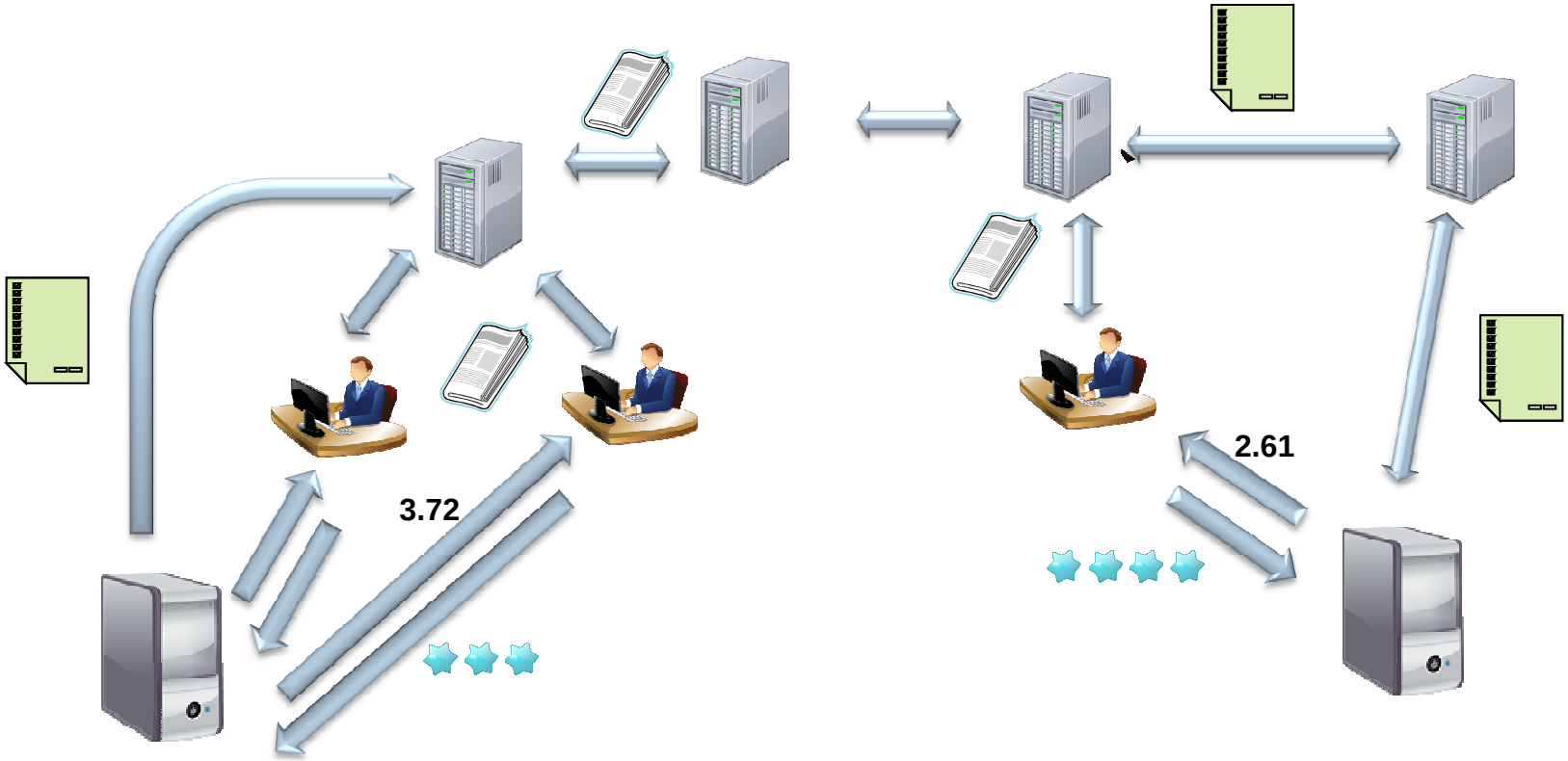
Problems of NetNews



Signal to noise ratio is too low

- Splitting bulletin board into newsgroups
- Moderated newsgroups
- News clients
  - Summary of the author and subject line
  - Display discussion threads together
  - String search facilities
  - Kill files

Modification to NetNews



Predicting Scores

- Score prediction system is robust to certain differences of interpretation of the rating scale
  - One user rates 3-5 and the other 1-3
  - One thinks 1 and the other 5 is best score

Predicting Scores

- Predictions can be modeled as matrix filling

Item #	Ken	Lee	Meg	Nan
1	1	4	2	2
2	5	2	4	4
3			3	
4	2	5		5
5	4	1		1
6	?	2	5	

Predicting Scores

- Assign similarities to each of the other people
- Compute over articles rated by both
- Pearson Correlation Coefficients
  - Between -1 and 1

$$r_{\text{Ken}-\text{Lee}} = \frac{\text{cov}(K, L)}{\sigma_{\text{Ken}} \cdot \sigma_{\text{Lee}}} = \frac{E[(K - \mu_{\text{Ken}})(L - \mu_{\text{Lee}})]}{\sigma_{\text{Ken}} \cdot \sigma_{\text{Lee}}}$$

σ_{Ken} = standard deviation of Ken

μ_{Ken} = average of Ken's ratings

Predicting Scores

- Correlation Coefficients of Ken

User	Correlation
Lee	-0.8
Meg	+1
Nan	0

#	Ken	Lee	Meg	Nan
1	1	4	2	2
2	5	2	4	4
3			3	
4	2	5		5
5	4	1		1
6	?	2	5	

Predicting Scores

- Weighted average of all ratings on article 6
- Ken's prediction is 4.56

$$K_{6_{\text{Prediction}}} = \mu_K + \frac{\sum_{J \in \text{Raters}} (J_6 - \mu_J) \cdot r_{\text{Ken}-J}}{\sum_{J \in \text{Raters}} |r_{\text{Ken}-J}|}$$

Scaling Issues

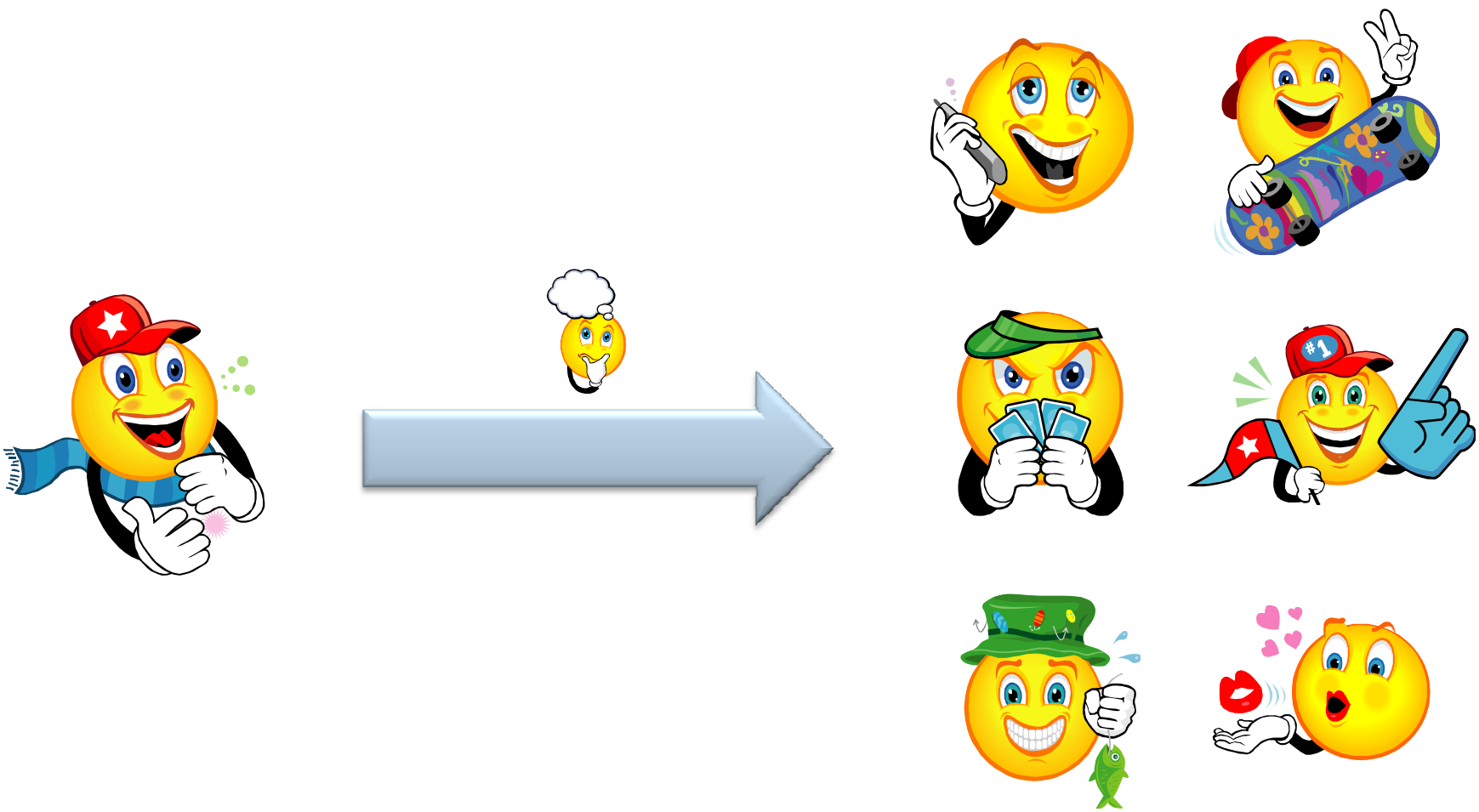
- Relevant performance measures
  - Prediction quality
  - Compute time and disk storage

- Rating is small, but each article may be rated by many users
- Volume of ratings could exceed volume of news

Scaling Issues

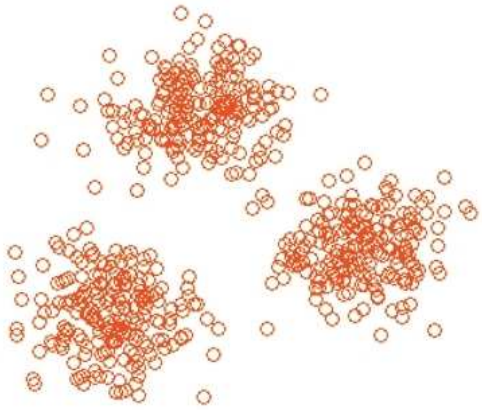
- Pre-fetching ratings and pre-computing predictions keeps user time constant
- High computation complexity
- Volume of all ratings may exceed the storage capacity
  - 100’000 users rate 10 articles per day. 100 bytes are required to store a rating. 1GB of storage required per 10 days.

Cluster Models



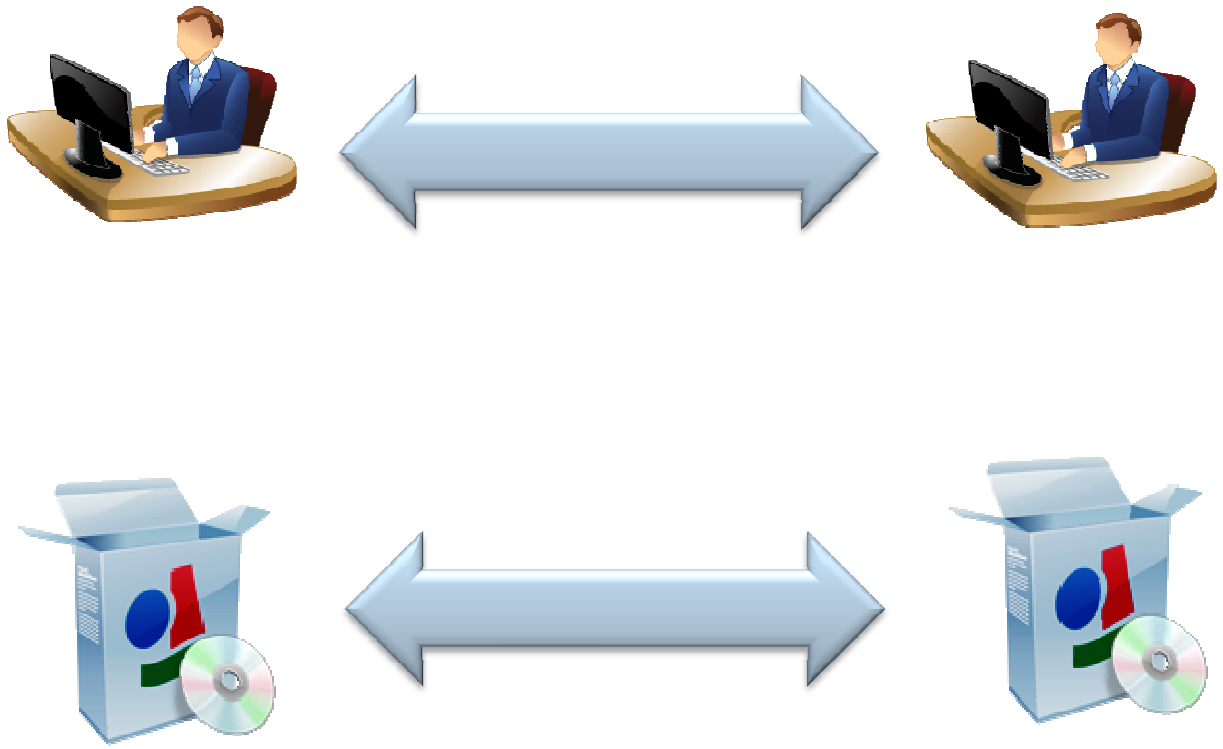
Cluster Models

- Better online scalability and performance than classical collaborative filtering
- Complex and extensive clustering is run offline
- Prediction quality gets reduced



amazon.com

Item-to-Item Collaborative Filtering

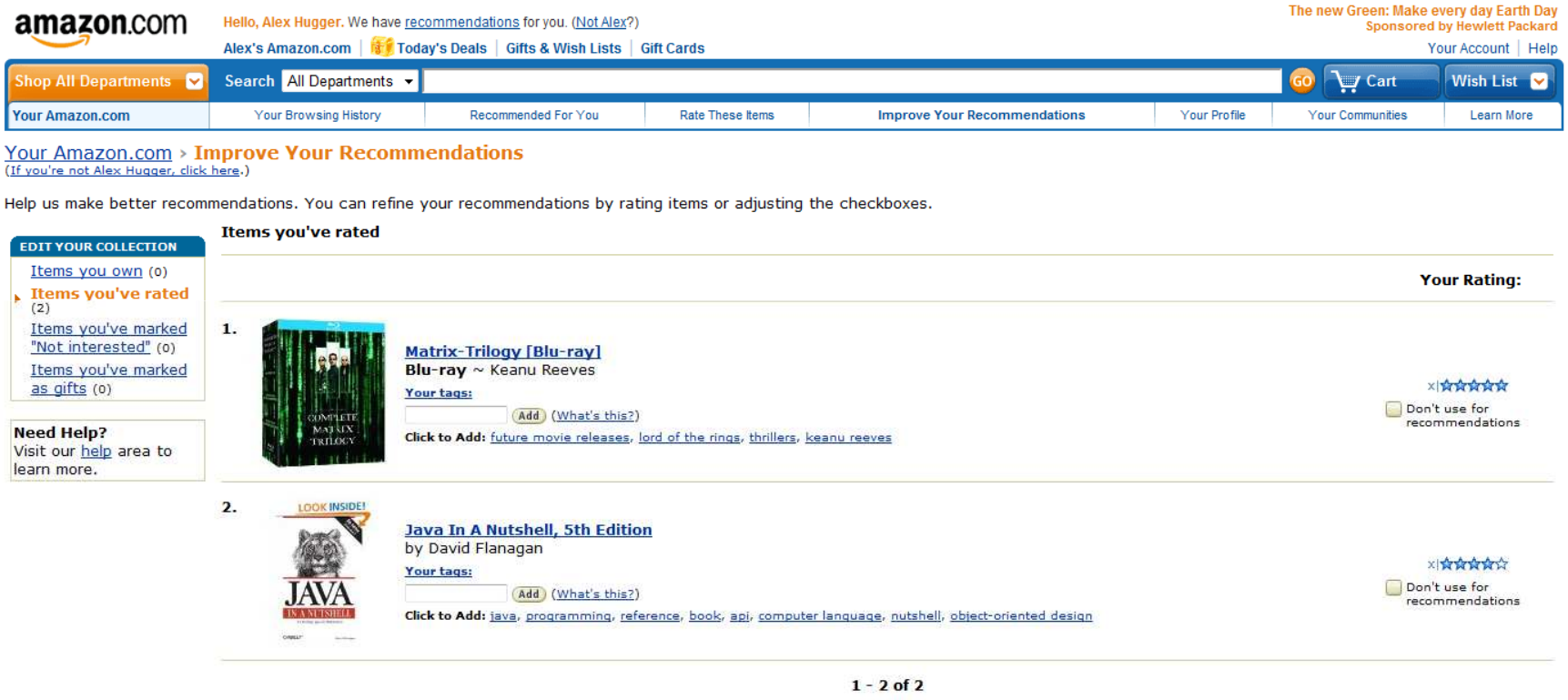


Item-to-Item Collaborative Filtering

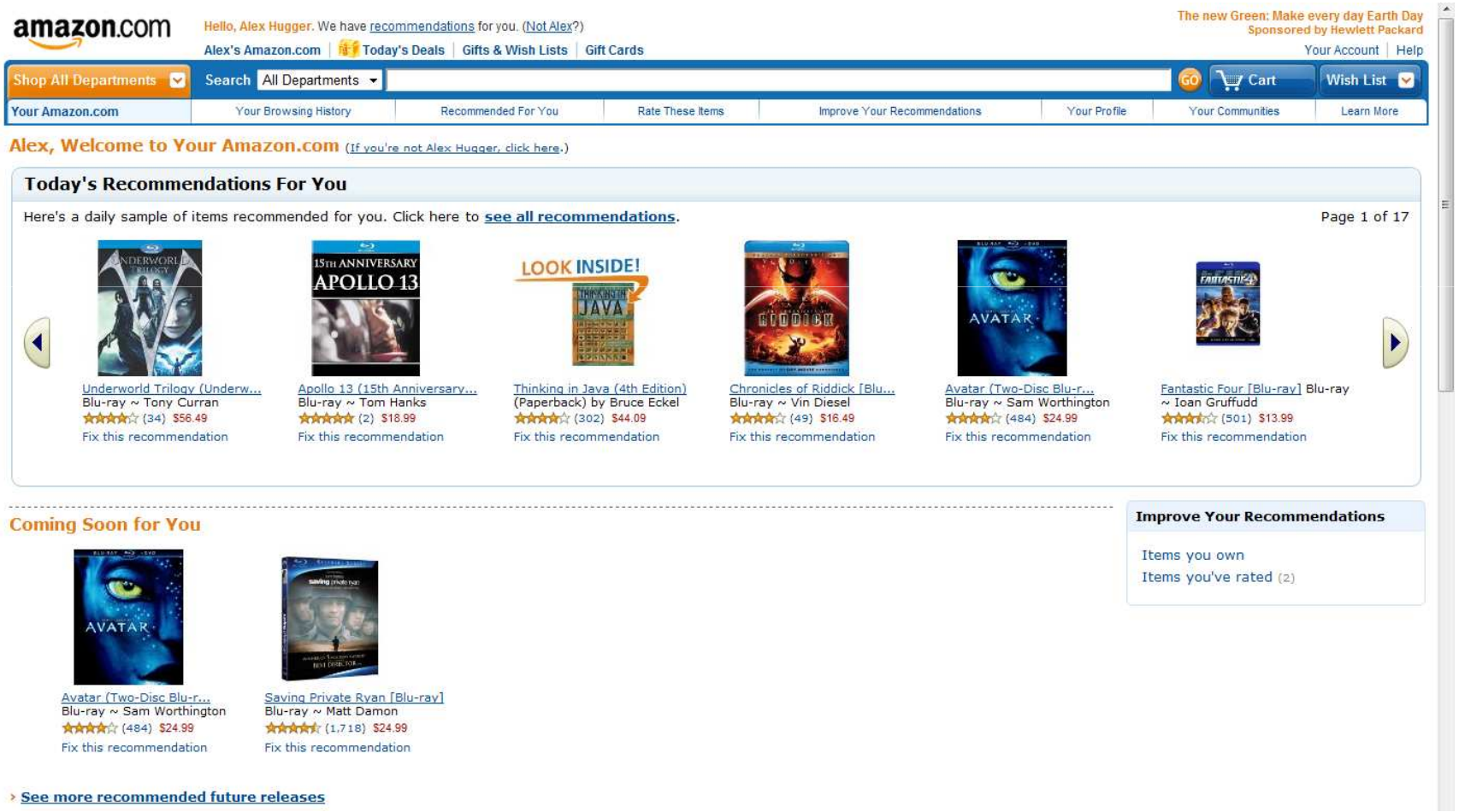
- Amazon.com extensively uses recommendation algorithms
- 10’000’000 products and customers

- Result returned in real-time (< 0.5s)
- Algorithm must respond immediately to new information

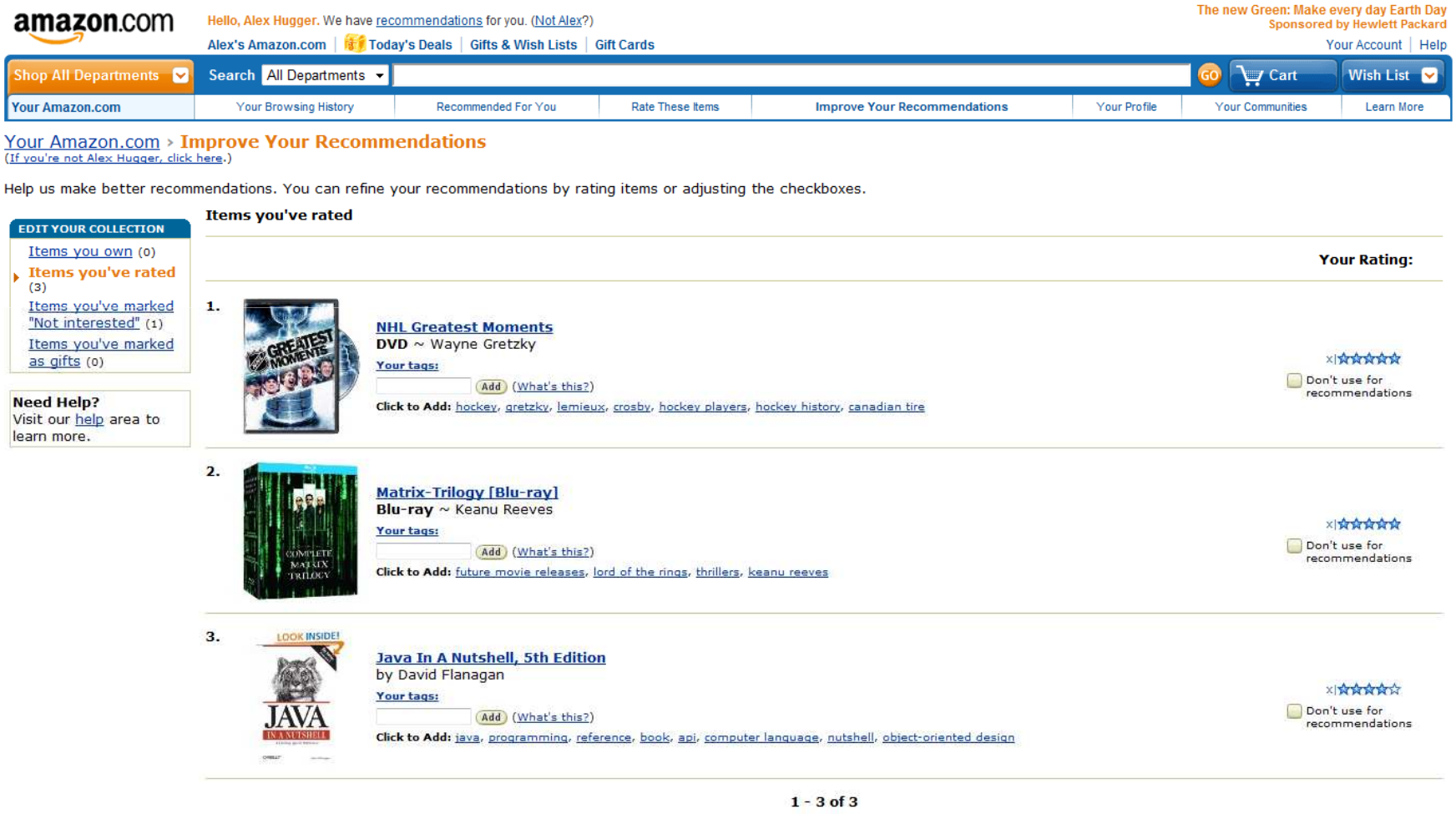
Amazon.com


amazon.com
Hello, Alex Hugger. We have recommendations for you. (Not Alex?)
The new Green: Make every day Earth Day
Sponsored by Hewlett Packard
Alex's Amazon.com | Today's Deals | Gifts & Wish Lists | Gift Cards
Your Account | Help
Shop All Departments
Search All Departments
GO
Cart
Wish List
Your Amazon.com
Your Browsing History
Recommended For You
Rate These Items
Improve Your Recommendations
Your Profile
Your Communities
Learn More
Your Amazon.com › Improve Your Recommendations
(If you're not Alex Hugger, click here.)
Help us make better recommendations. You can refine your recommendations by rating items or adjusting the checkboxes.
EDIT YOUR COLLECTION
Items you own (0)
Items you've rated (2)
Items you've marked "Not interested" (0)
Items you've marked as gifts (0)
Need Help?
Visit our help area to learn more.
Items you've rated
Your Rating:
1.
Matrix-Trilogy [Blu-ray]
Blu-ray ~ Keanu Reeves
Your tags:
Add (What's this?)
Click to Add: future movie releases, lord of the rings, thrillers, keanu reeves
Don't use for recommendations
2.
LOOK INSIDE!
JAVA IN A NUTSHELL
Java In A Nutshell, 5th Edition
by David Flanagan
Your tags:
Add (What's this?)
Click to Add: java, programming, reference, book, api, computer language, nutshell, object-oriented design
Don't use for recommendations
1 - 2 of 2

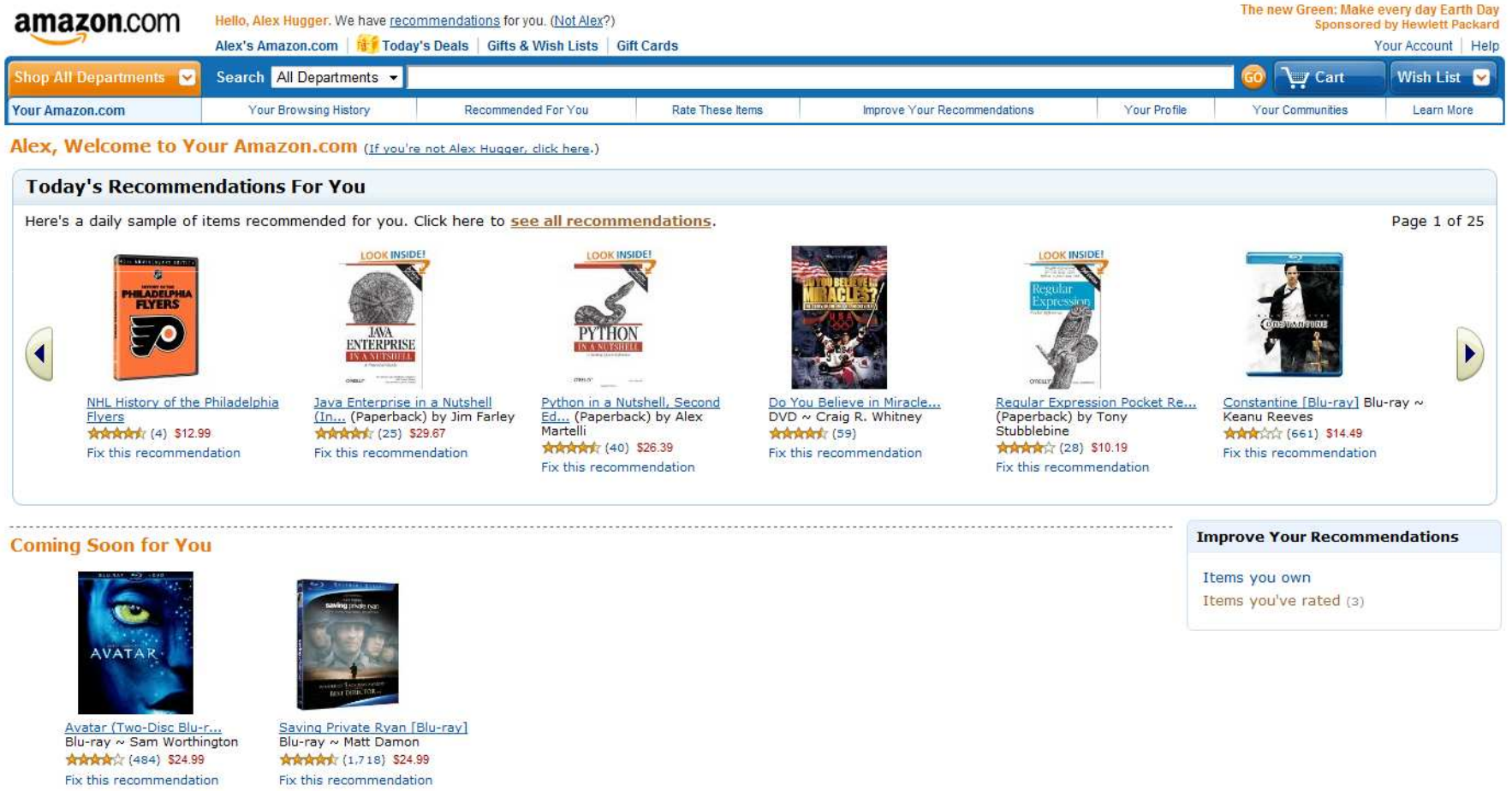
Amazon.com


amazon.com
Hello, Alex Hugger. We have recommendations for you. (Not Alex?)
Alex's Amazon.com | Today's Deals | Gifts & Wish Lists | Gift Cards
The new Green: Make every day Earth Day
Sponsored by Hewlett Packard
Your Account | Help
Shop All Departments
Search All Departments
GO
Cart
Wish List
Your Amazon.com
Your Browsing History
Recommended For You
Rate These Items
Improve Your Recommendations
Your Profile
Your Communities
Learn More
Alex, Welcome to Your Amazon.com (If you're not Alex Hugger, click here.)
Today's Recommendations For You
Here's a daily sample of items recommended for you. Click here to see all recommendations.
Page 1 of 17
Underworld Trilogy (Underw... Blu-ray ~ Tony Curran (34) $56.49 Fix this recommendation
Apollo 13 (15th Anniversary... Blu-ray ~ Tom Hanks (2) $18.99 Fix this recommendation
Thinking in Java (4th Edition) (Paperback) by Bruce Eckel (302) $44.09 Fix this recommendation
Chronicles of Riddick [Blu... Blu-ray ~ Vin Diesel (49) $16.49 Fix this recommendation
Avatar (Two-Disc Blu-r... Blu-ray ~ Sam Worthington (484) $24.99 Fix this recommendation
Fantastic Four [Blu-ray] Blu-ray ~ Ioan Gruffudd (501) $13.99 Fix this recommendation
Coming Soon for You
Avatar (Two-Disc Blu-r... Blu-ray ~ Sam Worthington (484) $24.99 Fix this recommendation
Saving Private Ryan [Blu-ray] Blu-ray ~ Matt Damon (1,718) $24.99 Fix this recommendation
Improve Your Recommendations
Items you own
Items you've rated (2)
See more recommended future releases

Amazon.com


amazon.com
Hello, Alex Hugger. We have recommendations for you. (Not Alex?)
Alex's Amazon.com | Today's Deals | Gifts & Wish Lists | Gift Cards
The new Green: Make every day Earth Day
Sponsored by Hewlett Packard
Your Account | Help
Shop All Departments
Search All Departments
GO
Cart
Wish List
Your Amazon.com
Your Browsing History
Recommended For You
Rate These Items
Improve Your Recommendations
Your Profile
Your Communities
Learn More
Your Amazon.com › Improve Your Recommendations
(If you're not Alex Hugger, click here.)
Help us make better recommendations. You can refine your recommendations by rating items or adjusting the checkboxes.
EDIT YOUR COLLECTION
Items you own (0)
Items you've rated (3)
Items you've marked "Not interested" (1)
Items you've marked as gifts (0)
Need Help?
Visit our help area to learn more.
Items you've rated
Your Rating:
1.
NHL Greatest Moments
DVD ~ Wayne Gretzky
Your tags:
Add (What's this?)
Click to Add: hockey, gretzky, lemieux, crosby, hockey players, hockey history, canadian tire
Don't use for recommendations
2.
Matrix-Trilogy [Blu-ray]
Blu-ray ~ Keanu Reeves
Your tags:
Add (What's this?)
Click to Add: future movie releases, lord of the rings, thrillers, keanu reeves
Don't use for recommendations
3.
LOOK INSIDE!
JAVA IN A NUTSHELL
Java In A Nutshell, 5th Edition
by David Flanagan
Your tags:
Add (What's this?)
Click to Add: java, programming, reference, book, api, computer language, nutshell, object-oriented design
Don't use for recommendations
1 - 3 of 3

Amazon.com


amazon.com
Hello, Alex Hugger. We have recommendations for you. (Not Alex?)
Alex's Amazon.com | Today's Deals | Gifts & Wish Lists | Gift Cards
The new Green: Make every day Earth Day
Sponsored by Hewlett Packard
Your Account | Help
Shop All Departments
Search All Departments
GO
Cart
Wish List
Your Amazon.com
Your Browsing History
Recommended For You
Rate These Items
Improve Your Recommendations
Your Profile
Your Communities
Learn More
Alex, Welcome to Your Amazon.com (If you're not Alex Hugger, click here.)
Today's Recommendations For You
Here's a daily sample of items recommended for you. Click here to see all recommendations.
Page 1 of 25
NHL History of the Philadelphia Flyers
(4) $12.99
Fix this recommendation
Java Enterprise in a Nutshell (In... (Paperback) by Jim Farley
(25) $29.67
Fix this recommendation
Python in a Nutshell, Second Ed... (Paperback) by Alex Martelli
(40) $26.39
Fix this recommendation
Do You Believe in Miracle... DVD ~ Craig R. Whitney
(59)
Fix this recommendation
Regular Expression Pocket Re... (Paperback) by Tony Stubblebine
(28) $10.19
Fix this recommendation
Constantine [Blu-ray] Blu-ray ~ Keanu Reeves
(661) $14.49
Fix this recommendation
Coming Soon for You
Avatar (Two-Disc Blu-r... Blu-ray ~ Sam Worthington
(484) $24.99
Fix this recommendation
Saving Private Ryan [Blu-ray] Blu-ray ~ Matt Damon
(1,718) $24.99
Fix this recommendation
Improve Your Recommendations
Items you own
Items you've rated (3)

amazon.com

How It Works - Offline

- Similar-items table
- Calculating similarity between a single product and all related products
  - Complexity: $O(mn^2)$ - in practice: $O(mn)$
    - m: number of users
    - n: number of items

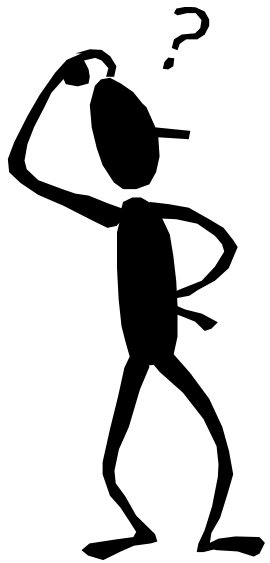
amazon.com

How It Works - Online

- Given a similar-items table
- Find all similar items to each of the users ratings and purchases
- Aggregate those items
- Recommend most popular and correlated items
- Number of users has no effect on performance

General difficulties

- Cold start
- Self-fulfilling prophecy
- Recommendations for groups
- Evaluation of recommendation systems



Conclusion

- Effective form of targeted marketing
- Mostly used in e-commerce business
  - But can always be used when signal to noise ratio is too low

Questions?

References

- GroupLens: An Open Architecture for Collaborative Filtering of NetNews
  - Published 1994
  - Paul Resnick, MIT Center for Coordination Science
  - Neophytos Iacovou, University of Minnesota
  - Mitesh Suchak, MIT Center for Coordination Science
  - Peter Bergstrom , University of Minnesota
  - John Riedl , University of Minnesota
- Amazon.com Recommendations
  - Published 2003
  - Greg Linden
  - Brent Smith
  - Jeremy York