

Introduction To Solr 4.x

Varun Thacker
Unbxd
@varunthacker



Solr Highlights

- Search Server based on Apache Lucene
- Provides REST-like API's
- Support various input and result formats (XML, JSON, CSV or binary results)
- DIH (DataImportHandler) - Imports data from your SQL store into Solr.
- MLT (More Like This)
 - Helps find similar documents
 - Can be used to detect plagiarism

Solr Highlights

- Faceted Searching
- Highlighted context snippets
- Lots of Analyzers
 - Support for over 30 languages.
 - <http://wiki.apache.org/solr/LanguageAnalysis>
- Really good Admin Panel

Extended DisMax Query Parser

- “Supports phrases”
- +requiredTerms
- -prohibitedTerms
- loose Terms

Extended DisMax Query Parser

- Query Fields (qf)
 - Matches each term against all qf fields
 - Each field can be given an individual boost
 - The score is associated with the highest field score and not the sum of the scores.
- Phrase Fields (pf)
 - Matches the entire query
 - Option to specify an optional slop factor

Demo Improvements

- Better search results for queries like c++ and .Net etc.
 - Define a custom Analyzer?
 - Define synonyms to deal with it
- Boost results based on Views, CreationDate, Score,
 - Use boost functions provided by eDismax
- And lots more!