

Introduction to Collaborative Filtering using



Recommendation Engines

- Amazon
- Stumbleupon
- Youtube
- Last.fm
- Netflix
- Digg
- Google News

Recommendation Engines

Why?

- For internet users → Useful, fun
- For e-commerce sites → Boosts sales
- For developers → Interesting, cool



Collaborative
Filtering



Clustering



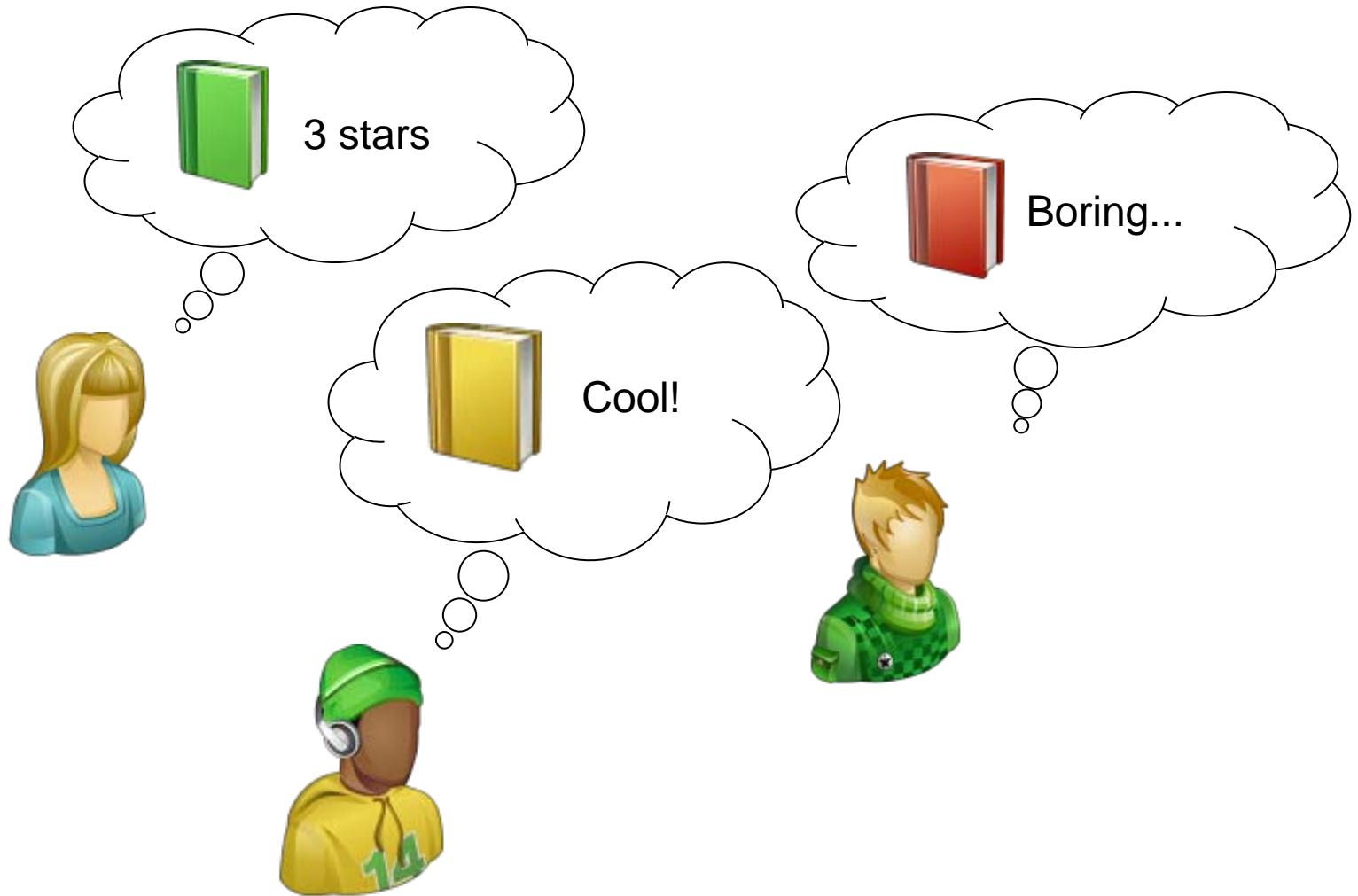
Is this SPAM?

Classification

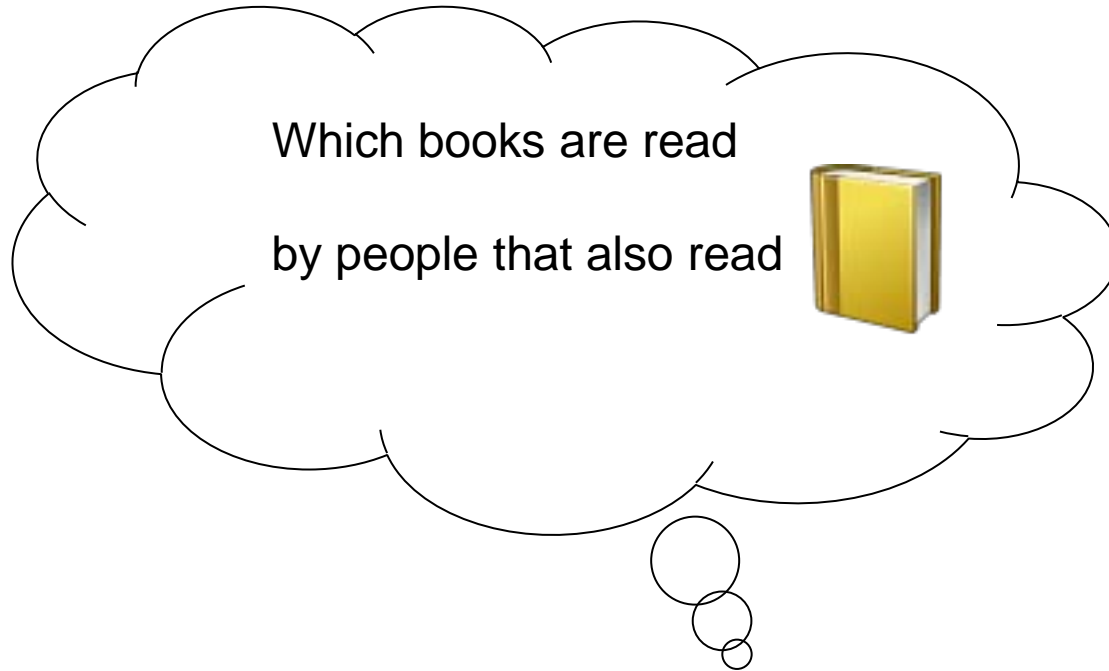
Users & Items



Preferences



Item-based recommendation

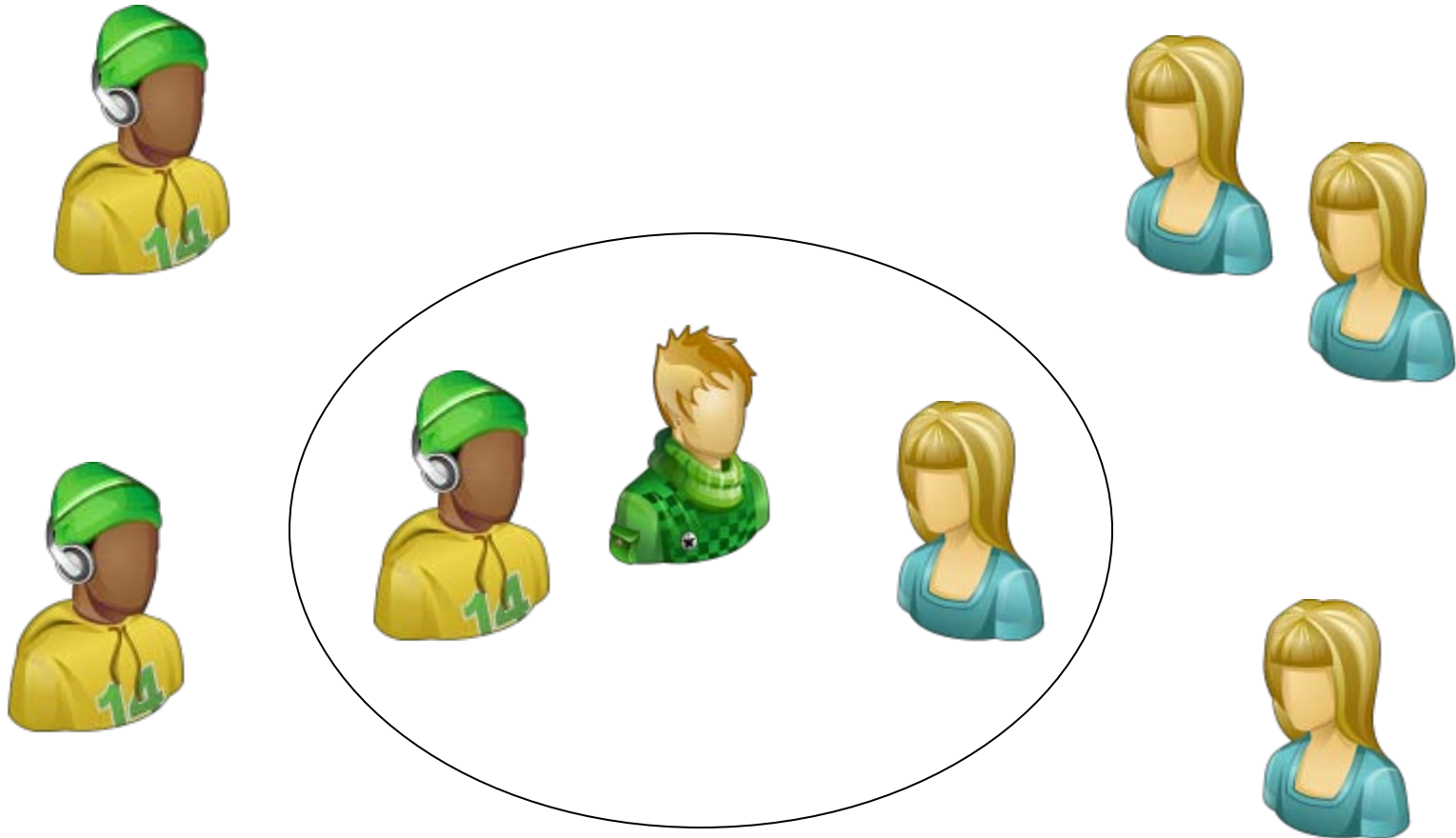


User-based recommendation

We've got similar tastes,
read any good books?



User neighbourhood



Demo

Movielens dataset (100K)

934



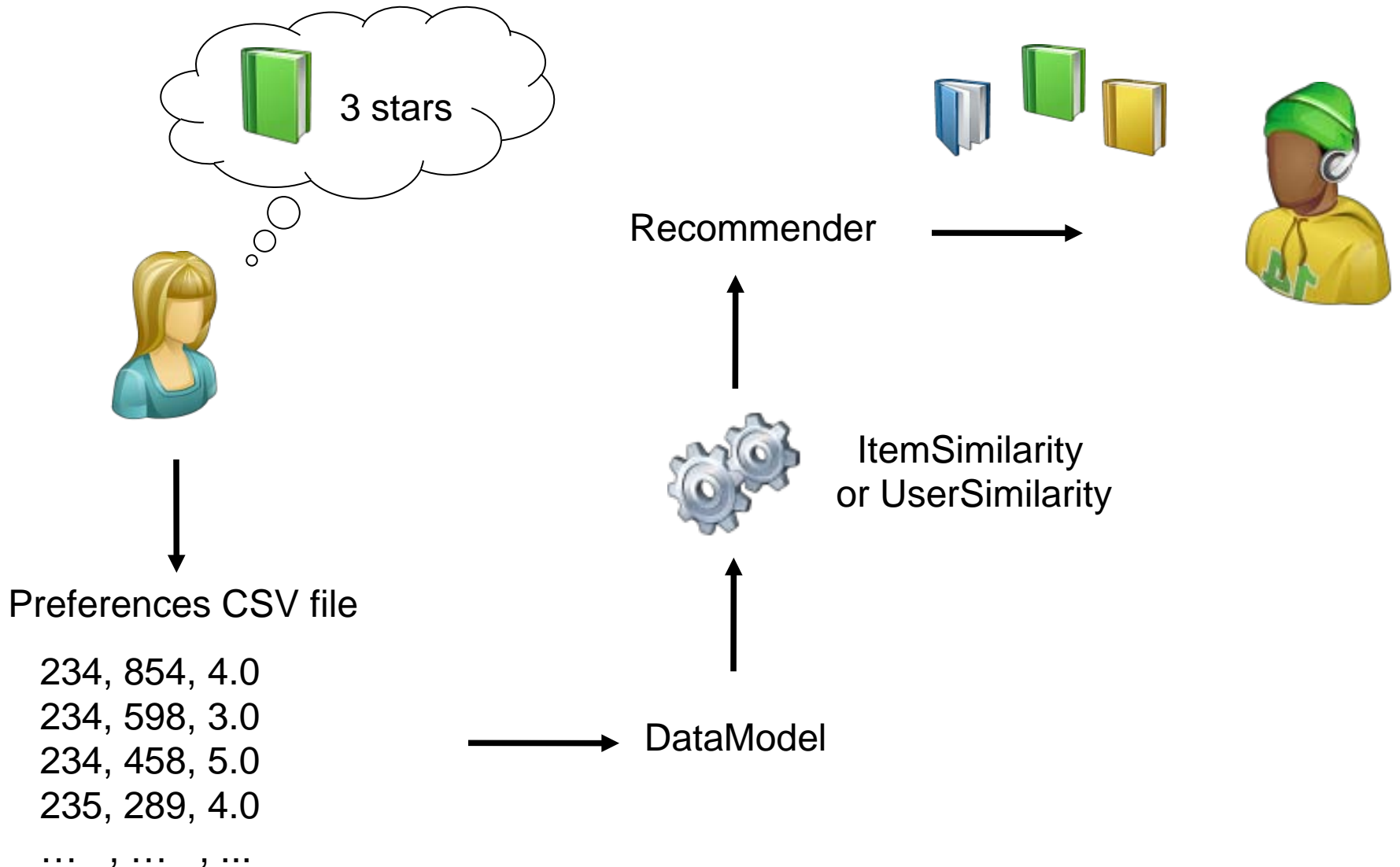
100.000



1682



Taste Architecture



Preferences

- Preference
 - long userId;
 - long itemId;
 - float value;
- PreferenceArray
- BooleanUserPreferenceArray & BooleanItemPreferenceArray

DataModels

- FileDataModel
- GenericJDBCDataModel
- MySQLDataModel

Similarity Algorithms

Class	Explicit?	Implicit?
EuclidianDistanceSimilarity	✓	
PearsonCorrelationSimilarity	✓	
UncenteredCosineSimilarity	✓	
SpearmanCorrelationSimilarity	✓	
TanimotoCoefficientSimilarity	✓	✓
LogLikelihoodSimilarity	✓	✓
Slope One	✓	

Precomputed Similarities

- MySQLJDBCItemSimilarity
- Generic*Similarity
 - GenericItemSimilarity.ItemItemSimilarity
 - GenericUserSimilarity.UserUserSimilarity

Recommenders

- User- and/or Item-based recommendation
- Refresh logic
- Access to DataModel
- “Recommended because”

References



Mahout in Action EAP



Mailinglist



<http://blog.jteam.nl>