Understanding Search

Xavier Morera
PASSIONATE ABOUT TEACHING

@xmorera www.xaviermorera.com

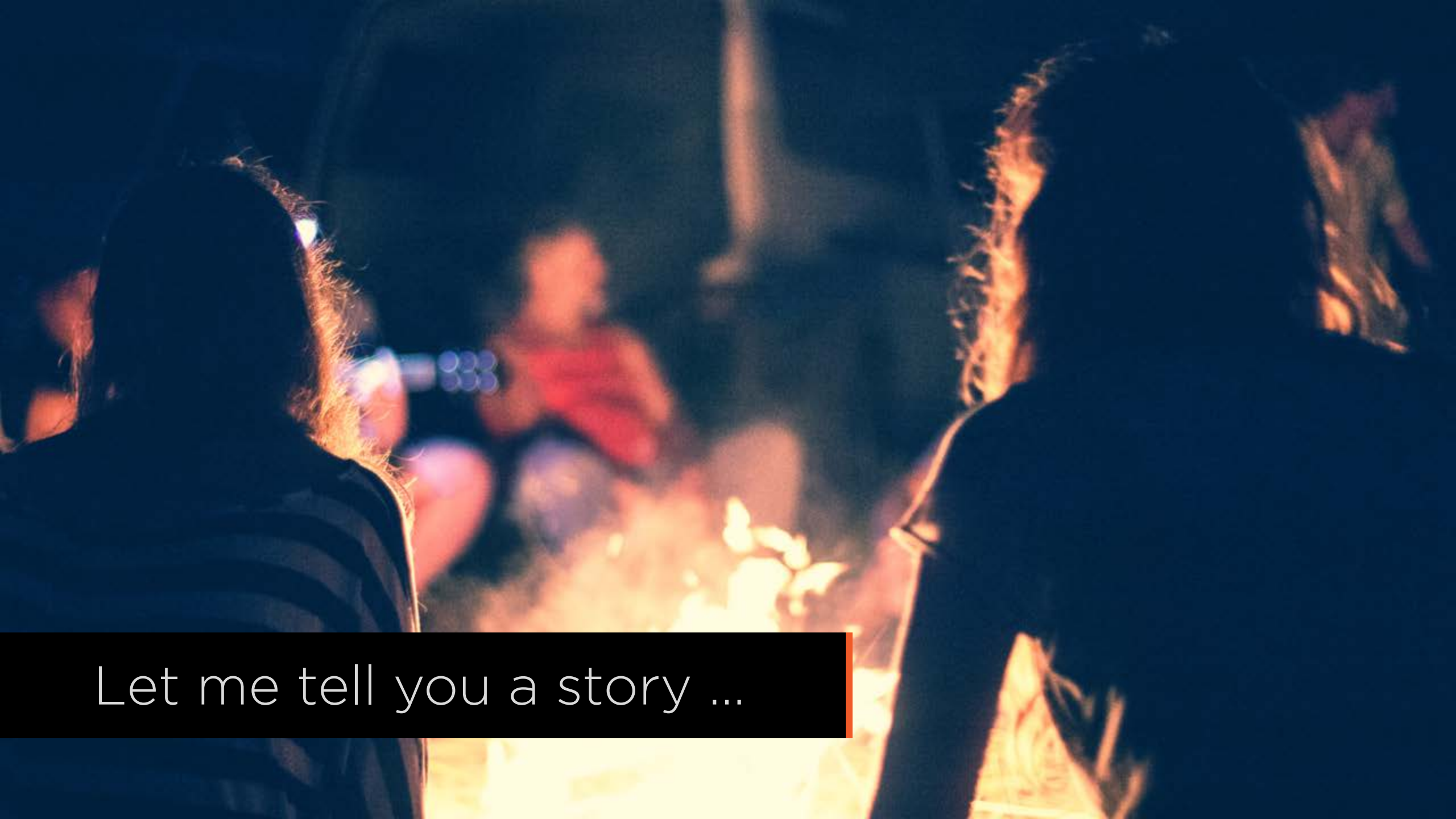
Act of Searching Exists for Many Years

Libraries

Books

Searching

- A problem humanity has found different ways to solve, usually manually
- Proliferation of digital media has changed the game (vs. printed media)
- Imagine looking for a file in your computer's drive without search?
  - Small and rather "limited" scenario
  - Finding the web page you need without a search engine?
- **Search is an absolute must**
- But a lot of people take it for granted


Let me tell you a story ...

Information Retrieval

Field that deals with representation, storage, organization and access to information in multiple different formats in such a way that it can then provide users easy access to the information of their interest

“I need to find the Pluralsight courses that will teach me the skills I need for my job”

It all starts with what a user needs

Problem Of/With Search

- Want to know about Learning Solr
  - Database thinking would query LIKE '%Learning Solr%'
  - Course called Learn Solr
    - You would not find it
  - Or LIKE '%Learning%' AND LIKE '%Solr%'
    - Hmmm... No
    - Instead LIKE '%Learning%' OR LIKE '%Solr%'
    - A bit better...
    - Might bring back results that are not totally accurate
      - i.e. Learning Git

Problem Of/With Search

- What if user types Learning the Solr
  - Sounds funny, but it is possible
  - Should ignore the (stop word)
- Want to know about Learning Solr
  - And Learning should match with Learn
- Or someone came to search for Learning Sowlar?
  - Pretty common for non native speakers
- If user searched for solr?
  - Case sensitive can affect results

Problem Of/With Search

- And many other scenarios that search engines need to cater for
  - Plenty of features to optimize search results
- Results returned very fast!
  - Possible because of inverted index
- Search engine must return results with the most relevant first, relevancy
- Returning the documents the user was looking for, known as precision
- Without leaving matching documents out of the results, known as recall

Getting Started with Enterprise Search Using Apache Solr

by Xavier Morera

Search is one of the most misunderstood functionalities in the IT industry. Apache Solr brings high quality Enterprise Search to the masses.

▶ Resume Course

Table of contents | Description | Transcript | Exercise files | Discussion | Learning Check | Recommended

Expand all

Course author

Xavier Morera

Xavier is an entrepreneur, project manager, technical author, trainer, Certified Scrum Professional & Scrum Master, and Certified Microsoft Professional who has spent a great deal of his career wor...

Course info

Level	Intermediate
Rating	★★★★★ (245)
My rating	★★★★★
Duration	3h 32m
Released	23 Jun 2014

Bookmark

Add to playlist

Share course

Functions of a Search Engine

Indexing

Search

Browse & Storage

Query Parsing

Relevance Ranking

Functions of a Search Engine

- Never Suggest Dead Ends
- Facets
- Corrections
- Query Suggestions
- Snippets & Highlighting

The Six Commandments of Search Implementation

Paul Nelson

Chief Architect at Search Technologies

I was recently asked to do a presentation on "Search Best Practices". Naturally, this is a subject that I think about all the time. What makes a good search system? What works? What doesn't?

Personally, I don't care much for the phrase "Best Practices". It's a squishy phrase which has become overused - often to justify what is nothing more than an opinion.

So let's step it up. What follows are "Search Commandments". These aren't just opinions. These are search system design rules which have proven themselves over and over again in countless search applications.

Commandment #1: Thou Shall Not Join

Search is fast. Millions of records can be searched in a fraction of a second. How is this possible? Because there are no

From the Chief Architect

How Are We Building Smarter Search Engines in the Big Data Age?

Engine Scoring and Predictive Analytics for Search Accuracy

Search Engine Commandments

1. Thou shall not join
2. Treat thy search engine not as thy RDMS
3. Reserve in thy search engine only what is most needed for thy search results
4. Be ye not careless as to what is a document
5. Let there be but a single purpose for each of thy search engine fields
6. Do not be desirous too much of thy engine

Architecture ft. Search – 10,000ft view

Applications

Services (Machine readable)

Application Layer

Search App API's

Search Engine

Search Features

Content | Security

Metadata | Indexes

Document Processing Pipelines (DPMS)

Connectors (Additions, Modifications, Deletions)

Tools Suite

Databases

WWW
Internet

File Systems

SOA, ESB
Web Service

Groupware and
Collaboration

SAP

Content
Management

As .NET Developers How Do We Implement Search?

Apache Solr

Great open source search engine

Mature and proven track record

Large and vibrant community

Learning resources

SolrNet

Solr from the comfort of .NET

Namely C# in this course

Multiple reasons that I will cover soon, mainly around efficiency and usability

Production proven

SolrNet is an Apache Solr client for .NET

Why Solr?

CEO

SysAdmin

Architect

Developer

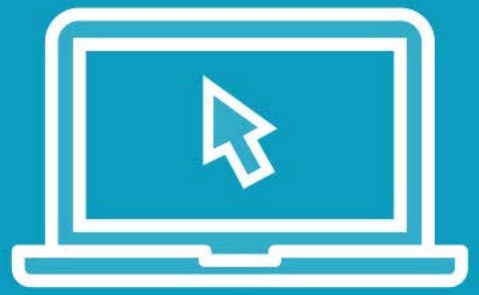
Search Engine: Apache Solr

- Advanced Full-Text Search Capabilities
- Five fundamental operations: query, index, delete, commit, and optimize
- Optimized for High Volume Web Traffic
- Standards Based Open Interfaces - XML, JSON and HTTP
- Comprehensive HTML Administration Interface
- Server statistics exposed over JMX for monitoring
- Near Real-time indexing
- Flexible and Adaptable with XML configuration and via API

Search Engine: Apache Solr

- Extensible Plugin Architecture
- A Real Data Schema, with Numeric Types, Dynamic Fields, Unique Keys
- Faceted Search and Filtering
- Highly Configurable and User Extensible Caching
- Performance
- Multiple search indices
- Scalable (SolrCloud)
- And more...

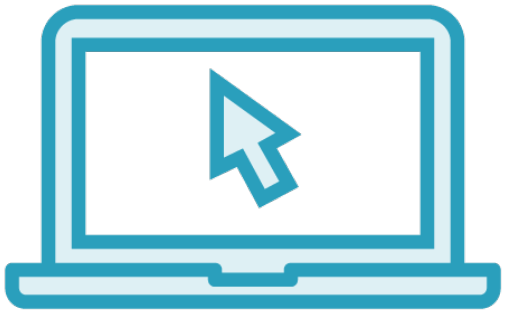
Demo



Simple Search Demo Using RESTSharp

M2 D1

Demo Summary



Solr exposes a REST API

Easy to use

REST means Response

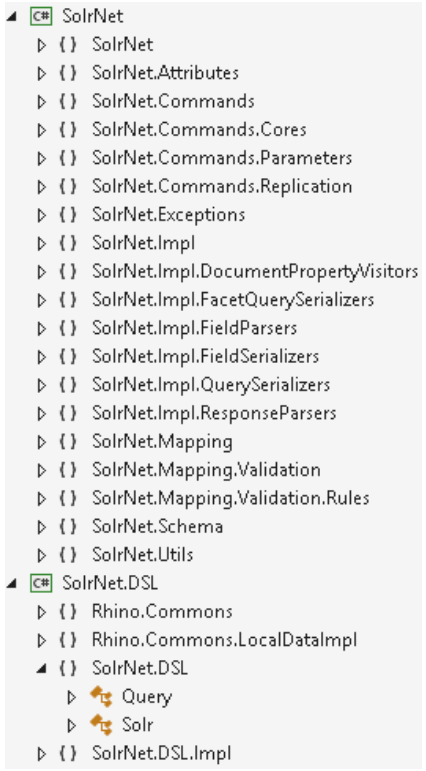
- JSON, XML, ...

Parse it and Understand it

Errors can occur

There has to be a better way...

Why SolrNet?



SolrNet abstracts Solr

Very complete functionality (Guides you)

Objects and classes

- POCOs, Functions, Intellisense, ...

Instead of

- Creating HTTP requests
- Parsing JSON/XML responses

Developers want out of the syntax business!

Equivalent Operations

More bang for your buck

Operation	Solr	SolrNet
Index	/update	.Add()
Binary index	/extract	.Extract()
Query	/select	.Query()

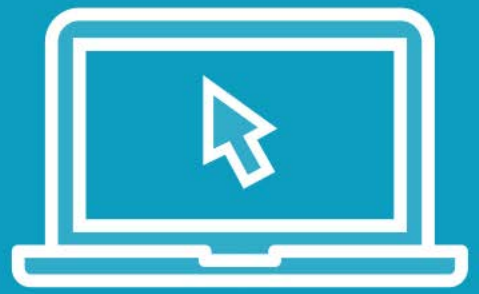
Plenty of SolrNet Functionality

AbstractSolrQuery.cs
AddParameters.cs
ClusterResults.cs
CollapsedDocument.cs
CollapseExpandResults.cs
CollapseResults.cs
DateFacetingResult.cs
ExtractField.cs
ExtractFormat.cs
. . .
The list goes on!

For a full list open SolrNet.sln!

SolrNet: Don’t Reinvent the Wheel

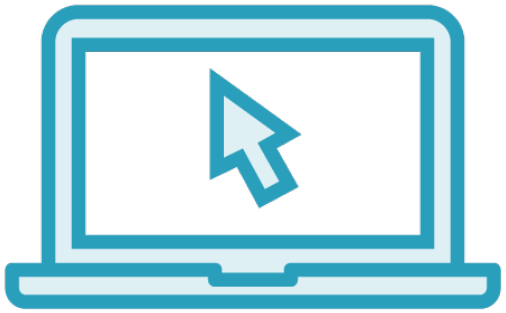
Demo



Simple Search Demo

M2 D2

Demo Summary



Using the REST API works

Lots of plumbing code

Instead, use SolrNet

SolrNet is succinct and efficient

It guides you

And makes the process easier and enjoyable

Takeaway

Need to search exists for many years

It all starts with a user need

Solved in multiple ways, usually manually

Proliferation of digital media

Hard problem to solve

Implement search using Solr & SolrNet

Great choice for CEOs, SysAdmins, Architects and Developers

Works via REST API

Much better with SolrNet