

SAN FRANCISCO MAY 25-26 2011

Building SaaS solutions with Apache Solr

Alberto Mijares, Canoo Engineering AG
alberto.mijares@canoo.com, 26/05/2011

Twitter: @lemaiol



LUCENE
REVOLUTION

Presented by

lucid
IMAGINATION

Apache
Solr 

Lucene

Bullet point time!

What I Will Cover

- Practical applications of Apache Solr and Apache Lucene: how to increase the time spent by a user in an website and do website “cross-selling”.
- Use case: how Canoo helped Axel Springer Switzerland to increased the page impressions, user permanence time and traffic in their financial online newspapers.
- Key concepts:
 - **How to achieve this using Lucene & Solr**
 - **How to profit from a SaaS business model**

Who I am

- Alberto Mijares
- Canoo Engineering AG
- Background in web applications and standards:
 - **Participated in W3C Semantic Web interest group (SWEO)**
 - **Led web standards compliance tools development in the past (Web Accessibility and Mobile Web)**
 - **Led enterprise information retrieval projects in the recent past**
 - **Actually coaching Google Web Toolkit projects' development**

Who is Canoo

- People:
 - **Dirk Koenig: Groovy founder**
 - **Andres Almiray: Griffon project lead and Java Champion**
 - **Hamlet D'Arcy: Groovy committer and enthusiast**
 - **... almost 40 more top software engineers**

- Products:
 - **WebTest: framework for web functional testing**
 - **RIA Suite (aka ULC): Java based RIA framework**
 - **FindIT: information retrieval and search tools**
 - **WMTrans: language analysis tools**

Canoo FindIT



<http://www.canoo.com/videos/FindIT.html>

Stop “bullet-pointing”!

The facts

Axel Springer group is a market leader

Bilanz, Handelszeitung and Stocks

In Switzerland financials are important!

Financial language is German

Online media is the future

The facts

Axel Springer group is a market leader

Bilanz, Handelszeitung and Stocks

In Switzerland financials are important!

Financial language is German

Online media is the future

The gap



Make the online versions more profitable

Make all newspapers “market leaders”

The gap



Make the online versions more profitable

Make all newspapers “market leaders”

The how



“Related articles”

“Cross-selling”

The how



“Related articles”

“Cross-selling”

The analysis

Use Lucene's "More like this"

Integrate back the suggestions

Implement a selection mechanism

Find a funding model

The analysis

Use Lucene's "More like this"

Integrate back the suggestions

Implement a selection mechanism

Find a funding model

The issues

“More like this” was “experimental”

Without “semantics” not always makes sense

Indexing full pages produces noise

Works out-of-the-box only in English

The issues

“More like this” was “experimental”

Without “semantics” not always makes sense

Indexing full pages produces noise

Works out-of-the-box only in English

The key

LEO Ein Online-Service der LEO GmbH 

   Englisch Forum Trainer	   Französisch Forum Trainer	  Spanisch Forum Trainer	  Italienisch Forum Trainer
--	---	--	---

  <>  
Erweiterte Suche
English version Ankündigungen 632.722 Einträge 7.888.324 Anfragen

canoonet

German Dictionaries and Grammar

[Dictionaries](#) | [Word Grammar](#) | [Word Formation](#) | [Sentence Grammar](#) | [Spelling \(D\)](#) | [Glossary](#) | [Index](#)

The key

LEO Ein Online-Service der LEO GmbH 

 Englisch Forum Trainer	 Französisch Forum Trainer	 Spanisch Forum Trainer	 Italienisch Forum Trainer
--	---	---	---

Erweiterte Suche 
English version Ankündigungen 632.722 Einträge 7.888.324 Anfragen Marketing & Sales

canoonet

German Dictionaries and Grammar

[Dictionaries](#) | [Word Grammar](#) | [Word Formation](#) | [Sentence Grammar](#) | [Spelling \(D\)](#) | [Glossary](#) | [Index](#)

The functional requirements



Discover and index articles

Extract only content

Simple and flexible query service

The functional requirements



Discover and index articles

Extract only content

Simple and flexible query service

The funding model



The business model



SaaS

The “other” requirements

Lucene-based analysis pipeline

Web oriented platform

Multi-application platform

Reliable, fast and scalable

Plan B?

The “other” requirements

Lucene-based analysis pipeline

Web oriented platform

Multi-application platform

Reliable, fast and scalable

Plan B?

The search

Wraps Lucene in a nice way

It is mature and Open Source

Supports scheduling, REST API, DIH,...

Scalability out-of-the-box

Well documented and has professional support

The search

Wraps Lucene in a nice way

It is mature and Open Source

Supports scheduling, REST API, DIH...

Scalability out-of-the-box

Well documented and has professional support

The plan



From POC to PROD in “80 days”

The plan



From POC to PROD in “80 days”

The results



Google analytics

The results



Google analytics

The conclusions



The Q&A



Thanks!

Sources

- Links
 - <http://people.canoo.com/share>
 - <http://www.canoo.com>
 - <http://www.canoo.net>
 - <http://www.leo.org>
 - <http://www.bilanz.ch>
 - <http://www.handelszeitung.ch>
 - <http://www.stocks.ch>

Contact

- Alberto Mijares
 - alberto.mijares@canoo.com
 - Twitter: [@lemaiol](https://twitter.com/lemaiol)

Architecture

Platform: **Apache Solr 1.4.1**

Architecture:

