

Contrary to popular belief, elephants do not provide good relevancy tests.

Nor do cats.



<http://daisythecurlycat.blogspot.com/2009/03/elephant-mancat.html>

Practical Relevancy Testing

Naomi Dushay

Stanford University Libraries
Code4Lib 2011





<http://bacolicio.us/>
<http://www.jesterartsillustrations.com/images/free/300/17066.jpg>



SUOLAIR MY ACCOUNT FEEDBACK ASK US

Stanford University Libraries SearchWorks

Search 6,473,314 resources in Stanford's libraries

Advanced Search
Select multiple facets and combine search terms across multiple fields

Looking for Articles?
Articles are not included in SearchWorks yet.

- See [Databases A-Z](#) to find the best database(s) to search for your topic.
- Try [xSearch](#) to search several databases at once.

Course Reserves

- Search the [Course Reserves Catalog](#) for print, media and e-resources set aside for classes.

Can't find it at Stanford?

- Request it through [interlibrary borrowing](#).
- [Ask us to purchase it](#) for the libraries.

SearchWorks News [See all SearchWorks news](#)

Advanced Search
Wed, 26 Jan 2011 18:43:54
Advanced search allows for selecting multiple facets to search, searchin...

Simple Search
Wed, 26 Jan 2011 18:41:56
The simple search box appears on the SearchWorks start page and in the r...

Icons
Wed, 26 Jan 2011 17:48:04
gallery view displays search results in a matrix of title/author an...

Access

- At the Library (5,827,054)
- Online (894,066)

Format

- Book (5,759,823)
- Journal/Periodical (295,975)
- Microformat (272,093)
- Conference Proceedings (193,766)
- Thesis (109,072)
- Music - Recording (93,403)
- Other (87,443)
- Music - Score (83,917)
- Map/Globe (59,945)
- Video (46,630)
- Manuscript/Archive (12,473)
- Newspaper (5,403)
- Computer File (3,652)
- Sound Recording (3,475)
- Image (3,081)

Call Number

Location

Language

powered by Blacklight

© Stanford University, Stanford, California 94305. (650) 725-1064. Terms of Use | Copyright Complaints

SUOLAIR MY ACCOUNT FEEDBACK ASK US

SearchWorks

bacon Search

Advanced search

Your Search: bacon Limited to: (no limits selected) Selected items (none)

< Previous | 1 - 20 of 5557 results | Next > Sort by relevance >>> 20 per page

Refine your search

Access

At the Library (4,621)
Online (1,229)

Format

Book (5,258)
Microformat (355)
Video (123)
Music - Recording (83)
Music - Score (36)
Thesis (32)
Conference Proceedings (24)
Journal/Periodical (19)
Manuscript/Archive (16)
Map/Globe (13)
Other (5)
Sound Recording (3)
Newspaper (1)

Author

Publication Year

Topic

Location

Call Number

Organization (as Author)

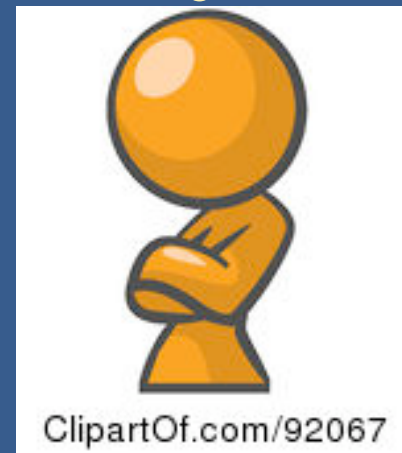
Region

Era

Language

- Bacon : portraits and self-portraits**
Title: Bacon, English
Author/Creator: Bacon, Francis, 1909-1992
Published: 1996
At the Library: Art > Stacks > ND1329.B27 A4 1996A F
Format: Book
- Bacon**
Author/Creator: Bacon, Francis, 1909-
Published: 1995
At the Library: SAL3 > Stacks > ND497 .B16 A4 1995 F
Format: Book
- Bacon**
Author/Creator: Bacon, Francis, 1909-1992.
Published: 2008
At the Library: Art > Stacks > ND497 .B16 A4 2008C F
Format: Book
- Bacon**
Author/Creator: Church, R. W. (Richard William), 1815-1890.
Published: 1920
At the Library: SAL > Stacks > 824.3 .B1282CA
Format: Book
- Bacon**
Author/Creator: Church, R. W. (Richard William), 1815-1890.
Published: 1901
At the Library: Green > Stacks > 820.92 .M864C V.2 V.2
Format: Book

< Previous 1 2 3 4 5 6 7 8 9 ... Next >



<http://bacolicio.us/>
<http://www.jesterartsillustrations.com/images/free/300/92067.jpg>

Solr has ...

Solr has ...

Shiny
Knobs!



http://www.aandhbrass.co.uk/products/door_furniture/mortice_knobs.htm



**NO changes
without tests!**

How Do You Test Search Result Relevancy?

What IS

Search Result Relevancy?



http://www.istockphoto.com/file_thumbview_approve/1457289-cherry-slot-machine.jpg



http://www.istockphoto.com/file_thumbview_approve/1457289-cherry-slot-machine.jpg



http://www.istockphoto.com/file_thumbview_approve/1457289-cherry-slot-machine.jpg

How Do You Evaluate the Relevancy of Search Results?



Bess Sadler

A: Use Complaints.

A: Use ~~Complaints~~.

... er ...

Feedback.

“... A search for (no quotes):

memoirs of a physician dumas

... book with that title by dumas is the third result; I would expect it to be the first.”



ClipartOf.com/92067

HOW Do You **TEST**

Search Result Relevancy?

Repeatable

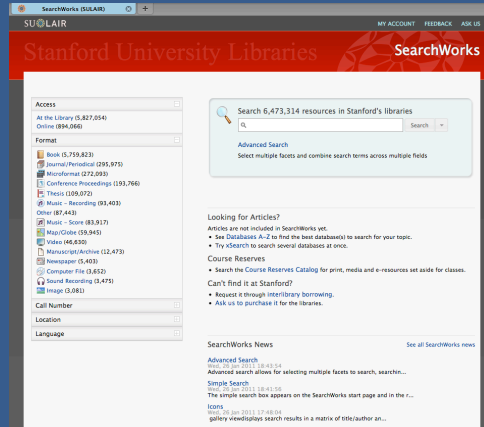
Automatable

Full Stack Test:

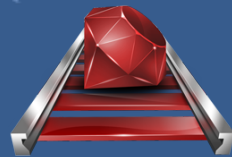
As if:

- Query entered in UI
- App gets form input
- App processes query
- App sends processed query to Solr
- Solr processes query
- App processes Solr results
- Results returned to UI

<http://discovery-grindstone.blogspot.com/>



blacklight



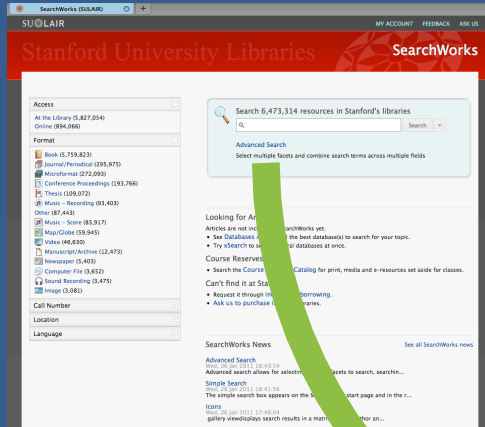
Ruby on Rails



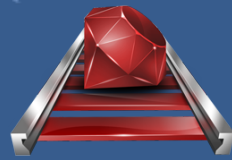
<http://cukes.info/>



Automatable!



blacklight



Ruby on Rails



“... A search for (no quotes):

memoirs of a physician dumas

... book with that title by dumas is the third result; I would expect it to be the first.”



ClipartOf.com/92067

(demo)

```
Feature: code4lib 2011 demo-  
  In order to prove cucumber works-  
  As a presenter-  
  I want the scenario below to run-  
  -  
  Scenario: "memoirs of a physician"-  
    Given a SOLR index with Stanford MARC data-  
    When I go to the home page-  
    And I fill in "q" with "memoirs of a physician"-  
    And I press "search"-  
    Then I should get at least 2 of these ckeys in the first 2 results: "2318407, 8790640"-
```


Feature: code4lib 2011 demo

In order to prove cucumber works

As a presenter

I want the scenario below to run

Scenario: "memoirs of a physician"

Given a SOLR index with Stanford MARC data

When I go to the home page

And I fill in "q" with "memoirs of a physician"

And I press "search"

Then I should get at least 2 of these ckeys in the first 2 results: "2318407,
8790640"

When I go to the home page

And I fill in "q" with "memoirs of a physician"

And I press "search"

Then I should get at least 2 of these ckeys in the first 2 results: "2318407, 8790640"

Use (a copy of?) **your full index for testing.**

Tests assert that searches, in context,
retrieve correct results.

VUF-779 Student

“... [searching for] Le Rayon Vert,
... Searchworks stupidly supplies
results for textiles, when the French
Rayon in question refers to
sunshine.”



ClipartOf.com/92067

New Test!

Feature: code4lib 2011 demo

In order to prove cucumber works

As a presenter

I want the scenario below to run

Scenario: Le Rayon Vert search should not return rayon (fabric)

Given a SOLR index with Stanford MARC data

And I go to the home page

When I fill in "q" with "Le Rayon Vert"

And I press "search"

Then I should get these ckeys in the first 3 results: "4522134, 729146"

And I should get no more than 20 results

And I should not see "industry" in the results

Feature: code4lib 2011 demo

Feature: code4lib 2011 demo

In order to prove cucumber works

As a presenter

I want the scenario below to run

Scenario: Le Rayon Vert search should not return rayon (fabric)

Given a SOLR index with Stanford MARC data

And I go to the home page

When I fill in "q" with "Le Rayon Vert"

And I press "search"

Then I should get these ckeys in the first 3 results: "4522134, 729146"

And I should get no more than 20 results

And I should not see "industry" in the results

And I should not see "industry" in the results

Feature: code4lib 2011 demo
In order to prove cucumber works
As a presenter
I want the scenario below to run

Scenario: Le Rayon Vert search should not return rayon (fabric)

Given a SOLR index with Stanford MARC data

And I go to the home page

When I fill in "q" with "Le Rayon Vert"

And I press "search"

Then I should get these ckeys in the first 3 results: "4522134, 729146"

And I should get no more than 20 results

And I should not see "industry" in the results

Scenario: Le Rayon Vert search should not return rayon (fabric)

Given a SOLR index with Stanford MARC data

And I go to the home page

When I fill in "q" with "Le Rayon Vert"

And I press "search"

Then I should get at least 2 of these ckeys in the first 5 results: "4522134, 729146"

And I should get no more than 20 results

And I should not see "industry" in the results

Scenario: Le Rayon Vert search should not return rayon (fabric)

Given a SOLR index with Stanford MARC data

And I go to the home page

When I fill in "q" with "Le Rayon Vert"

And I press "search"

Then I should get at least 2 of these ckeys in the first 5 results: "4522134, 729146"

And I should get no more than 20 results

And I should not see "industry" in the results

Feature: code4lib 2011 demo

In order to prove cucumber works

As a presenter

I want the scenario below to run

Scenario: Le Rayon Vert search should not return rayon (fabric)

Given a SOLR index with Stanford MARC data

And I go to the home page

When I fill in "q" with "Le Rayon Vert"

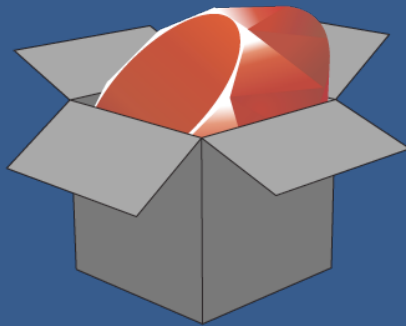
And I press "search"

Then I should get at least 2 of these ckeys in the first 5 results: "4522134, 729146"

And I should get no more than 20 results

And I should not see "industry" in the results

<http://cukes.info/>



rspec
rspec-rails
webrat

cucumber:

Regular Expressions

webrat:

Faking User Input to Web Pages

Step Definitions

Examples of html simulations via webrat

- **pulldown:**

And I select "____" from "____"

- **Link:**

And I follow "_____"

rspec:

Behavior Driven Development - a twist on unit testing:

`object.should (be_something | do_something)`

vs.

`assertXXX(object.method)`

Step Definitions

Localized:

“the first 4 results should be ...”

“record 777 should be before record 999”

Step Definitions

Localized:

Not too brittle!!



Accommodate ongoing changes to your data

Step Definitions

Stanford's step definitions:

<http://discovery-grindstone.blogspot.com/>

[http://www.stanford.edu/~ndushay/
code4lib2011/search_result_steps.rb](http://www.stanford.edu/~ndushay/code4lib2011/search_result_steps.rb)

But I USE (toasters) ...

<http://cukes.info/>

PHP:

<https://github.com/aslakhellesoy/cucumber/wiki/PHP>

Google: cucumber toasters

Weaknesses With This Approach

1. you don't know what you don't know.

- unreported problems
- poor recall
- good precision for some items only

2. some feedback is too vague.

3. context experts are too ...

- “Socrates shows 8 results, Searchworks 7, and there are some results in one but not the other and vice versa”
- “Great! Tell me what should be in the result set.”
- (crickets)

4. positive feedback for specific searches and results is rare.
- 4a. positive feedback gets little attention.

5. cucumber tests can be slow

Shiny Knobs!



http://www.aandhbrass.co.uk/products/door_furniture/mortice_knobs.htm

SUOLAIR MY ACCOUNT FEEDBACK ASK US

SearchWorks

Search bacon

Advanced search

Your Search: bacon Limited to: (no limits selected)

Selected items (none)

< Previous | 1 - 20 of 5557 results | Next > Sort by relevance >>> >>> 20 per page

Refine your search

Access

At the Library (4,621)
Online (1,229)

Format

Book (5,258)
Microformat (355)
Video (123)
Music - Recording (83)
Music - Score (36)
Thesis (32)
Conference Proceedings (24)
Journal/Periodical (19)
Manuscript/Archive (16)
Map/Globe (13)
Other (5)
Sound Recording (3)
Newspaper (1)

Author

Publication Year

Topic

Location

Call Number

Organization (as Author)

Region

Era

Language

1. Bacon : portraits and self-portraits
Title: Bacon, English
Author/Creator: Bacon, Francis, 1909-1992
Published: 1996
At the Library: Art » Stacks » ND1329.B27 A4 1996A F
Format: Book

2. Bacon
Author/Creator: Bacon, Francis, 1909-
Published: 1995
At the Library: SAL3 » Stacks » ND497 .B16 A4 1995 F
Format: Book

3. Bacon
Author/Creator: Bacon, Francis, 1909-1992.
Published: 2008
At the Library: Art » Stacks » ND497 .B16 A4 2008C F
Format: Book

4. Bacon
Author/Creator: Church, R. W. (Richard William), 1815-1890
Published: 1920
At the Library: SAL » Stacks » 824.3 .B1282CA
Format: Book

5. Bacon
Author/Creator: Church, R. W. (Richard William), 1815-1890.
Published: 1901
At the Library: Green » Stacks » 820.92 .M864C V.2 V.2
Format: Book

< Previous 1 2 3 4 5 6 7 8 9 ... Next >



<http://www.jesterartsillustrations.com/images/free/300/19089.jpg>

Zap ‘em back with superlove!

“Back in June, you reported that a search ... did not get the expected results.

[fix ...]

Thank you for reporting your specific example, especially the expected ckey -- we use the information to write test case(s) that, once fixed, must pass forever more.”

<http://discovery-grindstone.blogspot.com/>

The Big Guns



Whenever you test something ("manually") in SearchWorks, we would like to capture your expectations as a cucumber scenario so we can run the test repeatedly and automate it.

Benefits:

we won't have to keep asking you to check the same things over and over. **Imagine never having to perform a given test search again!**

We can ensure that applying a fix for one problem **won't inadvertently break something we've already fixed.**

We can automate running a large suite of tests nightly so **we keep checking that we haven't broken anything.**

As we add specific searches and expected results against our own (meta)data corpus, **we are accruing relevancy tests for our own data, based on human review of search results.**

<http://hubpages.com/hub/BodyBuilding-The-Science-behind-Creatine-A-Definitive-Guide>



KTHXBAI

<http://cukes.info/>

<http://rspec.info/>

<http://www.pragprog.com/titles/achbd/the-rspec-book>

<http://discovery-grindstone.blogspot.com/>

<http://www.stanford.edu/~ndushay/code4lib2011/>

Evaluating search result relevancy is difficult for any sizable amount of data, since human vetted ideal search results are essentially non-existent. This is true even for library collections, despite dedicated librarians and their familiarity with the collections. So how can we evaluate if search engine configuration changes (e.g. boosting, field analysis, search analysis settings) are an improvement? How can we ensure the results for query A don't degrade while we try to improve results for query B? Why yes, Virginia, automatable tests are the answer. This talk will show you how you can easily write these tests from your hidden goldmine of human vetted relevancy rankings.