



Optimizing
Solr performance and User
Experience for E-Commerce
15 April 2010

Thinking Lucene ▼ Think Lucid.

Agenda

Slides posted at the end of this presentation; full replay available within ~48 hours of live webcast

- ▼ **Introductions**
- ▼ **Apache Lucene/Solr and eCommerce - Grant**
 - ▼ Lucene/Solr Powered Sales
 - ▼ Common Scenarios
 - ▼ Search Needs for eCommerce
- ▼ **Use Case: Sheet Music Plus - Brian**
 - ▼ Leveraging Solr to sell Music online
 - ▼ Gauging search success through timely metrics

Introductions

▼ Grant Ingersoll

- ▼ Co-Founder of Lucid Imagination
- ▼ Lucene/Solr/Mahout committer



▼ Brian Doll

- ▼ Search Architect at Sheetmusicplus.com
- ▼ 12+ years experience in eCommerce, finance, retail and media



A few Lucene/Solr-Powered Commerce Sites



the web's most popular shoe store!



eCommerce Scenarios

- ▼ **Users can't buy it if they can't find it!**
 - ▼ Search/Discovery is often mission critical in eCommerce
 - ▼ Users don't know how to spell
 - ▼ Users often don't even know how to describe it
- ▼ **Online stores typically look like:**
 - ▼ SKUs (products): 1K to 500K typically, but we have customers with 10-50M+ SKUs in Solr
 - ▼ High Volume Search: Millions of queries/day (peaks @ Xmas in US)
 - ▼ Lots of Metadata with 1 or 2 text fields
 - ▼ Many product names are "tricky":
GPSGolfPro, iPod, Tchaikovsky, Garmin nuvi 255w

eCommerce Checklist for Search

- ▼ Keyword search
- ▼ High Quality relevance (precision @ < 10)
- ▼ Faceting/Discovery
- ▼ Flexible language analysis tools
 - ▼ Stemming, protected words, case changes, alpha-numeric
- ▼ Multilingual support
- ▼ Frequent Incremental Updates
 - ▼ Ratings, Reviews, New Products
- ▼ Best Bets

Instruments
Piano
Guitar
Voice

Narrow by keyword:
(search within results)

Go!

Narrow by type:

Instruments

- [Piano and Keyboard \(5299\)](#)
- [Strings \(3745\)](#)
- [Vocal \(3403\)](#)
- [Woodwinds \(3147\)](#)
- [Brass \(2386\)](#)
- [Percussion \(1180\)](#)
- [Guitar \(259\)](#)
- [Folk \(130\)](#)
- [see 1 more...](#)

Ensembles

- [Choir \(2516\)](#)

eCommerce Checklist for Search, continued

- ▶ Auto-suggest
- ▶ Did you mean?
- ▶ Related Searches/Items
- ▶ Editorial Relevance Controls
 - ▶ Sales, Margins, Inventory, Fixed results, Exclusions, Ratings
- ▶ Admin:
 - ▶ Scalability, Fault Tolerance, Easy Setup and Config
- ▶ Recommendations (See Mahout)
- ▶ Analytics and other Business Tools



eCommerce and Solr

- ▼ **Most of the checklist comes out of the box with Apache Solr**
- ▼ **Primary missing piece: Analytics + high level business tools**
 - ▼ But one could argue that is a feature, not a bug!
 - ▼ Many people already have their own reporting/analytics
 - ▼ Often quite easy to hook in Solr/Lucene to those solutions

Solr E-Commerce Case Study: Sheetmusicplus.com

Overview

- ▼ About Sheetmusicplus and our Solr infrastructure
- ▼ Onsite search: understanding the value
- ▼ Understanding your traffic, your search and your customers
- ▼ Setting search strategies
- ▼ Understanding Customer Value
- ▼ Q&A

About sheetmusicplus.com

- ▼ Largest selection of sheet music
- ▼ 12 year old online business
- ▼ 600k+ SKUs
- ▼ 1.5M+ Songs
- ▼ Faceted navigation and on-site search with Solr:
 - Up to 35 application requests per second this year/~3 million per day
 - Up to 11 solr queries per second this year/~1 million per day
- ▼ Ruby on Rails app servers running on Passenger / Apache
- ▼ MySQL / Memcached / VMWare
- ▼ Lots of needles in a huge haystack

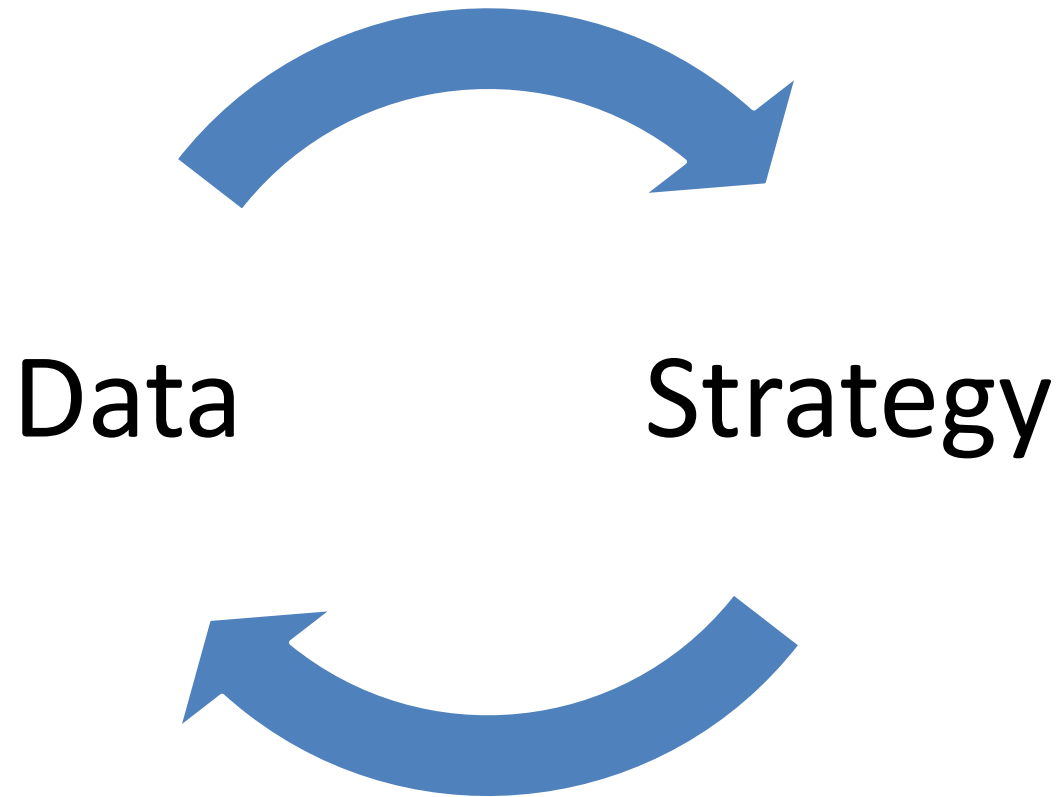
What's the value of on-site search?

- ▼ Value of a new prospect
- ▼ Value of an existing customer
- ▼ eCommerce conversion rate of the search page
- ▼ Dollar value of the search page
- ▼ Dollar value of top search terms

Why Solr?

- ▼ Strategies:
A set of patterns to match search operations with various subsets of use cases to optimize results for different users
- ▼ We love that the structure of your data is part of your Solr environment (schema.xml)
- ▼ And the way you query against that data is part of your app
- ▼ Solr provides an easy and flexible API that is easy to experiment with

Step 1



Get to know your visitors

Google Analytics

**on-site search terms are not as varied as we
had thought**

Surprisingly, lots of people search like this:

Google Analytics

Surprisingly, lots of people search like this:

Google Analytics

piano

Search!

While others search like this:

The Clash London Calling Guitar Tablature Medium Difficulty

Search!

Understanding your data

- ▼ **We have very diverse data**
 - ▼ Most product attributes are inconsistent
- ▼ **We receive product information in 136 different formats**
 - ▼ Product title
 - ▼ Contributor roles: Artist, Composer, Performer, Arranger, etc.
 - ▼ Musical Genres, instrumentation, difficulty level, format
 - ▼ Algorithmically applied facets based on all data elements
 - ▼ Some items have scores of facets, some have only a few

Analyzing your data

- ▼ How might we categorize all of our potential fields?
- ▼ What data elements might contain a person's name?
- ▼ What elements might refer to a musical instrument?
- ▼ What do we do about unstructured publisher descriptions?

- ▼ **How can you provide an effective search service with the data you have?**

A “one size fits all” search strategy may not be ideal.

piano

Search!

The Clash London Calling Guitar Tablature Medium Difficulty

Search!

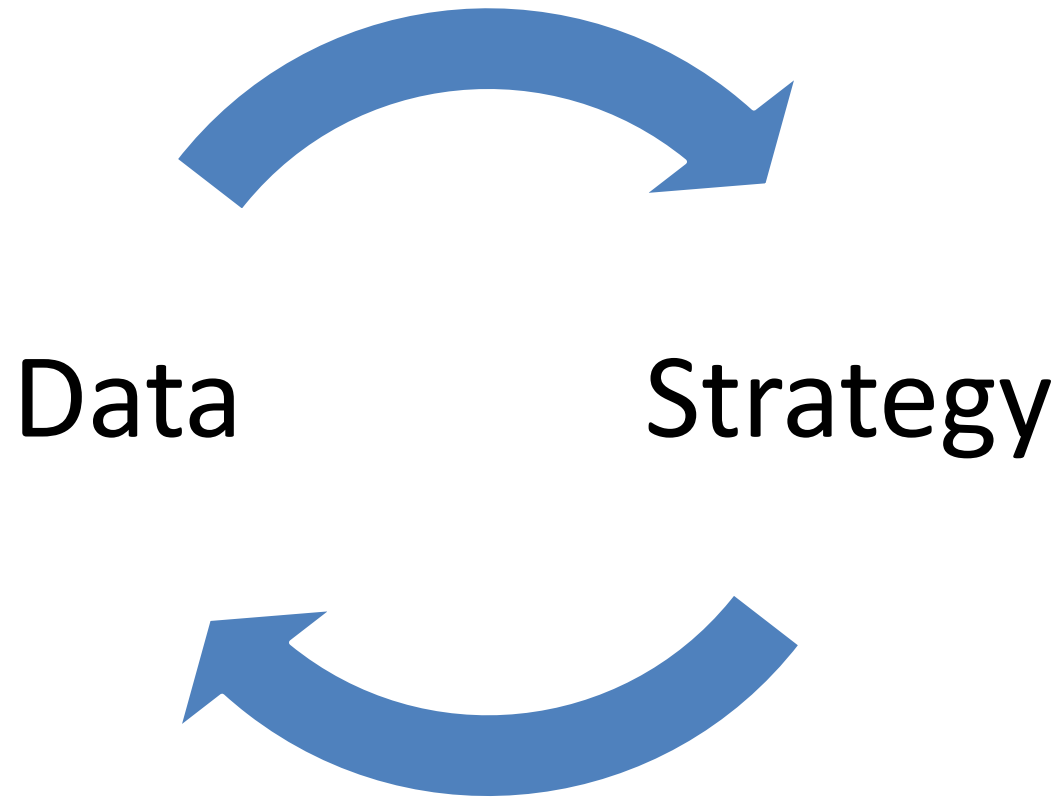
How we set up our search strategies

- ▼ **Each strategy must answer the following:**
 - ▼ A unique name and description to help with analytics
 - ▼ A hash of solr fields to search, containing:
 - The copy field name
 - The type of search (phrase or term searching)
 - The boost value
 - The slop value
 - ▼ The default sort order of that particular strategy

Grouping your data

- ▼ There are significant performance benefits in minimizing the number of fields to search against
- ▼ Group the fields that share common traits and have a similar level of influence on your search results to boost performance
 - ▼ Example:
“Contributors” are all equally important to our visitors. Many searches include an artist name, but just as often those names are composers, arrangers or popular performers. We can create a single copy field that aggregates all of those fields into one searchable Solr field

Step 2



Building a search strategy

- ▼ What (copy) fields do you search against?
- ▼ What boost is applied to each of those fields?
- ▼ What sort order do you default to?
- ▼ What other factors might you influence your results with?
 - ▼ (We occasionally influence results with sales ranking data)

- ▼ **Having more than one strategy is much easier than trying to get it exactly right**
- ▼ **It'll be unique to your site**
- ▼ **We have a classifier algorithm to help determine which strategy we should apply to an incoming search request**

- ▼ Implementing these strategies provided an **86% increase in per-search value.**
- ▼ We can now fine-tune individual categories of searches, as well as specific phrases.

Expanding your portfolio of search strategies

- ▼ How are customers searching your data?
- ▼ What categories of searches might you need to address?
- ▼ Sheetmusicplus.com examples:
 - ▼ **Artist search** {Pink Floyd, Haydn, Louis Armstrong}
 - ▼ **Instrument search** {Guitar, ukulele, oboe}
 - ▼ **Genre search** {Jazz, classical, standards}
 - ▼ **Title search** {"I left my heart in San Fernando's Hideaway"}
- ▼ Evaluate: How does this relate to metadata?

Monitor and refine your strategies

- ▼ Use Google Analytics (or a similar tool) to understand what your users are searching for
- ▼ Deliver search strategies to help them find what they want
- ▼ Don't bother chasing obscure queries that don't generate revenue
- ▼ Focus on delivering customer value
- ▼ Win!

Track customer value per strategy

- ▼ **Use Google Analytics to track the value of your strategies**
 - ▼ Add a dummy URL parameter to each of your search results indicating which strategy was used during that search
[/some/resulting/doc?s=4](#)
 - ▼ In Google Analytics, search your page views for 's=4'
 - Graph of page views over time, eCommerce value per page, etc

Summary

- ▼ Data --> Strategy
- ▼ Tuning your search algorithm is an endless game, and if you focus on pleasing customer X, it'll cost you.
- ▼ Don't try to please everybody. Please the people who make you money!

Q&A

Go to <http://bit.ly/solr-ecom> to download slides;
Full replay available within 48 hours of live webcast