

# SAN FRANCISCO MAY 25-26 2011

---

## Using **LWE/Solr/Lucene** for eCom

Grant Ingersoll, Lucid Imagination  
@gsingers



**LUCENE**  
**REVOLUTION**

Presented by  
**lucid**  
IMAGINATION

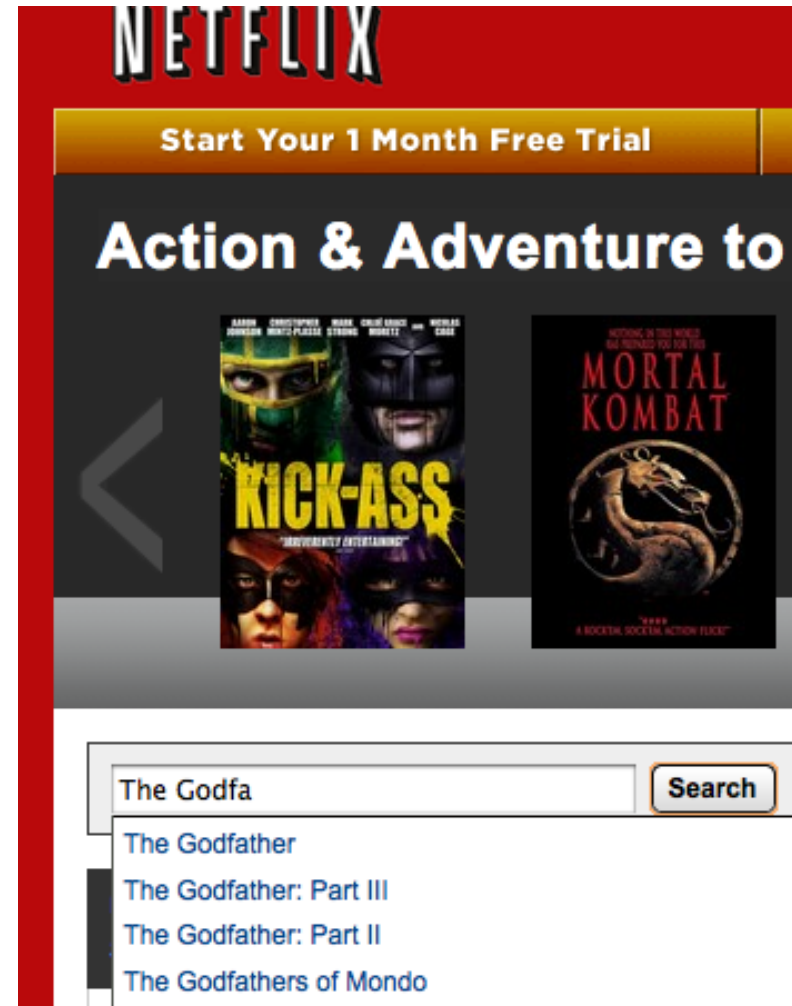
Apache  
**Solr** 

# Difference Makers

- Case Study 1: Relevance Matters
  - **Large Electronics Manufacturer**
  - **Top selling product on page 10 for a search by product name**
- Case Study 2: Don't Overthink it
  - **Large Online Retailer**
  - **Simply adding auto-suggest added millions to bottom line at very little cost**
- Case Study 3: Test, Test, Test
  - **Amazon Recommendation System**
  - **<http://glinden.blogspot.com/2006/04/early-amazon-shopping-cart.html>**

# Topics

- The Stack
- Knowing Users
- Search Features in LWE/  
Solr/Lucene
  - Minimum Features for  
eCom
  - Extended Features
- Not Just Search
- What's Missing?
- What's Next?



NETFLIX

Start Your 1 Month Free Trial

Action & Adventure to

KICK-ASS

MORTAL KOMBAT

The Godfa

Search

The Godfather

The Godfather: Part III

The Godfather: Part II

The Godfathers of Mondo

# eCom Stack Choices



## Lucid Works Enterprise

- Power of Apache Solr
- Easy Admin, Click Scoring, Reports



## Apache Solr

- Faceting, Easy Config
- Scale



## Apache Lucene

- Powerful Search Library
- Total Flexibility

# Users: Get to Know Them!

- Audience Poll:
  - How many of you are developers?
  - How many of the developers know what the top 10 queries are on your site?
  - How many of the non-developers know?
- Your users represent 100% of your opportunity to sell your products ;-)
  - **Shouldn't you know what they are searching for?**

# Search Analytics

- “If you can’t measure it, you can’t manage it”
  - **Attributed to Peter Drucker, however, see \***
- Ultimately, it’s all about conversion
  - **May not be the best measure for judging search**
- Is there One Right Answer on your Site or Multiple?
  - **Known Item search vs Keyword/Category**

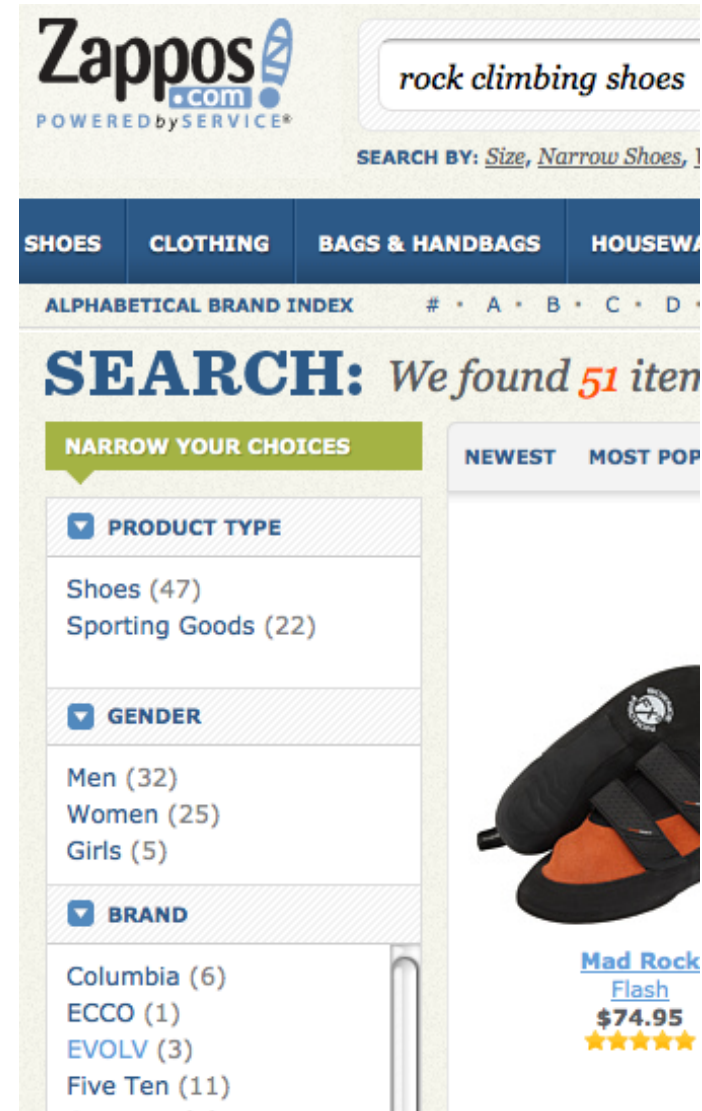
\*<http://edkless.com/2009/06/peter-drucker-and-time-sheets/>

# Useful Metrics

- **Mean Reciprocal Rank or Precision @ 10**
  - *Known Item vs. Keyword/Category*
- **“Show me the money” -- Top Product Analysis**
  - *Identity Search - If your top product is named X and someone searches for X, is X on the first page? Is it number 1?*
  - *Is a top product underperforming as it relates to search?*
- **Top X Queries and Query Terms**
- **Zero Results and % of Zero Results**
- **Avg. # of facets/filters/spellchecks clicked per session**
- **Avg # of searches per user session**
- **Auto-suggest usage**

# Minimum Search Features

- High Quality Relevance for keyword and known item search
  - **P@10 or MRR close to 1**
- Sub-second response time under load
- All achievable in LWE/Solr/Lucene



Zappos.com  
POWERED by SERVICE®

rock climbing shoes

SEARCH BY: [Size](#), [Narrow Shoes](#), 1

SHOES CLOTHING BAGS & HANDBAGS HOUSEWARE

ALPHABETICAL BRAND INDEX # · A · B · C · D ·

**SEARCH:** We found 51 items

NARROW YOUR CHOICES

NEWEST MOST POPULAR

PRODUCT TYPE

- Shoes (47)
- Sporting Goods (22)

GENDER

- Men (32)
- Women (25)
- Girls (5)

BRAND

- Columbia (6)
- ECCO (1)
- EVOLV (3)
- Five Ten (11)

**Mad Rock Flash**  
\$74.95  
★★★★★

# Faceting

## Pivot Facets

- LWE/Solr support faceting by:
  - **Field**
  - **Date/Number Ranges**
  - **Pivot (“what if” faceting)**
  - **Hierarchical**
    - *(via domain modeling)*
  - **Dynamic (via Carrot<sup>2</sup>)**
- Single and multi-select faceting supported
- Facet by Function In Development
  - <https://issues.apache.org/jira/browse/SOLR-1581>
- <http://wiki.apache.org/solr/SimpleFacetParameters>

[cat::electronics](#) (3)

[inStock::false](#) (2)

[inStock::true](#) (1)

[cat::connector](#) (2)

[inStock::false](#) (2)

[cat::music](#) (1)

[inStock::true](#) (1)

# More Features

- Extensible Language Analysis
  - **Multilingual Support**
  - **Synonyms**
  - **Overrides on a per-word basis**
  - **Pluggable Framework**
  
- Frequent/Incremental Updates
  - **How often do you update your index?**
  - **Near Real Time (`IndexReader.open()` )**
  - **Column Stride Fields (4.0)**

# Relevance Controls

- Function Queries
  - Ratings/Reviews
  - Margin/Inventory/Price/Location
  - Can Sort by Functions
  - `.../solr/browse?q=ipod&bf=price`
- Editorial Controls (QueryElevationComponent)
  - Fine grained controls
  - `.../solr/elevate?q=YYYY&enableElevation=true`
- Landing Pages (if done in search...)
  - Implement: Docs with field that is filtered on or a separate index/core
  - Editorial Controls
- Click Scoring (LWE only)
  - Popularity based ranking

# Beyond the Search Box

- Many eCom sites actually power all navigation by the search engine
- Many other tools in the Stack to help users discover content
  - **Auto Suggest**
  - **Spell Checking**
  - **More Like This**
  - **Spatial**

# Complementary Tools

- Apache Mahout
  - **Recommendation Systems**
    - *Crude Solr/Mahout Rec Integration at <https://github.com/gsingergs/ApacheCon2010>*
  - **Classifiers/Clustering**
    - *User Analysis, Content Analysis, etc.*
- Social
  - **BazaarVoice, etc.**
- Business Rules Engine
  - **Drools or others**

# What's Missing?

- UI Controls for non-devs:
  - **Synonyms (LWE has UI/REST support)**
  - **Facets (Field support in LWE)**
  - **Relevance Control (LWE REST API Support)**
  - **Business Rules Integration**
  - **Deeper Taxonomy Support**
  - **More performance reports (LWE has some)**
- Facet Management tools
  - **Labels**
  - **Sort order other than Count or Alphabetical**
  - **Editorial facet control**

# What's Next?

- Some sample code and more discussion at <http://www.lucidimagination.com/blog/2011/01/25/implementing-the-ecommerce-checklist-with-apache-solr-and-lucidworks/>

# Resources

- Principles for Effective Search in E-Commerce Design
  - <http://lucene.li/2T>
  - [grant@lucidimagination.com](mailto:grant@lucidimagination.com)
  - @gsingers
- <http://www.lucidimagination.com/search/?q=ecommerce>

<http://www.lucidimagination.com>