Faceted Browsing for Combined Access to a Digital Repository and a Library Catalog

Bess Sadler
Leslie Johnston
University of Virginia Library

DLF Fall 2007 Forum
What is Project Blacklight?

- Blacklight is a research project. Bess Sadler and Erik Hatcher started an experiment to index MARC data in Lucene/Solr in January 2007.
- The first prototype included ~3.7M MARC records, 320 Tang Dynasty Chinese poems in TEI, and 470 Digital Collections Repository objects.
- MARC indexing consulting came from Erin Stalberg (formerly UVA but now at NC State) and Ed Summers at Library of Congress.
What Blacklight Can Solve For Us

- Faceted browsing where we have none.
- Relevancy ranking where we have none.
- Facilitates the use of search and faceted browse together to make it easier to perform complex discovery operations without knowing the logic behind it.
- In addition to providing access to the entire catalog, we can create multiple additional interfaces to accommodate specialized browsing of different types of collections, such as music collections (audio and scores), that takes advantage of specialized use of MARC.
- We can mix in data that’s not explicit in the MARC record. In our music interface, we solved a user frustration where they couldn’t find music by century – we’re extrapolating at the time of indexing from the year that is in the metadata.
- We can federate indexes of MARC records with indexes of metadata for digital objects in our Repository for a single discovery method.
- We can potentially add additional applications into the mix, such as one where we will track our holdings that are included in Google Book Search that will supply link URLs for the interface.
Blacklight Technical Details

• The MARC records are indexed through a Ruby script, directly from MARC binary without transforming into MARCXML first.
• The implementation was accomplished using Ruby on Rails.
• The box it’s all running on is a development server, with four medium fast CPUs and 3.5Gb of RAM.
• Solr Flare is running through Jetty, and Blacklight is running on a pack of four mongrel instances, with apache doing load balancing out front.
• Sessions get written to a MySQL database, and we use Capistrano 2.0 for deployment and versioning.
• It sends us all an email every time an error gets generated.
Issues in Project Blacklight

• Our MARC data isn’t utf-8-compliant and that caused issues with the diacritics.
• Indexing in new ways always exposes new data inconsistencies.
• We have not yet identified a production workflow for keeping the catalog updated daily.
Blacklight Portal Entry

This is BLACKLIGHT, a research project of the University of Virginia Library.

Describe Blacklight...

- Entire Catalog
- Music Library
Combined Catalog with MARC Records and Repository Objects

This is BLACKLIGHT, a research project of the University of Virginia Library.

SOURCE
- Digital Library (10143)
- Library Catalog (153692)
- Music (13030)
- Other (111300)

FORMAT
- Archives (1683)
- Audio CD (1091)
- Book (215690)
- CD-ROM (307)
- Cassette (304)
- DVD (3)
- Digital Media (8)

Collection: UVA Library Text
Subject genre: Jefferson, Thomas, 1743-1826; University of Virginia; History
Composition era: 20th Century
Content type: TEL
Source: Digital Library
Language: English
Year: 1920
Date: [c1920-1922]
Creator: Bruce, Philip Alexander; Macmillan
Title: History of the University of Virginia, 1819-1919; the lengthened shadow of one man.
**Music Facets for Browsing**

<table>
<thead>
<tr>
<th>Title</th>
<th>Author(s)</th>
<th>Library</th>
<th>Language</th>
</tr>
</thead>
<tbody>
<tr>
<td>On trial, the soviet state versus &quot;abram tertz&quot; and &quot;nikolai arzhak.&quot;</td>
<td>Terletsky, Akram</td>
<td>Alderman, LAW</td>
<td>English</td>
</tr>
<tr>
<td>Bibliography of Ghana, 1957-1960</td>
<td></td>
<td>Alderman</td>
<td>English</td>
</tr>
<tr>
<td>The composition and origin of the terrestrial fauna of middle asia</td>
<td>Kryzhanovskii, D. O. L.</td>
<td>Brown SEL</td>
<td>English</td>
</tr>
<tr>
<td>A general catalogue of books, arranged in classes, offered for sale</td>
<td>Qwantch, Bernard</td>
<td>Alderman</td>
<td>English</td>
</tr>
<tr>
<td>The new universal gazetteer and geographical dictionary</td>
<td>Thomson, John</td>
<td></td>
<td>English</td>
</tr>
</tbody>
</table>
## More Music Facets for Browsing

<table>
<thead>
<tr>
<th>Library</th>
<th>Ivy Annex</th>
</tr>
</thead>
<tbody>
<tr>
<td>Topic form genre</td>
<td>Libraries</td>
</tr>
<tr>
<td>Subject geographic</td>
<td>California, Southern</td>
</tr>
<tr>
<td>Language</td>
<td>English</td>
</tr>
</tbody>
</table>

**Memorials of a quiet life**
- **Author:** Hare, Augustus J. C
- **Library:** Alderman
- **Language:** English

**Treaties and alliances of the world**
- **Library:** Alderman, LAW
- **Topic form genre:** Alliances, Treaties
- **Language:** English

### Recording Format
- CD (2837)
- Cassette (466)
- DVD (4)
- LP (3728)
- Open Reel Tape (452)
- VHS (2476)
- Video Disc (106)

### Language
- (19773)
- 1 (1)
- 536 (1)
- ??? (822)
- Afrikaans (119)
- African (Other) (3)
- Akladian (3)
- Albanian (13)
- Albanian (Other) (1)
- Amharic (1)
- Arabic (2576)
- Aramaic (4)
- Armenian (8)
- Assamese (12)
- Austro-Bavarian (Other) (6)
- Avestan (2)
- Awadhi (4)
- Aymara (1)
- Azerbaijani (1)
- Balinese (1)

### Instrument
- Alto voice (44)
- Bard (27)
- Baritone horn (3)
- Baritone voice (46)
- Bass clarinet (16)
- Bass voice (45)
- Bassoon (119)
- Bowed Strings, Other (2)
- Bowed Strings, Unknown (1)
- Bowed Strings, Unspecified (8)
And Even More

<table>
<thead>
<tr>
<th>TOPIC FORM GENRE</th>
</tr>
</thead>
<tbody>
<tr>
<td>History (28137)</td>
</tr>
<tr>
<td>Electronic journals (26681)</td>
</tr>
<tr>
<td>History and criticism (13926)</td>
</tr>
<tr>
<td>Law and legislation (7092)</td>
</tr>
<tr>
<td>Education (4541)</td>
</tr>
<tr>
<td>Research (3797)</td>
</tr>
<tr>
<td>Finance (3591)</td>
</tr>
<tr>
<td>Philosophy (3556)</td>
</tr>
<tr>
<td>Study and teaching (3254)</td>
</tr>
<tr>
<td>Architecture (3171)</td>
</tr>
<tr>
<td>Women (3041)</td>
</tr>
<tr>
<td>World War, 1939-1945 (2853)</td>
</tr>
<tr>
<td>Music (2780)</td>
</tr>
<tr>
<td>Social aspects (2661)</td>
</tr>
<tr>
<td>Environmental aspects (2455)</td>
</tr>
<tr>
<td>Economic aspects (2335)</td>
</tr>
<tr>
<td>Insurance, Flood (2327)</td>
</tr>
<tr>
<td>Management (2062)</td>
</tr>
<tr>
<td>Government policy (2196)</td>
</tr>
<tr>
<td>English language (1957)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>SUBJECT GEOGRAPHIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>United States (70780)</td>
</tr>
<tr>
<td>History (20614)</td>
</tr>
<tr>
<td>Politics and government (16587)</td>
</tr>
<tr>
<td>Great Britain (9518)</td>
</tr>
<tr>
<td>France (7580)</td>
</tr>
<tr>
<td>Italy (8735)</td>
</tr>
<tr>
<td>India (6690)</td>
</tr>
<tr>
<td>Germany (6341)</td>
</tr>
<tr>
<td>Virginia (6314)</td>
</tr>
<tr>
<td>Economic conditions (5963)</td>
</tr>
<tr>
<td>Foreign relations (5613)</td>
</tr>
<tr>
<td>Soviet Union (5309)</td>
</tr>
<tr>
<td>Description and travel (5053)</td>
</tr>
<tr>
<td>China (4158)</td>
</tr>
<tr>
<td>Europe (4154)</td>
</tr>
<tr>
<td>Social conditions (4154)</td>
</tr>
<tr>
<td>England (4102)</td>
</tr>
<tr>
<td>Social life and customs (3394)</td>
</tr>
<tr>
<td>Spain (3147)</td>
</tr>
<tr>
<td>Collection (3172)</td>
</tr>
</tbody>
</table>
# 20th-century Opera in English

<table>
<thead>
<tr>
<th>Title</th>
<th>Author(s)</th>
<th>Composition era</th>
<th>Library</th>
<th>Topic form genre</th>
<th>Recording format</th>
<th>Language</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bluebeard's castle [opera in one act]</td>
<td>Bartok, Bela</td>
<td>20th Century</td>
<td>Ivy Annex</td>
<td>Operas</td>
<td>LP</td>
<td>English</td>
</tr>
<tr>
<td>Death in Venice</td>
<td>Britten, Benjamin</td>
<td>20th Century</td>
<td>Music</td>
<td>Operas</td>
<td>CD</td>
<td>English</td>
</tr>
</tbody>
</table>

Be my love: the golden voice of Mario Lanza
View a Record

Death in Venice

Author: Britten, Benjamin,
Composition era: 20th Century
Library: Music
Topic form genre: Operas
Recording format: CD
Language: English

Click for Virgo.
call number location library format
CD 10462 MU-CIRCUSK MUSIC MUSIC-CD

LEADER 001876 2 2003299 4000
001 53913605
001 ocma88708235
003 O00LC
005 20030109611177
007 sl fongm—as
007 sl fongm—as
007 sl fongm—as
008 0166493991974xunp di eng d
019 5 34976845
024 1 5 028943566927
028 02 5 425 669-2 St London
028 06 5 425 670-2 St London
BlacklightDL

• A corollary project to replace the XPAT indexes and search interface for our Digital Collections Repository (on top of Fedora).
• Indexes full text and metadata for over 10,000 TEI texts, almost 4,000 EAD, and over 20,000 images.
• Work done by OpenSource Connections in conjunction with Bess Sadler and Matt Mitchell at the UVA Library.
• BlacklightDL will be folded into Blacklight.
BlacklightDL Search
Plus Subject Facet Filter
19th-century Books in English with Women as a Subject
Advanced Search in BlacklightDL
The Advanced Search Results
Plus a Subject Facet Filter
BlacklightDL Image Object View
BlacklightDL Image Object View with Menu Options
BlacklightDL Image Object View with Metadata
Questions?

Contact us:

bess@virginia.edu
johnston@virginia.edu