

BnNetAiService (IndexerService) mit OCR-Texterkennung

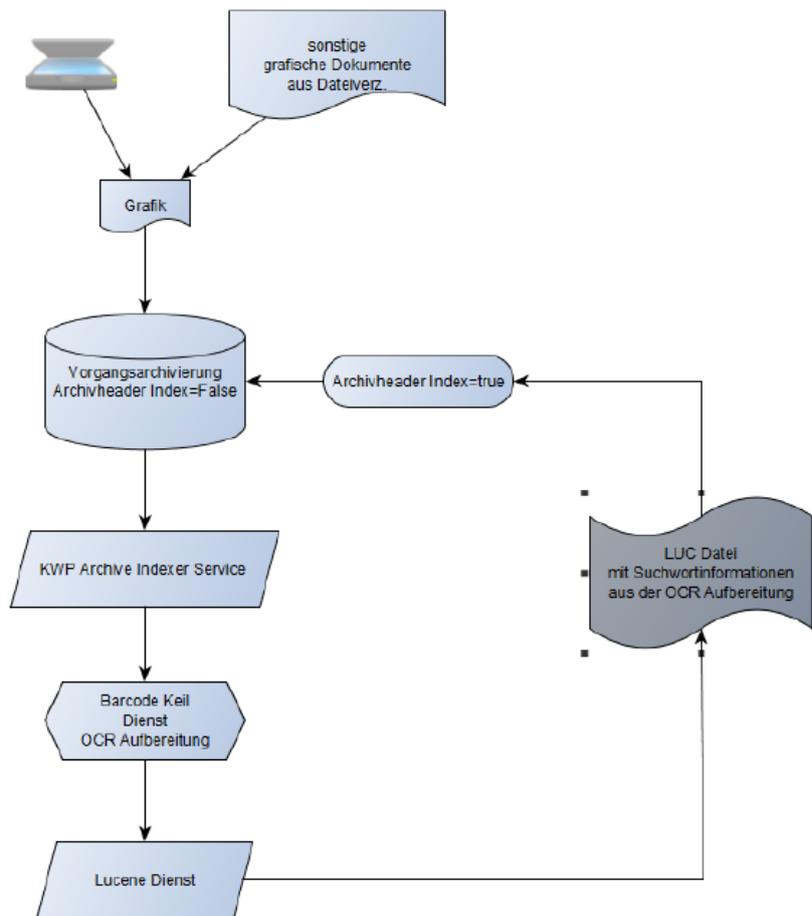
Der neue IndexerService ist eine Ergänzung zur bereits bestehenden OCR-Texterkennung im Archiv, dabei können Dokumente die nicht durch die OCR-Texterkennung erfasst wurden (z.B. KWPBox) nachträglich indiziert werden.

Der BnNetAiService durchsucht dabei über ein Flag die archivierten PDF-Dateien automatisiert nach Schlüsselwörtern und bereitet dadurch einen Index zur optimierten Textsuche dokumentenübergreifend auf.

Voraussetzungen zur Installation des BnNetAiService:

- Modul Vorgangsarchivierung
- Keil Barcode OCR KWP Plus Edition
- Anpassungen am Barcode OCR-Dienst
- Installierter, funktionierender KWP Lucene-Dienst
- Einstellungen in der bncontrl.ini

Die Funktion lässt sich folgendermaßen beschreiben:

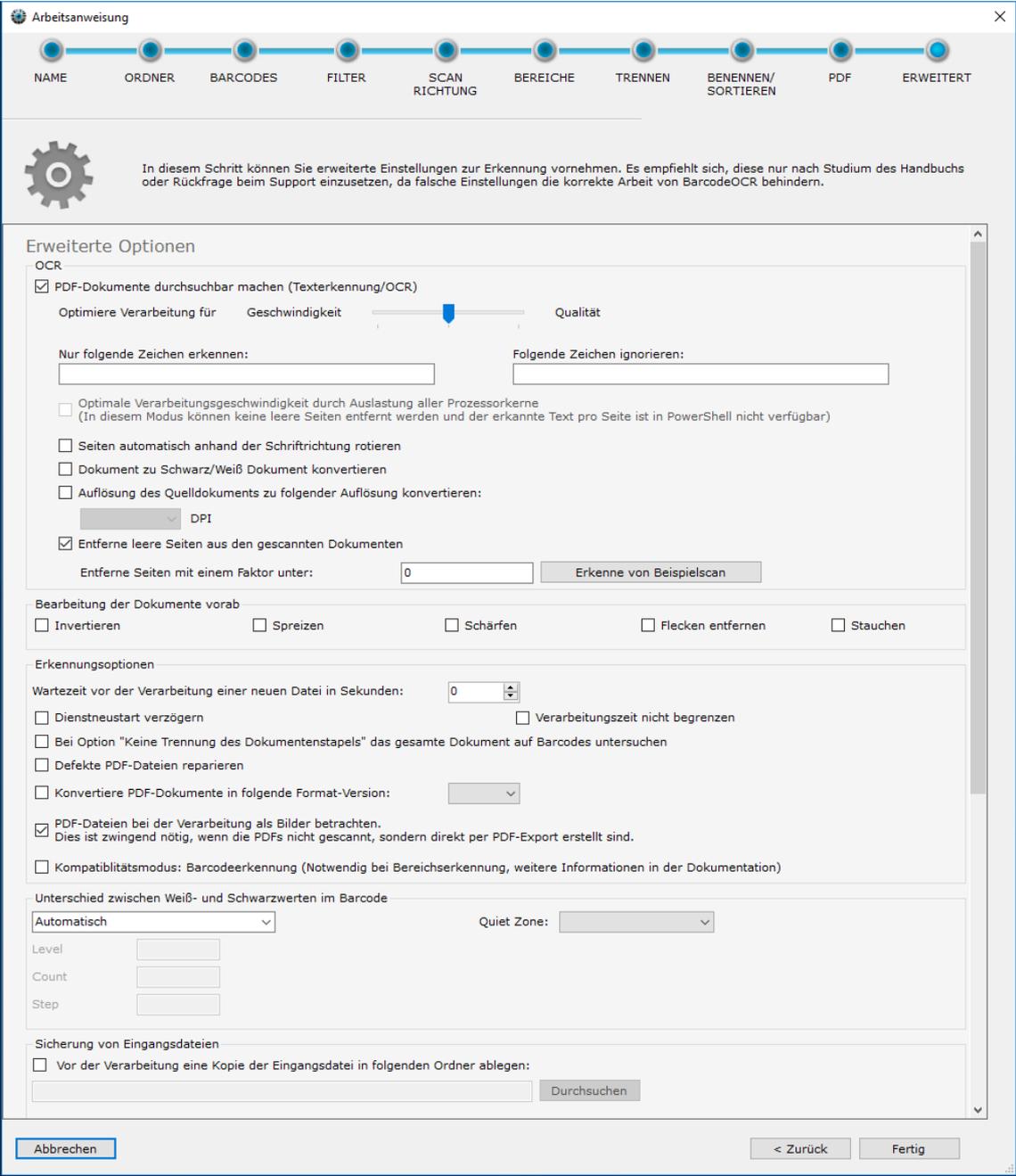


BnNetAiService (IndexerService) mit OCR-Texterkennung

Dokumente und Grafiken mit Text werden zuerst eingescannt und wie gewohnt archiviert. Liegen diese Dokumente nun im Archiv, prüft der IndexerService, ob es zu diesem Archiveintrag schon einen aufgebauten Index gibt. Falls nicht, wird dieser Index mit Hilfe der Barcode OCR-Aufbereitung erstellt.

1. Anpassung Barcode OCR

Um die OCR-Texterkennung korrekt und zuverlässig aufbauen zu können, muss dafür eine eigene Arbeitsanweisung mit **separaten IN/OUT/FEHLER Ordnern** im Barcode-OCR der Firma Keil mit folgenden zusätzlichen Eigenschaften eingestellt werden.



Tel. +49 (0)871 7008 - 0 Fax +49 (0)871 7008 - 34 E-Mail office@kwpsoftware.de Internet www.kwpsoftware.de



BnNetAiService (IndexerService) mit OCR-Texterkennung

2. Einrichtung des KWP-IndexerService

Es wird empfohlen den bnNetAiService unter dem Pfad "C:\Program Files (x86)\KWP Informationssysteme GmbH" zu installieren.

Zu finden ist dieser Service im WebPortal oder unter "H:\Sonderprogramme\bnNetAiService".

Sollte es mehrere Mandanten in einer Firma geben, so muss pro Mandant eine eigene Indexer-Instanz installiert werden. Idealerweise sollten die Bezeichnung des IndexerService und die Mandantenbezeichnung übereinstimmen. Beispielsweise heißt der 1. Mandant hier ABC.

Dieser PC > Windows (C:) > Programme (x86) > KWP Informationssysteme GmbH			
Name	Änderungsdatum	Typ	Größe
bnKatPflegeDienst	20.11.2019 09:48	Dateiordner	
bnNetAiService_ABC	25.11.2019 15:02	Dateiordner	
bnNetAiService_XYZ	25.11.2019 15:02	Dateiordner	
Licensing Service	16.08.2017 10:44	Dateiordner	

Das Logging erfolgt über die jeweilige Nlog.Config. (befindet sich im jeweiligen Indexer-Verzeichnis) und kann dort editiert werden.



BnNetAiService (IndexerService) mit OCR-Texterkennung

3. Anpassung der Konfigurationsdatei bnNetAiService.exe.config

Da der Dienst keine grafische Benutzeroberfläche besitzt, muss die Konfiguration in der **bnNetAiService.exe.config** über den Editor vorgenommen werden. Bei mehreren Mandanten muss unbedingt auf die korrekte Mandantenbezeichnung geachtet werden. Die korrekte Bezeichnung kann - falls nötig - aus der Datei **bnSettings.xml** aus dem bnwin Verzeichnis abgelesen werden.

```
<mandanten>
  <mandant id="default">
  <mandant id="sql">
  <mandant id="ABC">
  <mandant id="71">
```

Die **bnNetAiService.exe.config** muss wie folgt angepasst werden, wenn **mit Mandanten** gearbeitet wird (in unserem Fall der Mandant ABC):

```
<add key="ServiceName" value="KwpArchiveIndexerService_ABC"></add>
<add key="ServiceDescription" value="KWP Archive Indexer Service ABC"></add>
<add key="ServiceDisplayName" value="KWP Archive Indexer Service ABC"></add>

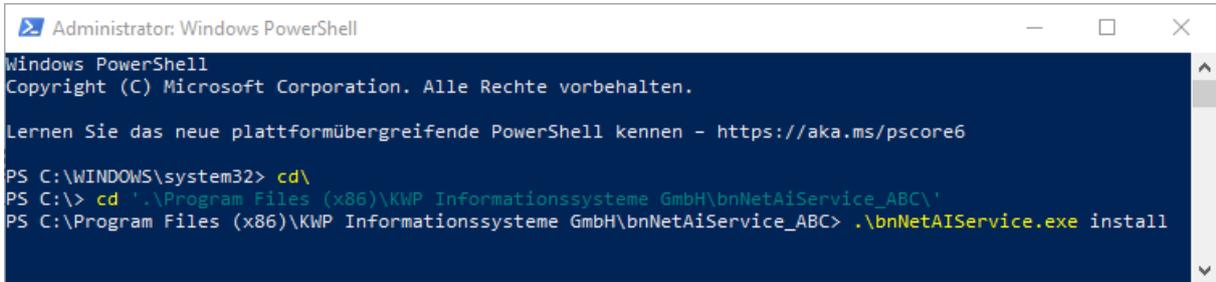
<add key="BnWinDirectory" value="\\SERVER\KWP\bnWin"/>
<add key="BnWinBenutzer" value="ADMIN"/>
<add key="BnWinMandant" value="ABC"/>
<add key="BnWinEigeneGrunddaten" value="false"/>
<add key="OcrScanDirectory" value="\\SERVER\KWP\bnWin\OCR\in"/>
<add key="OcrErrorDirectory" value="\\SERVER\KWP\bnWin\OCR\error"/>
<add key="OcrResultDirectory" value="\\SERVER\KWP\bnWin\OCR\out"/>
```

Falls **keine Mandanten** verwendet werden, besteht nur der „default“-Knoten und in der bnNetAiService.exe.config werden lediglich die Pfadanpassungen benötigt. Schlüssel die sich lediglich auf Mandanten beziehen, können dabei entweder leer gelassen oder gelöscht werden.



BnNetAiService (IndexerService) mit OCR-Texterkennung

4. Installation des Dienstes über die PowerShell



```
Administrator: Windows PowerShell
Windows PowerShell
Copyright (C) Microsoft Corporation. Alle Rechte vorbehalten.

Lernen Sie das neue plattformübergreifende PowerShell kennen - https://aka.ms/pscore6

PS C:\WINDOWS\system32> cd\
PS C:\> cd '.\Program Files (x86)\KWP Informationssysteme GmbH\bnNetAiService_ABC\'
PS C:\Program Files (x86)\KWP Informationssysteme GmbH\bnNetAiService_ABC> .\bnNetAiService.exe install
```

PowerShell als Administrator starten

Befehle zur Installation:

Wechsel in das root Verzeichnis:

```
PS C:\>cd\
```

Navigieren in das korrekte Verzeichnis zur Installation mit dem Mandanten ABC:

```
PS C:\> cd '.\Program Files (x86)\KWP Informationssysteme GmbH\bnNetAiService_ABC\'
```

Installation des bnNetAiService:

```
PS C:\Program Files (x86)\KWP Informationssysteme GmbH\bnNetAiService_ABC>
.\bnNetAiService.exe install
```

Dies muss pro Mandanten durchgeführt werden.

Gestartet werden kann der Dienst über die **bnNetAiService.exe**



BnNetAiService (IndexerService) mit OCR-Texterkennung

Troubleshooting

1. Lucene und IndexerService stoppen.
2. Barcode OCR-Einstellungen prüfen und neu starten.
3. Lauschordner + Luceneordner überprüfen.
4. Alle Lucene- und Indexer-Log-Dateien löschen.
5. „False“-Eintrag in SQL-DB überprüfen.
6. IndexerService neu starten.

Dienst starten via Powershell als Administrator:

```
PS C:\Program Files (x86)\KWP Informationssysteme GmbH\bnNetAiService_ABC > start-service -display "kwp archive indexer service"
```

Dienst stoppen:

```
PS C:\Program Files (x86)\KWP Informationssysteme GmbH\bnNetAiService_ABC > stop-service -display "kwp archive indexer service"
```

Danach kann über die PowerShell geprüft werden, welche der Dienste installiert wurden:

```
get-service -display "kwp*"
```

Dies sieht dann ungefähr wie folgt aus:

Status	Name	DisplayName
Running	KwpArchiveIndex...	KWP Archive Indexer Service

Hier wird pro installiertem IndexerService eine Zeile angezeigt.

Im Beispiel des Mandanten „ABC“ gibt es noch eine weitere Zeile mit dem DisplayName „KWP Archive Indexer Service ABC“.

Verarbeitung via Lucene

Im eingestellten Lucene-Austauschordner, z. B. K:\bnwin\Database\Lucene\Austausch\ wird nach dem Verarbeiten einer Datei über den bnNetAiService eine LUC-Datei erzeugt. In dieser Datei sind sämtliche Suchfragmente enthalten. Im letzten Abschnitt der LUC-Datei wird das Datenbankverzeichnis angezeigt.

```
"LTextOcrDokument", "Store": true}}], "LuceneDatenbankVerzeichnis": "\\KWP-Kandler\KWP\bnWin\Database\Lucene\ABC\Archiv", "Typ": 1}
```

Hier muss das Mandantenverzeichnis eingetragen sein, ansonsten ist die Weiterverarbeitung nicht möglich. Dies wurde aber mit der Konfiguration der bnNetAiService.exe.config schon berücksichtigt.



BnNetAiService (IndexerService) mit OCR-Texterkennung

Prüfen, welche Dateien über den IndexerService verarbeitet werden:

Dafür kann folgende Abfrage am SQL Server verwendet werden:

```
SELECT TOP(200) ID, Betreff, Bemerkung, ErstelltAm, Indiziert
FROM ArchivHeader
```

ID	Betreff	Bemerkung	ErstelltAm	Indiziert
9	Telekom_Pascal...	NULL	2018-12-04 13:3...	True
10	Reisekosten_08...	NULL	2018-12-04 13:5...	True
11	Finanzamt	fa	2018-12-20 09:2...	True
12	finanzamt	NULL	2018-12-20 15:1...	True
13	fa_Anlauf.pdf	NULL	2018-12-20 15:3...	True
14	FA_Backhaus.pdf	NULL	2018-12-20 15:5...	True
15	Hotze1.pdf	NULL	2018-12-20 16:1...	True
19	VSB Küche.pdf	TEST MIT BASTI	2019-02-14 08:2...	True
20	db.JPG	NULL	2019-02-14 08:3...	True
21	Küchenfachma...	test basti zum 2.	2019-02-14 08:4...	False

Alle Datensätze, die mit Indiziert = False gekennzeichnet sind, werden über den IndexerService weiterverarbeitet. Einträge mit True sind bereits bearbeitet worden. Der gestartete IndexerService hat ein Wiederholungsintervall von 10-15 Minuten.

Es wird eine Log-Datei in den Ordner
 C:\Program Files (x86)\BnServerDienste\bnNetAiService_ABC\log
 geschrieben.

