

Dublettenprüfung mit dem Autoindex – angelehnt an KBA-36873

automatische Nummerierung

<https://support.docuware.com/de-DE/knowledgebase/article/KBA-36861>

Dublettenprüfung mit Autoindex

<https://support.docuware.com/de-de/knowledgebase/article/KBA-36873>

Testsystem: DocuWare 7.9 mit MySQL DB und Intelligent Indexing v2 on premises

Ich möchte einen „allgemein gültigen“ Autoindex zur Duplikatprüfung bauen – Die Ansätze die ich bisher gesehen habe, sind sehr „Kunden/System Abhängig“. Einen DocuWare Workflow möchte ich dazu nicht verwenden.

Der Ablagedialog enthält den Wert „Neu“ im Statusfeld, da ich dies so vorbelegt habe.

The screenshot shows the 'Autoindex' configuration page in DocuWare. The main heading is 'Dublettenprüfung' (Duplicate Check). Below it are tabs for 'Auslöser' (Trigger), 'Indexdaten' (Index Data), 'Datenzuweisung' (Data Assignment), and 'Zurückschreiben' (Overwrite). The 'Auslöser' tab is active, showing the configuration for triggering the process when data is added to a specific archive.

Auslöser für das Hinzufügen von Indexdaten festlegen

Daten hinzufügen in: Archiv1
Autoindex greift mit dem Benutzer „dwadmin“ auf das Archiv zu. [Benutzer wechseln](#)

Archivereignis | Dateiereignis | Zeitsteuerung

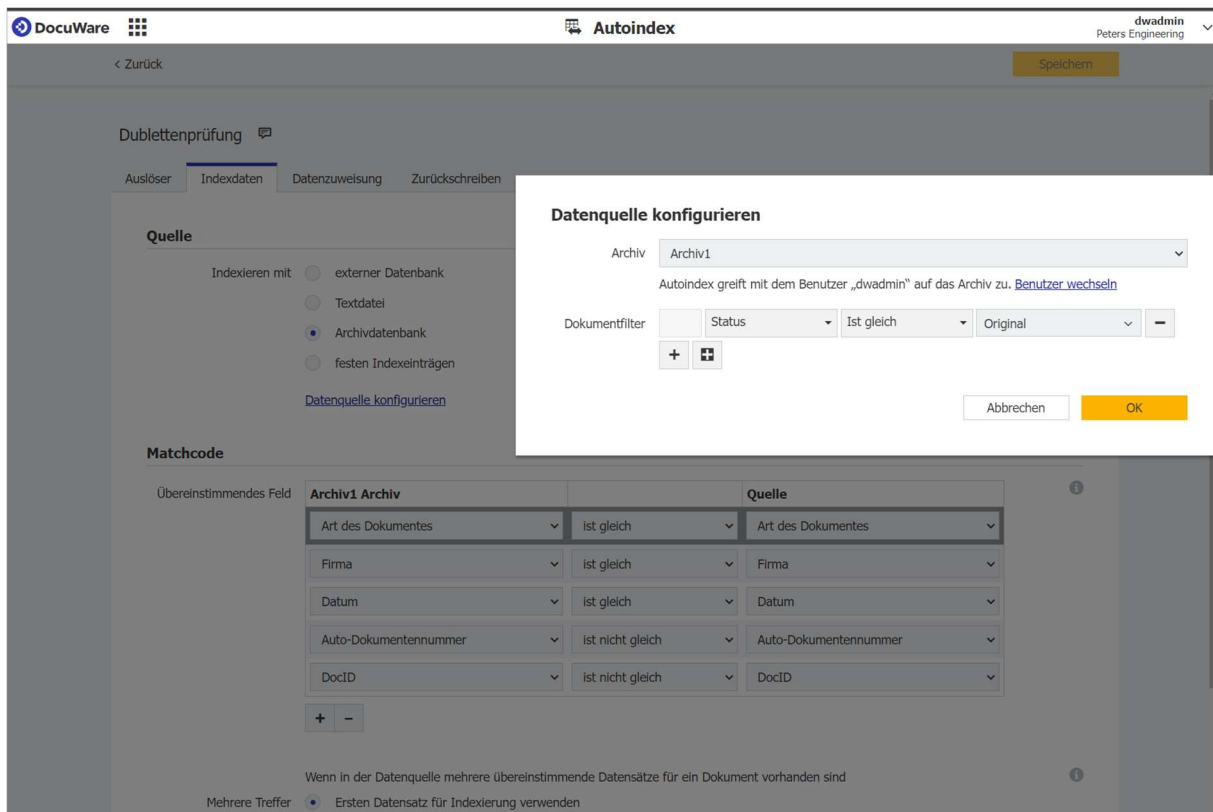
Ausführung für neue Dokumente
 wenn Indexeinträge von bestehenden Dokumenten geändert wurden
Nur Dokumente und Indexeinträge mit folgenden Kriterien ändern

Vor der Änderung Es sind noch keine Regeln definiert. + Neue Regel + Neue Gruppe von Regeln

Nach der Änderung

UND	Status	Ist gleich	Neu	-
	Status	geändert		-

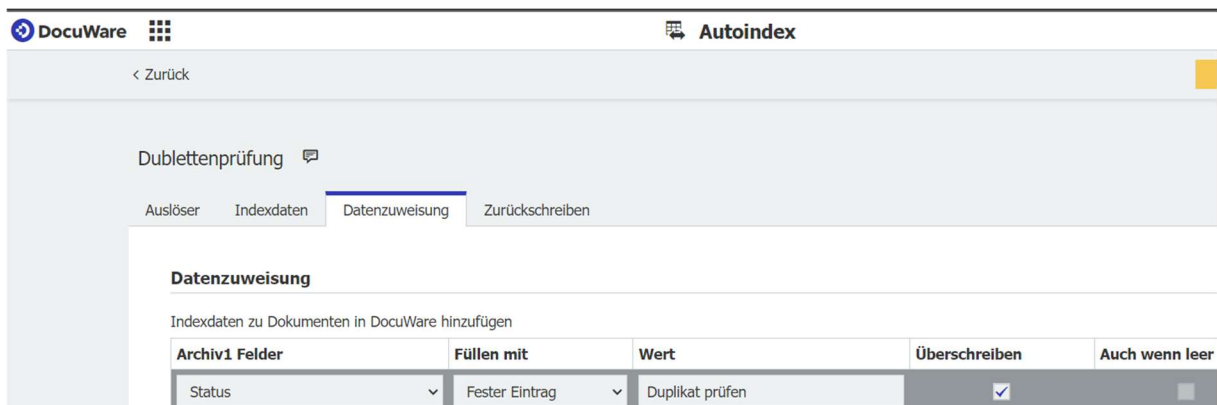
+ +



Ich bin mir nicht sicher wie das Feld Dokumentennummer im Beispiel KBA-36873 konfiguriert und „genutzt“ werden sollte...

In meinem Testarchiv gibt es ein weiteres Feld Dokumentennummer aber das Feld ist nicht immer korrekt (durch OCR und Intelligent Indexing v2 oder manuell) befüllt worden. Daher dient dieses nicht besonders gut zur Duplikatprüfung.

Wenn kein Duplikat anhand der im Matchcode definierten Werte gefunden wird – Wird das Indexfeld auf „Original“ geändert.



Wenn ein Duplikat im Archiv gefunden wird, wird das Feld Status auf „Duplikat prüfen“ geändert.

Bisher habe ich für den Test nur eine Liste erstellt, die alle Dokumente mit dem Status „Duplikat prüfen“ enthält – So kann man die Duplikate manuell löschen.

Fragestellung:

Da ich auch die Workflow Advanced Videos zum Thema Duplikatprüfung gesehen habe, bin ich jetzt nicht sicher, wie ich es am geschicktesten konfigurieren kann. Dort ist die Logik etwas erweitert.

Mein Ansatz war – Dass ich das Feld Checksum (im Archiv aktiviert „Integrität vom Dokument mit Checksumme prüfen“ <https://help.docuware.com/#/home/62385/1/1>) hinzuziehen. Die Checksumme (also der Hashwert auf den Inhalt der Datei) bezieht sich nicht auf die Indexdaten, sondern damit könnte man die exakt selbe PDF Datei – die schon im Archiv abgelegt wurde erkennen.

Der SQL Select dazu sieht in etwa so aus:

```
WITH DuplicateChecksums AS (  
    SELECT CHECKSUM  
    FROM [dwdata].[dbo].[SAMPLEDE_SEC]  
    WHERE CHECKSUM IS NOT NULL  
    GROUP BY CHECKSUM  
    HAVING COUNT(*) > 1  
)  
SELECT s.*  
FROM [dwdata].[dbo].[SAMPLEDE_SEC] s  
INNER JOIN DuplicateChecksums d  
ON s.CHECKSUM = d.CHECKSUM  
ORDER BY s.CHECKSUM;
```

Disclaimer: Der Befehl sollte nicht dazu ermutigen, Datei-Duplikate im Archiv zu löschen – Rechnungskopien etc. haben auch Ihre Daseinsberechtigung.

Duplikate werden über den SQL Befehl erkannt. DocuWare errechnet BEIM ABLEGEN einmal eine **Checksumme** auf den **Inhalt** des Dokumentes und speichert diesen in der Datenbank.

Beim Veränderungen Prüfen Button wird diese mit einer neu errechneten Prüfsumme (wo wird diese gespeichert?) verglichen.

Wenn man das Dokument mit DocuWare verändert (Stempel verschmelzen lassen) wird nicht erkannt, dass das Dokument geändert wurde, weil die Checksumme nicht neu gebildet wird. Daher würde mein SQL Befehl auch das Dokument weiterhin als Duplikat erkennen.