History of Blacklight

Blacklight Summit, Princeton University
November 4, 2015

Robert Cartolano
Associate VP, Digital Programs and Technology Services
Columbia University
Why Blacklight?

Let’s go back ten years...
Back in 2005...

- **Google**
  - Google Books, Maps, Scholar, Video; Google went public in 2004
  - Indexed over 8 billion documents
  - Search was simple, fast, enjoyable

- **Amazon**
  - Made it easy and fun to find books
  - Introduced “search inside” feature

- **Libraries**
  - Usually outsourced search and discovery to vendors
  - Vendor solutions did not offer features that consumers now expected
Expectations

“Our users expect simplicity and immediate reward and Amazon, Google, and iTunes are the standards against which we are judged. Our current systems pale beside them.”

- “Rethinking How We Provide Bibliographic Services for the University of California”, UC Libraries Final Report: Dec. 2005

“Our Systems still suck”

- “Googlezon, Episode VI: Return of the Librarians”, Roy Tenant, Sep. 2005
  Keynote at LITA National Forum
What Was Needed?

- Relevance ranking
- Stemming
- Field weighting
- Spell-checking
- Refining original queries
- Support for popular query operators
- Boolean bag o’ goods
- Flexible default query processing
- In-line query limiters

- Duplicate detection
- Sort Flexibility
- Character Sets (Unicode UTF-8)
- Faceting
- Advanced search
- Easily customized search results
- Best bets
- Search logging and reporting
- Well-rounded administrative interface

- “How OPACs Suck” Parts 1, 2 & 3, Karen G. Schneider, ALA TechSource, March, April, May 2006
Libraries Needed...

- Simplicity and speed
- Enhanced search and retrieval
- Unified discovery (catalog and so much more)
- Continuous improvement

- “Rethinking How We Provide Bibliographic Services for the University of California”, UC Libraries Final Report: Dec. 2005
Many Efforts

- 2005 - NCSU builds separate search index using Endeca
- 2007 - University of Rochester builds Extensible Catalog
- 2010 - Villanova University releases VuFind Beta

Meanwhile, at University of Virginia…

- 2005 - NINES - Jerome McGann, Bethany Nowviskie
- 2007 - COLLEX - Bethany Nowviskie, Erik Hatcher

leading to...
Project Blacklight

or…

“Adapting an Open-Source Scholarly Web 2.0 System for Findability in Library Collections (or: “Frankly, Vendors, We Don’t Give a Damn.”)

Blacklight Design Goals

- Relevance ranking
- Faceted browsing
- Open source design principles
- Ability to include siloed materials
- Customizable interfaces
- Re-mixable data
  - Project Blacklight: a next generation library catalog at a first generation university, Elizabeth (Bess) Sadler, 2008
Blacklight Adoption

- 2008 - University of Virginia
- 2009 - Stanford, Agriculture Network, Northwest Digital Archive
- 2010 - NCSU, WGBH Open Vault, Wisconsin-Madison

In 2008 - Project Hydra formed - Hull, Stanford, Virginia founders
Blacklight Community Achievements

- Search capabilities that meet user expectations
- Speed
- Flexibility to customize to meet local requirements
- Unified discovery (beyond catalog)
- Vibrant community - continuous improvement
Univ. Virginia - Blacklight Wall of Fame

- Bethany Nowviskie - NINES, COLLEX, Blacklight
- Erik Hatcher - COLLEX, Blacklight, Solr, Flare
- Bess Sadler - Blacklight
- Prof. Jerome McGann - NINES
- Bob Haschart - SolrMARC
- Mellon Foundation - funded NINES
- University of Virginia

Honorable Mention: Karen Schneider, Roy Tenant
Blacklight Today

- Installations as diverse as our community
- Meeting both local and consortial needs
- Accelerating technical collaboration - we are going further together!
- Spirit of “Radical Collaboration”
  - partnership on solving significant problems
  - efforts to standardize where we have common needs
  - Shared code, shared experiences
Blacklight Development - Nov. 2015

- 2,801 commits (2,135 as of April 2014)
- 65 contributors (49 as of April 2014)
- 14,389 lines of code (10,855 as of April 2014)

[Diagram]

https://www.openhub.net/p/blacklight
Blacklight Community

- 469 Members (328 in April 2014)
  - blacklight-development@googlegroups.com
- Over 100 Organizations (estimated)
- Part of the Hydra Project

200 HydraNauts from 60 institutions at Hydra Connect 2015
https://wiki.duraspace.org/display/hydra/Hydra+Connect+2015
Blacklight/Solr Features

- Relevance ranking ✔
- Stemming ✔
- Field weighting ✔
- Spell-checking ✔
- Refining original queries ✔
- Support for popular query operators ✔
- Boolean bag o’ goods ✔
- Flexible default query processing
- In-line query limiters
- Duplicate detection
- Sort Flexibility ✔
- Character Sets (Unicode UTF-8) ✔
- Faceting ✔
- Advanced search ✔
- Easily customized search results
- Best bets ✔
- Search logging and reporting ✔
- Well-rounded administrative interface
Blacklight Versions

- 2011 - Blacklight 3 - Rails 3
- 2012 - Blacklight 4 - Bootstrap 2, dropped Ruby 1.8
- 2014 - Blacklight 5 - Bootstrap 3, Rails 4, Schema.org

Continuous Improvement!

- https://github.com/projectblacklight/blacklight/releases
Rapid Growth
The Future of Blacklight

- Collaboration through vibrant communities
- Leverage existing tools and standards
- Improving & Extending Search
- Beyond Solr - Elastic Search, others
Vibrant Communities

● Communities support sustainable project
● Communities share development efforts
● Communities help debug and maintain over time
● Communities help with obtaining external funding for targeted projects

● Help Blacklight! Contribute! Participate!
  ○ http://projectblacklight.org/
  ○ http://projecthydra.org/
Leverage Tools, Standards

- Portland Common Data Model (PCDM)
- International Image Interoperability Framework (IIIF)
- New and Emerging web, content and technology standards
Improving Search

● Improving user experience
  ○ Mobile
  ○ Accessibility
  ○ Consumer expectations for web content

● Leveraging Fair Use for Improved Search & Discovery
  ○ Search Inside (full-text & keyword search)
  ○ Book preview
Extending Search

- Blacklight and Hydra
  - Digital Collections - image, media
  - Archives and Special Collections
  - Institutional Repository
  - Electronic Theses and Dissertations (ETDs)
- GeoBlacklight
- Spotlight
- Arclight
- Time-Based Media (Avalon)
Beyond Solr

- Consistent search interface
  - Unified discovery for aggregated data sources
  - Custom interfaces for specific constituencies

- Examples
  - Elastic Search (e.g. DP.LA)
  - Cloud Computing (e.g. Amazon AWS)
Discussion
References


