

Architecture of an Enterprise Search Application

Xavier Morera

@xmorera

www.searchtechnologies.com



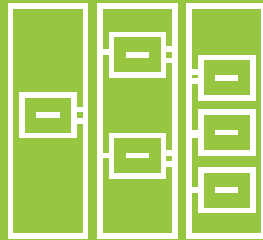
pluralsight 
hardcore dev and IT training

Architecture

A white rounded rectangle containing the word "Find" in a bold, green, sans-serif font.

Find

Beauty &
Simplicity



Architecture

The Lucene logo, featuring a stylized white bird-like icon to the left of the word "Lucene" in a white, italicized, sans-serif font.

Lucene
The Engine



Search
Applications

**Do Not Place Anything
in This Space**

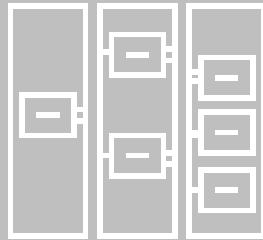
(Add watermark during
editing)

Note: Warning will not appear
during Slide Show view.

Architecture : Beauty & Simplicity

Find

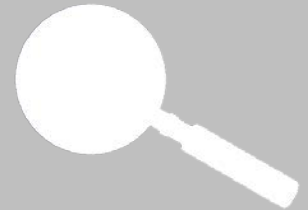
Beauty &
Simplicity



Architecture

Lucene

Lucene
The Engine



Search
Applications

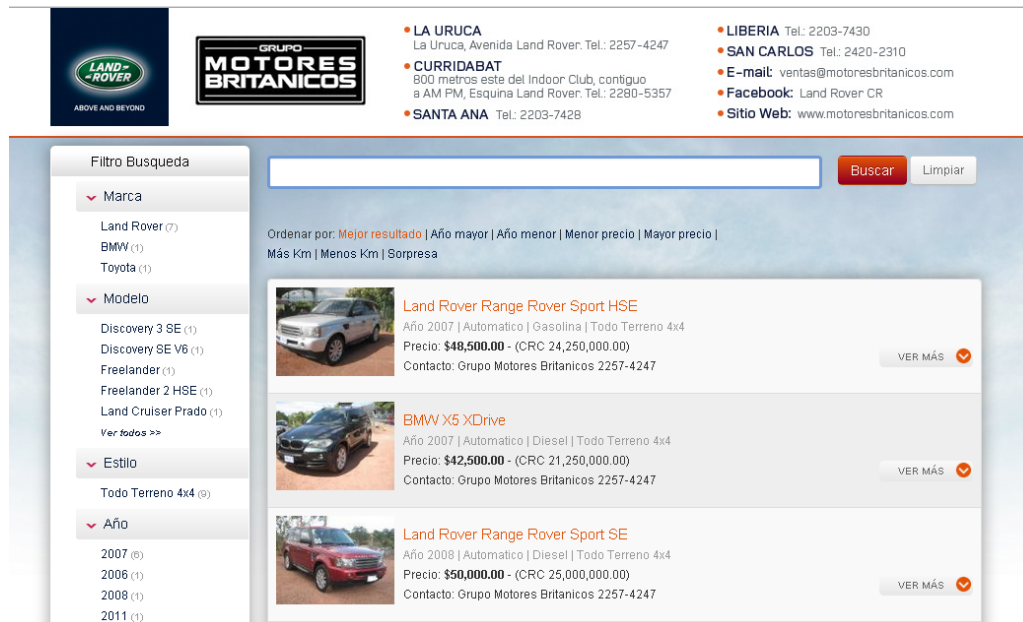
**Do Not Place Anything
in This Space**

(Add watermark during
editing)

Note: Warning will not appear
during Slide Show view.

Search Applications

- Beauty is in the eye of the beholder
- Users usually don't care about the application
- They care about getting accurate and fast results
 - They don't care about the "how"



**Do Not Place Anything
in This Space**

(Add watermark during
editing)

Note: Warning will not appear
during Slide Show view.

Search Application

- And users love simplicity and accuracy



Google Search

I'm Feeling Lucky

- And simplicity is even patented!



US00D599372S

(12) **United States Design Patent**
Pelczarski et al.

(10) Patent No.: **US D599,372 S**

(45) Date of Patent: **** Sep. 1, 2009**

(54) **GRAPHICAL USER INTERFACE FOR A
DISPLAY SCREEN OF A COMMUNICATIONS
TERMINAL**

D419,972	S	*	2/2000	Johnson et al.	D14/486
D420,993	S	*	2/2000	Decker	D14/486
D436,580	S	*	1/2001	Navano et al.	D14/486

- But simplicity is complex!

**Do Not Place Anything
in This Space**

(Add watermark during
editing)

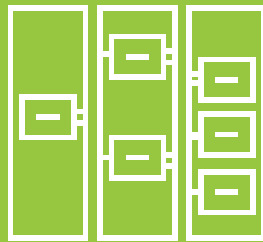
Note: Warning will not appear
during Slide Show view.

Architecture

A white rounded rectangle containing the word "Find" in a bold, sans-serif font.

Find

Beauty &
Simplicity



Architecture

The Lucene logo, featuring a stylized white bird-like icon to the left of the word "Lucene" in a white, italicized, sans-serif font.

Lucene

Lucene
The Engine



Search
Applications

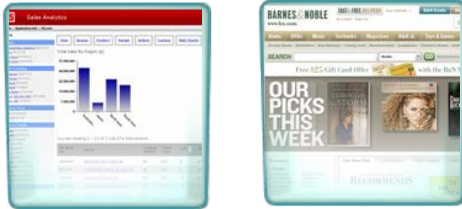
**Do Not Place Anything
in This Space**

(Add watermark during
editing)

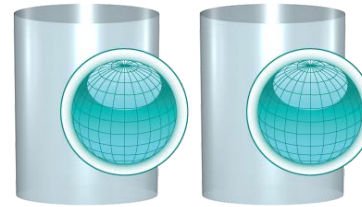
Note: Warning will not appear
during Slide Show view.

Search – 10,000ft view

Applications



Services (Machine readable)



Application Layer

Search Application API's



Search Engine
Search Features

Content
Metadata
Security
Indexes

**Tools
Suite**

Document Processing Pipelines (DPMS)

Connectors (Additions, Modifications, Deletions)



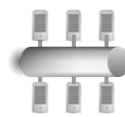
Databases



WWW
Internet



File Systems



SOA, ESB,
Web Service



Groupware and
Collaboration



SAP



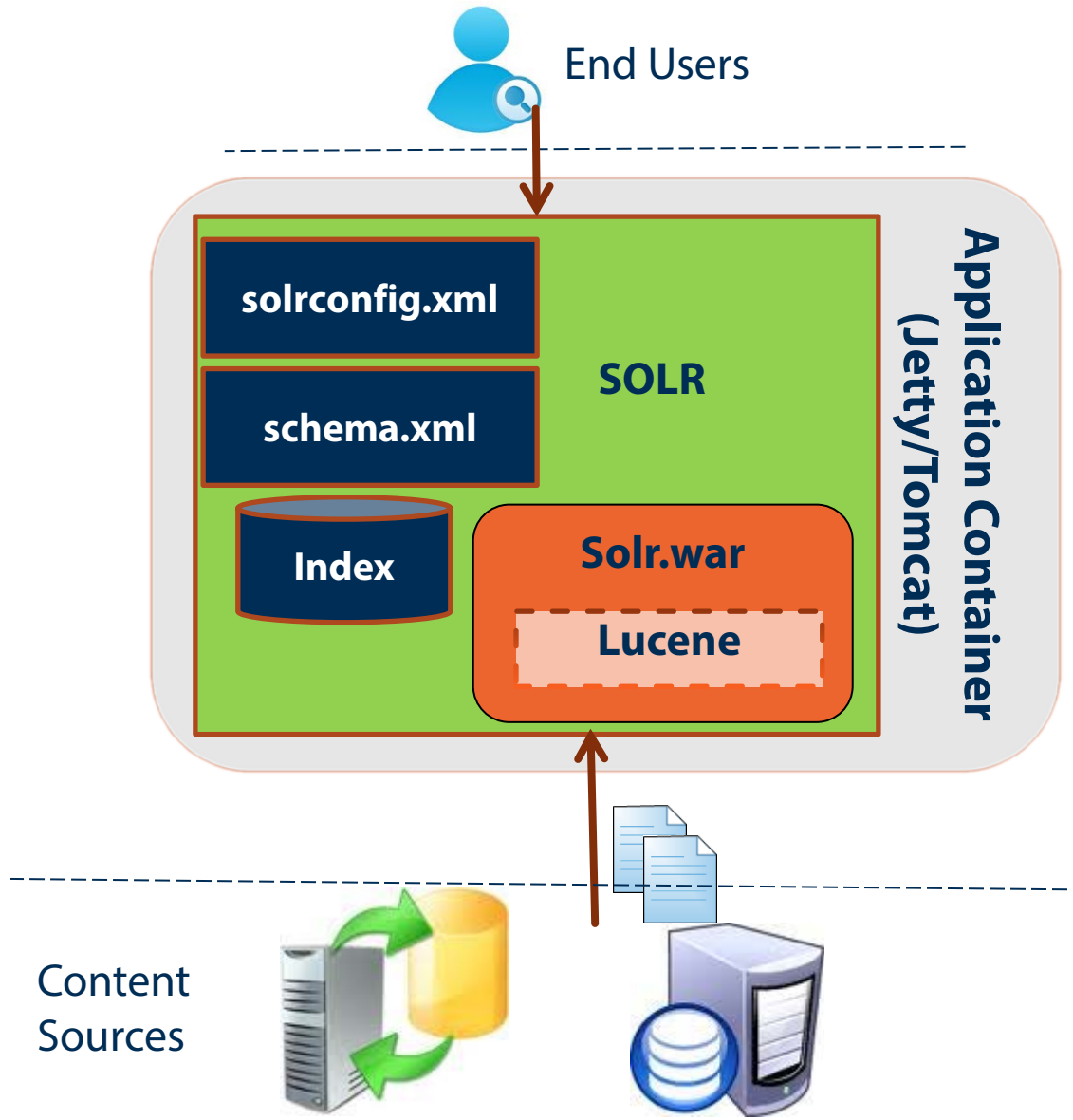
Content
Management

**Do Not Place Anything
in This Space**

(Add watermark during
editing)

Note: Warning will not appear
during Slide Show view.

SOLR Architecture Overview

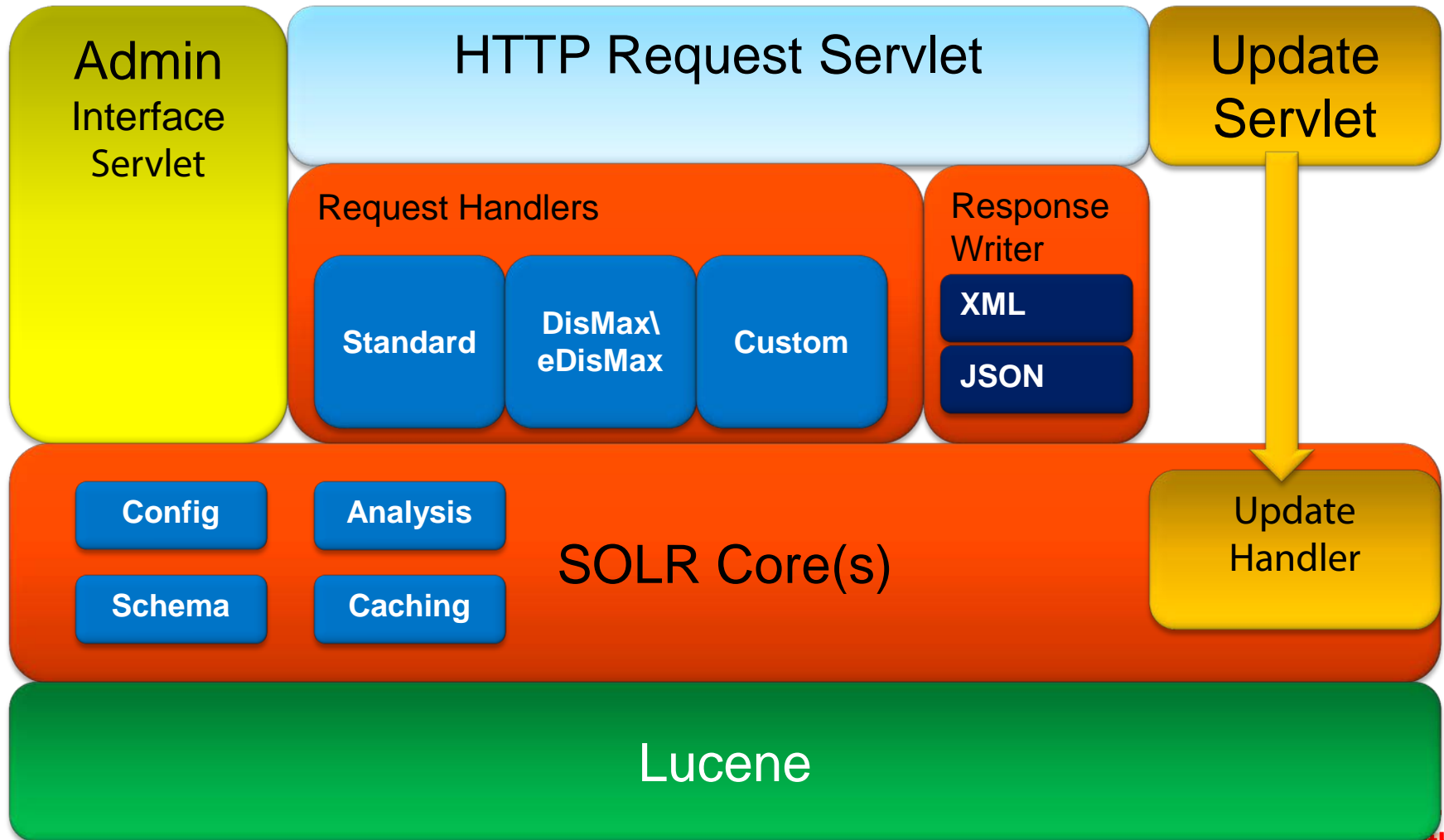


Do Not Place Anything in This Space

(Add watermark during editing)

Note: Warning will not appear during Slide Show view.

SOLR Architecture Overview

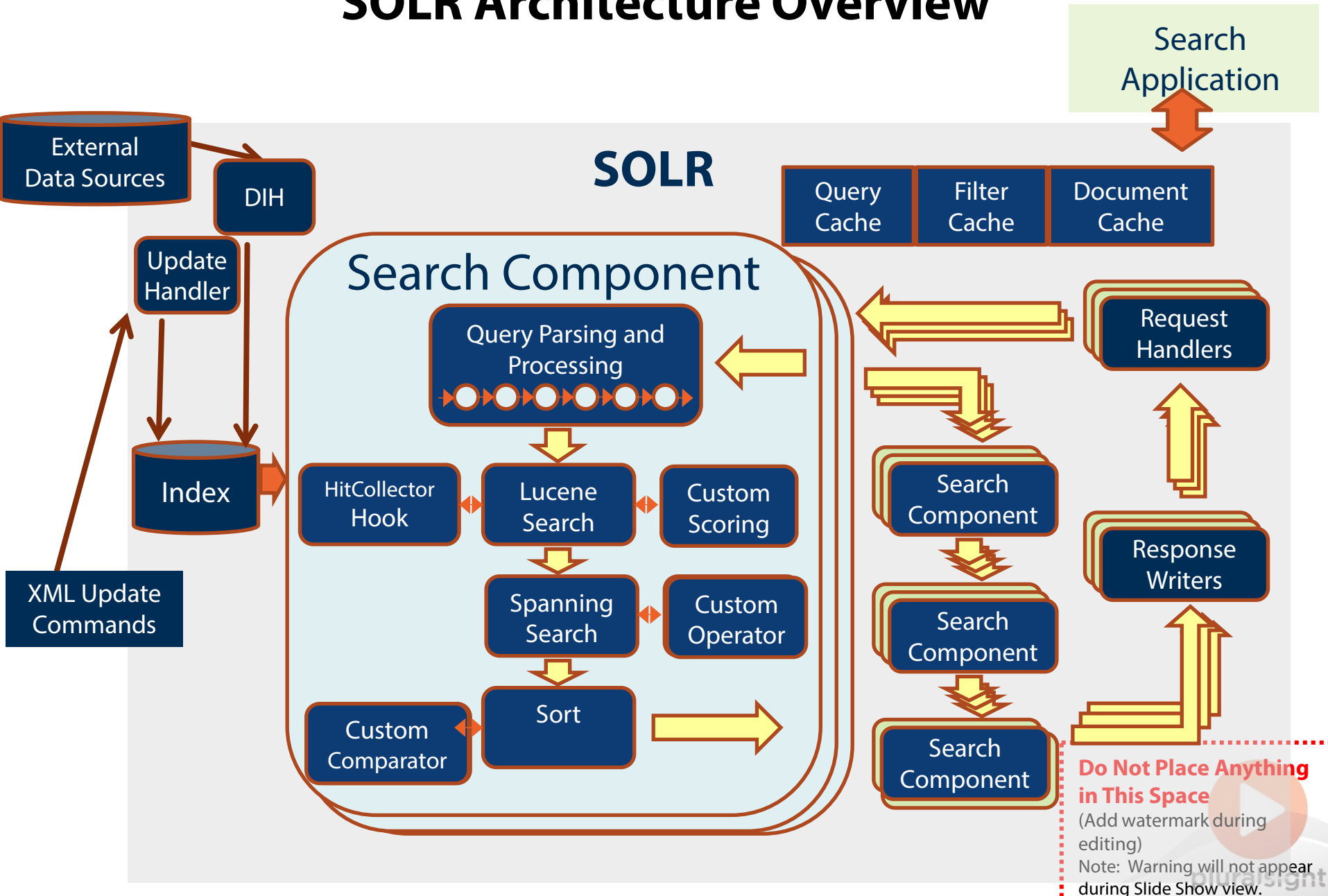


**Do Not Place Anything
in This Space**

(Add watermark during
editing)

Note: Warning will not appear
during Slide Show view.

SOLR Architecture Overview

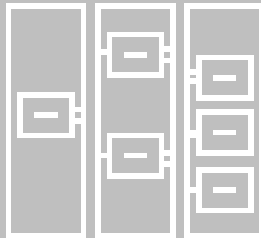


Architecture: Lucene



Find

Beauty &
Simplicity

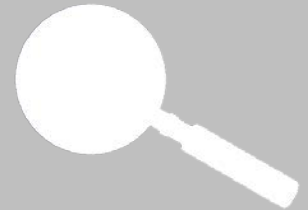


Architecture



Lucene

Lucene
The Engine



Search
Applications

**Do Not Place Anything
in This Space**

(Add watermark during
editing)

Note: Warning will not appear
during Slide Show view.

Solr & Lucene

□ Solr

- “Solr is serverization of Lucene”
- Open-source enterprise search platform
- Flexible and customizable
- RESTful
- Multiple search features
- Supports replication and shards



■ Lucene

- open-source search and indexing engine
- can work with any document which has fields that can be extracted and indexed
- It has a specific syntax



**Do Not Place Anything
in This Space**

(Add watermark during
editing)

Note: Warning will not appear
during Slide Show view.

What's Lucene

- Open code API for information Retrieval
- Originally developed in Java
 - Heavily ported to other languages (C#, C++, Python, etc)
- Main purpose: Indexing and Search
- High performance, scalable



**Do Not Place Anything
in This Space**

(Add watermark during
editing)

Note: Warning will not appear
during Slide Show view.

What Lucene “Ain’t”

- An “end to end” solution
- An application
- Web crawler
- Parser/filter for “any existent document type”



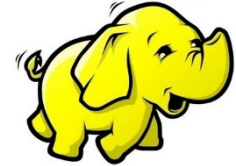
**Do Not Place Anything
in This Space**

(Add watermark during
editing)

Note: Warning will not appear
during Slide Show view.

Other Apache Projects

hadoop



- [Apache Hadoop](#): platform to create distributed systems

- [Apache Lucene.Net](#): Port of Lucene to MS .Net (C#)

- [Apache Lucy](#): Port of Lucene to C



- [Apache Mahout](#): Machine learning library



- [Apache Nutch](#): Web crawler



- [Apache OpenNLP](#): Machine learning library for processing natural language

- [Apache Tika](#): Parser to detect and extract meta data



**Do Not Place Anything
in This Space**

(Add watermark during
editing)

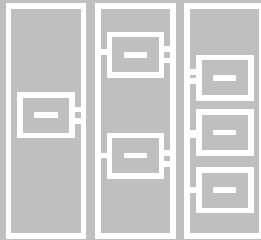
Note: Warning will not appear
during Slide Show view.

Architecture: Search Applications

A white rounded rectangle containing the word "Find" in a bold, sans-serif font.

Find

Beauty &
Simplicity



Architecture

The Lucene logo, featuring a stylized white bird-like icon to the left of the word "Lucene" in a white, italicized, sans-serif font.

Lucene
The Engine



Search
Applications

**Do Not Place Anything
in This Space**

(Add watermark during
editing)

Note: Warning will not appear
during Slide Show view.

User Point of View: Search Application

- Many different ways of creating search applications
- It all depends on user's requirements
- Usually Google comes to mind when talking about search
 - But that is only one (highly popular and effective) example
- Search is much more
- Let's look at a few aspects of searching
 - Some are personal or third party opinions
 - Not written in stone



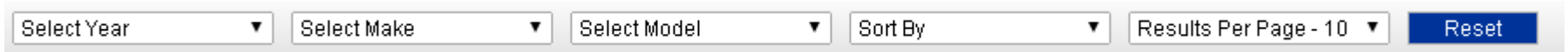
**Do Not Place Anything
in This Space**

(Add watermark during
editing)

Note: Warning will not appear
during Slide Show view.

Searching

- **Developers like one input box per field**



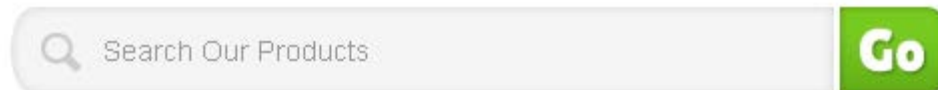
Select Year ▼ Select Make ▼ Select Model ▼ Sort By ▼ Results Per Page - 10 ▼ Reset

- This makes their life easier. But that is not the point. And repeat after me

*“I am here to make the user’s life easier, not mine.
Simplicity is the ultimate sophistication”*

- **When possible, the user should start with a single input box.**

- They should not make the heavy lifting
 - Users crave simplicity



Search Our Products Go

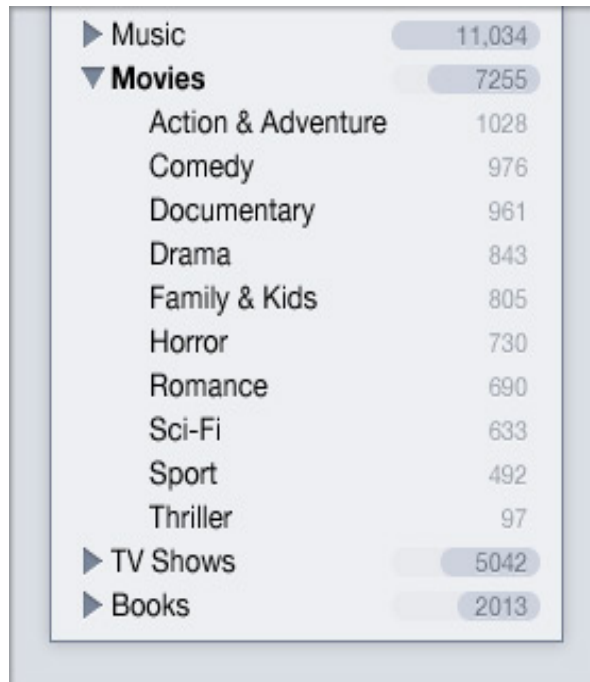
**Do Not Place Anything
in This Space**

(Add watermark during
editing)

Note: Warning will not appear
during Slide Show view.

Refiners/Navigators/Facets

- Allow drill down before or after the search
- Act as summaries and queries at the same time
- Big help to users



▶ Music	11,034
▼ Movies	7,255
Action & Adventure	1,028
Comedy	976
Documentary	961
Drama	843
Family & Kids	805
Horror	730
Romance	690
Sci-Fi	633
Sport	492
Thriller	97
▶ TV Shows	5,042
▶ Books	2,013



Guided Navigation

- Materials
 - Stainless (945)
 - Aluminum (540)
 - Brass (82)
 - More...
- Preferred Suppliers
 - Arnault (316)
 - Smithson (194)
 - Trinity (79)
 - More...
- Mean Time To Failure
 - 10 (674)
 - 15 (330)
 - 20 (213)
 - More...

Inner Diameter (mm)

0 16 32 48 64 96 128

Max Torque (Nm)

0 125 250 375 500 625 750

Screw (ISO 4762/12.9)

More ▼

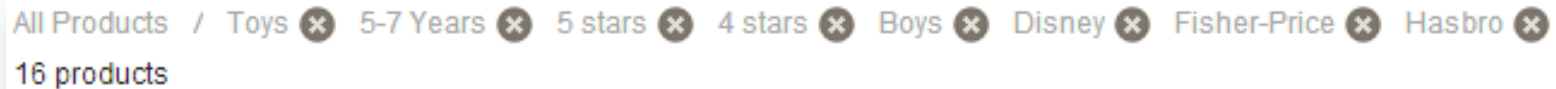
**Do Not Place Anything
in This Space**

(Add watermark during
editing)

Note: Warning will not appear
during Slide Show view.

Breadcrumbs

- Visual feedback to users on what they have selected



All Products / Toys ✕ 5-7 Years ✕ 5 stars ✕ 4 stars ✕ Boys ✕ Disney ✕ Fisher-Price ✕ Hasbro ✕
16 products

- Allow them to unselect a specific refinement
 - Expand result set when results not meet criteria
- Can contain both query and refinements

Showing results matching:

ladybug

Teacher Products ✕

Grade K / Ages 5-6 ✕

Bulletin Board Set ✕

**Do Not Place Anything
in This Space**

(Add watermark during
editing)

Note: Warning will not appear
during Slide Show view.

Other Aids

- And include other functionalities to help users

- Did you mean?

Showing results matching:

ladybgu

Did you mean? [ladybug](#)

- Synonyms
 - Spell checking
 - Autosuggestions
 - Lemmatize
 - Phonetic search
 - And more...



**Do Not Place Anything
in This Space**

(Add watermark during
editing)

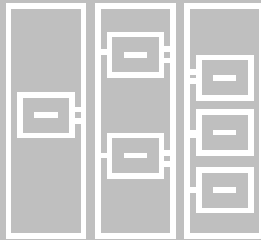
Note: Warning will not appear
during Slide Show view.

Architecture

A white rounded rectangle containing the word "Find" in a bold, sans-serif font.

Find

Beauty &
Simplicity



Architecture

The Lucene logo, featuring a stylized white bird-like icon followed by the word "Lucene" in a white, italicized, sans-serif font.

Lucene

Lucene
The Engine



Search
Applications

**Do Not Place Anything
in This Space**

(Add watermark during
editing)

Note: Warning will not appear
during Slide Show view.

Takeaway

- In Search make user's life easier, simplicity is king
- Architecture of Search applications is complex
 - But manageable with some effort
- Solr is the serverization of Lucene
- Lucene is the actual search engine under the hood
- Aim to help the user find their results and fast
- Input box with the aid of refiners/navigators/facets & breadcrumbs
- Additional aids include “did you mean”, synonyms, autosuggest, spell check, phonetic search, lemmatize and more



**Do Not Place Anything
in This Space**

(Add watermark during
editing)

Note: Warning will not appear
during Slide Show view.