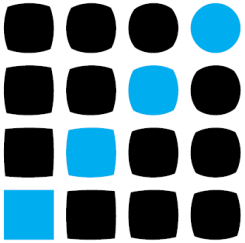


adaptTo()

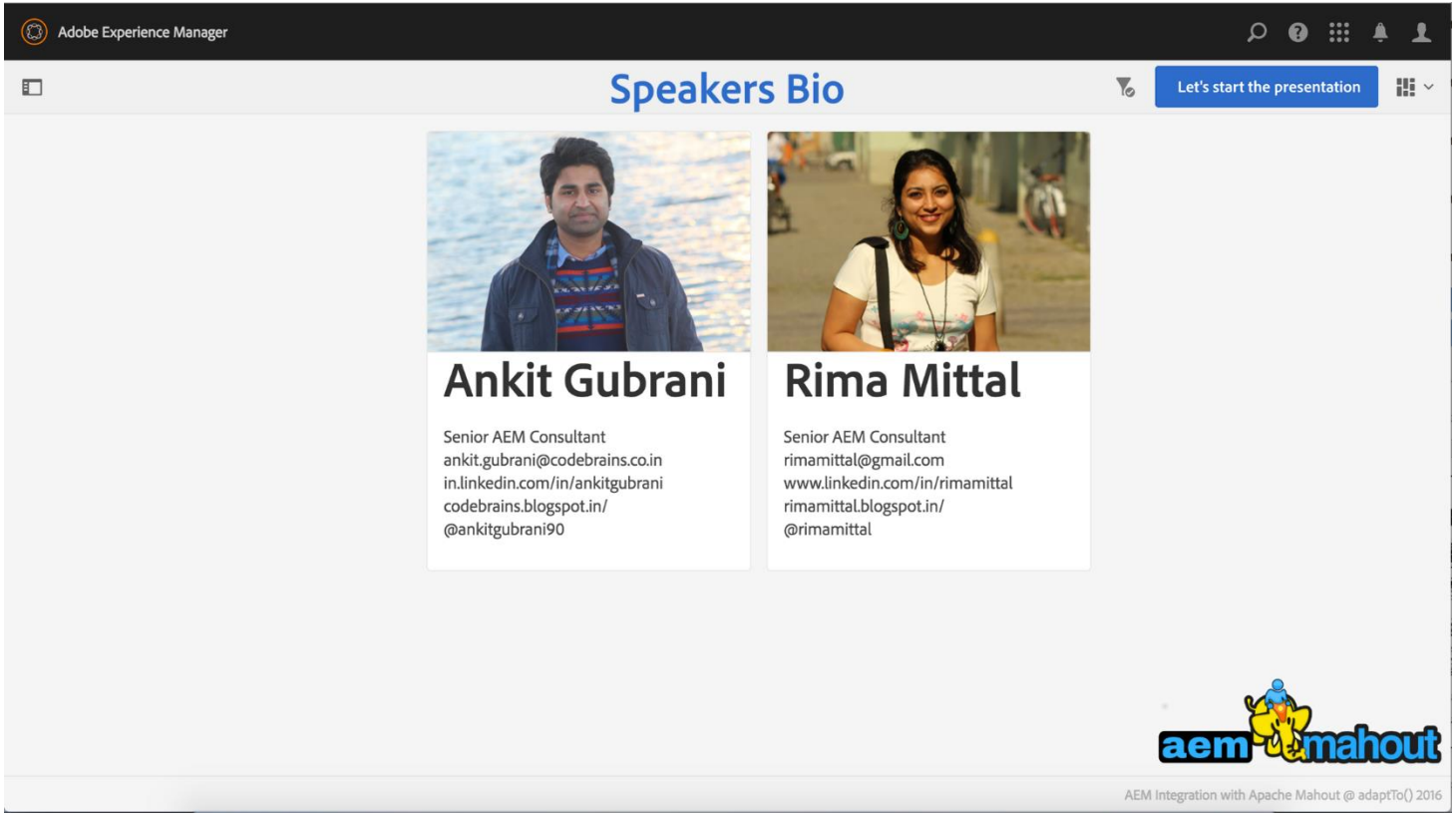
APACHE SLING & FRIENDS TECH MEETUP

BERLIN, 26-28 SEPTEMBER 2016

Integrating Apache Mahout with AEM

Ankit Gubrani & Rima Mittal

Speakers Bio



Agenda

Agenda

- Introduction to Apache Mahout
- Machine Learning
- Recommendations
- AEM with Apache Mahout
- Demo
- Extension Points

Introduction to Apache Mahout

What is Apache Mahout?

- Project of the Apache Software Foundation.
- Producing free implementations of scalable machine learning algorithms, written in Java.

History

- Started as a Lucene sub-project.
- Became Apache TLP in April 2010.
- Latest version – 0.12.2 – Released on 13th June 2016.

Why Apache Mahout?

- Increasing volume of data!
- Traditional Data mining algorithms struggle to process very large datasets.
- Apache Mahout to the rescue!

Traditional Machine Learning

Machine Learning with Mahout

Applications

- Adobe, Facebook, LinkedIn, Twitter and Yahoo use Mahout internally.
- Twitter uses Mahout for interest modelling.
- Yahoo! Uses Mahout for pattern mining.

Machine Learning

Machine Learning

- Programming computers to optimize a Performance Criterion using Example Data or Past Experience
  - Branch of Artificial Intelligence.
  - Computers evolve behavior based on Empirical data.

Techniques

- Supervised Learning
  - Use Labelled training data to create a classifier that can predict output for unseen inputs.
- Unsupervised Learning
  - Use Unlabeled training data to create a function that can predict output.

Machine Learning with Apache Mahout

- Data Science use cases Mahout supports:
  - Collaborative Filtering
  - Clustering
  - Classification

Collaborative Filtering

- User behavior mining to make product recommendations.

Clustering

- Organizing items into naturally occurring groups, such that items belonging to same group are similar to each other

Classification

- Learning from existing categorizations and assigning unclassified items to the best category

Recommendations

Apache Mahout Recommendation Engine

- Helps users find items they might like based on historical behavior and preferences.
- Mahout provides a rich set of components from which a customized recommender system can be constructed using a selection of Algorithms.

Architecture

Top Level Packages

- DataModel
- UserSimilarity
- ItemSimilarity
- UserNeighboorhood
- Recommender

AEM with Apache Mahout

Checklist

- AEM 6.2
- Mahout as a Maven Dependency

```
<dependency>
        <groupId>org.apache.mahout</groupId>
        <artifactId>mahout-mr</artifactId>
        <version>0.10.0</version>
</dependency>
```

JCRDataModel

- DataModel
  - Implementations representing a repository of information about users and their associated preferences.
    - AbstractDataModel, JDBCDataModel, FileDataModel, GenericBooleanPrefDataModel, GenericDataModel.
    - AEM - **JCRDataModel.**

Code (1/2)

- Using The AEM JCRDataModel

```
public JSONArray getUserBasedRecommendations(ResourceResolver
resourceResolver, String userId, int numberOfRecommendations) {

    //Creating JCRDataModel to fetch information from JCR
    DataModel model = JCRDataModel.createDataModel(resourceResolver);

}
```

Code (2/2)

- AEM-Mahout Recommendation steps

```
UserSimilarity userSimilarity = getSimilarity(model);

UserNeighborhood neighborhood = getNeighbourHood(N_NEIGHOBUR_HOOD,
userSimilarity, model);

GenericUserBasedRecommender recommender = new
GenericUserBasedRecommender(model, neighborhood, userSimilarity);

recommendations = recommender.recommend(userIdHash,
numberOfRecommedations, null, false);
```

AEM Product Recommendation

- User Based Recommendation
  - Takes user ratings into consideration
  - Based on PearsonCorrelationSimilarity
  - Uses NearestNUserNeighborhood

Configuring JCRDataModel

- Configurations
  - User Generated Content Path
    - /content/usergenerated/asi/jcr
  - Product Path
    - /etc/commerce/products/geometrixx-outdoors
  - Rating Resource Type
    - Defaults to social/tally/components/response

Demo

Appendix

Appendix

- https://mahout.apache.org/
- http://www.slideshare.net/VaradMeru/introduction-to-mahout-and-machine-learning
- https://www.youtube.com/watch?v=iMAMYzfRiS4

Clone the code!

Code Repository

- https://github.com/rimamittal/AEMMahout.git

Questions.

Thank you.