

Search Tools In PHP

How They Match Up

Rob Young
IPC Media

<http://www.roryoung.co.uk>

Types of Search Tool



Types of Search Tool



Hosted

Pros

- Easy setup
- Low maintenance
- Scalable

Cons

- Low re-index interval
- Lack of control
- Lack of features

What To Look For



- Ease
- Performance
- Features

- Ease includes installation, integration, configuration and also how easy it is to find support.
- Performance includes speed and quality
- A popular measure of result quality is precision and recall http://en.wikipedia.org/wiki/Information_retrieval#Performance_measures

How They Match Up



- Easy to install
- Easy to integrate
- Support everywhere
- Fairly slow
- Few features

•See: <http://dev.mysql.com/doc/refman/5.0/en/fulltext-search.html>

How They Match Up



- Easy installation
- Some features
- ZF mailing list
- Poor performance

•See: <http://framework.zend.com/manual/en/zend.search.lucene.html>

How They Match Up

- Fast
- Features
- Pretty busy mailing list
- Needs installation



•See: <http://xapian.org/>

How They Match Up

- Very fast
- Features
- Forum with quick response
- Needs installation
- Needs configuration

 Sphinx

•See: <http://www.sphinxsearch.com/>

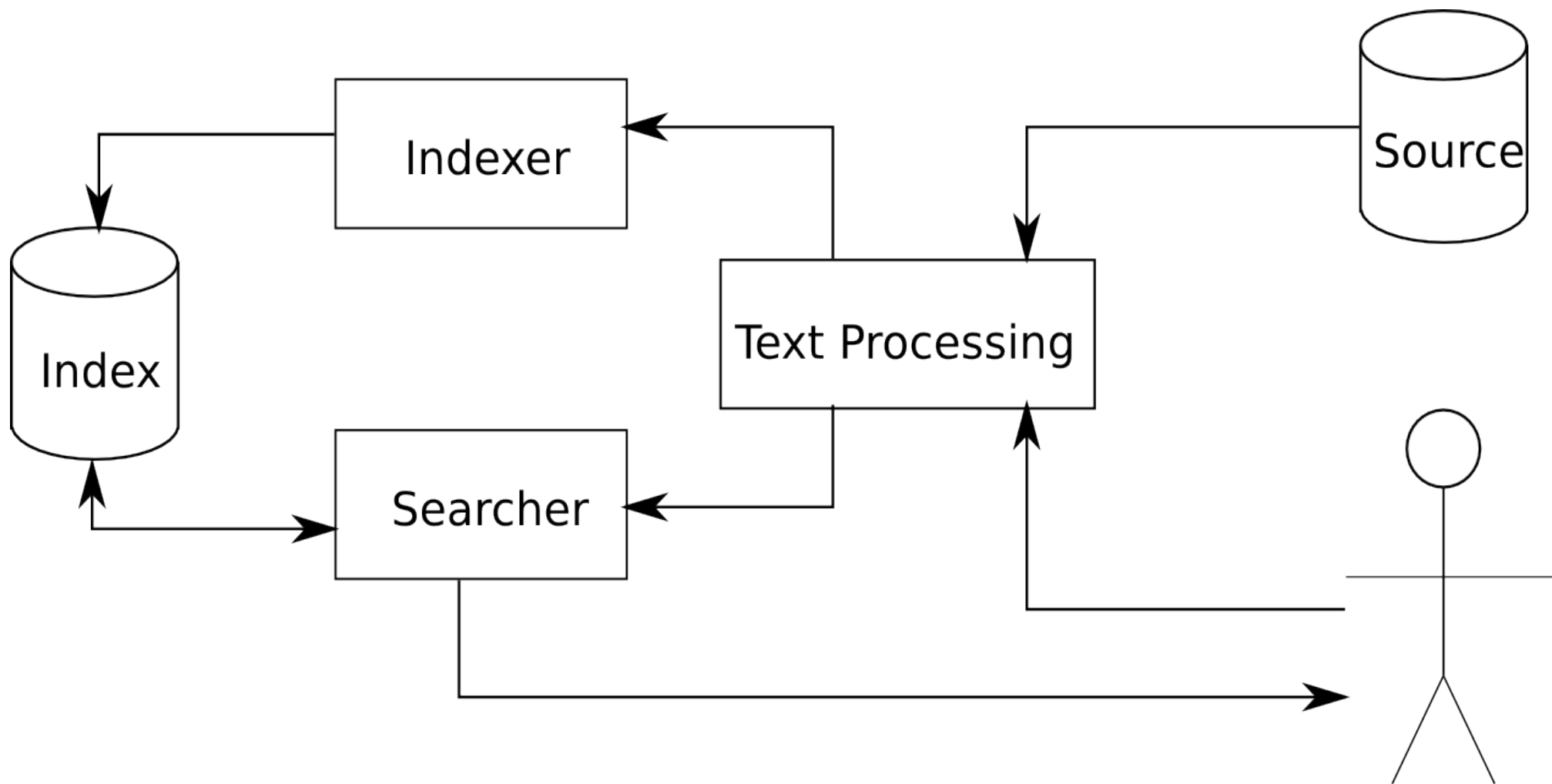
How They Match Up

- Very Fast
- Packed with features
- Very busy mailing list
- Needs installation
- Needs configuration

Solr

- See: <http://lucene.apache.org/solr/>
- Related projects: <http://lucene.apache.org/>
- Notably Tika: <http://incubator.apache.org/tika/>

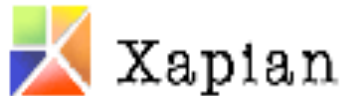
Architecture



- Source may be many things a database; a crawler or spider; a remote source or data feed for example, or even a combination of many sources.
- Searcher would include a query parser, ranking model and searching interface.

Text Processing

- Stop word removal

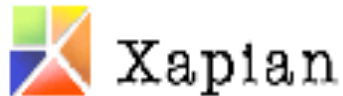


• This is usually done at index time as well as search time to optimize the index but it doesn't have to.

• “a word which occurs in 80% of the documents in the collection is useless for purposes of retrieval” - Baeza-Yates, Ribeiro-Neto

Text Processing

- Stop word removal
- Stemming



• English stemming is usually provided by the Porter2 algorithm.

– <http://tartarus.org/~martin/PorterStemmer/>

– <http://snowball.tartarus.org/algorithms/english/stemmer.html>

• For information about stemming with Zend_Search_Lucene see <http://www.codeangel.org/article/stemming-in-zend-search-lucene>

Text Processing

- Stop word removal
- Stemming
- Synonyms



•All these text processing features are language specific. This makes multi-lingual search across the same index very difficult. Separate indexes are usually a simpler solution.

Boosting

- Field level boosting



•The ability to say that one part of a document is more important than another. For example, you may wish the title to be more important than the body.

Boosting

- Field level boosting
- Document level boosting



- The ability to say that one document is more or less important than others.

Boosting

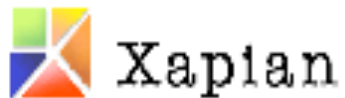
- Field level boosting
- Document level boosting
- Term level boosting



•The ability to say that a term (or word) in a query is more or less important than others.

Other Features

- Prefix Querying





information ar	
information architecture	25,000,000 results
information architect	5,480,000 results
information architects	2,700,000 results
information architecture institute	8,100,000 results
information arts	61,800,000 results
information architecture for the world wide web	2,070,000 results
information architecture summit	201,000 results
information around us	397,000,000 results
information architecture diagram	341,000 results
information architect institute	7,960,000 results
	close

[Advanced Search](#)
[Preferences](#)
[Language Tools](#)



Other Features

- Prefix Querying
- Spell checking



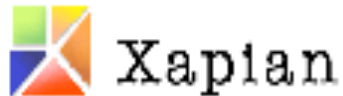
Other Features

- Prefix Querying
- Spell checking
- Suggested terms



Other Features

- Prefix Querying
- Spell checking
- Suggested terms
- Faceted Navigation





You can do it. We can help.™

SHOPPING CART | ORDER STATUS | MY LIST | MY ACCOUNT | SIGN IN



Appliances | Bath | Building Materials | Décor | Doors & Windows | Electronics | Flooring | Kitchen | Lighting & Fans | Outdoors | Paint | Storage | Tools & Hardware

Enter Keyword or SKU

SEARCH

Gift Cards | Gifts | Promotions | Know-How | Pro | Credit | Home Services | Weekly Ad | Store Finder | Help

You are here: HOME > Tools & Hardware > Hand Tools > Cutting Tools

10% OFF Appliances \$299 or More [Shop Now](#)

Shop Products

To View Products & Prices

Enter Zip Code:

GO

YOU HAVE SELECTED:

Category

▶ Cutting Tools

CONTINUE NARROWING BY:

Category

- ▶ Bolt Cutters (3)
- ▶ Hacksaws (2)
- ▶ Hand Saws & Miter Boxes (17)
- ▶ Knives (21)
- ▶ Punches & Nail Setters (1)
- ▶ Scissors (2)
- ▶ Sharpening Stones (2)
- ▶ Snips & Sidecutters (19)
- ▶ Utility Knives & Blades (44)

Price

- ▶ Less than \$50 (80)
- ▶ \$100 - 200 (3)
- ▶ \$200 - 400 (1)

Brand

▶ Buck Knives (6)

Cutting Tools

Print Page

84 Products

Sort By: Top Sellers

Results per page: 12

View: Grid | List

Products: Online | In-Store | All Products

1 2 3 4 5 ▶

Select up to 4 items to compare. COMPARE

Select to compare



Husky
Folding Lock-back
Utility Knife

Model 211113C

\$7.98

ADD TO CART

Check Your Local Store

Select to compare



Wiss
Metalmaster Straight
Aviation Snip

Model M3RS

\$12.97

ADD TO CART

Check Your Local Store

Select to compare



Stanley
100 Pack Heavy Duty
Blades

Model 11-921K

\$11.99

ADD TO CART

Check Your Local Store

Select to compare



Husky
Lock Back Knife
w/Blade Storage

Model 12120

\$14.97

ADD TO CART

Check Your Local Store

Yea, OK, but what should I use?

- If you use an abstraction layer like Forage you can develop to one engine which is easy to install and then switch to a more advanced engine when you need to.

- <http://code.google.com/p/forage>

Where to Learn More

<http://www.findability.org>

<http://flickr.com/photos/morville/collections/72157604060564791/>

<http://arnoldit.com/wordpress/>

<http://www.boxesandarrows.com/search?q=search>

<http://code.google.com/p/forage>