IMPROVING DISCOVERY & PATRON EXPERIENCE THROUGH DATA MINING

Florida International University
Boylan (Keven) Guan – Lead Developer, GIS Center
Jamie Rogers – Assistant Director, Digital Collections Center
1. Introduction
2. Problem statement
3. Overview of the recommendation system concept
4. Technical design and implementation
5. Sharing the technology
6. Questions?
MAIN PROBLEMS

- Limited Discoverability
  - 12,376 (24.9%) out of the total 49,737
- Lack of assistance to explore the system
  - 700 clients out of 1,127 with a specific record
- Drop off visits
  - 978,814 visits (58.8%) out of total 1,664,813 visits
INTRODUCTION | PROBLEM | OVERVIEW | DESIGN & IMPLEMENTATION | INSTITUTIONS
FIU Sound and Image

The Sound & Image Collections consist of audiovisual resources intended to support the teaching needs of the faculty, as well as supplementing the academic research needs of students while providing cultural awareness and enrichment to all.
FIU Sound and Image

The Sound & Image Collections consist of audiovisual resources intended to support the teaching needs of the faculty, as well as supplementing the academic research needs of students while providing cultural awareness and enrichment to all.
Variation of Target Users

- Variation of User Regions
- UI Design
- Huge amount of contents
- Enter point of the system
- Correlation & Separation of collection topics
- Variation of User behaviors
What do the users **WANT** to know?
What do the users **NEED** to know?

- Get to know the users & their region information
- Awareness about the content
- Awareness about the system condition (usages, statistic, user interactions)

**Data Mining !!**  \[ \rightarrow \]  **Recommendation System**
Discoverability
Usability
Drop-off Visits

UI Interactions
Failed

Key Factors
Identified

Intelligent Recommendation System

Cannot acquire additional user information
Minimize the Side Effect
Low impact on the performance
OVERVIEW OF THE RECOMMENDATION SYSTEM CONCEPT
• **200** research articles were published
• **62** proposed methods

**RECOMMENDERS**

- **Content Based Filtering (CBF)**: 55%
- **Collaborative Filtering (CF)**: 16%
- **Graph-based Recommendations**: 11%
- **Other Recommendation Concepts** (Stereotyping, Item-centric Recommendations, and Hybrid Recommendations): 18%
ADDITIONAL METHODS

GLOBAL RELEVANCE

LOCATION BASED

QUERY/TERM SUGGESTION (QS / TS)
DPANTHER ARCHITECTURE

STEREOTYPING

- CBF
- CF
- Global Relevance
- TS/QS
- Location Based

Content Based Filtering (CBF) | Collaborative Filtering (CF) | Query/Term Suggestion (QS / TS)

INTRODUCTION | PROBLEM | OVERVIEW | DESIGN & IMPLEMENTATION | INSTITUTIONS
TECHNICAL DESIGN AND IMPLEMENTATION
## RECOMMENDATION STRATEGY

<table>
<thead>
<tr>
<th></th>
<th>CBF</th>
<th>CF</th>
<th>Global Relevance</th>
<th>QS / TS</th>
<th>Location Based</th>
</tr>
</thead>
<tbody>
<tr>
<td>Internal Users</td>
<td>👎🏻</td>
<td>👎🏻</td>
<td>💖</td>
<td>🌻</td>
<td>🌻</td>
</tr>
<tr>
<td>Academic Users</td>
<td>💖</td>
<td>😊</td>
<td>🍊</td>
<td>🍊</td>
<td>🍊</td>
</tr>
<tr>
<td>Digital Library Users</td>
<td>😊</td>
<td>😊</td>
<td>😊</td>
<td>😊</td>
<td>😊</td>
</tr>
<tr>
<td>Active Users</td>
<td>😊</td>
<td>😊</td>
<td>😊</td>
<td>😊</td>
<td>😊</td>
</tr>
<tr>
<td>Passive Users</td>
<td>😊</td>
<td>😊</td>
<td>😊</td>
<td>😊</td>
<td>😊</td>
</tr>
</tbody>
</table>

*Content Based Filtering (CBF) | Collaborative Filtering (CF) | Query/Term Suggestion (QS / TS)*
BUILD USERS VECTOR

Stereotyping ➔ Classification ➔ Building the vector for users

Web Server Log + Database Log

<table>
<thead>
<tr>
<th>Collections</th>
<th>Items</th>
<th>Search Terms</th>
<th>ID Involved?</th>
<th>Geo-Location</th>
<th>User Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>{number}</td>
<td>{number}</td>
<td>{number}</td>
<td>{True / False}</td>
<td>{Nominal}</td>
<td>(Nominal)</td>
</tr>
</tbody>
</table>
STEREOTYPING DESIGN – SAMPLE DECISION TREE

Use 3 out of 5 dimensions

Need to use a data mining classifier

TOO MANY VARIATIONS

IP Address

Metadata

Geo-Locations

Search Terms

User who haven’t review any record

Exist?

Yes

No

Search Terms

Exist?

Yes

No

Does identifier shows in search term

Search Terms

Library Users

Active Users

Passive Users

On Statists Page

On Item Default Page

Recommendation

• Popular in areas
• Popular items

Recommendation

• CBF

Recommendation

• New Added

Recommendation

• CBF

Recommendation

• CBF

Recommendation

• CF

On Statists Page

Recommendation

• Popular Items
• Popular in areas

Internal Users

Academic Users

Hide recommendation on start page

Default

Default
USER DATA CLASSIFICATION

- 15,126 Distinct Users
- 5,583 Distinct Search Terms
- 12,376 Items Visited
- 92 Collections
- 169 Countries from the World
- 335 Counties in the US

Label data (training data)

Run Tree Based Classifier (e.g. J48)
**RECOMMENDATION FILTERING**

**Collaborative filtering (want to know)**

\[
\text{sim}(i, j) = \frac{\sum_{u \in U} (R_{u,i} - \bar{R}_u)(R_{u,j} - \bar{R}_u)}{\sqrt{\sum_{u \in U} (R_{u,i} - \bar{R}_u)^2} \sqrt{\sum_{u \in U} (R_{u,j} - \bar{R}_u)^2}}
\]

Where the set \( U \) means the group of users who rate both item \( i \) and item \( j \), \( R_{u,i} \) or \( R_{u,j} \) means the score for such the item by the user, \( \bar{R}_u \) is the average of the \( u \)-th user’s ratings.

**Content Based Filtering (should know)**

\[
\text{Jaccard Similarity } (a, b) = \frac{| \text{Intersection } (A, B) |}{| \text{Union } (A, B) |}
\]

Where \( a, b \) are any two items in dPanther repository system \( A \), \( B \) are the set of metadata fields and value of item \( a \) and \( b \).
Query / Term Suggestion (want to know)

Global Relevance / Location Based Recommendation (should know)

\[ J(a, b) = \frac{|DS_a \cap DS_b|}{|DS_a \cup DS_b|} \]

- Where \( a \) is the input search term
- \( b \) is the subject key words value of any item in dPanther

GR Score \( \rightarrow A_C \propto \text{(numbr of hits of the item)} \rightarrow A_C \)

Where \( A_C \) is the location information for collection any collection C in dPanther.
<table>
<thead>
<tr>
<th></th>
<th>CBF</th>
<th>CF</th>
<th>Global Relevance</th>
<th>QS / TS</th>
<th>Location Based</th>
</tr>
</thead>
<tbody>
<tr>
<td>Internal Users</td>
<td>🌟</td>
<td>🌟</td>
<td>⭐️</td>
<td>🌟</td>
<td>🌟</td>
</tr>
<tr>
<td>Academic Users</td>
<td>🌟</td>
<td>🌟</td>
<td>🌟</td>
<td>🌟</td>
<td>🌟</td>
</tr>
<tr>
<td>Digital Library Users</td>
<td>🌟</td>
<td>🌟</td>
<td>🌟</td>
<td>🌟</td>
<td>🌟</td>
</tr>
<tr>
<td>Active Users</td>
<td>🌟</td>
<td>🌟</td>
<td>🌟</td>
<td>🌟</td>
<td>🌟</td>
</tr>
<tr>
<td>Passive Users</td>
<td>🌟</td>
<td>🌟</td>
<td>🌟</td>
<td>🌟</td>
<td>🌟</td>
</tr>
</tbody>
</table>

*Content Based Filtering (CBF) | Collaborative Filtering (CF) | Query/Term Suggestion (QS / TS)*
Internal Users

INTRODUCTION | PROBLEM | OVERVIEW | DESIGN & IMPLEMENTATION | INSTITUTIONS
Internal Users

CBF & CF are hidden

Global Relevance
<table>
<thead>
<tr>
<th></th>
<th>CBF</th>
<th>CF</th>
<th>Global Relevance</th>
<th>QS / TS</th>
<th>Location Based</th>
</tr>
</thead>
<tbody>
<tr>
<td>Internal Users</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Academic Users</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Digital Library Users</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Active Users</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Passive Users</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- CBF: Content-based filtering
- CF: Collaborative filtering
- Global Relevance: How relevant is the content globally?
- QS / TS: Quality and trustworthiness
- Location Based: How relevant is the content based on location?
Academic Users

The Florida International University Panther digital repository hosts unique digital collections of cultural, historical, scholarly, educational, and community resources, including rare books, manuscripts, maps, photographs, and oral histories held by the university as well as partner institutions.

INTRODUCTION | PROBLEM | OVERVIEW | DESIGN & IMPLEMENTATION | INSTITUTIONS

QS / TS are Hidden

Location Based & Global Relevance are Hidden
Academic Users
<table>
<thead>
<tr>
<th></th>
<th>CBF</th>
<th>CF</th>
<th>Global Relevance</th>
<th>QS / TS</th>
<th>Location Based</th>
</tr>
</thead>
<tbody>
<tr>
<td>Internal Users</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Academic Users</td>
<td>▼</td>
<td>▼</td>
<td><img src="image" alt="Heart" /></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Digital Library Users</td>
<td><img src="image" alt="Smiley" /></td>
<td><img src="image" alt="Smiley" /></td>
<td><img src="image" alt="Smiley" /></td>
<td><img src="image" alt="Smiley" /></td>
<td><img src="image" alt="Smiley" /></td>
</tr>
<tr>
<td>Active Users</td>
<td><img src="image" alt="Smiley" /></td>
<td><img src="image" alt="Smiley" /></td>
<td><img src="image" alt="Smiley" /></td>
<td><img src="image" alt="Smiley" /></td>
<td><img src="image" alt="Smiley" /></td>
</tr>
<tr>
<td>Passive Users</td>
<td><img src="image" alt="Smiley" /></td>
<td><img src="image" alt="Smiley" /></td>
<td><img src="image" alt="Smiley" /></td>
<td><img src="image" alt="Smiley" /></td>
<td><img src="image" alt="Smiley" /></td>
</tr>
</tbody>
</table>
Digital Library Users

INTRODUCTION | PROBLEM | OVERVIEW | DESIGN & IMPLEMENTATION | INSTITUTIONS
Digital Library Users

INTRODUCTION | PROBLEM | OVERVIEW | DESIGN & IMPLEMENTATION | INSTITUTIONS
Sample Use Case
A bill is to be introduced in the Florida legislature

- Essay about the history of Seminoles in Florida

\[ \text{The efforts being made to reclaim the Everglades (a broad expanse of prairie country covered with water) and the attendant "booming" by land companies and speculators, have materially changed the conditions under which the Seminoles formerly lived. The construction of the various drainage canals have made that part of the State more accessible, and thousands of people have either gone to southern Florida or invested money in "well-watered" tracts that were offered for sale through the medium of alluring literature.} \]

\[ \text{While the influx of these homeseekers and speculators had the effect of compelling the Seminoles to flee from the approach of the white man's civilization rather than risk a repetition of their bitter experience from previous contact with it, public attention has also been sharply called to the wrongs and neglect from which they have suffered and to the importance. Indeed, the necessity--of all lovers of justice demanding that the miserable treatment of these Red Patriots shall be stoned for in some measure before it is too late.} \]

\[ \text{"Why have the Seminoles of Florida been continually denied} \]
The Florida Seminoles and their rights in the Everglades
- The Florida Seminoles and their rights in the Everglades.
INTRODUCTION | PROBLEM | OVERVIEW | DESIGN & IMPLEMENTATION | INSTITUTIONS
POSSIBLE USE BY OTHER INSTITUTIONS
QUESTIONS?