

4-21-2017

Improving Discovery and Patron Experience Through Data Mining

Boyuan Guan

Florida International University, bguan@fiu.edu

Jamie Rogers

Florida International University, rogersj@fiu.edu

Follow this and additional works at: <https://digitalcommons.fiu.edu/glworks>



Part of the [Databases and Information Systems Commons](#), [Library and Information Science Commons](#), [Systems Architecture Commons](#), and the [Theory and Algorithms Commons](#)

Recommended Citation

Guan, Boyuan and Rogers, Jamie, "Improving Discovery and Patron Experience Through Data Mining" (2017). *Works of the FIU Libraries*. 64.

<https://digitalcommons.fiu.edu/glworks/64>

This work is brought to you for free and open access by the FIU Libraries at FIU Digital Commons. It has been accepted for inclusion in Works of the FIU Libraries by an authorized administrator of FIU Digital Commons. For more information, please contact dcc@fiu.edu.



IMPROVING DISCOVERY & PATRON EXPERIENCE THROUGH DATA MINING



Florida International University
Boyuan (Keven) Guan – Lead Developer, GIS Center
Jamie Rogers – Assistant Director, Digital Collections Center

OUTLINE

- 1. Introduction**

- 2. Problem statement**

- 3. Overview of the recommendation system concept**

- 4. Technical design and implementation**

- 5. Sharing the technology**

- 6. Questions?**



INTRODUCTION



1971 University Park Groundbreaking Ceremony
Florida International University Archives



dPanther Repository

The Florida International University dPanther 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.

Learn more →



Advanced Search ▾ Press "enter" from keyboard or type "*" and click search button to retrieve all items

COLLECTIONS

Partners/Contributors Collections



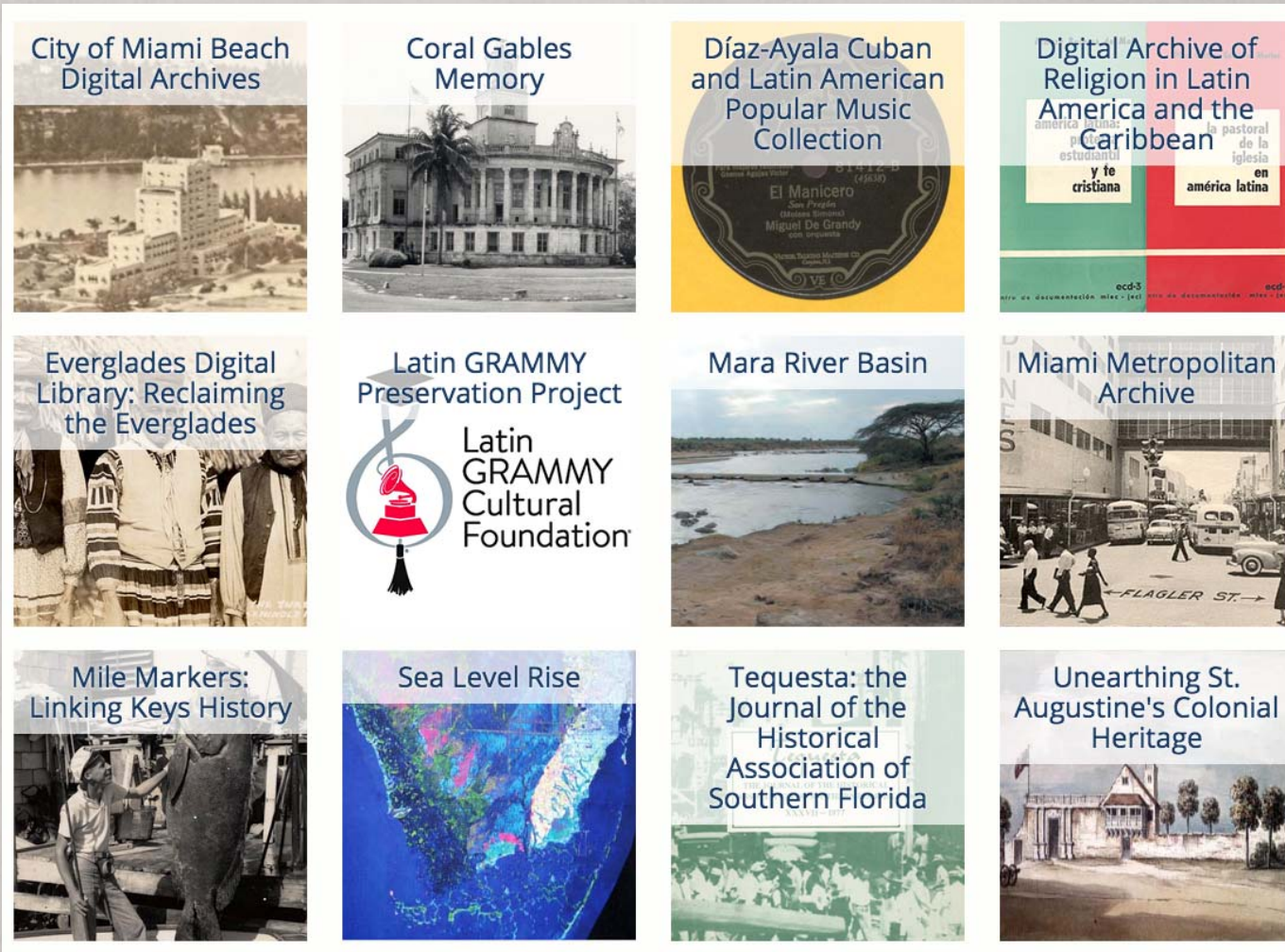
City of Miami Beach

Disaster Risk Reduction

FIU Government Documents

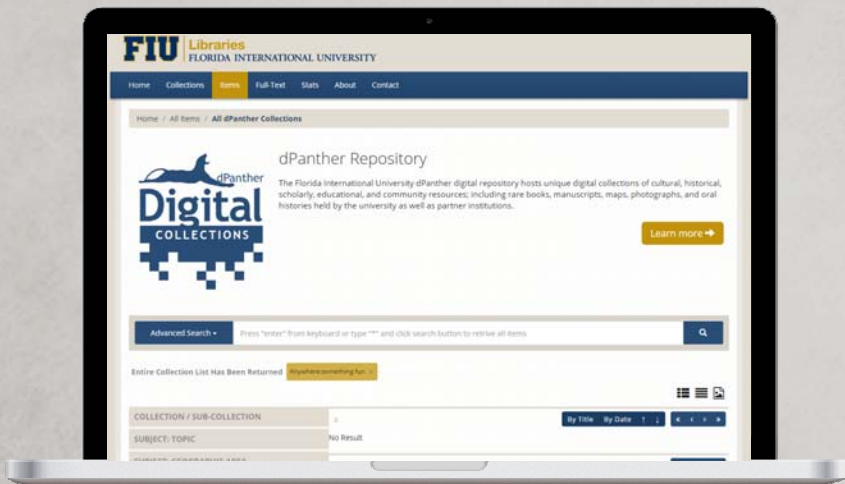
FIU Sound and Image







PROBLEM STATEMENT



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



Editor suggested item

repository hosts unique digital collections of cultural, historical, scholarly, educational, and community resources; including rare books, manuscripts, maps, photographs, and well as partner

Item search

Learn more →



Advanced Search ▾

Press "enter" from keyboard or type "*" and click search button to retrieve all items



Collection Selection COLLECTIONS

Partners/Contributors Collections

City of Miami Beach



Disaster Risk Reduction



FIU Government Documents



FIU Sound and Image





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.

Advanced Search -

Press "enter" from keyboard or type "*" and click search button to retrieve all items



Refine By Collection: sai



COLLECTION / SUB-COLLECTION

1-20 of 38426

By Title By Date ↑ ↓ < >

Díaz-Ayala Cuban and Latin American Popular Music Collection **37423**

• Latin GRAMMY Preservation Project **717**

FIU Sound and Image **38422**

[Show More...](#)

SUBJECT: TOPIC

Title: "Adieu donc belle France". Romance de l'Opéra Maria Stuart.

Volume: FIS1001653 / 00001

Author: Niedermeyer, Louis.

PublishYear:

Type: Notated Music



Advanced Search - Press "enter" from keyboard or type "*" and click search button to retrieve all items

Refine By Collection: sai

COLLECTION / SUB-COLLECTION

- Diaz-Ayala Cuban and Latin American Popular Music Collection **37423**
- Latin GRAMMY Preservation Project **717**
- FIU Sound and Image **38422**
- Brazilian Scores **90**
- Flute Scores **425**
- Piano Scores **78**
- Sergio Roberto de Oliveira **121**
- Vocal Music Scores **142**
- Latin GRAMMY Preservation Project **717**

[Show Less...](#)

SUBJECT: TOPIC

- Latin pop (music) **9150**
- bolero **6038**
- Popular music **4752**
- guaracha **1802**
- tango **1431**
- Digital media **867**
- canción **781**
- rumba **592**
- vals **575**
- bolero-son **485**
- son **473**
- mambo **436**
- pasodoble **327**
- danzón **325**
- maracaibo **300**

1-20 of 38426

By Title By Date ↑ ↓

Title: "Adieu donc belle France". Romance de l'Opéra Maria Stuart.

Volume: FISI001653 / 00001

Author: Niedermeyer, Louis.

PublishYe..

Type: Notated Music

Title: "Alpine Violets." (Alpenveilchen.) Idyl in Ländler Style. Op. 100.

Volume: FISI000364 / 00001

Author: André, Ludwig

PublishYe.. 1915

Type: Notated Music

Title: "Ecco ridente il Cielo" Cavatina, from Rossini's Opera "Il Barbiere di Siviglia", Arranged for the Flute, with an Accompaniment for the Piano Forte, by Tulou.

Volume: FISI000652 / 00001

Author: Tulou, J. L.

PublishYe..

Type: Notated Music

Title: "El Trovador Tropical"

Volume: FIDA026699 / 00001

Author: Gonzalez, Berto (performer)
Farres, Osvaldo (composer)
Hernandez, Rafael (composer)
D.R.A. (composer)
Delfin, Eusebio (composer)
Greever, Maria (composer)
Flores, Pedro (composer)

NO IMAGE AVAILABLE

FIU Libraries
FLORIDA INTERNATIONAL UNIVERSITY

Home Collections **Items** Full-Text Stats About Contact

Home / collections / sai

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.

Advanced Search - Press "enter" from keyboard or type "*" and click search button to retrieve all items

Collections:
FIU Sound and Image

Search for:

flute scores in Anywhere

and in Title

and in Creator

and in Subject Keyword

Search

By Title By Date ↑ ↓

Title: "Adieu donc belle France". Romance de l'Opéra Maria Stuart.

Volume: FISI001653 / 00001

Author: Niedermeyer, Louis.

PublishYe..

Type: Notated Music

Title: "Alpine Violets." (Alpenveilchen.) Idyl in Ländler Style. Op. 100.

Volume: FISI000364 / 00001

Author: André, Ludwig

PublishYear: 1915

Type: Notated Music

Title: "Ecco ridente il Cielo" Cavatina, from Rossini's Opera "Il Barbiere di Siviglia", Arranged for the Flute, with an Accompaniment for the Piano Forte, by Tulou.

Volume: FISI000652 / 00001

Author: Tulou, J. L.

PublishYe..

Type: Notated Music

Title: "El Trovador Tropical"

Volume: FIDA026699 / 00001

Author: Gonzalez, Berto (performer)
Farres, Osvaldo (composer)
Hernandez, Rafael (composer)
D.R.A. (composer)
Delfin, Eusebio (composer)
Greever, Maria (composer)
Flores, Pedro (composer)

SUBJECT: GEOGRAPHIC AREA

- Brazil **26**
- South America **26**
- Argentina **1**





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.

Advanced Search -

Press "enter" from keyboard or type "*" and click search button to retrieve all items



Did you mean: [Score reading and playing](#), [Scores](#), [Piano scores](#)

Refine By: **Aggregation:fls** x



COLLECTION / SUB-COLLECTION

1-20 of 425

By Title By Date ↑ ↓ < >

FIU Sound and Image **425**

- Brazilian Scores **12**
- Flute Scores **425**

[Show More...](#)

SUBJECT: TOPIC

- 1892 **1**
- Flute Music **1**

SUBJECT: GEOGRAPHIC AREA

AUTHOR / CREATOR

Tulou, J. L. **29**

Title: "Alpine Violets." (Alpenveilchen.) Idyl in Ländler Style. Op. 100.

Volume: FIS1000364 / 00001

Author: André, Ludwig

PublishYear: 1915

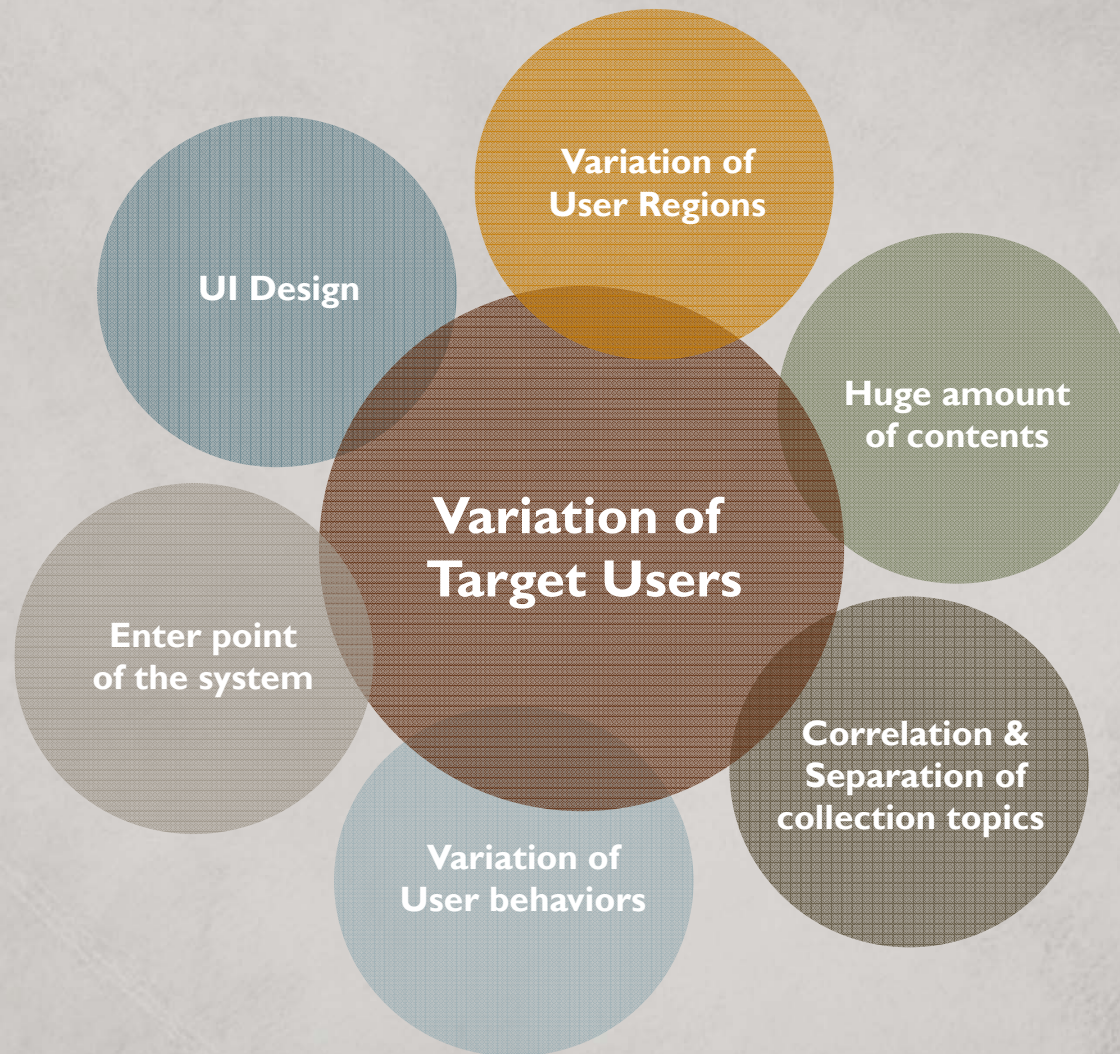
Type: Notated Music



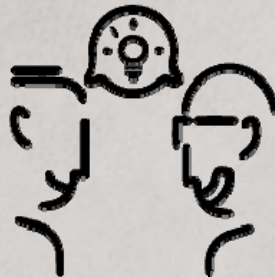
Save

Title: "Ecco ridente il Cielo" Cavatina, from Rossini's Opera "Il Barbiere di Siviglia", Arranged for the Flute, with an Accompaniment for the Piano Forte, by Tulou.





What do the users **WANT** to know?
What do the users **NEED** to know?



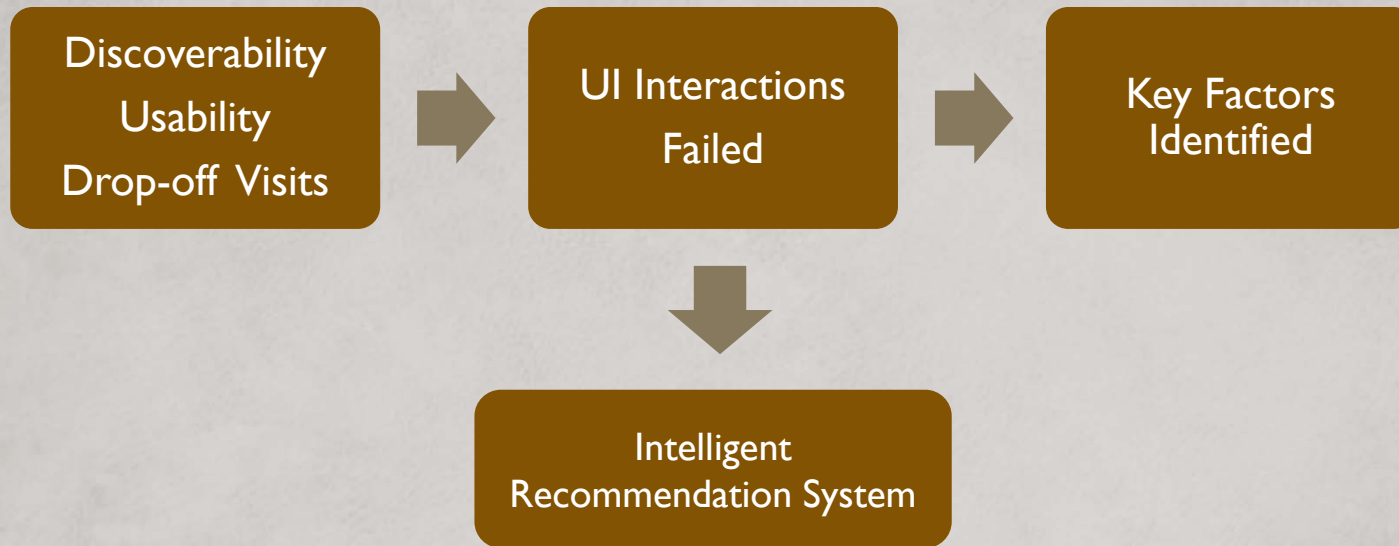
Get to know the users &
their region information

Awareness about the
content

Awareness about the
system condition (usages,
statistic, user interactions)

Data Mining !!

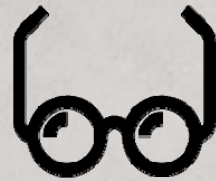
Recommendation System



**Cannot acquire
additional user
information**

**Minimize the
Side Effect**

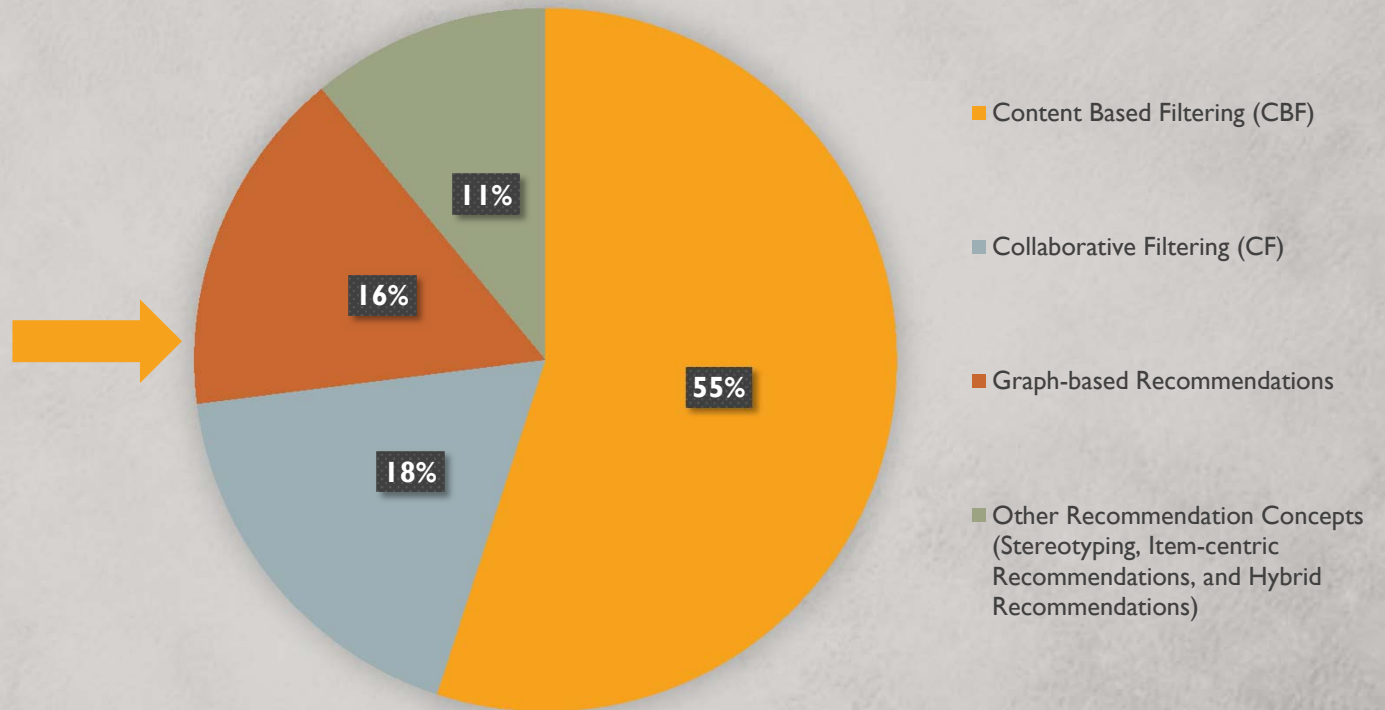
**Low impact
on the
performance**



OVERVIEW OF THE RECOMMENDATION SYSTEM CONCEPT

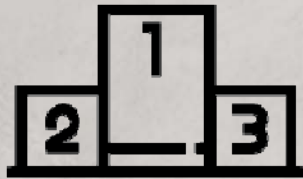
RECOMMENDERS

- **200** research articles were published
- **62** proposed methods



ADDITIONAL METHODS

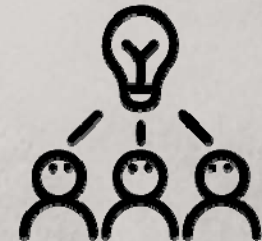
GLOBAL RELEVANCE



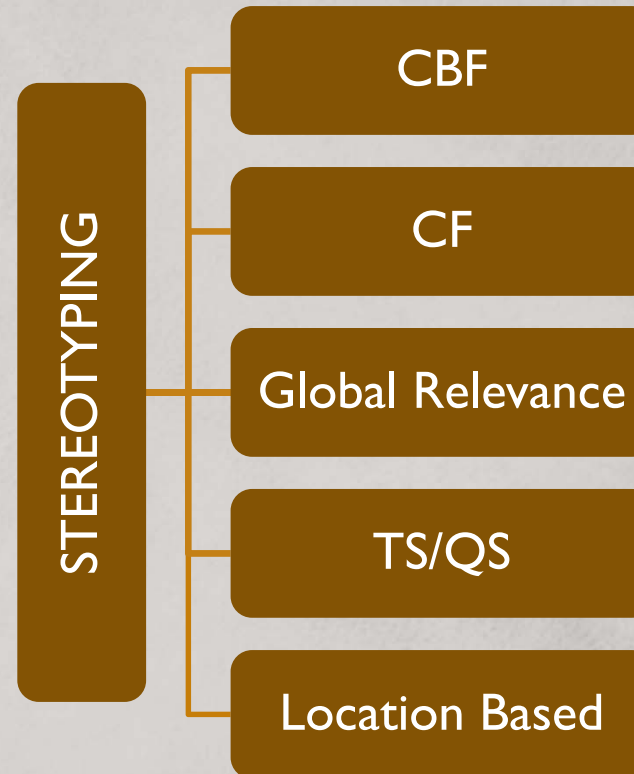
LOCATION BASED



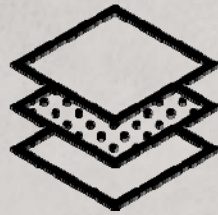
QUERY/TERM SUGGESTION
(QS / TS)



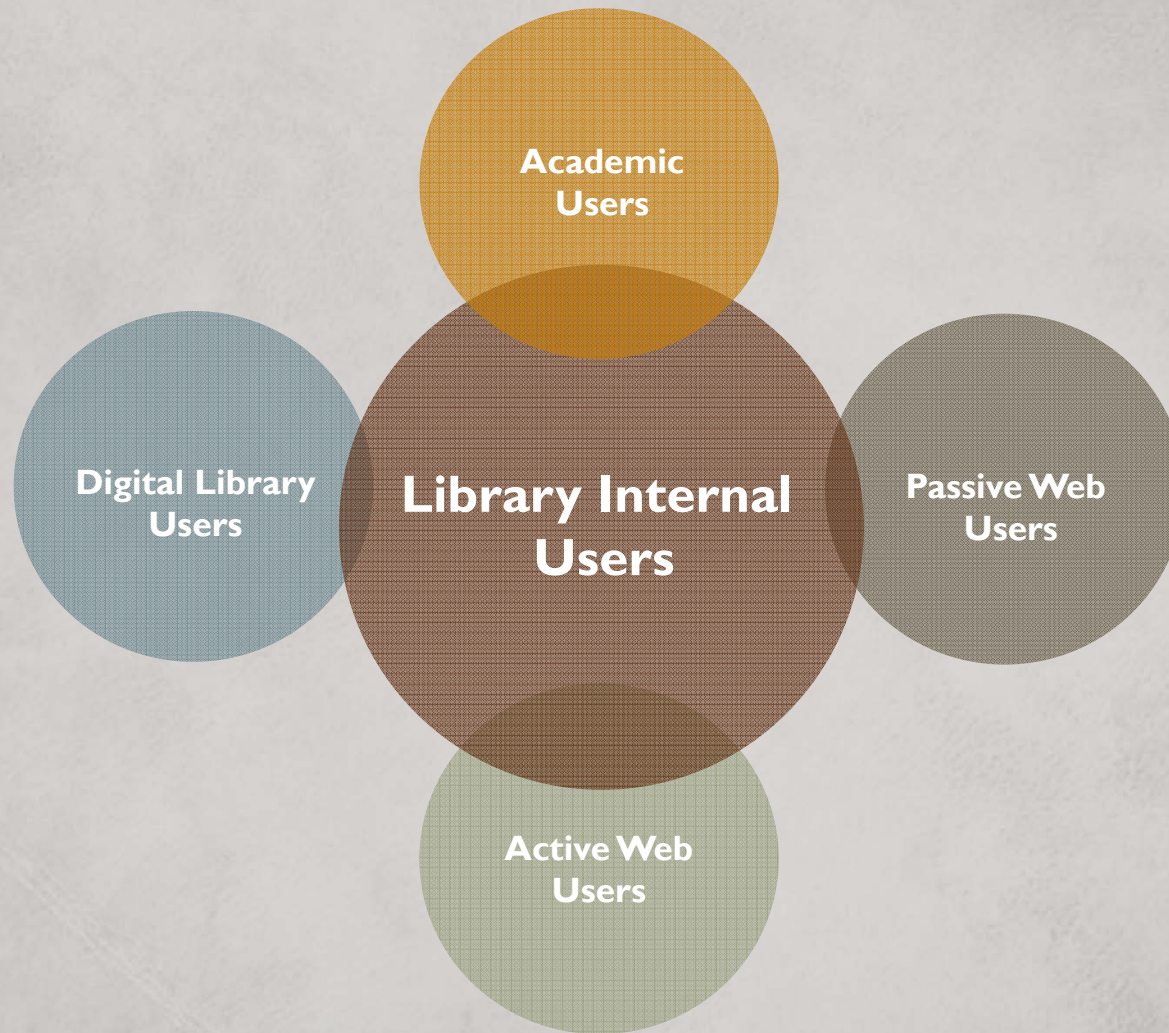
DPANTHER ARCHITECTURE




























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



TECHNICAL DESIGN AND IMPLEMENTATION



RECOMMENDATION STRATEGY

| | CBF | CF | Global Relevance | QS / TS | Location Based |
|-----------------------|---|---|---|---|---|
| Internal Users |  |  |  |  |  |
| Academic Users |  |  |  |  |  |
| Digital Library Users |  |  |  |  |  |
| Active Users |  |  |  |  |  |
| Passive Users |  |  |  |  |  |

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

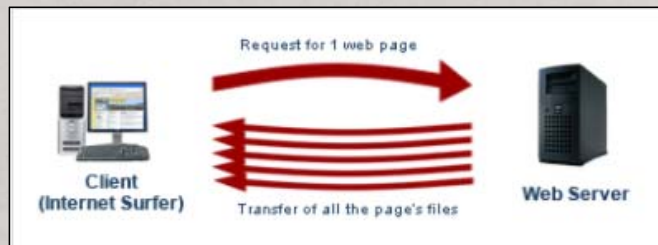
BUILD USERS VECTOR

Stereotyping → Classification → Building the vector for users



Web Server Log

Database Log



Users →

| Collections | Items | Search Terms | ID Involved? | Geo-Location | User Type |
|-------------|----------|--------------|----------------|--------------|-----------|
| {number} | {number} | {number} | {True / False} | {Nominal} | {Nominal} |

STEREOTYPING DESIGN – SAMPLE DECISION TREE

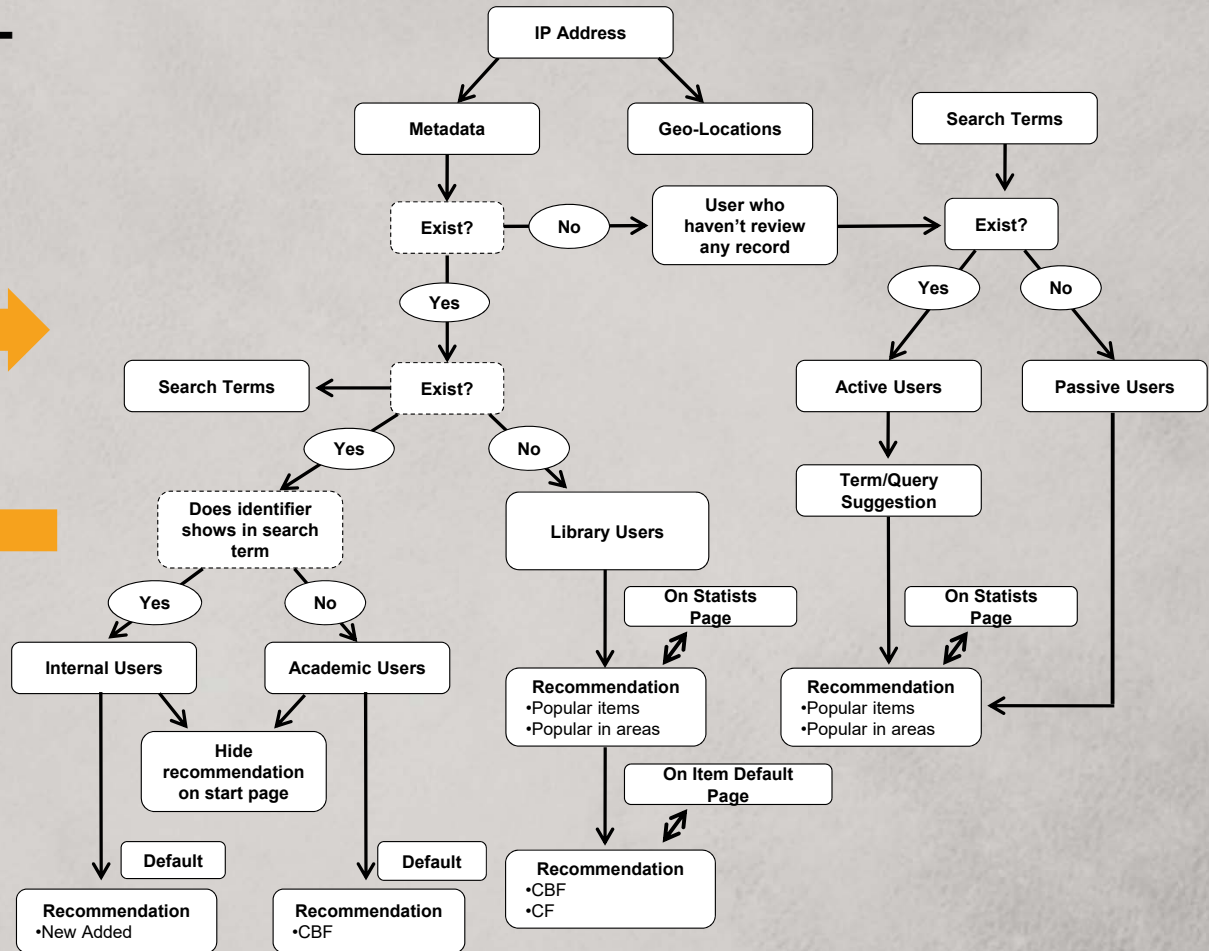
Use 3 out of 5 dimensions



Need to use a data mining classifier



TOO MANY VARIATIONS



USER DATA CLASSIFICATION

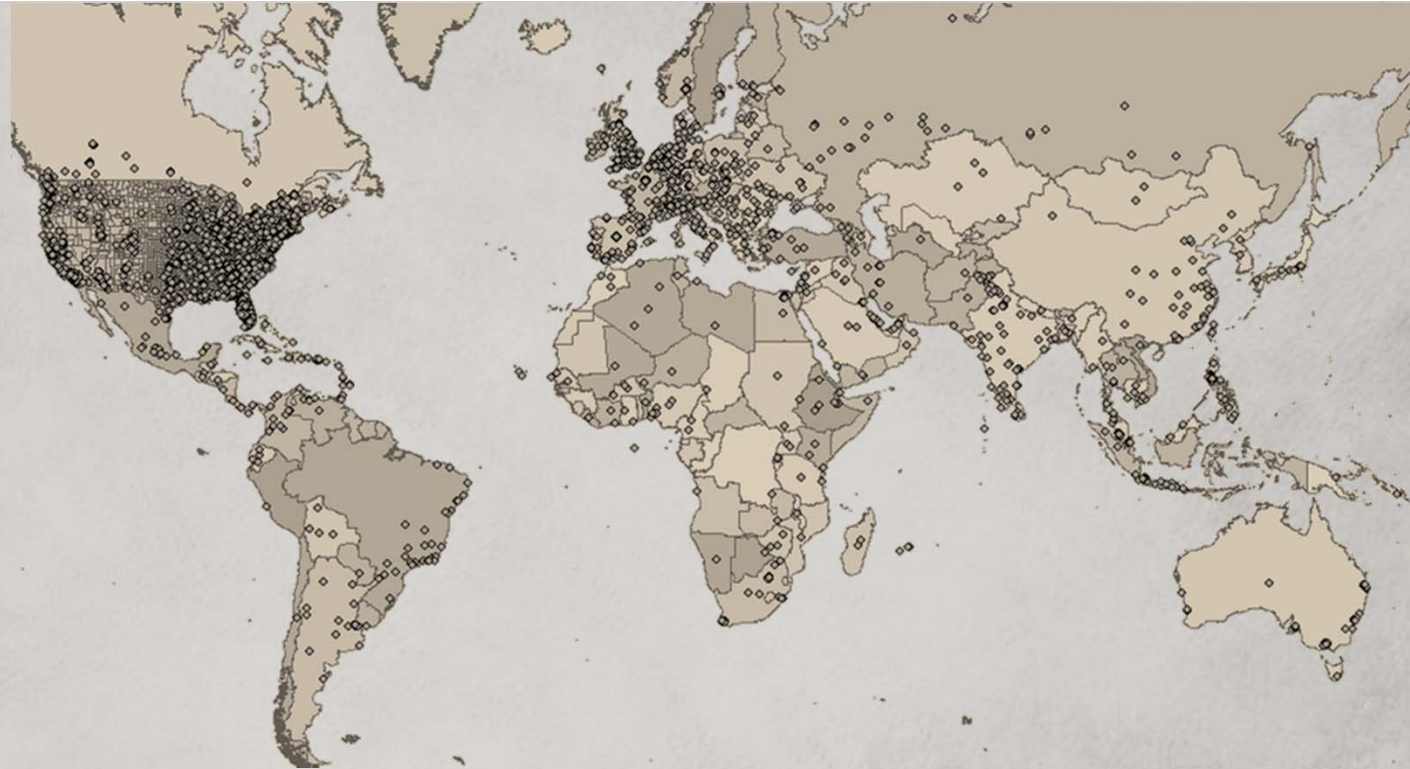
- 15,126 Distinct Users
- 5,583 Distinct Search Terms
- 12,376 Items Visited
- 92 Collections
- 169 Countries from the World
- 335 Counties in the US



Label data (training data)

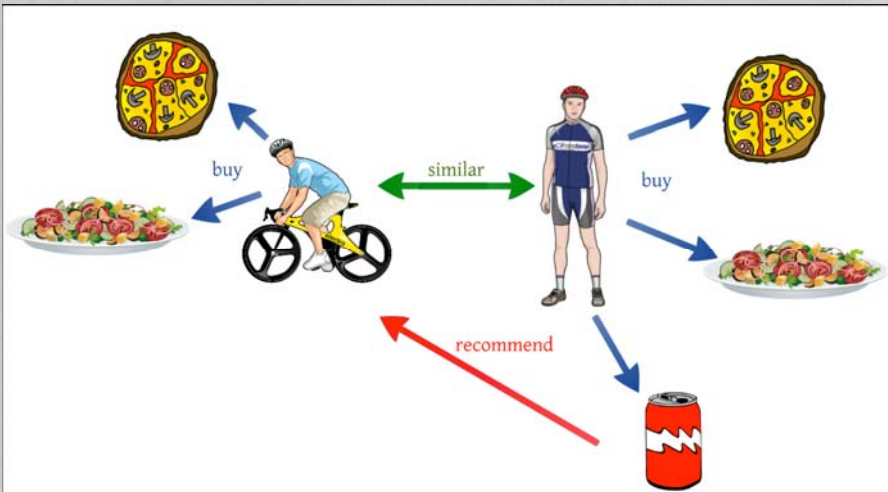


Run Tree Based Classifier
(e.g. J48)



RECOMMENDATION FILTERING

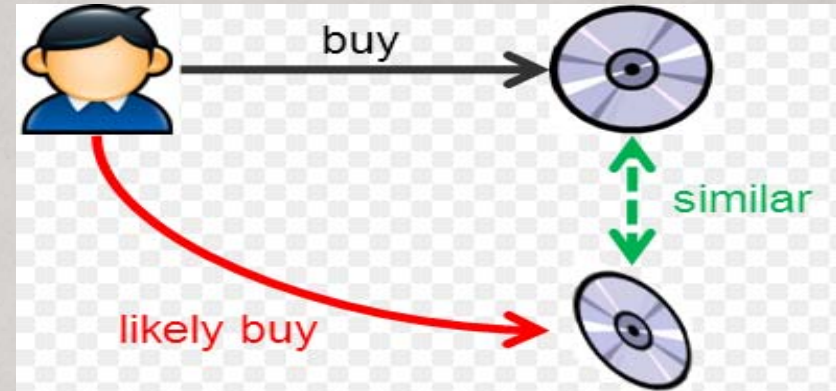
Collaborative filtering (want to know)



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

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)

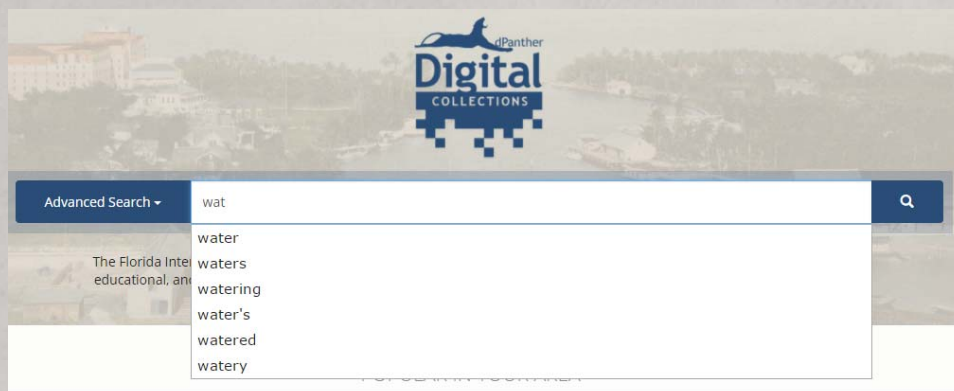


[subject_keywords, publication date, aggregation code, holding locations, title, type, publisher]

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

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)



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


























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

Global Relevance / Location Based Recommendation (should know)



$$GR\ Score \rightarrow A_c \propto (\text{numebr of hits of the item}) \rightarrow A_c|$$

Where A_c is the location information for collection any collection C in dPanther.

| | CBF | CF | Global Relevance | QS / TS | Location Based |
|--------------------------|---|---|---|---|---|
| Internal Users |  |  |  |  |  |
| Academic Users |  |  |  |  |  |
| Digital Library Users |  |  |  |  |  |
| Active Users |  |  |  |  |  |
| Passive Users |  |  |  |  |  |

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

Internal Users

The screenshot displays the Digital Collections website interface. At the top, the logo for Digital Collections is visible. Below it, an "Advanced Search" box contains the text "wat" and a search button. A dropdown menu lists suggestions: "water", "waters", "watering", "water's", "watered", and "watery". Below the search box is a map of Florida with a pop-up window listing items: "The Beacon, Vol. 25, Issue 83", "Miami 1920's photo album", and "[Collection of negatives of official events and sites in Miami Beach, 1986]". At the bottom, a "FEATURED ITEMS" section shows two items: "Tequesta: The Journal of the Historical Association of Southern Florida. Volume 1, number 63." and "The Beacon, Vol. 25, Issue 83".

QS / TS

Location Based

Global Relevance

Internal Users

Home / items / itemdetail

FILE VIEWER

The Beacon, Vol. 25, Issue 83
• PDF1

The screenshot shows a digital newspaper page titled 'THE BEACON' with the subtitle 'A Forum for Free Student Expression of Florida International University'. The main headline is 'FIU degree ranked second biggest waste of money'. The sidebar on the right contains a 'File Viewer' menu with options: 'Record Information', 'Maps / Geo-Reference', 'Item Usage', 'RELATED ITEMS +', 'OTHERS LIKED +', 'NEWLY ADDED ITEMS -', and 'PARTNERS'. The 'NEWLY ADDED ITEMS -' section is highlighted with a red box and contains a list of items related to convocation events. Below the sidebar are social media icons for Facebook, LinkedIn, Twitter, and Google+, along with 'Print' and 'Top' buttons.

File Viewer


























- Record Information
- Maps / Geo-Reference
- Item Usage
- RELATED ITEMS +
- OTHERS LIKED +
- NEWLY ADDED ITEMS -
 - 1983 Convocation Keynote Address Robert McAfee Brown
 - 1982 Convocation President Gregory Wolfe Remarks
 - Florida International University Convocation Program 1983
 - Florida International University Convocation Invitation 1982
 - Florida International University Library Clipping Index
- PARTNERS

f in t g+

Print Top

CBF & CF are hidden

Global Relevance

| | CBF | CF | Global Relevance | QS / TS | Location Based |
|--------------------------|---|---|---|---|---|
| Internal Users |  |  |  |  |  |
| Academic Users |  |  |  |  |  |
| Digital Library Users |  |  |  |  |  |
| Active Users |  |  |  |  |  |
| Passive Users |  |  |  |  |  |

Academic Users

The screenshot shows the top section of the dPanther Digital Collections website. At the top center is the logo for "dPanther Digital COLLECTIONS". Below the logo is a search bar with a dropdown menu set to "Advanced Search" and the search term "WATER". A red box highlights the search bar. Below the search bar is a paragraph of text: "The Florida International University dPanther 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. [Learn more...](#)". Below this text are two more red boxes: one containing "POPULAR IN YOUR AREA +" and another containing "FEATURED ITEMS +". Below these is a section titled "COLLECTIONS" with two tabs: "Partners/Contributors" and "Collections". Under the "Collections" tab, there are four collection cards: "City of Miami Beach" (with a photo of a building), "Disaster Risk Reduction" (with a photo of a damaged boat), "FIU Government Documents" (with a photo of a document), and "FIU Sound and Image" (with a photo of musical notation).

QS / TS are Hidden

Location Based &
Global Relevance
are Hidden

Academic Users

Home / items / Itemdetail

FILE VIEWER

Tequesta: The Journal of the Historical Association of Southern Florida. Volume 1, number 63.

- Tequesta: The Journal of the Historical Association of Southern Florida. Volume 1, number 63



File Viewer
Record Information
Maps / Geo-Reference
Item Usage

RELATED ITEMS -

- Tequesta: The Journal of the Historical Association of Southern Florida. Volume 1, number 48.
- Tequesta: The Journal of the Historical Association of Southern Florida. Volume 1, number 39.
- Tequesta: The Journal of the Historical Association of Southern Florida. Volume 1,

OTHERS LIKED -

- Sinfonia Festiva (Para comemorar os 500 anos do Descobrimento do Brasil e o Aniversário de São Petesburgo)
- SW 7th and SW 8th Corridor, Little Havana ; Economic Impact Analysis Final Report
- Photographs depicting Everglades scenes in Miami-Dade County, 1920.


























NEWLY ADDED ITEMS +

RECORD INFORMATION

CBF

CF

Global Relevance

| | CBF | CF | Global Relevance | QS / TS | Location Based |
|--------------------------|---|---|---|---|---|
| Internal Users |  |  |  |  |  |
| Academic Users |  |  |  |  |  |
| Digital Library Users |  |  |  |  |  |
| Active Users |  |  |  |  |  |
| Passive Users |  |  |  |  |  |

Digital Library Users

The screenshot shows the 'Digital Collections' interface. At the top, there is a search bar with the text 'wat' and a dropdown menu listing suggestions: 'water', 'waters', 'watering', 'water's', 'watered', and 'watery'. Below the search bar is a map of Florida with a popup window listing items: 'The Beacon, Vol. 25, Issue 83', 'Miami 1920's photo album', and '[Collection of negatives of official events and sites in Miami Beach, 1986]'. At the bottom, there is a 'FEATURED ITEMS' section with three items: 'Tequesta: The Journal of the Historical Association of Southern Florida. Volume 1, number 63.', 'The Beacon, Vol. 25, Issue 83', and another instance of 'The Beacon, Vol. 25, Issue 83' with a thumbnail and metadata.

QS / TS

Location Based

Global Relevance

Digital Library Users


Home / items / itemdetail

FILE VIEWER

Tequesta: The Journal of the Historical Association of Southern Florida. Volume 1, number 63.

- Tequesta: The Journal of the Historical Association of Southern Florida. Volume 1, number 63.


FI08060963.pdf 1 / 97



Tequesta
THE JOURNAL OF THE
HISTORICAL ASSOCIATION OF SOUTHERN FLORIDA
2003
LVIII

RECORD INFORMATION

Permanent Link: <http://dpanther.fiu.edu/dpService/dpPurlService/purl/FI08060963/00001>



File Viewer

- Record Information
- Maps / Geo-Reference
- Item Usage

RELATED ITEMS -

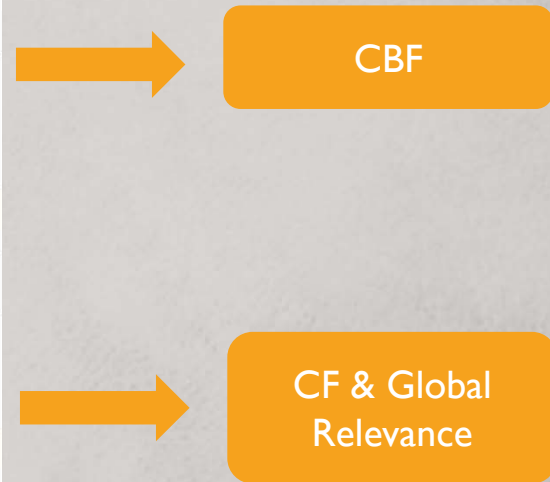
- Tequesta: The Journal of the Historical Association of Southern Florida. Volume 1, number 48.
- Tequesta: The Journal of the Historical Association of Southern Florida. Volume 1, number 39.
- Tequesta: The Journal of the Historical Association of Southern Florida. Volume 1, number 30.

OTHERS LIKED -

- Sinfonia Festiva (Para comemorar os 500 anos do Descobrimto do Brasil e o Aniversário de São Petesburgo)
- SW 7th and SW 8th Corridor, Little Havana : Economic Impact Analysis Final Report
- Photographs depicting Everglades scenes in Miami-Dade County, 1920.
- Taxidermy Trail

NEWLY ADDED ITEMS -

- Tequesta: The Journal of the Historical Association of Southern Florida. Volume 1, number 13.
- Tequesta: The Journal of the Historical Association of Southern Florida. Volume 1, number 12.
- Tequesta: The Journal of the Historical Association of Southern Florida. Volume 1, number 11.

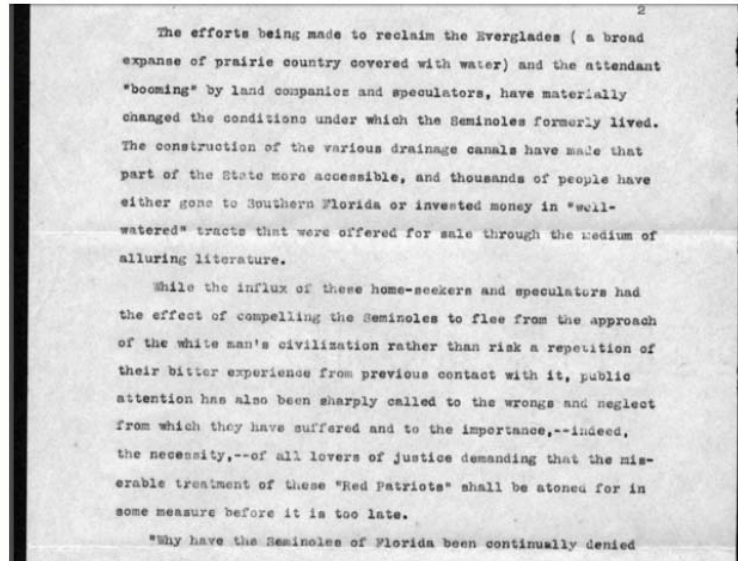


Sample Use Case

FILE VIEWER

A bill is to be introduced in the Florida legislature

- Essay about the history of Seminoles in Florida



RECORD INFORMATION

Permanent Link: <http://dpanther.fiu.edu/dpService/dpPurlService/purl/MW00660001/00001>

File Viewer

Record Information

Maps / Geo-Reference

Item Usage

RELATED ITEMS -

- The Florida Seminoles and their rights in the Everglades
- The Florida Seminoles and their rights in the Everglades
- The cry of the Seminole
- The cry of the Seminole

OTHERS LIKED +

NEWLY ADDED ITEMS -

- 2012 Annual Report of the Water Quality Monitoring Project for the Water Quality Protection Program of the Florida Keys National Marine Sanctuary
- 2013 Annual Report of the Water

PARTNERS

**HISTORY
MIAMI**



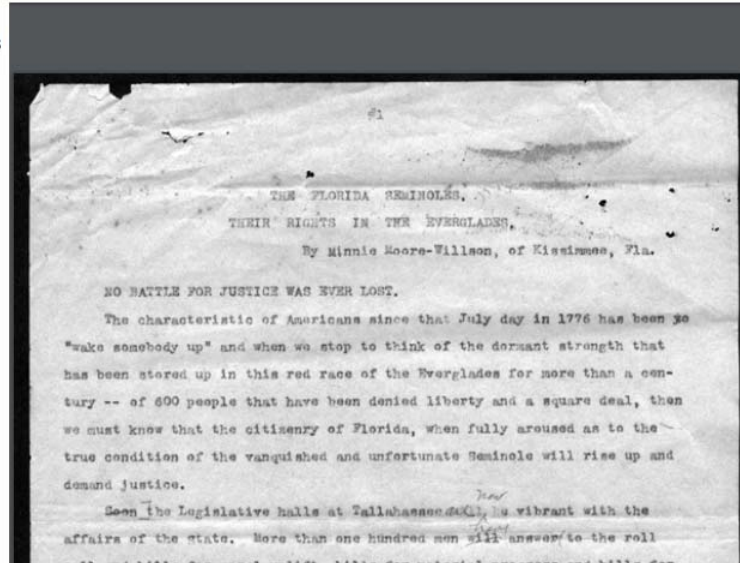
Print

↑ Top

FILE VIEWER

The Florida Seminoles and their rights in the Everglades

- The Florida Seminoles and their rights in the Everglades.



File Viewer

Record Information

Maps / Geo-Reference

Item Usage

RELATED ITEMS -

- The Seminoles of Florida and their rights in the Everglades
- The Seminoles of Florida and their rights in the Everglades
- A bill is to be introduced in the Florida legislature

OTHERS LIKED +

NEWLY ADDED ITEMS -

- 2012 Annual Report of the Water Quality Monitoring Project for the Water Quality Protection Program of the Florida Keys National Marine Sanctuary
- 2013 Annual Report of the Water

PARTNERS

**HISTORY
MIAMI**



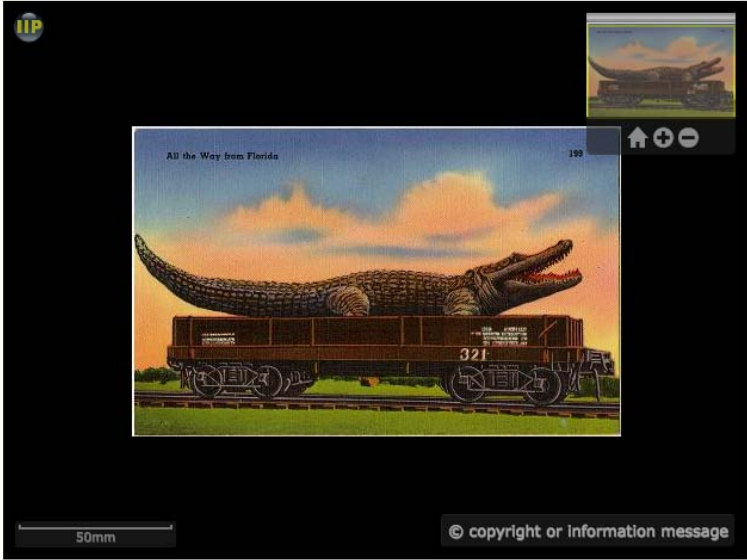
RECORD INFORMATION

Permanent Link: <http://dpanther.fiu.edu/dpService/dpPurlService/purl/MW00050011/00001>

dPanther Home Collections **Items** Full-Text Stats About Contact

All the way from Florida

- Front
- Back



50mm © copyright or information message

RECORD INFORMATION

Permanent Link: <http://dpanther.fiu.edu/dpService/dpPurlService/purl/FI07050838/00001>

File Viewer

Record Information
Maps / Geo-Reference
Item Usage

RELATED ITEMS -

- The Florida Gator
- The Florida Gator
- Having a snapping good time in Florida
- Having a snapping good time in Florida

OTHERS LIKED +

NEWLY ADDED ITEMS -

- 2012 Annual Report of the Water Quality Monitoring Project for the Water Quality Protection Program of the Florida Keys National Marine Sanctuary
- 2013 Annual Report of the Water

PARTNERS

f in t g+

Print ↑ Top

Having a snapping good time in Florida

- Front
- Back



File Viewer

Record Information

Maps / Geo-Reference

Item Usage

RELATED ITEMS -

- White Tiger and his pet
- All the way from Florida
- All the way from Florida
- Big George, largest alligator in captivity 14 feet 7 inches-Greets you in Florida

OTHERS LIKED +

NEWLY ADDED ITEMS -

- 2012 Annual Report of the Water Quality Monitoring Project for the Water Quality Protection Program of the Florida Keys National Marine Sanctuary
- 2013 Annual Report of the Water

PARTNERS

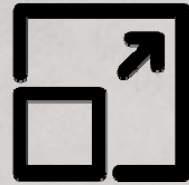


Print

Top

RECORD INFORMATION

Permanent Link: <http://dpanther.fiu.edu/dpService/dpPurlService/purl/FI07050837/00001>



POSSIBLE USE BY OTHER INSTITUTIONS



QUESTIONS?