

Lucene/Solr Case Studies

Presented by Erik Hatcher

March 25, 2009

ApacheCon EU

Amsterdam

erik.hatcher@lucidimagination.com

Lucene Development with Ant

```
<index index="{index.base.dir}/index"  
    xmlns="antlib:org.apache.lucene.ant">  
  <fileset dir="{files.dir}"/>  
</index>
```

Rossetti Archive

- Purpose: Dramatically improve findability and serendipitous discovery of Rossetti materials
- Data Source: TEI-like XML
- Challenges: case sensitive search, academic relevance tuning
- Technologies: Lucene, Tapestry
- <http://www.rossettiarchive.org/rose/>

recent searches:

Rossetti Archive Search

free form search

[help](#) case sensitive:

structured searching

title search

[help](#)

boolean search

[help](#) case sensitive:

phrase search

[help](#) case sensitive:

genre search

[help](#) case sensitive:

name search

[help](#)

Rossetti Archive Search

recent searches:

free form search

+wombat +archivetype:rap

case sensitive:

search

help

new search

1 - 2 of 2 results (page 1 of 1 pages)

1: **Death of a Wombat** 1869 November 6 - pen and ink - 1 reproduction | [More like this...](#)



I never reared a young **Wombat** To glad me with his pin-hole eye, But when he most was sweet & fat And tail-less; he was sure to die!

2: **Mrs. Morris and the Wombat** 1869 (circa) - pen and ink - 1 reproduction | [More like this...](#)



Collex

- Purpose: Build search/discover/share platform for scholarly objects, starting with 19th century domain (NINES) but aiming general purpose.
- Data Sources: RDF and MARC
- Challenges: Tagging update speed
- Technologies: Solr, Ruby on Rails
- <http://www.collex.org>



Home

My 9s

Search

Tags

Exhibits

News

About

LIMIT BY RESOURCE

Select a resource with which to narrow your search.

▼ **PROJECTS (93 objects)**

The Rossetti Archive (15 objects)

→ **The William Blake Archive (78 objects)** [remove]

LIMIT BY GENRE

Select the genres to include in your search.

Poetry (78 objects)

→ **Visual Art (78 objects)** [remove]

LIMIT TO FREE CULTURE

'Free Culture' resources are freely available in their full form on the Internet.

Free Culture Only (78 objects)

SEARCH QUERY : blakesnakes

Keyword	serpent	AND	[remove]
Genre	Visual Art	AND	[remove]
Resource	The William Blake Archive	AND	[remove]
Keyword	<input type="text" value="click here to add new search term"/>	AND	Add



SEARCH RESULTS: 78 OBJECTS FOUND

Displaying objects per page.

[select all] [collect selected] [tag selected] [expand all items]

1 2 3 4 5 6 7 8 >>



Blake's Illustrations of Dante, (object 6)

The Serpent Attacks Buoso De' Donati, He ey'd the serpent and the serpent him

Genre: Visual Art; Poetry

Tags: [Collect this item to add tags](#)

Blake's Illustrations of Dante, (object 6)

The Serpent Attacks Buoso De' Donati, He ey'd the serpent and the serpent him

Gen
Colle
Tags
[more
Add

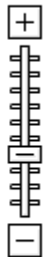
Save Cancel

collect exhibit

WHAT ARE TAGS? A tag is simply a word you use to describe an object you find in NINES. Tags make it easy to organize and browse large collections. To join two words together use a hyphen (*oil-painting*) or an underscore (*oil_painting*).

filter tags

1795 1808 1915_SF_Exposition aesthetics affinities **african** alastor
america american **american-art** americana and angel **animals** annuals
antebellum architect **arnold art art-criticism** art-history arthur arthurian
article atr baby bachelor beard **beatrice binding**
binding_commentary **blake borgia** botticelli **bronte** browning
byron caricature casson cather cenci christina-carroll-rossetti citation
commentary **comrade corelli** corpus criticism crucifix **crystal-palace**
culture dante death **decorative-arts dgr dickinson digital-**





NINES

nineteenth-century studies online

[Exhibit List](#)

signed in as erikhatcher | [my account](#) | [sign out](#) | [admin](#)

This exhibit is not peer reviewed.

An Introduction to D.G. Rossetti

Jerome J. McGann

The Life



Dante Gabriel Rossetti was born in London on 12 May 1828 and he died on Easter Day, 9 April 1882. He spent nearly his entire working life in the city of his birth, and indeed he only left Great Britain three times, in each case but the first quite briefly. Though his work is steeped in Italian traditions (both poetical and pictorial), Rossetti never visited Italy. He is first and always an English - more, a London - writer and artist.

His father was the celebrated (and controversial) Dante scholar and Italian political exile Gabriele Rossetti (1783-1854). His mother Frances (1800-1886), much younger than her husband, was Anglo-Italian, Polidori on her father's side. (Her brother, Dr. John Polidori, was Byron's doctor and companion during the first part of his exile from England in 1816.) Rossetti had three siblings, two younger than himself. All were remarkable. His sister Christina (1830-1894) became as distinguished a poet as her brother. His brother William Michael (1829-1919), a writer himself, edited his brother's work after the latter's death and served as the first archivist and historian of the Pre-Raphaelite movement. His other sister was the oldest child, Maria Francesca (1827-1876); she published a commentary on Dante and became an Anglican nun.

Blacklight

- Purpose: Open source scalable clean next generation library discovery interface.
- Data Sources: MARC, Fedora, EAD, ... anything
- Challenges: academia, competitors
- Technologies: Solr, Ruby, Rails, Java indexer(SolrMarc)
- "A process, not a product"
 - <http://code4lib.org/node/177>
 - <http://blacklightopac.org/>

Catalog

Music

All Fields

Results per page

10

Search

Limit Search by...

Library

Format

Subject

Language

Call number

Region

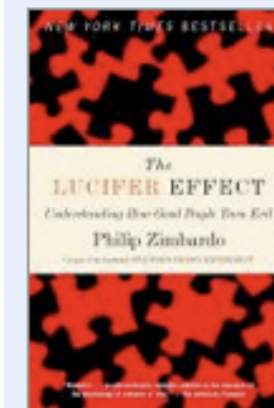
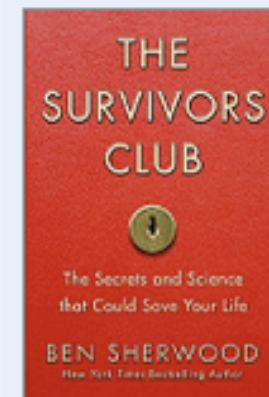
Digital collection

Source

Series title

You don't have any searches or limits applied. To get started, type a search above or click one of the limits on the left side of the screen.

Recently Added Items



Catalog

Music

amsterdam

All Fields

Results per page

10

Search

Displaying all 3 items for the search **amsterdam** in items with **INSTRUMENT** **Guitar** (Start over?)

Sort by Relevancy

Get updates with **RSS**

Limit Search by...

Format

Recordings and scores

Recording format

Instrument

Author

Library

Language

Source

Subject

1 Interlude for guitar, 1984

Author Leeuw, Ton de

Format Musical Score

Publication Date c1985

Call Number M127 .L43 I5 1985

Availability [check availability](#)

Cover
Not
Available

2 The well-tempered guitars

Author: Castelnuovo-Tedesco, Mario

Format LP; Musical Recording

Call Number LP 12929

Availability [check availability](#)

Cover
Not
Available

Catalog

Music

lucene

All Fields

Results per page

10

Search

Easy information retrieval: building a search tool based on lucene api.



Add to:  [Delicious Bookmarks](#)

Author(s) / Editor(s) [Andersen, Christopher](#)

Format Thesis/Dissertation | Book

Published 2008

Availability



Recalls and other requests

< Previous | Next >

Blacklight story

- Bethany Nowviskie, Bess Sadler, and Erik Hatcher. “Adapting an Open Source, Scholarly Web 2.0 System for Findability in Library Catalogs.” *Library 2.0 Initiatives in Academic Libraries*. (Laura Cohen, ed.). Association of College and Research Libraries: Chicago, 2008.

Flare

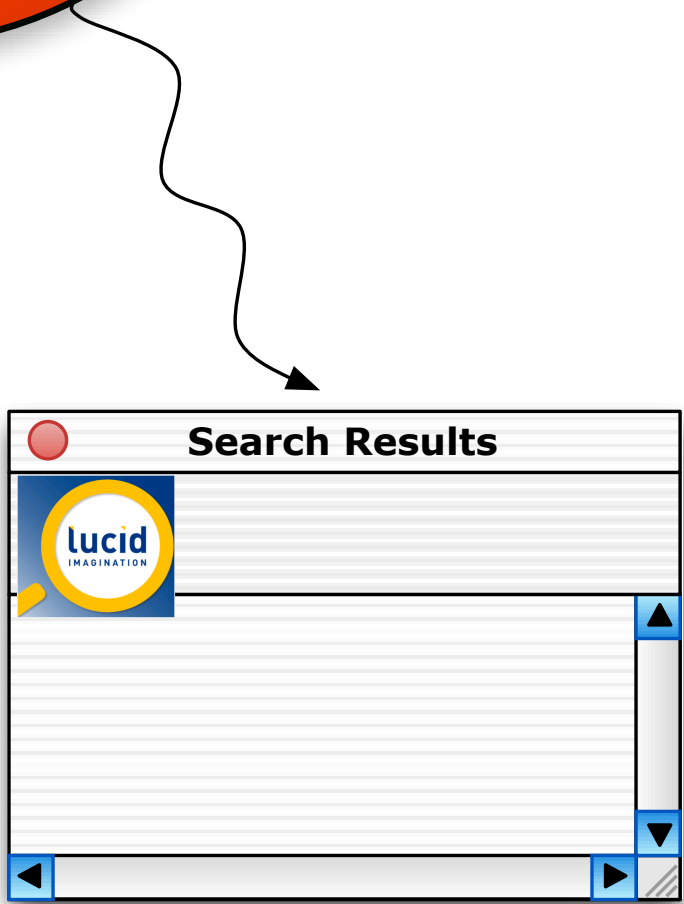
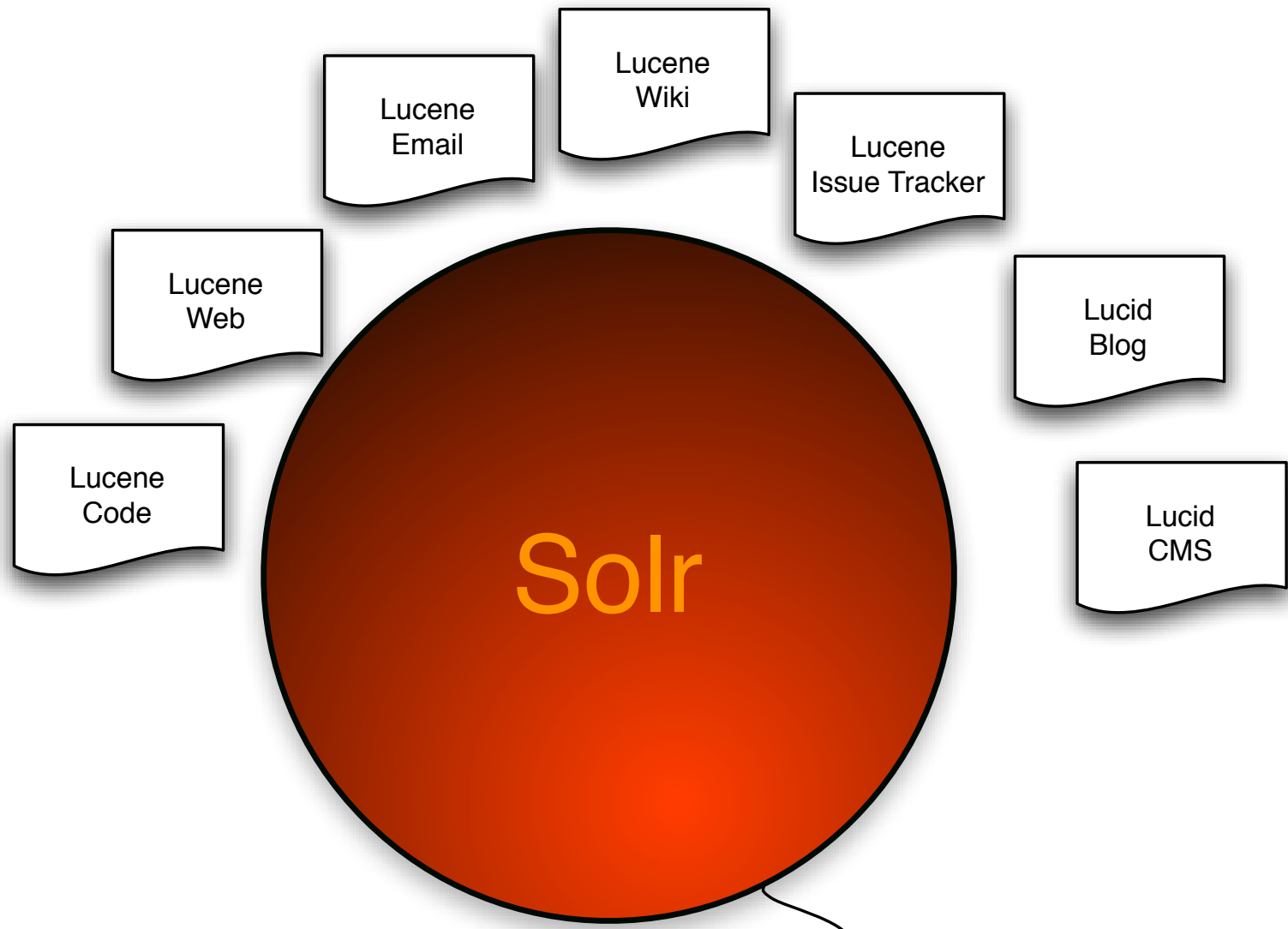
- Distilled from Blacklight development, proof-of-concept Rails plugin
- Features: suggest, saved searches, pie chart faceting, Simile Timeline/Exhibit integration
- <http://wiki.apache.org/solr/Flare>

Solritas

- Light-weight Velocity templated Solr output
- rapid prototyping
- <http://wiki.apache.org/solr/Solritas>

LucidFind

- Purpose: Company technology showcase, community focused service. Indexed `lucene.apache.org/*`: wiki, web, code, issues, e-mail, nice UI
- Challenges: None to speak of
- Technologies: Solr, PHP, Ant
- <http://www.lucidimagination.com/search>



Options

PROJECT

[clear facet](#)

- Lucene (68849)

- Solr (35512)

- Nutch (25838)

- Tika (2040)

- Mahout (3304)

- Droids (356)

- PyLucene (223)

- Lucene.Net (3024)

- Lucy (246)

- Lucene4c (132)

Search within: [Lucid site](#) [Lucene wikis and web](#) [Lucene issues, email, and code](#) [Lucid site, Lucene wikis and web](#)



BETA

[Start new search](#)

SOURCE

[clear facet](#)

- Lucid (102)

- Wiki (2608)

- Apache Lucene Web (80)

- Email (108527)
 - user (76667)
 - dev (31860)

- Issues (26841)
 - open (5134)
 - closed (21707)
 - main issue (3904)
 - comment (22937)

- Code (1451)

Powered by Ant

```
$ ant -p  
Buildfile: build.xml
```

Main targets:

<code>archive-focus-logs</code>	Archive rolling log files for posterity
<code>commit</code>	Commit to Solr
<code>delete-source</code>	Delete specified source from index
<code>index-code</code>	Index Lucene projects code
<code>index-issues</code>	Index JIRA issues and comments
<code>index-lia</code>	Index Lucene in Action (1st edition)
<code>index-lucid</code>	Index Lucid site and articles
<code>index-mail</code>	Index mail
<code>index-web</code>	Index Lucene web content
<code>index-wiki</code>	Index Lucene wiki
<code>optimize</code>	Optimize Solr index

Questions?

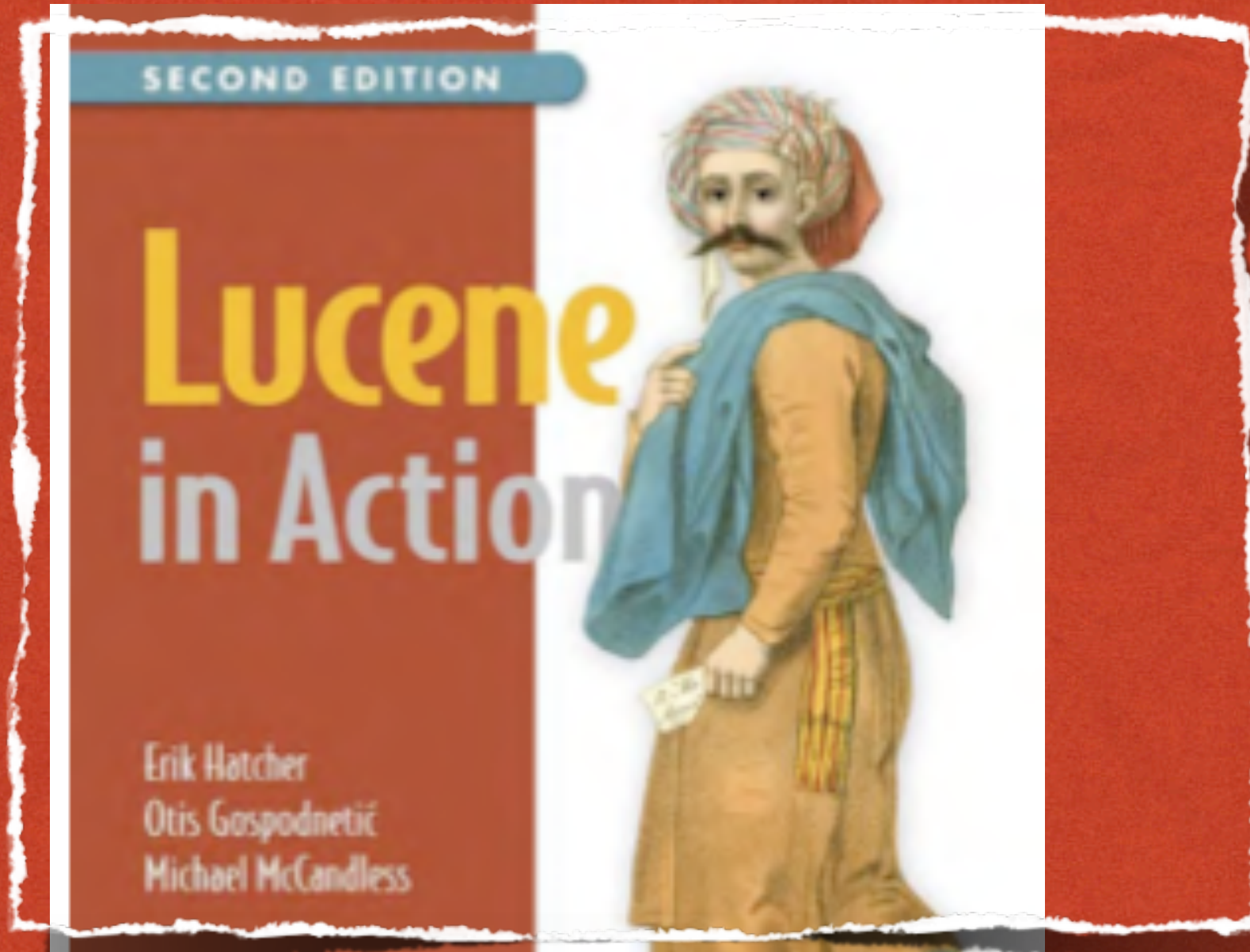
Answer:

"it depends"

lucid
IMAGINATION

Thinking Lucene ▼ Think Lucid.

lucidimagination.com



e-book now available!

Print coming this summer

<http://www.manning.com/hatcher3>