OPENTEXT[™]

intentive



Apache Solr & OTWSM

Hintergründe. Technik. Möglichkeiten.



Inhalt

- Einführung Solr
- → Integration OTDS & Solr
- → Beispiele & Ideen





"Solr is the popular, blazing-fast, enterprise search platform built on Apache Lucene™. It runs in a Java servlet container such as Tomcat."





"It powers some of the most heavily-trafficked websites and applications and has been tuned to handle the world's largest sites."

AT & T Instagram Adobe
Disney Netflix SAP Hybris
MTV Networks Internet Archive Immonet
The Guardian Nasa CNET





offers everything but the kitchen sink.

- Spell-Checking
- Auto-Complete
- ▶ Typeahead-Search
- Language-Detection
- ▶ Facettierte Suchen
- ▶ Hit-Highlighting...

- Geobasierte Suche
- Multiple Such-Indizes
- Multilingual anpassbare Text-Analyse
- ▶ Plugin-Architektur
- **...**





is powerful and easy to use.

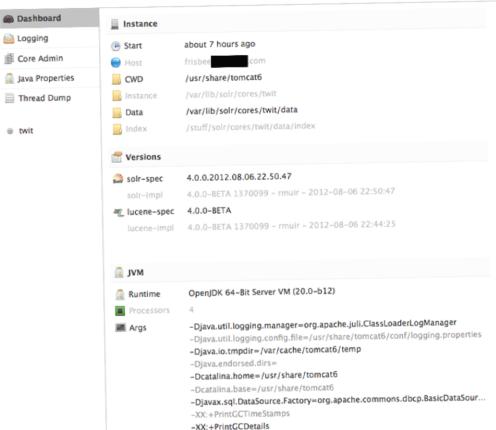
- ▶ Lucene[™]-basierte Volltextsuche
- Near-Real-Time-Indexing
- Umfangreiches Monitoring
- Standardschnittstellen auf Basis von XML, JSON & HTTP / REST

- ▶ Zentrale Cluster-Konfiguration
- Automatisches Loadbalancing & Failover
- Anpassbares Caching
- Umfangreiche Admin-Oberfläche









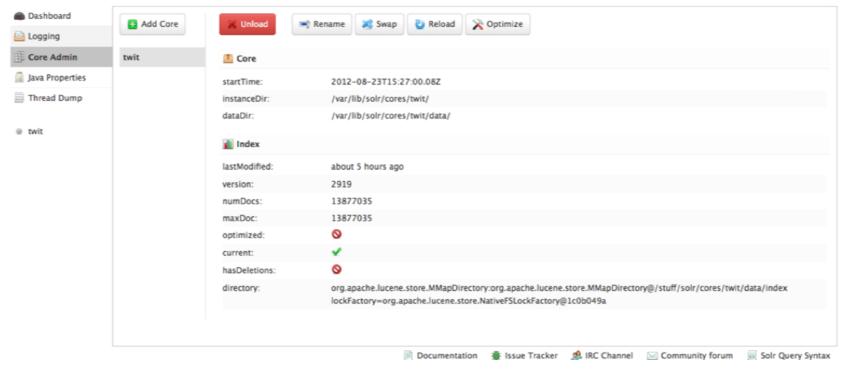
-verbose:gc



7





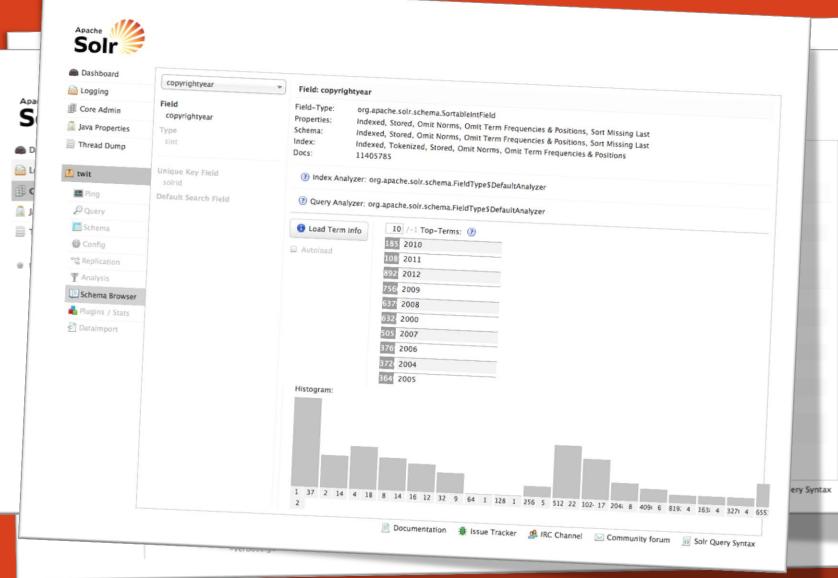


-XX:+PrintGCTimeStamps

-XX:+PrintGCDetails

-verbose:gc





a





Integration OpenText Delivery Server & Apache Solr







- NOT packaged with WSM
- No License required
- ▶ Since Delivery Server 10.1 SP 2 HF6





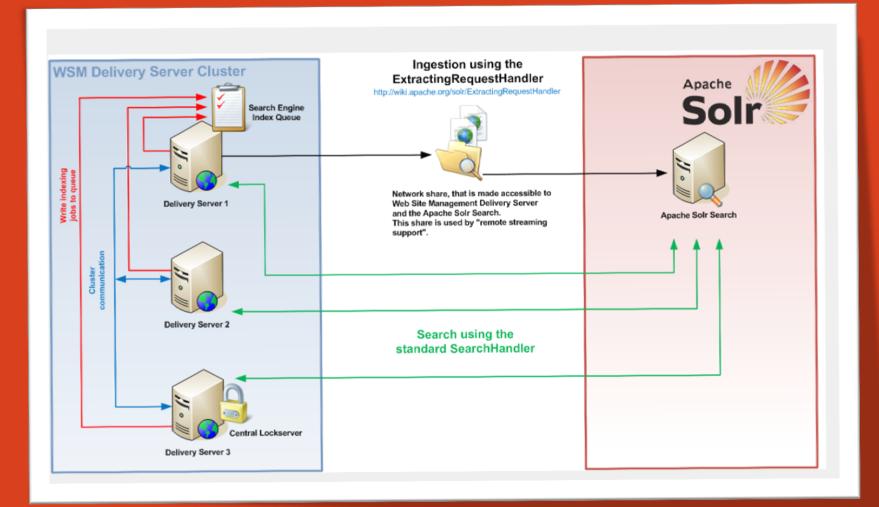
Documentation

- Configuration see Delivery Server Installation Guide
- Delivery Server Connector Guide, Delivery Server DynaMent Guide
- ▶ Solr Reference Guide: https://www.apache.org/dyn/closer.cgi/lucene/solr/ref-guide/





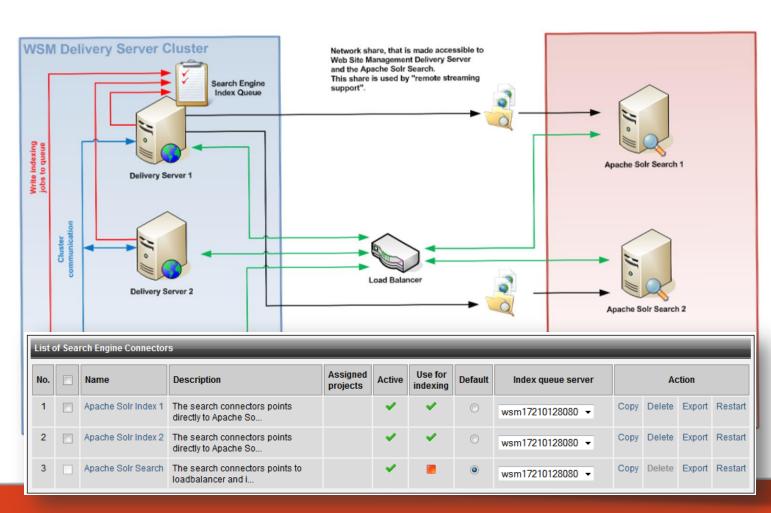
Architecture







Cluster deployment











- Search indexes are created based on default WSM Apache Solr Templates
 - Must be installed on Solr server
 - ▶ Shipped at ...\tools\apache-solr\...
 - ▶ Different Templates, Solr version 1.4, 3.x, 4.x, 4.8
- Templates for languages English, German, French, ..., undefine
- Default Template undefine





Configuration

- Stop Word List
 - "the", "and", "is" are Not indexed and Not searchable
 - ▶ Configuration file, e.g.: ..\default-en\stopwords.txt
- Thesaurus/Synonym
 - ▶ Configuration file, e.g.: ..\default-en\synonym.txt
 - ▶ Example: car => auto, automobile, machine, motorcar
- Content Metadata (Attribute)
 - > String, Date, Int, Long, Float, Double and Text
 - Dynamic fields definition (schema.xml): "*_str", "*_float", "*_int", ...





Supported Formats

- Apache Tika http://tika.apache.org/
 - Microsoft Office and Acrobat formats
 - ▶ Metadata, e.g: title, content_type



The Apache Tika™ toolkit detects and extracts metadata and text from over a thousand different file types (such as PPT, XLS, and PDF). All of these file types can be parsed through a single interface, making Tika useful for search engine indexing, content analysis, translation, and much more. You





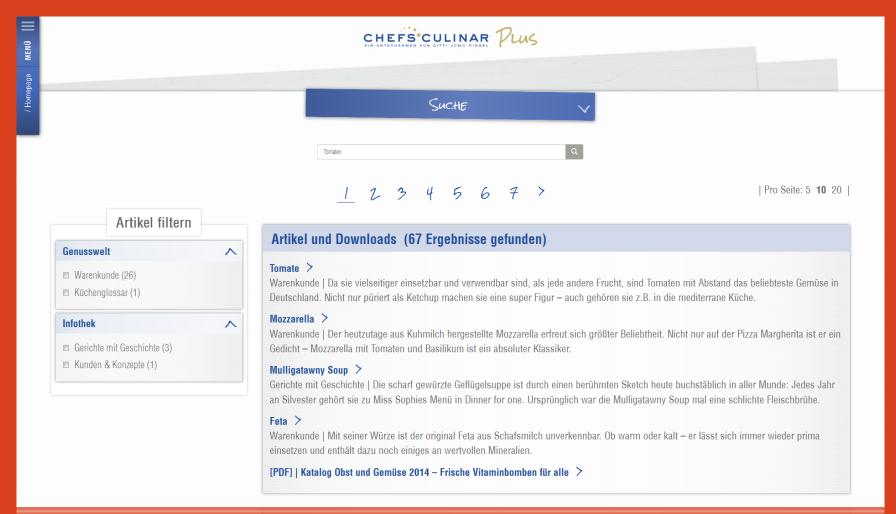


Beispiele & Ideen

Wie und wofür lässt sich Apache Solr in der Praxis nutzen?

Facet and Drill Down

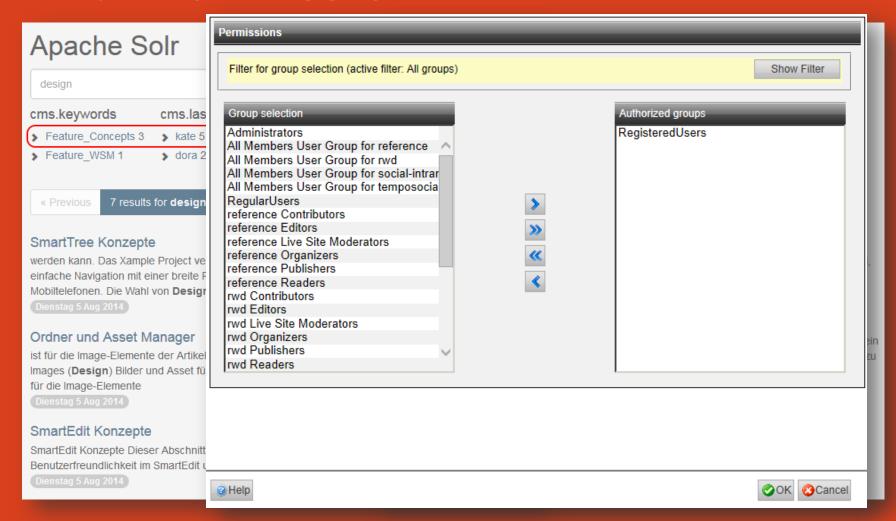






Access control (Constraint) with Permission

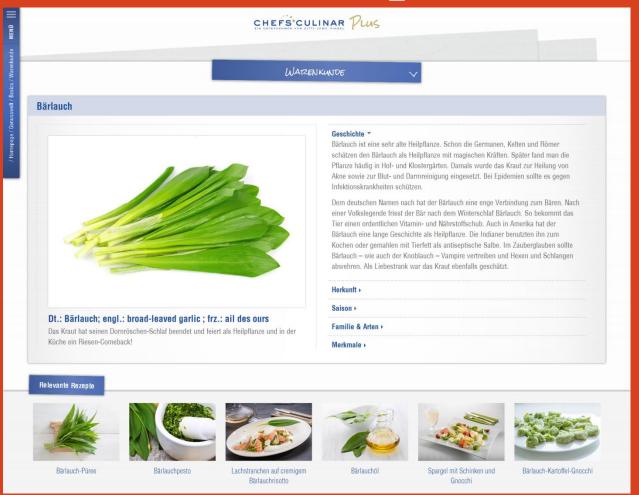






Kontextsensitive Suchteaser Beispiel 1







Kontextsensitive Suchteaser Beispiel 2



CHEFS	INAR PLUS	
	GRUNDREZEPT BATZLAUCH—PUTZEE *******(0) Koch: René Ludwig, CHEFS CULINAR Nährwerte (pro 100 g) Energie: 173 koal / 711 kJ Fett: 16,8 g Kohlentydrate: 4 g Eiweiß: 1,50 g Produktionsmenge: 0,5 kg	
Rezept für 0,5 kg		
Zubereitung & Zutaten Nährwerte & Gesundheit Warenkunde		
VOIZTEILE NUTZEN + Portionen ändern + passende Weine zum Rezept + Rezeptbuch anlegen Jetzt anmeld	en V	
Bärlauch-Püree		Alles aufklappen
Bärlauch		~
Petersilie		~
Knoblauch		~



Kontextsensitive Suchteaser Beispiel 3





14.04.2015

KÖSTLICHE SPATZGELZEIT

Endlich Frühling – und Genießer kommen wieder voll auf ihre Kosten!



Frühlingsluft, Sonnenschein – und überall Spargeldämme auf den Feldern: Für Genießer beginnt jetzt die schönste Zeit des Jahres. Die Spargelsaison ist eröffnet! Bis zum 24. Juni können Sie nun den Gaumen Ihrer Gäste mit dem Gemüse aus heimischen Regionen verwöhnen. Ob klassisch mit Schinken, zu Fisch, Lamm oder als Suppe – die gesunden Stangen sind immer ein Genuss.

Liebhaber sprechen, wenn sie Spargel meinen, nicht umsonst vom Königsgemüse: Es tut nämlich einiges für unsere Gesundheit. Die grünen Stangen punkten mit wenigen Kalorien, haben kaum Fett, wirken entwässernd und beeinflussen kaum den Blutzuckerspiegel, sodass auch Diabetiker gerne zugreifen können. Weißer Spargel bringt zudem wichtige



Spargelmousse mit Schnittlauch-Vinaigrette < >



Did you mean?



Apache Solr

helle

Search

Did you mean holle hölle holte halte hecke

No results for helle

```
chunk="[#request:chunk#1#]"
chunksize="[#request:chunksize#5#]"
highlight="yes"
item-tag="Content"
tag="search-results"
cachingtime="0"
metainfo="title;CIS TYPE;"
dateformat-out="EEEE d MMM yyyy, hh:mm"
include-mode="metainfo,meta;"
customresult="//lst[@name='spellcheck'];" >
<rde-dm:query-cis>[#request:q#]</rde-dm:query-cis>
    <rde-rd:param name="spellcheck">true</rde-rd:param>
    <rde-rd:param name="spellcheck.onlyMorePopular">true</rde-rd:param>
    <rde-rd:param name="spellcheck.dictionary">default</rde-rd:param>
    <rde-rd:param name="spellcheck.count">5</rde-rd:param>
</rde-rd:params>
```

```
-<search-results chunk="1" chunksize="5" connector="Apache Solr Search" con
 external-only="ignore" facet="cms.keywords;cms.lastmodifier;cms.releasedby"
 constraints="PERMISSION" include-mode="metainfo,meta;facet" lastchunk="
 CIS MODIFIED; CIS TYPE" project="xample" searchengine="Apache Solr">
    <rde-dm:query-cis>helle</rde-dm:query-cis>
  - < lst name="spellcheck">
    - < lst name="suggestions">
      - < lst name="helle">
          <int name="numFound">5</int>
          <int name="startOffset">0</int>
          <int name="endOffset">5</int>
        -<arr name="suggestion">
             <str>holle</str>
             <str>hölle</str>
             <str>holte</str>
             <str>halte</str>
             <str>hecke</str>
          </arr>
        </lst>
      </lst>
    </lst>
 </search-results>
```

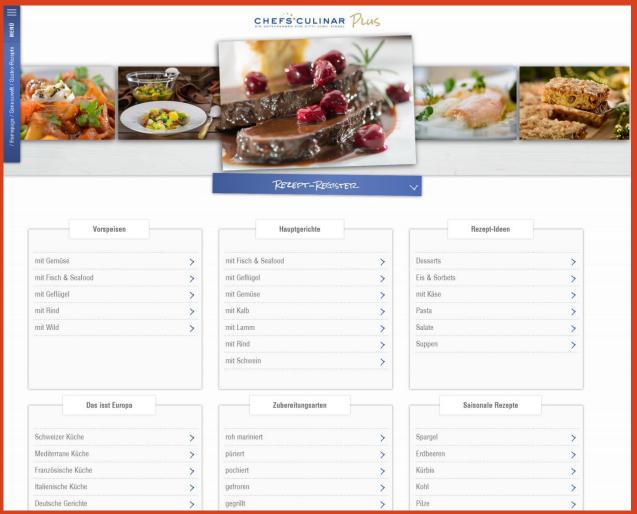


</rde-dm:query>



Vordefinierte Suchen Beispiel: "Rezeptregister"

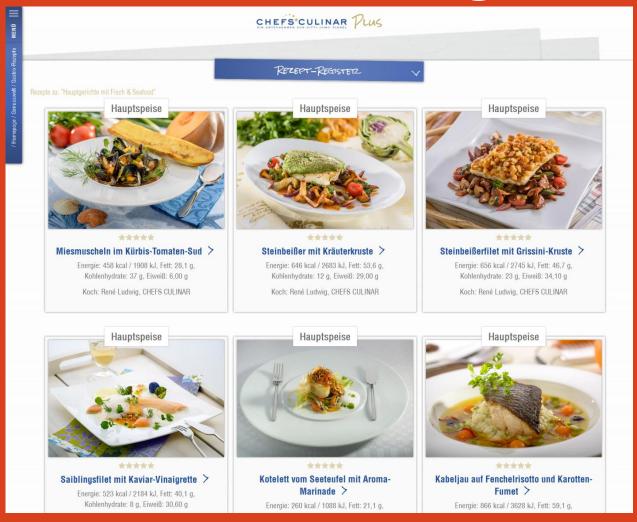






Vordefinierte Suchen Beispiel: "Rezeptregister"







Hit-Highlighting

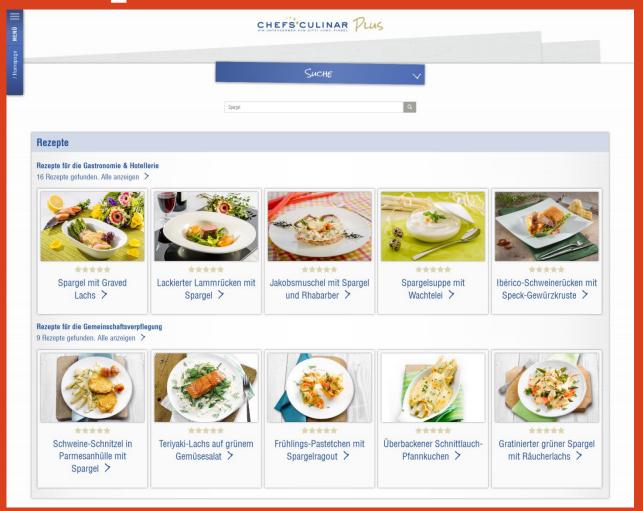


Apache Solr				
haus			Search	
cms.keywords	cms.lastmodifier	cms.releasedby		
> Feature_Concepts 1	> kate 1	> kate 1		
« Previous 1 results for ha	aus, page 1 / 1 Next »			
Berechtigungen und für Sprachen. Globales Berechtigungspaket Gebäude Als Grundlage fungiert ein globales Berechtigungspaket, daß für alle Benutzer unterhalb der SiteBuilder-Rolle nur lesenden Zugriff erlaubt. Grundsätzlich wird immer empfohlen Mittwoch 6 Mai 2015				



Darstellung Suchergebnis Beispiel 1







Darstellung Suchergebnis Beispiel 2



Suchen		
Backofen	> Suchen	
hre Suche nach Backofe	<mark>en</mark> ergab <mark>297</mark> Treffer.	Anzahl Treffer pro Seite: 10 V
hre Suche nach Backofe Alle (297)	en ergab 297 Treffer. Produkte im Suchergebnis	Anzahl Treffer pro Seite: 10 < <u>Vorherige Seite 1 2 3 4 30 Nächste Seite</u>
	Produkte im Suchergebnis Dampfgarer n Backofen Hav	Vorherige Seite 1 2 3 4 30 Nächste Seite it Backofen DGC 6400 Dampfgarer mit



Tag-Cloud mit Suchwörtern

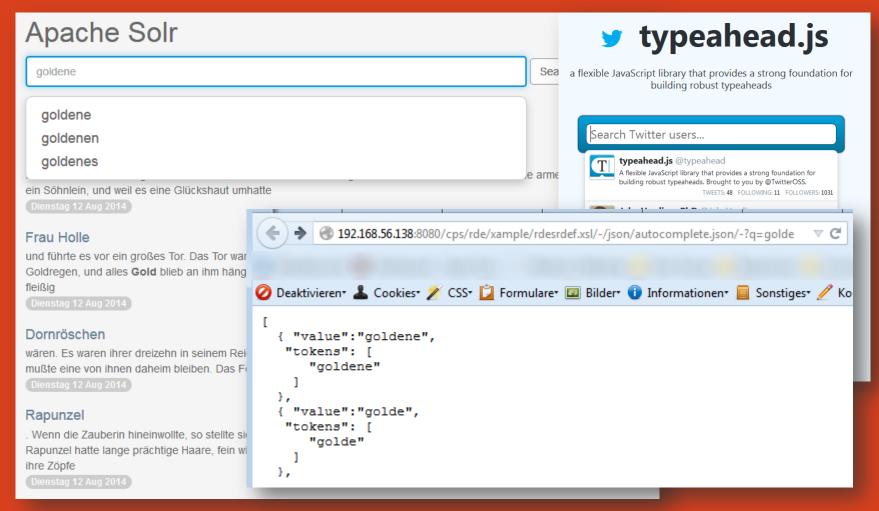


Henkel Excellence is ou	r Passion	Henkel Deutschland ▶ wechseln
▶ Über Henkel		Services
▶ Henkel Aktuell		Services
Produkte & Marken		Q Suche ▶Go
Innovation		
Nachhaltigkeit	Beratung Events Geschäftsbericht 2013 Henkel Aktien	Download & Services
Presse & Media Relations	Hygienespüler Innovation Innovationen Jobs Loctite	□ FAQs
Investor Relations		Download-Sammlung (0)
▶ Karriere	nachhaltige Produkte Nachhaltigkeitsberichts 2013 Persil	(0)
		Marken
		Auswählen ▼ ▶Go
	✓ Zurück zur Übersicht	
	Suche	
	Suche Suche Suchbereich Gesamte Webseite ✓	
	Suchbereich Gesamte Webseite ▼	



Autosuggestion oder Autocomplete







Synonyme & Wortstämme

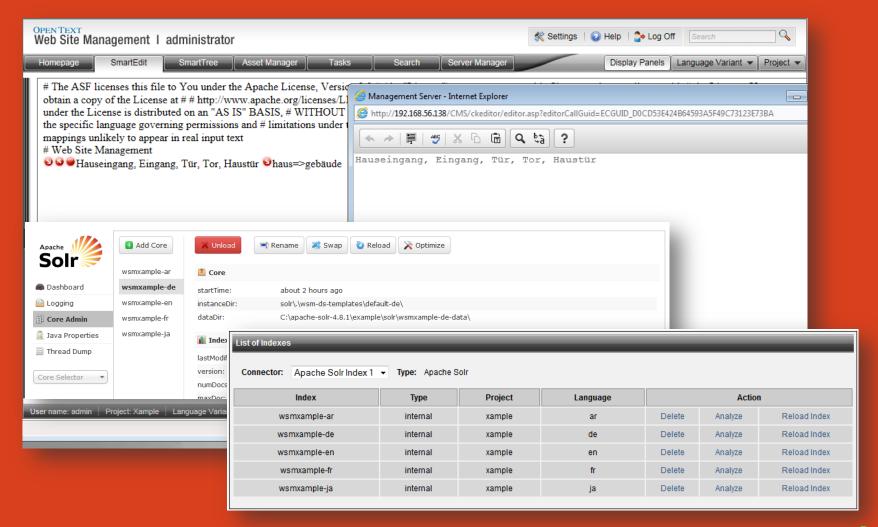


Messe Düsseldorf		Englisch Ko	ontakt Sitemap	Mobil App Suche	suchen
Messen Service	Presse	Messegelände	Unternehmen	International	Anreise & Aufenthalt
	Messe Düsseldorf -> Suche Nach: im Bereich: Allgemeine Akkrediti GmbH Ausweises von Jugen behält sich der Messe mehr dazu TV- & Radio-Beiträge Messevents wissen Blick hinter die Kulissen mehr dazu Download für PC, Han Düsseldorf GmbH	messe Gesamte Webse suchen erungsrichtlinien - Me dpresseorganisationen. everanstalter die weite - Messe Düsseldorf G wollten: Hintergrundberi	esse Düsseldorf Im Übrigen ere GmbH chte, Aktuelles und ein	Unsere Services für Sie Unsere freundlichen und kompetenten Mitarbeiter sin immer für Sie da. Unsere Services für: Besucher Aussteller Journalisten	d d



Pflege der Synonyme im SmartEdit

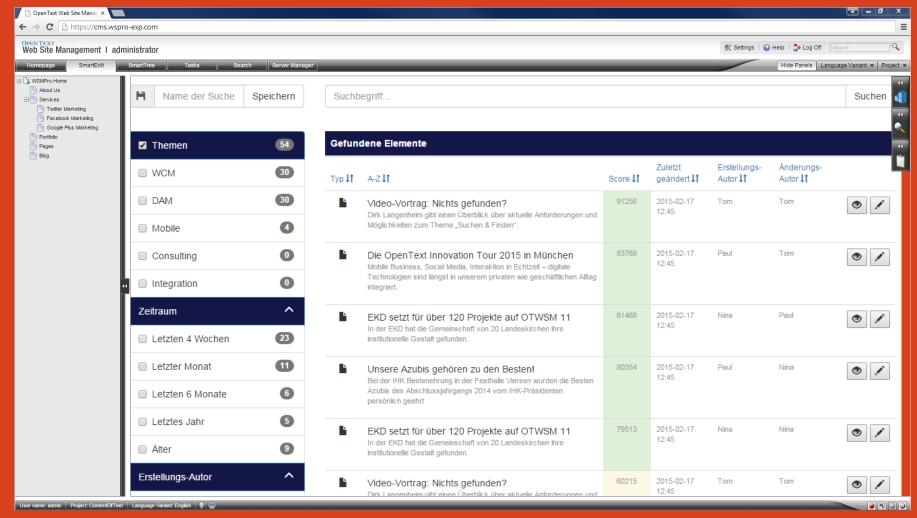






Suchfunktion im SmartEdit über publizierte Seiten







Vielen Dank für ihre Aufmerksamkeit!





Markus Heckmann

+49 (0) 441 93578-747 Markus.heckmann@opentext.com

<u>OPENTEXT</u>™

Dirk Langenheim

+49 (0) 2162 816156-312 dirk.langenheim@intentive.de

