



Whitepaper Lucene 2.0

Inhalt

Was ist Lucene 2.0?.....	2
Systemvoraussetzungen für Lucene 2.0.....	2
Aktivierung von Lucene 2.0 nach Update von 5.3/5.4 auf 5.5x.	3
Allgemeine Einstellungen zu Lucene 2.0	3
Arbeiten mit Lucene.....	6
Fehlersuche/Analyse	7
Konfiguration Lucene 2.0 per Konfig XML Datei (bwlucene.config.xml)	8

Ab der Version 5.56.100 (2015) BüroWARE und WEBWARE 1.56.100 (2015) liefert SoftENGINE bei **Neuinstallationen** standardmäßig mit der Volltextsuche Lucene 2.0 aus. Bei Updateinstallation von BüroWARE 5.3/5.4 muss Lucene 2.0 manuell eingerichtet werden. *Wenn Lucene 2.0 nicht aktiviert wird, wird die „alte“ Volltextsuche (Fulltextsrv.exe) verwendet.*

Was ist Lucene 2.0?

Apache Lucene ist eine Programmbibliothek zur Volltextsuche und ein Projekt der Apache Software Foundation.

Apache Lucene ist eine freie Software. Im World Wide Web realisierte Lucene beispielsweise bis 2014 die Suchfunktion der Wikipedia. Aktuell ist Twitter ein Beispiel für die Leistungsfähigkeit und Skalierbarkeit von Lucene.

Quelle: http://de.wikipedia.org/wiki/Apache_Lucene

Systemvoraussetzungen für Lucene 2.0

Um BüroWARE/WEBWARE mit der Volltextsuche Lucene nutzen zu können, muss auf dem System .NET Framework 3.5 installiert/aktiviert sein.

Die Lucene Suche ist ab der BüroWARE Version 5.5x integriert und ab der Version 5.56.xxx ist Lucene 2.0 bei Neuinstallationen standardmäßig aktiviert.

ACHTUNG:

Bei Updateinstallationen auf 5.5x wird Lucene NICHT automatisch aktiviert. Hierzu müssen Zugriffsrechte geschaltet werden. Nähere Informationen finden Sie im nachfolgenden Punkt.

Hinweise zu .NET

.NET Framework WIKI:

<http://de.wikipedia.org/wiki/.NET>

.NET Aktivierung unter Windows 8:

<http://msdn.microsoft.com/de-de/library/hh506443%28v=vs.110%29.aspx>

Aktivierung von Lucene 2.0 nach Update von 5.3/5.4 auf 5.5x.

Zugriffsrechtbereich: VOLLTEXT

502 BüroWARE Lucene COM-Komponente 2.0 verwenden (aktivieren/Standard aktiv)

106 Volltextsuche mit Lucene (aktivieren/Standard aktiv)

Wenn ZGR 106 deaktiviert ist, ist die alte Volltextsuche (Fulltextsrv.exe) aktiv.

109 Index Aktualisierung bei Lucene mit Cache (aktivieren/Standard aktiv)

Wenn dieses Recht aktiv ist, werden Index Update Aufträge gesammelt und alle 5 Sekunden an einen Unterthread übergeben und abgearbeitet. Bei großen Datenmengen kann der BüroWARE/WEBWARE-Hauptthread ohne Unterbrechungen weiterarbeiten und sorgt dadurch für mehr Performance. (Bei Updateinstallationen von 5.4 auf 5.5 kann u.U. das ZGR 108 noch angezeigt werden, dies hat jedoch keine Funktion mehr!)

Bei Updateinstallationen von 5.4 auf 5.5 kann u.U. das ZGR 108 noch angezeigt werden, dies hat jedoch ab den S-i-E Versionen keine Funktion mehr!

Hinweis:

Bei Aktivierung von Lucene 2.0 müssen Index Ordner und Index Dateien der alten Volltextsuche (Fulltextsrv.exe) manuell entfernt werden. Dafür muss der IDX Ordner im BüroWARE Datenpfad (Mandant) geleert werden.

Danach ist eine Volltextreorganisation im Datenbankassistenten notwendig. Diese Reorganisation kann im laufenden Betrieb vorgenommen werden.

Allgemeine Einstellungen zu Lucene 2.0

Die Volltextsuche Lucene 2.0 im **Mailsystem** kann mit folgendem Zugriffsrecht aktiviert werden:

[VOLLTEXT, Einstellungen Volltextindizierung (V50), Nr: 500 - Fulltext-Server durch Lucene ersetzen insbesondere Mail(V55)]

Danach ist eine Volltextreorganisation im Datenbankassistenten notwendig. Das Zugriffsrecht ist standardmäßig aktiviert.

Mit folgendem Zugriffsrecht können führende Leerzeichen in Primärindex Feldern berücksichtigt werden, sodass diese bei der Suche trotzdem gefunden werden:

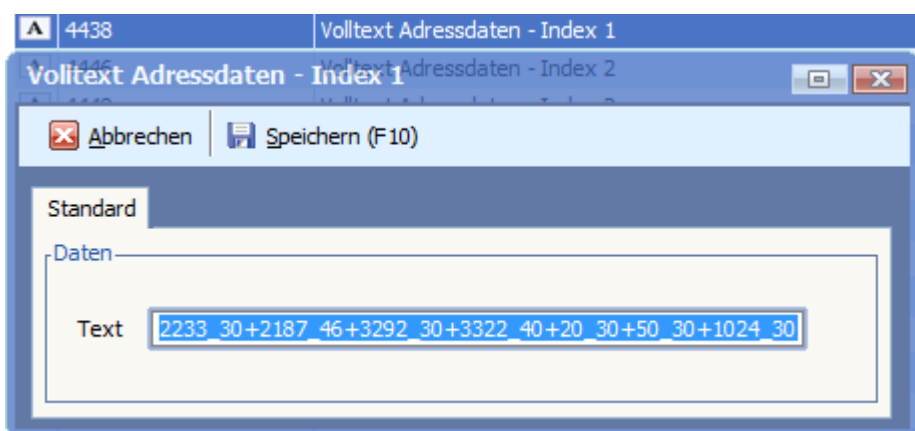
[FEHLER5, Nr: 052 - Fehlerbereinigung V, Berücksichtige führende Leerzeichen in Key-Felder für Lucene Index]

Das Zugriffsrecht ist im Standard deaktiviert.

Indizierung von beliebigen Feldern für die Volltextsuche Lucene 2.0

Standardmäßig sind nicht alle Felder der Datenbank indiziert.

Die Indizes der Volltextsuche LUCENE 2.0 werden ausschließlich über die Internen Texte über den Volltextindex 1 - 5 gebildet (zB. Index 4438 für Adressdatenindex 1).



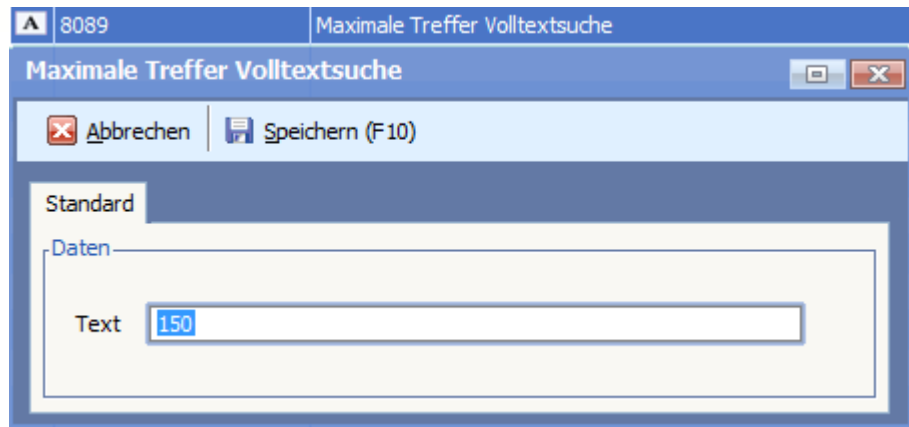
angeben.

Dies bedeutet, wenn Sie ein Feld in die Volltextsuche aufnehmen wollen, welches im Standard nicht indiziert wird, dann müssen Sie die Position und Länge des Feldes in den Internen Texten

Danach ist eine Volltextreorganisation im Datenbankassistenten notwendig.

Anzahl der Treffer

Im Internen Text (Index 8089 – Maximale Treffer Volltextsuche) kann die maximale Anzahl der zu findenden Datensätze angegeben werden.



Arbeiten mit Lucene

STRG+S (in WEBWARE STRG+F)

Mit der Tastenkombination [STRG]+[S] aktivieren Sie das Suchbegriff-Eingabefeld oberhalb der Tabelle. Mit [ENTER] wird die Suche gestartet und alle gefundenen Datensätze werden in der Tabelle angezeigt. Zusätzlich erscheint in der Tabellenüberschrift die Anzahl der Treffer. Das Suchergebnis kann über die Tastenkombination STRG+F4 gelöscht werden.

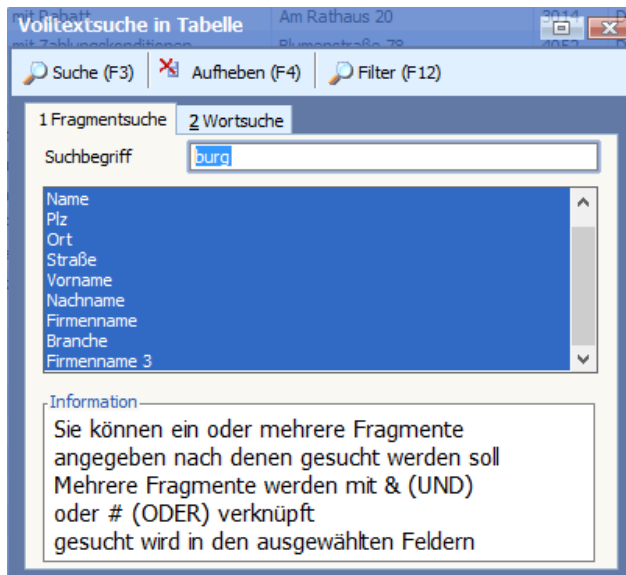
Es werden nur die Indizes durchsucht, welche auch als Spalten in der Tabelle eingebunden sind (kann auch Status "Gesperrt" sein).

STRG+S + STRG+S (2 mal hintereinander STRG+S)

Mit dieser Tastenkombination wird eine eigene Suchmaske aufgerufen, in welcher sämtliche Felder des Index aus den Internen Texten gezeigt und markiert werden.

Beispiel:

Es wurde ein Feld in den Internen Texten hinterlegt und indiziert welches nicht in der Tabelle eingebunden ist. Durch diese Art der Suche kann auch in diesen Feldern gesucht werden. Als weitere Option können gewünschte Felder per Mausklick von der Suche ausgenommen werden. Karteireiter 1 und 2 unterscheiden sich durch Wort- oder Fragment suche.



Fehlersuche/Analyse

Zur Fehlersuche/Protokollierung kann über die Datei bwlog.ini im Programmpfad der BüroWARE/WEBWARE das Debugging aktiviert werden.

Beispiel für eine bwlog.ini:

[categories]

bw.lucene.com=trace,**bw.lucene.txt**

(**bw.lucene.txt** ist der Name der Fehlerprotokolldatei. Dieser Dateiname darf keinesfalls bw.lucene.log heißen, da dieser von der Lucene Komponente verwendet wird.)

[output]

date=1

time=1

level=1

user=1

category=1

Hier ist anzugeben, was am Anfang einer Meldungszeile angezeigt werden soll.

Beispiel:

Datum ;Uhrzeit ;Level; Meldung

30.04.2015;08:21:26.123804;INF; GR 630 : GET_RELATION 630 SE0114

Konfiguration Lucene 2.0 per Konfig XML Datei (bwlucene.config.xml)

Mittels dieser XML Datei kann das Suchverhalten von Lucene 2.0 speziell angepasst werden.

Beispiel:

Suche nach "aa-bb"

Ergebnis ohne XML Datei: Alle Datensätze welche "aa" & "bb" beinhalten.

Ergebnis mit XML Datei: Die Suche wird als komplettes Wort gesehen und das Ergebnis zeigt nur Datensätze, welche "aa-bb" zusammenhängend enthalten.

Die XML-Datei "bwlucene.config.xml" muss im BüroWARE Programmpfad liegen.

Über diese Einstellung wird ein eigener "Analyzer" bei der Suche verwendet, so dass Worttrenner nicht beachtet werden.

Das XML-Schema ist wie folgt:

```
<xsd:schema xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xmlns:xs="http://www.w3.org/2001/XMLSchema"
xmlns:xsd="http://www.w3.org/2001/XMLSchema"
attributeFromDefault="unqualified" elementFormDefault="qualified">
<xs:element name="AnalyzeConfig">
<xs:complexType>
<xs:sequence>
<xs:element minOccurs="0" name="Aktiv" type="xs:boolean" />
<xs:element name="Assemblyname" type="xs:string" />
<xs:element name="Analyzerclass" type="xs:string" />
</xs:sequence>
</xs:complexType>
</xs:element>
</xsd:schema>
```

Beispiel für den Inhalt der bwlucene.confing.xml:

```
<AnalyzeConfig xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xmlns:xsd="http://www.w3.org/2001/XMLSchema">
<Assemblyname>Lucene/BWAnalyzer.dll</Assemblyname>
<Analyzerclass>BWAnalyzer.BWAnalyzer</Analyzerclass>
</AnalyzeConfig>
```