

d.velop

d.3 search: Administrieren

Inhaltsverzeichnis

1. d.3 search: Administrieren	3
1.1. Einleitung	3
1.1.1. Über d.3 search	3
1.1.2. Voraussetzungen	3
1.1.3. Internationalisierung	3
1.1.4. Informationen zu d.3 search	3
1.1.5. Architektur	4
1.2. Installation	8
1.2.1. Systemvoraussetzungen	8
1.2.2. Systemempfehlungen	9
1.2.3. Update	10
1.2.4. Installieren und Aktualisieren von d.3 search	11
1.2.5. d.3 gateway - Anbindung	12
1.2.6. d.3 SHAFT-Agent	13
1.2.7. Cluster	15
1.2.8. Integration im d.ecs monitor	28
1.2.9. SSL-Verschlüsselung	28
1.3. Konfiguration	29
1.3.1. Konfiguration per Weboberfläche	29
1.3.2. Konfiguration des d.3 search Servers	51
1.3.3. Konfiguration einer Dokumentensammlung	55
1.4. Technischer Hintergrund zum d.3 search Server	83
1.4.1. Verzeichnisstruktur	83
1.4.2. Die d.3 search engine	85
1.5. Wartung	93
1.5.1. Hinweise zur Optimierung von d.3 search	93
1.5.2. Hinweise zur Optimierung von d.3 search auf einem RAID5-System	94
1.5.3. Hinweise zur Wiederherstellung einer Dokumentensammlung	95
1.5.4. Reindexierung einer Dokumentensammlung	98
1.6. Fehler und Probleme	109
1.6.1. d.3 search	109
1.6.2. d.3 search engine	113
1.7. Weitere Informationsquellen und Impressum	117

1. d.3 search: Administrieren

1.1. Einleitung

In diesem Kapitel finden Sie allgemeine Produktinformationen und Konventionen der Produktdokumentation.

1.1.1. Über d.3 search

Das Produkt d.3 search erweitert das d.3 System um die Fähigkeit zur Volltext-Recherche.

Das Produkt besteht aus den beiden Komponenten d.3 search Server und der Suchmaschine d.3 search engine (Shaft-Engine).

Der d.3 search server Server ist dabei das Bindeglied zwischen dem d.3 server und der Suchmaschine d.3 search engine.

1.1.2. Voraussetzungen

Dieses Handbuch richtet sich an den d.3 Systemadministrator und erläutert den Aufbau, die Installation und die Konfiguration von d.3 search.

Kenntnisse in der d.3 Administration sowie in der Administration des d.ecs rendition servers werden vorausgesetzt.

1.1.3. Internationalisierung

d.3 search unterstützt folgende Sprachen:

- Deutsch
- Englisch
- Sprachen aus dem Paket d.3 national
- Japanisch
- Chinesisch (vereinfacht)

Anmerkung

Beachten Sie die Hinweise zu den [Delimitermode=4](#).

Anmerkung

Weitere mögliche Sprachen erfragen Sie bei der d.velop AG.

1.1.4. Informationen zu d.3 search

Allgemein

d.3 search erweitert das d.3 System um die Fähigkeit zur Volltext-Recherche.

Die Entscheidung, welche Daten in die Suchmaschine aufgenommen werden, wird mit einem Regelwerk am d.3 server administrativ bestimmt.

Die Volltext-Informationen (OCR) gelangen ebenfalls vom d.3 server über den d.3 search server in die Suchmaschine d.3 search engine (SHAFT). d.3 search unterstützt die Verwaltung von mehreren Dokumentensammlungen, so dass in den Regelwerken bestimmt werden kann, welche Volltext-Information in welche Dokumentensammlung landet. Dies kann von Nutzen sein, um bei der Nutzung der semantischen Suche sinnvolle Recherche-Vorschläge zu erhalten.

Auf Wunsch können bei einer nachträglichen Installation von d.3 search bereits vorhandene Volltext-Informationen über den Reindizierungs-Assistenten in einer Dokumentensammlung aufgenommen werden.

Führt der Benutzer am Client eine Volltext-Recherche aus, nimmt der d.3 server die Suchanfrage entgegen und leitet diese über den d.3 search server an die Suchmaschine d.3 search engine (SHAFT) weiter. Anschließend nimmt der d.3 server das Ergebnis der Suchmaschine entgegen und bereitet es für Auslieferung an den Benutzer auf.

Lizenzierung

Zu dem Produkt d.3 search existieren drei Typen von Lizenzen.

- **d.3 search basic:** Mit dieser Lizenz werden nur die Eigenschaftswerte von Dokumenten und Akten in den Index aufgenommen. Dokument-Inhalte aus der eigentlichen Datei werden beim Import verworfen. Somit ist eine Volltext-Recherche nur über Eigenschaftswerte möglich.
- **d.3 search file-content addon:** Erweitert die basic Lizenz um den Dateiinhalt. Es werden alle Informationen (Eigenschaften und Dokumentinhalte), die der d.3 server anliefert, in den Index aufgenommen.
- **d.3 search enterprise addon:** Mit dieser Lizenz werden die Features **Clustering** und **Skalierung** freigeschaltet. Zum einen kann mit mehreren d.3 search-Instanzen ein Aktiv/Aktiv-Cluster betrieben werden und zum anderen kann eine Dokumentensammlung auf mehrer d.3 search engines (SHAFT) skaliert werden.

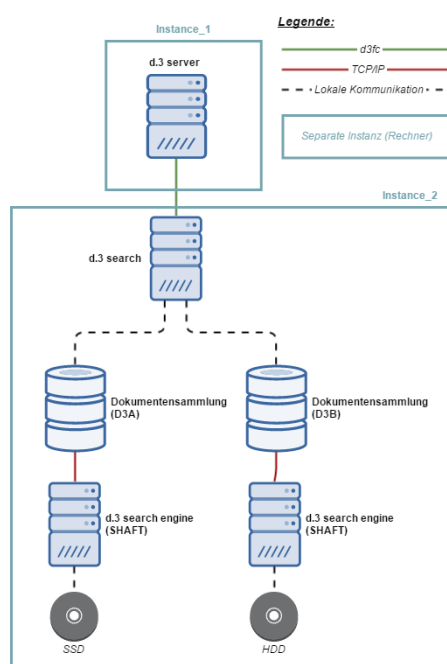
1.1.5. Architektur

Im folgenden Kapitel werden die verschiedenen Architekturvarianten von d.3 search vorgestellt.

Einfache Installation

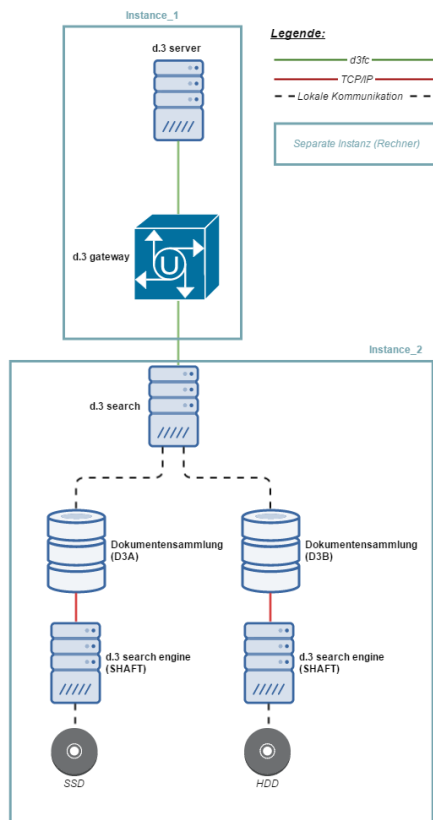
Einfache Installation ohne d.3 gateway

Das folgende Diagramm zeigt eine einfache Installation von d.3 search ohne Verwendung des d.3 gateway. Im Beispiel sind zwei Dokumentensammlungen angelegt worden. Die Dokumentensammlung **D3A** soll eine größere Dokumentensammlung darstellen, welche aus Performance-Gründen auf eine SSD-Festplatte gespeichert wird. Die Dokumentensammlung **D3B** ist hingegen eher klein, daher wird hier eine HDD-Festplatte verwendet. Beide Dokumentensammlungen werden vom d.3 server angesprochen.



Einfache Installation mit d.3 gateway

Das folgende Diagramm zeigt eine einfache Installation von d.3 search mit Verwendung des d.3 gateway. Wie im obigen Beispiel werden zwei Dokumentensammlungen angelegt. Beide Dokumentensammlungen werden vom d.3 server über den d.3 gateway angesprochen.

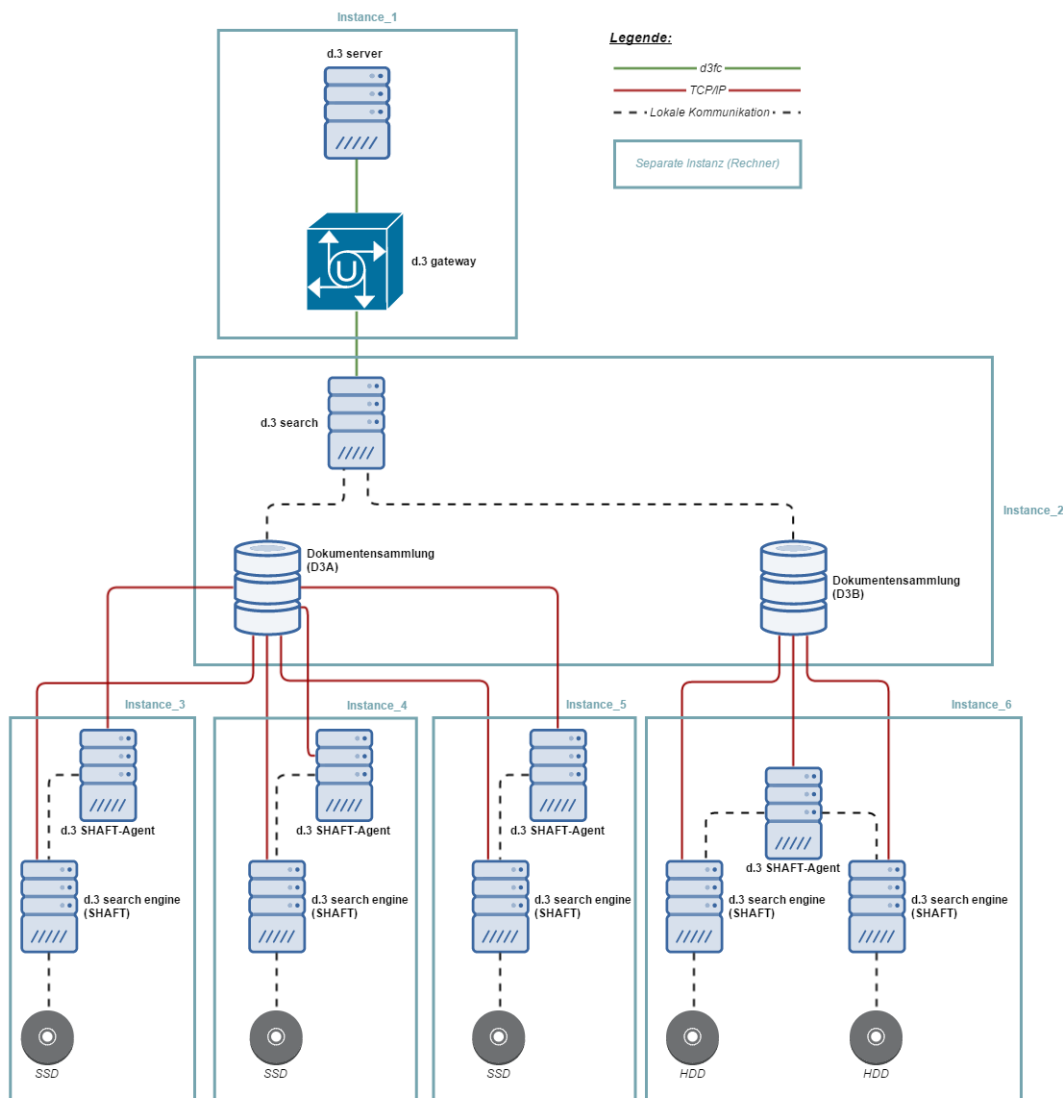


Skalierte Installation

Bei der skalierten Installation werden pro Dokumentensammlung mehrere d.3 search engines (SHAFT) erzeugt. Um die Performance zu erhöhen, können diese dann auf separate Instanzen ausgelagert werden. Um die Prozesse auf den separaten Instanzen zu kontrollieren, wird auf diese der d.3 SHAFT-Agent installiert.

In diesem Beispiel wurde die Dokumentensammlung D3A auf drei Instanzen und auf drei d.3 search engines (SHAFT) aufgeteilt. Es wäre auch denkbar, dass eine Aufteilung auf sechs oder mehr d.3 search engines (SHAFT) vorgenommen wird, diese aber vorerst auf eine Instanz installiert werden. Dies würde eine nachträgliche Skalierung bzw. Performancegewinnung ermöglichen. Die Dokumentensammlung D3B wurde auf zwei d.3 search engines (SHAFT) aufgeteilt. Beide Prozesse befinden sich aber auf einer Instanz, da die Dokumentensammlung kleiner ist als D3A und daher keine weitere Skalierung benötigt.

Die Skalierung bietet den Vorteil, dass eine große Dokumentensammlung auf mehrere d.3 search engines (SHAFT) aufgeteilt wird. Dies hat zur Folge, dass Suchen und Importe schneller durchgeführt werden können. Zudem können bei der Skalierung separate Festplatten verwendet werden, was die Performance noch einmal erhöht.



Die Architektur der skalierten Installation wurde hier mit dem d.3 gateway dargestellt. Eine direkte Kommunikation zwischen d.3 server und d.3 search wäre auch denkbar.

Cluster Installation

Bei der Cluster Installation werden mehrere, in diesem Fall drei, redundante d.3 search Instanzen installiert. Alle Instanzen kommunizieren über den d.3 gateway mit dem d.3 server. Zudem überwachen sich die d.3 search Instanzen gegenseitig, um bei einem Ausfall reagieren zu können. Wie bei den vorherigen Beispielen wurden wieder die Dokumentensammlungen **D3A** und **D3B** verwendet.

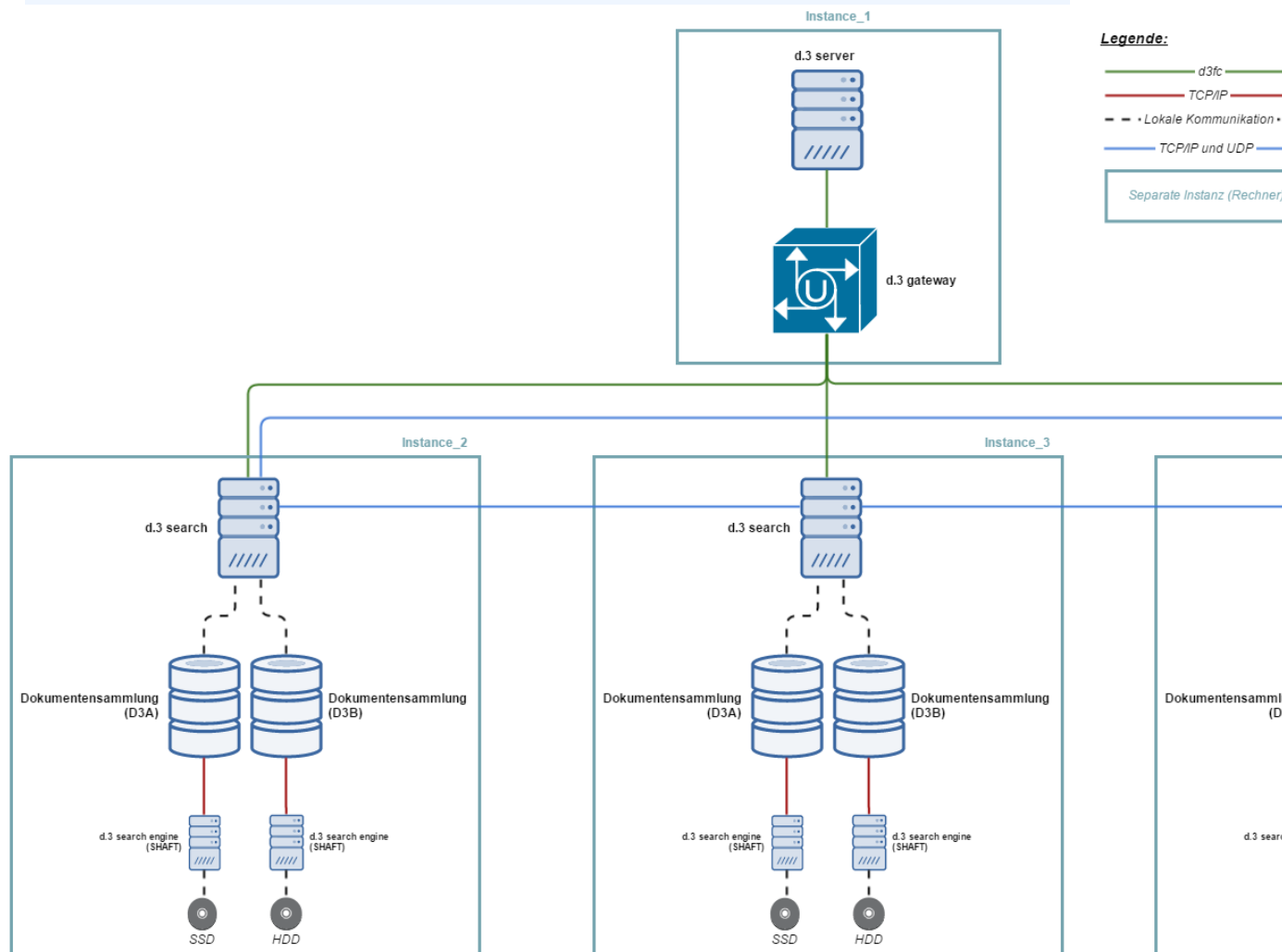
Ein Clusterbetrieb bietet den Vorteil, dass, solange es noch eine aktive Instanz gibt, auf der d.3 search installiert ist, abfragende Aufgaben weiterhin verarbeitet werden. Das Hinzufügen bzw. Ändern von Dokumenten im Suchindex ist während eines Ausfalls jedoch nicht möglich, da ansonsten inkonsistente Zustände in den Datenbeständen entstünden.

Anmerkung

Fallen die Instanzen Zwei und Vier weg, kann die Instanz Drei weiterhin die Suchanfragen vom d.3 server beantworten. Dokumentimporte von Nutzer würden währenddessen nicht durchgeführt.

Anmerkung

Bei der Einrichtung eines Clusters ist es zwingend notwendig, dass die Kommunikation von d.3 server nach d.3 search über den d.3 gateway erfolgt. Wird dies nicht berücksichtigt, ist eine Ausfallsicherheit nicht gewährleistet.



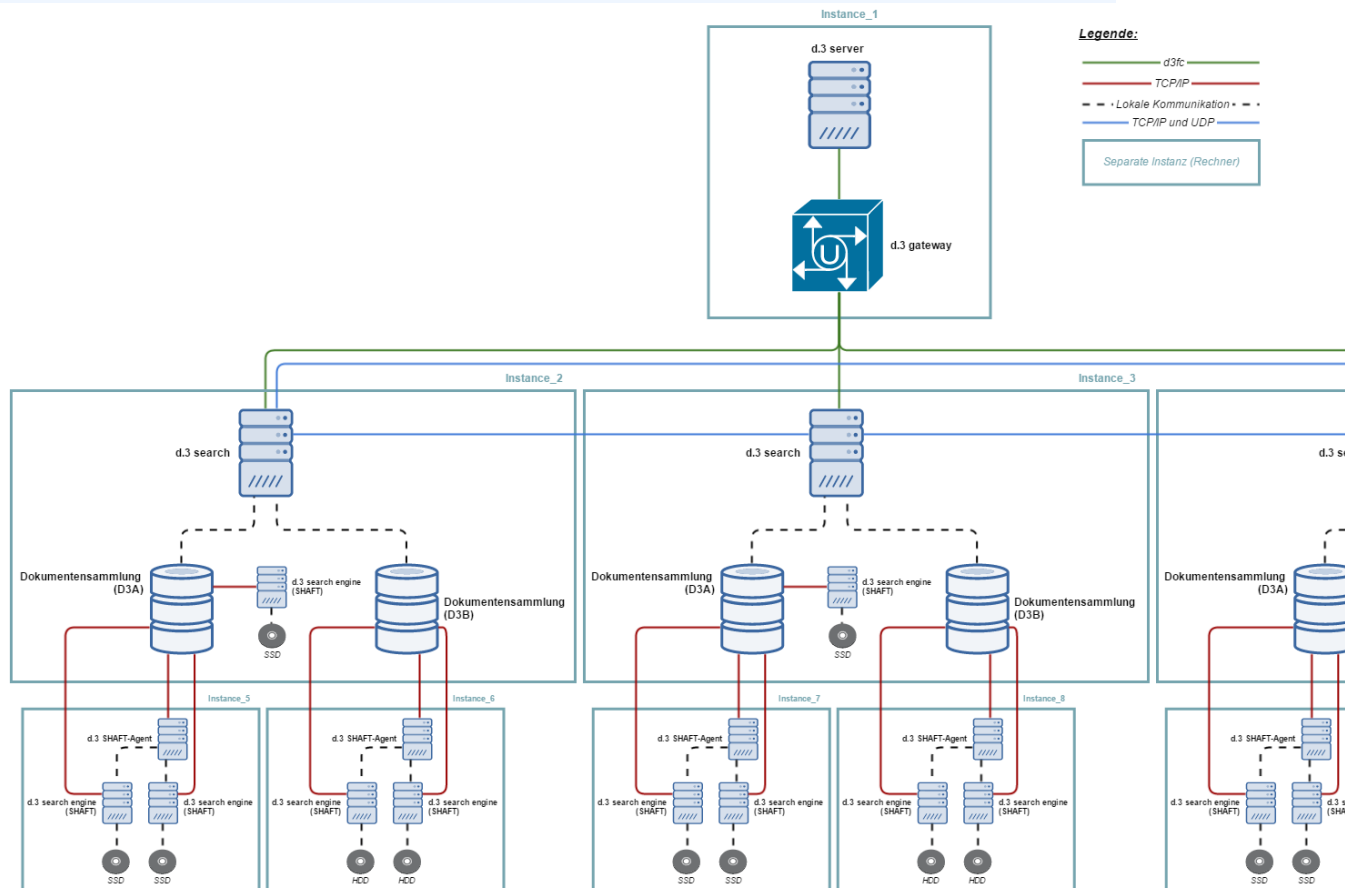
Skalierte und Cluster Installation

Bei der skalierten Cluster Installation werden mehrere, in diesem Fall drei, redundante d.3 search Instanzen installiert. Jede d.3 search Instanz wiederum weist skalierte Dokumentensammlungen auf. Bei der Dokumentensammlung D3A wurde eine d.3 search engine (SHAFT) auf der selben Instanz installiert, wie d.3 search selbst. Die übrigen zwei d.3 search engines (SHAFT) wurden, mittels des d.3 SHAFT-Agent, auf einer separaten Instanz eingerichtet. Die d.3 search engines (SHAFT) der Dokumentensammlung D3B befinden sich auf einer separaten Instanz. Alle d.3 search Instanzen kommunizieren über den d.3 gateway mit dem d.3 server. Zudem überwachen sich die d.3 search Instanzen gegenseitig, um bei einem Ausfall reagieren zu können.

In diesem Fall werden die Suchanfragen gleichmäßig durch den d.3 gateway auf die d.3 search Instanzen verteilt. Die Dokumentensammlung teilt sich auf die jeweils verfügbaren d.3 search engines (SHAFT) auf. Dieses Modell bietet sowohl eine optimale Ausfallsicherheit, als auch eine verbesserte Performance.

Anmerkung

Bei der Einrichtung eines Clusters ist es zwingend notwendig, dass die Kommunikation von d.3 server nach d.3 search über den d.3 gateway erfolgt. Wird dies nicht berücksichtigt, ist eine Ausfallsicherheit nicht gewährleistet.



1.2. Installation

Mit der d.3 search Installation wird die Verbindung vom d.3 Server über den d.3 search server zur Suchmaschine geschaffen, die letztlich die Textdaten verwaltet. Die Prozesskette erfordert die Installation des d.3 search servers, die Einrichtung einer neuen Dokumentensammlung für ein d.3 Archiv und die Aktivierung der d.3 search Unterstützung auf Seiten des d.3 servers.

Bitte beachten Sie die Hinweise für den Betrieb von d.3 search im Cluster.

Eine Beschreibung dieser Einzelschritte finden Sie in den folgenden Abschnitten.

1.2.1. Systemvoraussetzungen

Beachten Sie die zentralen [Systemvoraussetzungen für d.velop-Produkte \(On-Premises\)](#). Abweichende oder weiterführende Systemvoraussetzungen finden Sie in dieser Dokumentation.

d.3 search engine

Pro d.3 search engine-Instanz sollten für eine Installation mindestens 256 MB physikalischer Hauptspeicher vorhanden sein. Die Menge an benötigtem Platz ist von der Textmenge abhängig. Der Index ist ca. 1,4 mal so groß wie die Textmenge und benötigt entsprechend viel Speicherplatz. Um eine gute Performance zu gewährleisten, benötigt die Engine möglichst viel Cache und eine schnelle Festplatte. Nach Möglichkeit sollte der Cache mindestens ein Viertel der Größe des Index betragen. Optimal ist, wenn der Cache eine Trefferrate von 98 % erreicht. Für die Festplatte ist der IOPS-Wert bei wahlfreiem Zugriff auf

4 KB-Blöcke relevant. Wenn dieser Wert zwischen 150 und 200 liegt, reicht das System für maximal eine Million Dokumente. Bei der Wahl des Systems für einen Korpus sollten Sie den notwendigen IOPS-Wert nach folgender Formel berechnen: Mindestwert IOPS = Millionen Dokumente x 200. Für 50 Millionen Dokumente benötigen Sie somit ein System, das mindestens 10.000 IOPS erreicht.

1.2.2. Systemempfehlungen

Einleitung

Im Folgenden werden Systemempfehlungen für die Einrichtung einer Umgebung für d.3 search gegeben.

Anmerkung

Allgemein muss gesagt werden, dass eine konkrete Empfehlung aufgrund der verschiedenen Gegebenheiten (Anzahl Suchanfragen / Anzahl Dokumente / Größe der Dokumente etc...) nicht im Rahmen einer Dokumentation gegeben werden kann.

In diesem Kapitel soll Ihnen vermittelt werden, worauf Sie achten müssen und welche Größen ungefähr zu erwarten sind. Eine konkrete Planung kann nur für den Einzelfall betrachtet werden.

Grundsätzlich können die Ressourcen folgendermaßen absteigend priorisiert werden:

1. **Festplatte:** Die meiste Zeit wird bei Suchmaschinen für das Abrufen und Schreiben der Daten auf die Festplatte benötigt. Daher sollte hier eine dedizierte Festplatte verwendet werden, die im optimalen Fall nur von d.3 search verwendet wird. Um so performant wie möglich zu arbeiten, sollte als Typ der Festplatte eine SSD genutzt werden.
2. **Arbeitsspeicher:** Um die Geschwindigkeit der Festplatten zu überbrücken, sollte der Cache der d.3 search engine (SHAFT), mindestens 1/3 der Datenbankgröße betragen.
3. **CPU:** Die CPU wird durch d.3 search nicht außergewöhnlich stark beansprucht. Der Prozessor sollte prinzipiell auf dem aktuellen Stand der Technik sein und mindestens 4 Kerne aufweisen.

Bei der Einrichtung von d.3 search sollte zudem beachtet werden, welche Verzeichnisse Bedarf an Speicherplatz haben. Folgende Ablageorte benötigen Festplattenkapazität:

1. **Installationsverzeichnis:** Eine Installation von d.3 search benötigt einen Speicherplatz von ~116MB. Dazu zählen sämtliche Programmdateien, sowie die Dokumentation und der Uninstaller.
2. **Datenbankverzeichnis:** Das Datenbankverzeichnis enthält alle Dateien der Datenbank und ist von folgenden Faktoren abhängig:
 - a. **Anzahl der Dokumente:** Die Anzahl und Größe der jeweiligen Dokumente
 - b. **Inhalt der Dokumente:** Menge an OCR-Informationen eines Dokumentes
 - c. **Zusätzliche Speicherung der Volltextinhalte:** Wenn Sie die Volltextinhalte zusätzlich abspeichern, um Funktionen wie Phrasensuche, Semantische Suche, Textzusammenfassung zu ermöglichen, benötigt dies zusätzliche Festplattenkapazität
 - d. **Transaktions-Logs:** Das Speichern der Transaktionen in einem Log wird bei der Wiederherstellung des Indizes verwendet, wenn der Prozess unerwartet beendet worden ist. Die Größe des Logs ist abhängig von der Anzahl der Transaktionen zwischen den Vollbackup-Zeiten.
 - e. **Archive-Logs:** Mit Hilfe des Archive-Logs kann ein Index schneller wieder aufgebaut werden. Die Größe ist abhängig von der Anzahl der Transaktionen zwischen den Vollbackup-Zeiten.
 - f. **Automatische Reparatur:** Bei der automatischen Reparatur eines Indizes wird die Größe des aktuellen Indizes zusätzlich benötigt.
 - g. **Automatische Defragmentierung:** Für die Defragmentierung eines Index wird ein zusätzlicher Speicherplatz in Größe der Dateien `link.dat` und `words.dat` benötigt.
3. **Backupverzeichnis:** Für das Backupverzeichnis wird der Speicherplatz benötigt, die der Suchindex zum Zeitpunkt des Backups benötigt.

4. **Job-Verzeichnis:** Die Größe des Job-Verzeichnisses ist abhängig von der Anzahl der Dokumente, die in der Queue liegen. In diesem Ordner werden sowohl OCR-Informationen, als auch Informationen zu den Jobs hinterlegt.

Anmerkung

Im folgenden wird eine Beispiel-Installation von d.3 search gezeigt.

Das Beispiel enthält zum größten Teil kaufmännische Dokumente.

Folgende Eigenschaften sind gegeben:

Eigenschaft	Wert
Anzahl der Dokumente im Index	2.015.988
Anzahl der unterschiedlichen Worte im Index	51.049.071
Größe des Index	16,3 GByte
Davon Größe der zusätzlichen Speicherung der Volltextinhalte	4,4 GByte
Anzahl aktiver Benutzer	~ 200 Mitarbeiter

Empfohlene Architektur

Bei der Auswahl der Architektur gibt es zwei Ausrichtungen. Zum einen die Sicherheit, die durch ein Aktiv/Aktiv-Cluster erreicht werden kann und zum anderen die Skalierung.

Ein d.3 search System sollte im Clusterverbund betrieben werden, wenn Sie die Verfügbarkeit erhöhen wollen. Zudem hat ein Cluster den Vorteil, dass die Suchanfragen gleichmäßig auf alle Instanzen im Cluster verteilt werden, wodurch in gewissem Maße eine Lastverteilung stattfindet. Der Hauptaspekt eines Clusters sollte jedoch die Verfügbarkeit der Volltextrecherche sein.

Die Skalierung einer Dokumentensammlung ist spätestens erforderlich, wenn ein Index größer als 50 GByte ist. Ein Index mit 50 GByte ist, mit entsprechend leistungsstarker Hardware, das äußerste, was die d.velop für eine Dokumentensammlung empfiehlt. Am obigen Beispiel hoch gerechnet entspricht dies ca. 6 Mio. Dokumente.

Bei der Erstellung einer Dokumentensammlung sollten Sie sich bereits Gedanken über die mögliche Anzahl der Dokument in nächster Zeit machen. Es wird empfohlen einen Index nicht größer als 20 GByte werden zu lassen, um mit aktueller Hardware eine entsprechende Performance zu erreichen. Da die spätere Umstellung zu Skalierung einige Zeit in Anspruch nimmt (mehrere Tage), sollten Sie bereits bei der Erstellung einer Dokumentensammlung die Skalierung eher höher einstellen als benötigt.

Anmerkung

Sie haben ein jährliches Dokumenten-Aufkommen von 2 Mio. Dokumenten. Innerhalb von 10 Jahren wird Ihr Index, gemessen an den obigen Beispiel, ca. 163 GB groß sein. Um in den nächsten 10 Jahren die Skalierung nicht mehr ändern zu müssen empfiehlt es sich eine Skalierung auf 8 d.3 search engines (SHAFT) vorzunehmen.

1.2.3. Update

Sofern Sie d.3 search mit einer Version kleiner als 2.1 verwenden, beachten Sie bei einem Update folgendes:

Der Aufbau der Job-Dateien hat sich mit d.3 search Version 2.1 geändert. Der neue d.3 search server kann alte Job-Dateien nicht mehr interpretieren und verarbeiten.

Warnung

Vor einer Update-Installation muss also das Job-Verzeichnis des d.3 search servers in diesem Fall komplett abgearbeitet sein, da es keine Möglichkeit gibt, die alten Job-Dateien auf die neue Version zu migrieren.

Update bei bestehendem ClusterAktiv/Aktiv-Cluster

Wenn Sie bereits einen Aktiv/Aktiv-Cluster mit d.3 search aufgebaut haben, gehen Sie bei einem Update folgendermaßen vor:

- Fahren Sie eine d.3 search Instanz über den d.3 process manager herunter.
- Updaten Sie das ausgewählte System.
- Starten Sie die d.3 search Instanz über den d.3 process manager.
- Führen Sie die o.g. Schritte für alle weiteren d.3 search Instanzen durch.

Während des Updates kann die Suchfunktion von d.3 search weiterhin verwendet werden. Lediglich Änderungen der Datenbasis, an den Dokumentensammlungen, werden währenddessen nicht verarbeitet.

Aktiv/Passiv-Cluster

Wenn Sie bereits einen Aktiv/Passiv-Cluster mit d.3 search aufgebaut haben, gehen Sie bei einem Update folgendermaßen vor:

- Beenden Sie den d.3 process manager der aktuell **passiven** Instanz.
- Updaten Sie das ausgewählte System.
- Starten Sie den d.3 process manager.
- Beenden Sie den d.3 process manager der aktuell **aktiven** Instanz - die passive Instanz wird nach kurzer Zeit die aktive Rolle übernehmen.
- Updaten Sie das nun ausgewählte System.
- Starten Sie den d.3 process manager.

Während des Updates können alle Funktionen von d.3 search verwendet werden.

1.2.4. Installieren und Aktualisieren von d.3 search

Sie installieren die Software ausschließlich mit d.velop software manager. Wird eine Anwendung für verschiedene Produkte benötigt, werden die zugehörigen Softwarepakete ebenfalls automatisch installiert.

Weitere Informationen zum Installieren der Software finden Sie im d.velop software manager-Handbuch.

Bei der ersten Installation müssen Sie das Setup (**setup-dvelopSoftwareManager.exe**) aus dem d.velop service portal starten, um ein Produkt zu installieren. Sobald Sie d.velop software manager auf Ihrem Server installiert haben, können Sie weitere Änderungen vornehmen, indem Sie die Anwendung auf dem Server im Startmenü unter **d.velop > d.velop software manager** starten.

Anmerkung

Bei der Installation werden immer folgende Systemeinstellungen durch d.velop software manager auf dem Server angepasst:

- Anpassung des IPv4-Portbereichs für das TCP-Protokoll von 10.000 bis 55.000.
- Anpassung des IPv4-Portbereichs für das UDP-Protokoll von 10.000 bis 55.000.
- Anpassung des IPv6-Portbereichs für das TCP-Protokoll von 10.000 bis 55.000.
- Anpassung des IPv6-Portbereichs für das UDP-Protokoll von 10.000 bis 55.000.
- Anpassung des Registrierungsschlüssels **HKEY_LOCAL_MACHINE\SYSTEM\CurrentControlSet\Services\Tcpip\Parameters\TcpTimedWaitDelay** auf den Wert **30 (0x1e)**.

So geht's

1. Wählen Sie einen Feed aus.
2. Geben Sie die Adresse von d.ecs license server ein, falls der Lizenzserver nicht gefunden wurde. Sie können d.ecs license server auch direkt installieren, jedoch ist ein Neustart von d.velop software manager erforderlich.
3. Wählen Sie **Installieren** aus.
4. Wählen Sie die Produkte aus, die Sie installieren möchten.
5. Aktivieren Sie optional die Einstellungen für automatische Aktualisierungen.
6. Überprüfen Sie die Zusammenfassung und klicken Sie auf **Ändern**.

Sie aktualisieren die Software nur noch mit d.velop software manager.

Weitere Informationen zum Aktualisieren finden Sie im d.velop software manager-Handbuch.

1.2.5. d.3 gateway - Anbindung

Einleitung

Ab der Version 3.0.0 von d.3 search kann das Produkt über das d.3 gateway angesprochen werden. Das d.3 gateway verteilt die Anfragen nach dem Round-Robin Verfahren an die angemeldeten d.3 search Instanzen.

Sobald eine d.3 search Instanz ausfällt, erkennt dies das d.3 gateway und sendet kommende Anfragen nur noch zu den erreichbaren d.3 search Instanzen. Sobald die d.3 search Instanz wieder erreichbar ist, sendet das d.3 gateway Anfragen wieder zu der Instanz. Somit wird die Ausfallsicherheit durch das d.3 gateway verbessert.

Einstellungen in d.3 search

Damit sich d.3 search an das d.3 gateway anmeldet, müssen Einstellungen im Webinterface von d.3 search vorgenommen werden.

- Öffnen Sie dazu das Webinterface, welches standardmäßig unter der Adresse <http://localhost:6133> erreichbar ist. Sie können die Verknüpfung d.3 search Webinterface aus dem d.3 search Startmenü aufrufen.
- Öffnen Sie nun die Seite **Einstellungen | Global**.
- Auf dieser Seite richten Sie dann die folgenden drei Konfigurationsparameter, die sich in der Sektion **Gateway** befinden, ein (siehe auch Kapitel Konfiguration des d.3 search Servers):

HandlerCount	Gibt an, wie viele Verbindungen zum d.3 gateway aufgebaut werden sollen.
Host	Definiert den Hostnamen des d.3 gateways.
Port	Gibt den Port an, über dem der d.3 gateway erreichbar ist.

Der folgende Screenshot zeigt die zu setzenden Einstellungen im Webinterface:

Gateway

G HandlerCount	6	
G Host	d3-sql-712	
G Port	3400	

- Anschließend speichern Sie die Einstellungen über den Button **Speichern**.

Damit die Einstellungen gültig werden, muss der d.3 search server neu gestartet werden. Um die Verbindung über den d.3 gateway zu verwenden tragen Sie diesen in den Einstellungen vom d.3 server (d.3 admin) ein.

Anmerkung

Der d.3 search server ist weiterhin über den gängigen Port 3406 erreichbar. Die Anbindung an das d.3 gateway ist zusätzlich zu sehen.

Einstellungen im d.3 server (d.3 admin)

Um die Anfragen vom d.3 server über das d.3 gateway zu leiten müssen Sie folgende Einstellungen vornehmen:

- Starten Sie dazu den d.3 admin und öffnen Sie den Punkt **d.3 Komponenten | d.3 Config | Volltext-Suche (d.search)**.
- Tragen Sie nun unter dem Punkt **d.3 search Server** den Hostnamen des d.3 gateway ein.
- Schließlich geben Sie unter dem Punkt **TCP/IP Port für d.3 search** den Port, über den das d.3 gateway erreichbar ist, an. (Standard: 3400)

Nachdem die Einstellungen gespeichert wurden, ist d.3 search über das d.3 gateway erreichbar.

1.2.6. d.3 SHAFT-Agent

Einleitung

Ist Ihre aktuelle d.3 search Instanz in Hinsicht auf CPU und Arbeitsspeicher ausgelastet, haben Sie die Möglichkeit, mit dem d.3 SHAFT-Agent, die d.3 search engine (SHAFT) auf Remote-Recher auszulagern. Dadurch werden Ressourcen der aktuellen Instanz weniger in Anspruch genommen, was wiederum dazu führt, dass Anfragen schneller verarbeitet werden können.

Die CPU-Auslastung wird verringert, da die Verarbeitung der Anfragen von der d.3 search engine (SHAFT) auf den anderen Rechner vorgenommen werden und der verwendete Arbeitsspeicher geht zurück, da die d.3 search engine (SHAFT) den Suchindex auf dem separaten Rechner im Arbeitsspeicher hält. Somit sorgt die Verlagerung von Dokumentensammlungen auf einen separaten Rechner für geringere Anforderungen an die CPU und den Arbeitsspeicher auf der d.3 search Instanz.

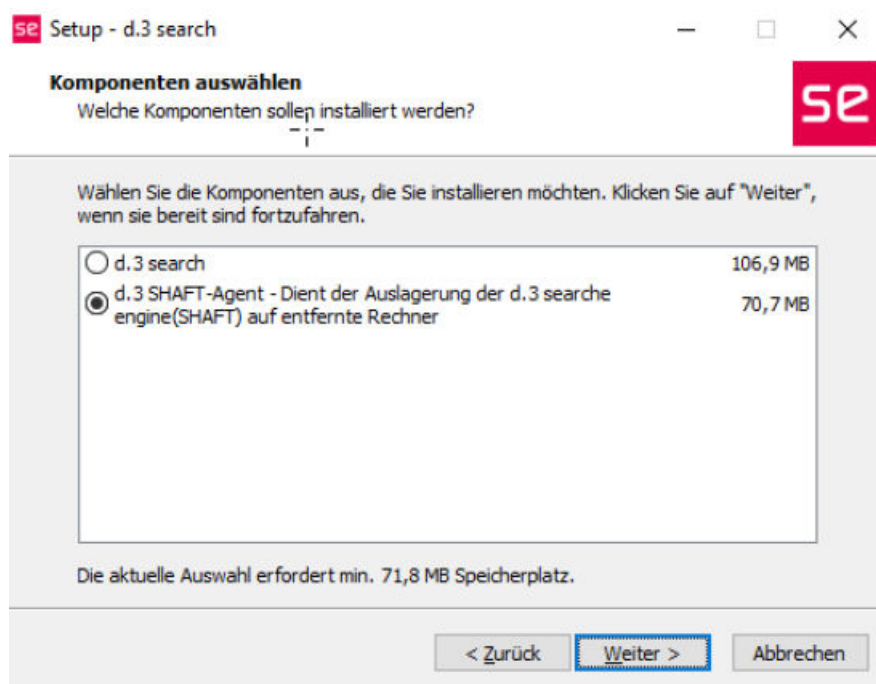
Der d.3 SHAFT-Agent dient bei der Auslagerung als Schnittstelle zwischen d.3 search engine (SHAFT) und dem d.3 search server und übernimmt die Remote-Steuerung der Prozesse und Datenübertragungen des Suchindex.

Sie sollten in folgenden Situationen eine Auslagerung der d.3 search engine (SHAFT) in Erwägung ziehen:

- die CPU-Auslastung der d.3 search Instanz ist ständig über 60%
- es ist kaum freier Arbeitsspeicher zur Verfügung und aus den PerfLogs geht hervor, dass die Cache-Ratio unter 80% liegt.

Installation und Einrichtung

- Starten Sie das Setup von d.3 search, um das Auslagern einer d.3 search engine (SHAFT) vorzubereiten.
- Auf der Seite **Komponenten auswählen** markieren Sie den Punkt **d.3 SHAFT-Agent**.
- Schließen Sie dann die Installation normal ab.



- Stellen Sie sicher, dass der d.3 SHAFT-Agent ordnungsgemäß im d.3 process manager eingetragen worden ist und gestartet wurde.
- Öffnen Sie nun das Webinterface auf dem Rechner, dessen d.3 search engine (SHAFT) Sie auslagern wollen. Dieses erreichen Sie standardmäßig unter der Adresse <http://localhost:6133/> oder rufen Sie die Verknüpfung d.3 search Webinterface aus dem d.3 search Startmenü auf.
- Dort wählen Sie den Menüpunkt **Einstellungen | Dokumentensammlungen | <CorpusId> | corpus.ini**, um die Konfiguration der Dokumentensammlung zu bearbeiten.
- Im letzten Schritt ändern Sie die Angaben in der Sektion **SHAFT1** auf die für Sie passenden Werte und speichern Sie diese schließlich ab.

Anmerkung

Bei der Auslagerung werden sämtliche Inhalte aus dem Datenbank-Verzeichnis der d.3 search engine (SHAFT) übertragen. Dazu zählt u.a. auch das Backup und die Transaktions- und Archivelogs. Sie können den Kopiervorgang beschleunigen, in dem Sie vor der Auslagerung ein manuelles Backup durchführen. Dadurch werden die Transaktions- und Archivelogs entfernt und müssen nicht übertragen werden.

Die Ausfallzeit kann durch eine manuelle Entfernung oder Verschiebung des Backups verkürzt werden. Verschieben Sie dazu das Backup-Verzeichnis außerhalb des Datenbank-Verzeichnis der d.3 search engine (SHAFT). Nach der Auslagerung können Sie das Backup-Verzeichnis manuell auf den neuen Rechner verschieben, oder ein neues Backup erstellen und das alte Entfernen.

Nach der Änderung der Einstellungen werden Sie dazu aufgefordert, d.3 search neu zu starten.

Der Start der Software kann nun, je nach Größe der Dokumentensammlung, einige Minuten dauern. Sobald die Dokumentensammlung an den d.3 SHAFT-Agent gesendet worden ist, ist d.3 search wieder voll funktionsfähig.

Den aktuellen Status des Sendevorgangs entnehmen Sie bitte aus dem d.3 log.

Konfiguration

In der SHAFTAgent.ini können Sie Einstellungen für den d.3 SHAFT-Agent vornehmen.

Sektion	Eintrag	Bedeutung
[Global]	Corpora-Path	Standard: \corpora Gibt den Pfad an, unter dem die Dokumentensammlungen zu finden sind.
	SHAFTPath	Standard (64-Bit Betriebssystem): shaft\64Bit Standard (32-Bit Betriebssystem): shaft Gibt den Pfad an, in dem die shaft.exe zu finden ist. Die gewählte shaft.exe wird für alle Dokumentensammlungen verwendet. Es ist nicht Möglich die shaft.exe pro Dokumentensammlung zu definieren.
	TCPPort	Standard: 6133 Gibt den TCPPort an, unter dem der d.3 SHAFT-Agent erreichbar ist.
	LogLevel	Standard: 6 Gibt das LogLevel für den d.3 SHAFT-Agent an. LogLevel=0: nur Fehler werden im Logviewer ausgegeben LogLevel=6: normale Protokollierung von d.3 search API Calls LogLevel=9: alle Meldungen (inkl. Debug-Ausgaben)

1.2.7. Cluster

Ab der Version 3.0.0 von dem Produkt d.3 search werden zwei verschiedene Clusterarten unterstützt. Cluster bieten die Möglichkeit die Ausfallsicherheit und die Skalierbarkeit eines Systems zu verbessern.

Unter Ausfallsicherheit ist bei d.3 search die Erhaltung der Suchfunktionalität zu verstehen.

Für den Aufbau eines Clusters existieren zwei Alternativen:

Aktiv/ Aktiv- Cluster	Dieser Aufbau eines Clusters bietet Ausfallsicherheit, sowohl beim Abbruch eines Prozesses/Rechners, als auch bei Fehlern in der Datenbasis, den Dokumentensammlungen. Der Benutzer kann während einen solchen Ausfalls weiterhin die Suchfunktion nutzen. Lediglich die Änderungen an der Datenbasis werden nicht verarbeitet bis alle Mitglieder des Cluster wieder aktiv sind. Bei einem Aktiv/Aktiv-Cluster werden Suchanfragen auf die jeweiligen Mitglieder verteilt. Dadurch wird eine Skalierung bei der Verarbeitung der Suchanfragen erreicht. Ein Aktiv/Aktiv-Cluster umfasst beliebig viele Mitglieder. Für weitere Informationen siehe Aktiv/Aktiv - Cluster
Aktiv/ Passiv- Cluster	Mit einem Aktiv/Passiv-Cluster erreichen Sie eine grundlegende Ausfallsicherheit, wenn ein Rechner bzw. der d.3 process manager nicht mehr erreichbar ist. Die Datenbasis kann beim Ausfall einer Instanz weiterhin geändert werden. Ein automatischer Schwenk findet nicht statt, wenn nur der d.3 search Prozess nicht mehr erreichbar ist. Zudem bietet dieser Aufbau keine Ausfallsicherheit, wenn die Datenbasis defekt ist. Die Clusterart umfasst immer genau zwei Mitglieder. Für weitere Informationen siehe Aktiv/Passiv - Cluster

Warnung

Da in einem Aktiv/Passiv-Cluster Fehler am Datenbestand zu einem Stillstand des Systems führen, wird die Nutzung eines Aktiv/Aktiv-Clusters empfohlen.

Um einen Cluster aufzubauen, können Sie eine bereits vorhandene d.3 search Instanz verwenden und mit wenigen Schritten für den Cluster vorbereiten.

Bevor man eine weitere Instanz in den Cluster aufnimmt, sollte man alle Konfigurationsarbeiten an der ersten Instanz abgeschlossen haben, damit die Einstellungen an der neuen Instanz direkt aktiv sind.

Aktiv/Aktiv - Cluster

Durch die Nutzung eines Aktiv/Aktiv-Clusters kann das d.3 search System ausfallsicher gestaltet werden. Dabei bietet der Aufbau eines d.3 search Clusters folgende Vorteile:

- Ein Cluster verteilt Suchanfrage gleichmäßig an alle beteiligten Rechnern; Änderungen an der Datenbasis, den Dokumentensammlungen, werden von jeder d.3 search Instanz synchron verarbeitet.
- Fällt ein Mitglied des Clusters aus, wird das Cluster inaktiv - dies verhindert Veränderungen an dem Datenbestand, bis das Cluster wieder vollständig aufgebaut ist.
- Die Suche kann während eines Ausfalls weiterhin verwendet werden, sodass die Funktionalität der Clients nicht behindert wird.

Hierbei gilt ein Mitglied des Clusters unter zwei Bedingungen als ausgefallen:

1. Die d.3 search engine (SHAFT) ist nicht gestartet.
2. Eine Instanz ist nicht erreichbar - Der Master erwartet alle fünf Sekunden eine Rückmeldung von dem Slave. Sobald die Rückmeldung zehn Sekunden überfällig ist, wird die Instanz als nicht erreichbar markiert.

Mitglieder eines Clusters befinden sich immer in einem der folgenden Modi:

Master	Im laufenden Clusterbetrieb gibt es immer einen Master. Dieser kontrolliert die Erreichbarkeit und den Status des Clusters und synchronisiert die Einstellungen mit allen Mitgliedern. Fällt der Master-Rechner aus, weist das Cluster automatisch dem nächsten Mitglied im Cluster den Master-Modus zu. Im Webinterface können Einstellungen nur über das Webinterface des Masters vorgenommen werden. Bei allen anderen ist der Speicher-Button deaktiviert und es erscheint eine Hinweismeldung.
Slave	Alle funktionsfähigen Mitglieder eines Clusters, die nicht den Modus Master haben, sind als Slave notiert. Ein Slave ist eine normal funktionsfähige d.3 search Instanz. Auf dem Slave kann das Webinterface für lokale Informationen ebenfalls aufgerufen werden. Für Einstellungsänderungen muss jedoch auf das Webinterface des Masters gewechselt werden.

Unbekannt	Ein Mitglied mit dem Modus Unbekannt ist zurzeit heruntergefahren, oder antwortet aus anderen Gründen nicht auf die Anfragen des Masters. Sobald ein Mitglied diesen Status annimmt ist das Cluster inaktiv und es werden keine Änderungen an der Datenbasis mehr vorgenommen. D.h. alle Jobs vom d.3 async werden mit dem Fehlercode 4100011 beantwortet und müssen anschließend durch den d.3 async job inspector manuell wieder angestoßen werden.
Wird synchronisiert	In diesem Modus befinden sich Rechner unter zwei Bedingungen: Im ersten Fall wurde die Clusterfunktion gerade aktiviert, bzw. die Instanz startet gerade. Solange der Rechner in diesem Modus steht ist er dabei, im Netzwerk nach einem möglichen Cluster zu suchen. Dieser Vorgang dauert maximal 15 Sekunden. Im zweiten Fall wechselt ein Mitglied in diesen Modus, wenn es von einem Master im Cluster synchronisiert/hinzugefügt wird. Dabei werden sämtliche Konfigurationen und Datenbestände übertragen.

Die jeweiligen aktuellen Modi der Rechner können Sie jederzeit über das d.3 search Version 3.0.0 Webinterface unter dem Menüpunkt **Cluster** einsehen. Dieses ist standardmäßig unter der Adresse `http://localhost:6133/` erreichbar.

Bestehende d.3 search Instanz für das Cluster einrichten

Um eine bestehende d.3 search Instanz für ein Cluster einzurichten führen Sie die folgenden Schritte durch:

- Rufen Sie auf Ihrer bestehenden d.3 search Instanz das d.3 search Webinterface auf. Dieses finden Sie unter der Adresse `http://localhost:6133` oder rufen Sie die Verknüpfung d.3 search Webinterface aus dem d.3 search Startmenü auf.

Anmerkung

Die Zugangsdaten zum Webinterface müssen bei allen Clustermitgliedern übereinstimmen.

- Stellen Sie sicher, dass die d.3 gateway Anbindung eingerichtet ist (siehe Kapitel [d.3 gateway - Anbindung](#)).

Warnung

Geben Sie als Namen für den Gateway-Host den Hostnamen des Computers und niemals `localhost` an!

- Aktivieren Sie nun das Cluster im Webinterface unter **Einstellungen | Global**.
- In der Sektion **Cluster** setzen Sie nun die Checkbox bei **Active**. Hier müssen Sie einen Gruppenname vergeben. Computer, die gemeinsam ein Cluster bilden sollen, müssen zwingend den gleichen Gruppennamen besitzen. Auch der Port muss bei jedem Computer des Clusters identisch gesetzt sein.

Weitere Information über die Einstellungsmöglichkeiten finden Sie im Kapitel [Konfiguration des d.3 search Servers](#).

Nach diesem Schritt muss d.3 search neu gestartet werden. Dies wird ebenfalls vom Webinterface gefordert.

Anmerkung

In Hinsicht auf den Gruppennamen ist zu bedenken, dass dieser möglichst einmalig sein sollte. Rechner mit aktivierter Clusterfunktion und gleichem Gruppennamen können sich automatisch über einen UDP-Broadcast finden und fügen sich automatisch zum bestehenden Cluster hinzu. Dabei werden alle Suchindices übertragen, die möglicherweise sensible Daten enthalten.

Cluster

L Active	☒	
G Group	Testgroup	
G Port	6134	

Nach der Aktivierung des Clusters ist nun im Webinterface der Menüpunkt **Cluster** gelistet.

- Rufen Sie über diesen die Clusterinformationen ab. Zu diesem Zeitpunkt sollte in der Sektion **Cluster-Server** Ihr lokaler Computer mit dem Modus Master gelistet sein.

Einstellungen

Automatische Aktualisierung ☒

Cluster-Server

IP-Adresse	Rechnername	Modus	Letzte Meldung	Information	Fehlermeldungen	
10.3.4.91	wathe01	Slave	16.02.2016 11:26:25			Lösch
10.95.0.80	d3-sql-712	Master	16.02.2016 11:26:25			Lösch
						Hinzufügen

Werden die **Cluster Informationen** richtig angezeigt, haben Sie erfolgreich Ihre bestehende Instanz cluster-fähig konfiguriert.

Nun fehlt noch das Hinzufügen von einer oder mehreren neuen Instanzen (siehe Kapitel [Neue d.3 search Instanz für das Cluster einrichten](#)).

Warnung

Das Konfigurieren und Starten der neuen Instanzen muss unbedingt nach dem Konfigurieren und Starten der bestehenden d.3 search Instanz geschehen.

Die Reihenfolge, in der die Server dem Cluster hinzugefügt werden, entscheidet über die Rollen, die die Instanzen einnehmen.

Neue d.3 search Instanz für das Cluster einrichten

Für jede neue Instanz ist das Vorgehen gleich und ähnelt dem Konfigurieren der bestehenden Instanz.

- Öffnen Sie auf der neuen Instanz das d.3 search Webinterface unter der Adresse `http://localhost:6133` oder über das Programm d.3 search Webinterface aus dem d.3 search Startmenü.

Anmerkung

Die Zugangsdaten zum Webinterface müssen bei allen Clustermitgliedern übereinstimmen.

- Wählen Sie nun den Menüpunkt **Einstellungen | Global**.
- Aktivieren Sie die Clusterfunktionalität, indem Sie in der Sektion **Cluster** den Punkt **Active** bestätigen.
- Geben Sie dann den Gruppennamen des Clusters an.
- Überprüfen Sie, ob der Port identisch dem der bestehenden Instanz ist.
- Im letzten Schritt überprüfen Sie nun, ob die Clusterrechner sich schon automatisch verbunden haben.
- Rufen Sie hierfür den Menüpunkt **Cluster** auf. In der Sektion **Cluster-Server** sollte die neue Instanz mit dem Modus **Wird synchronisiert** gelistet sein.

Cluster-Server

IP-Adresse	Rechnername	Modus	Letzte Meldung	Information	Fehlermeldungen
10.95.4.231	dsearcht2	Master	11.05.2017 11:21:48		
10.95.5.146	dsearcht1	Slave	11.05.2017 11:21:48		
10.95.2.225	dsearchdc1	Slave	11.05.2017 11:21:48		• 11.05.2017 11:21:48

In den ersten fünfzehn Sekunden nach Start der neuen Instanz versucht diese, eine schon konfigurierte, bestehende d.3 search Instanz zu finden und automatisch den Austausch der Konfigurationen und Dokumentensammlungen zu starten.

Warnung

Für diese automatische Synchronisation muss zwingend ein UDP-Broadcast zwischen der neuen Instanz und der bestehenden d.3 search Instanz gesendet werden können.

Ist dies nicht möglich, muss die zweite Instanz manuell hinzugefügt werden. Dies erreichen Sie, indem Sie das Webinterface Ihrer bestehenden Instanz aufrufen und dort unter dem Menüpunkt **Cluster** in der Sektion **Cluster-Server** den Button **Hinzufügen** betätigen. Geben Sie nun entweder die IP-Adresse oder den Hostnamen der neuen Instanz ein und bestätigen Sie Ihre Eingabe mit der **<Return>**-Taste.

Nutzungsanweisungen**Ausfall**

Eine d.3 search Instanz gilt als ausgefallen:

- wenn sie nicht über den TCP-Port des Clusters angesprochen werden kann.
- wenn eine oder mehrere d.3 search engines (SHAFT) nicht gestartet sind.

Einen Ausfall können Sie an den folgenden Kriterien erkennen:

- Der d.3 object monitor meldet:

The Cluster [**<Cluster Gruppennamen>**] is offline. Error servers: **<IP-ADDRESS>** (**<HOSTNAME>**)

- Im d.3 Log befindet sich eine der folgenden Meldungen:

```

! 21.08 09:49:56,519 DSEARCH 260C2F20 > The Cluster [SearchCluster] is offline. Error servers: d3-re
! 21.08 09:49:56,519 DSEARCH 260C2F20 > n-sto-712 (d3-ren-sto-712)

x 21.08 09:51:15,857 D3Server 00000FA0 d.search returned an error (4100011)
x 21.08 09:51:15,860 D3Server 00000FA0 Error on adding OCR data for 'P000000147' to d.search (4100011)

```

- Im Webinterface von d.3 search unter **Home** steht für das Cluster auf **Aktiv = false**:

Cluster information

Activ	false
Cluster name	SearchCluster
Server count	2
Status	Is_Master

- Im Webinterface von d.3 search unter **Cluster** befinden sich im Log die folgenden Meldungen:

```

21.08.2014 10:45:20: [SEER]: d3-ren-sto-712 - Timeout (IAMALIVEMASTER)
21.08.2014 10:45:32: [SEER]: d3-ren-sto-712 - Timeout (IAMALIVEMASTER)

```

Bei einem Ausfall des Masters wird automatisch ein neuer Master im Cluster ermittelt, der anschließend für Änderungen an der Konfiguration zuständig ist.

Wenn eine Instanz ausfällt, können keine Jobs von d.3 search verarbeitet werden. D.h., innerhalb dieser Zeit finden keine Änderungen an der Datenbasis, den Dokumentensammlungen, statt.

Fällt eine Instanz für längere Zeit (mehrere Tage) aus, empfiehlt es sich, die Instanz aus dem Cluster zu entfernen, um die Jobs nicht aufstauen zu lassen. Weitere Informationen dazu finden Sie im Kapitel [Clustermitglied entfernen](#).

Sobald das Cluster nicht aktiv ist, sollte der Administrator die Instanz möglichst rasch lauffähig bekommen. Mögliche Fehler und Lösungsansätze finden Sie im Kapitel [Aktiv/Aktiv-Cluster ist inaktiv](#).

Wenn Sie das defekte Mitglied aus dem Cluster entfernt haben (siehe Kapitel [Clustermitglied entfernen](#)) und anschließend wieder zu dem Cluster hinzufügen wollen, gehen Sie bitte vor wie im Kapitel [Neue d.3 search Instanz für das Cluster einrichten](#) beschrieben.

Sobald alle Clustermitglieder wieder aktiv und funktionsfähig sind, schaltet sich das Cluster aktiv und die Jobs werden verarbeitet.

Sie erkennen die Statusänderung an folgenden Meldungen:

- Der d.3 object monitor meldet:

```

d.3 search is active since <Datum>. <Anzahl Calls> calls handled. Pending
jobs(<Anzahl Add-Jobs> add, <Anzahl Update-Jobs> update, <Anzahl Delete-
Jobs> delete)

```

- Im d.3 Log werden Sie diese Meldung vorfinden:

```

i 21.08 10:49:47,761 DSEARCH 34682F20 > Cluster-Status changed: False ->True (One cluster member is
i 21.08 10:49:47,763 DSEARCH 34682F20 > not active)
i 21.08 10:49:47,763 DSEARCH 34682F20 The Cluster is Online

```

- Im Webinterface von d.3 search unter **Home** steht das Cluster auf **Aktiv = true**:

Cluster Informationen

Aktiv	true
Cluster-Name	SearchCluster
Anzahl Server	2
Status	Is_Master

- Im Webinterface von d.3 search unter **Cluster** befinden sich im Log die folgenden Meldungen:

21.08.2014 11:34:31: [SEOK]: d3-ren-sto-712 - IAMALIVEMASTER

21.08.2014 11:34:38: [SEOK]: d3-ren-sto-712 - IAMALIVEMASTER

- Änderungen an der Datenbasis, den Dokumentensammlungen, werden wieder vorgenommen - Jobs von d.3 server werden erfolgreich verarbeitet.

Die im d.3 async auf Fehler gelaufenen Jobs können Sie über den d.3 async job inspector im d.3 admin wieder starten.

Gehen Sie dazu wie folgt vor:

- Starten Sie den d.3 admin.
- Öffnen Sie den Punkt **d.3 Komponenten | d.3 async job inspector**.
- Geben Sie bei den Suchvorgaben folgende Werte an:

Suchvorgaben:					
Job-Nummer:		Status-ID:		Fehler:	4100011
Besitzer:		Job-Typ-ID:	OCR001 - (Verschlagwortung)	Dokument-ID:	

- Wählen Sie den Button **Suchen**.
- Markieren Sie alle gefundenen Jobs
- Wählen Sie anschließend den Button **Erneut Senden**.

Clustermitglied entfernen

Wenn Sie ein Mitglied aus dem Cluster entfernen wollen, gehen Sie wie folgt vor:

- Beenden Sie den d.3 search Prozess, der aus dem Cluster entnommen werden soll, über den d.3 process manager.
- Öffnen Sie das Webinterface der d.3 search Instanz, die aktuell den Modus Master hält. Das Webinterface erreichen Sie unter der Adresse `http://localhost:6133/` oder Sie können die Verknüpfung d.3 search Webinterface aus dem d.3 search Startmenü aufrufen.
- Navigieren Sie zu dem Punkt **Cluster**.
- Wählen Sie bei der entsprechenden Instanz **Löschen**. Sie können nur Instanzen entfernen, die heruntergefahren sind.

Cluster-Server

IP-Adresse	Rechnername	Modus	Letzte Meldung	Information
10.95.0.80	D3-SQL-712	Master	21.08.2014 11:47:24	<button>Löschen</button>
10.95.0.79	d3-ren-sto-712	Unbekannt		<button>Löschen</button>

- Deinstallieren Sie die heruntergefahrte d.3 search Instanz über **Software | Programme und Funktionen | d.3 search**.

Warnung

Die aus dem Cluster entfernte Instanz darf nicht wieder neu gestartet werden, da ansonsten ein inkonsistenter Status vorliegen würde.

Daher löschen Sie auf jeden Fall die entsprechende Instanz und führen Sie eine Deinstallation durch.

Aktiv/Passiv - Cluster

Erstellen Sie für d.3 search ein Aktiv/Passiv-Cluster, um die Ausfallsicherheit bei Softwareaktualisierungen und Rechner-Abstürze zu gewährleisten. Ein Aktiv/Passiv-Cluster besteht immer aus zwei d.3 search Instanzen, welche sich mithilfe zweier d.3 process manager gegenseitig kontrollieren und einer zentralen Netzwerkfreigabe, auf der der Datenbestand gespeichert wird. Wird nun eines der Systeme, z.B. für Updates, heruntergefahren, übernimmt automatisch die andere Instanz.

Beide d.3 search Instanzen greifen dabei auf den gleichen Datenbestand zu und müssen somit eine gemeinsame Netzwerkfreigabe haben.

Wichtig hierbei ist, dass immer nur eine Instanz zur selben Zeit laufen darf.

Anmerkung

Das Aktiv/Passiv-Cluster ist eine grundlegende Form des Clusters. Ein Aktiv/Aktiv-Clusters bietet sehr viel mehr Stabilität und Funktionen. Falls Sie planen, ein Cluster aufzubauen, empfehlen wir Ihnen die Aktiv/Aktiv-Struktur zu verwenden.

Einrichtung

Für die Einrichtung werden zwei Rechner, ein primärer und ein sekundärer, mit installiertem d.3 search vorausgesetzt. Beide Instanzen müssen in jeweils einem d.3 process manager integriert sein. Die Dokumentensammlung befindet sich optimaler Weise auf einem externen Datenvolumen und muss von beiden Instanzen aus erreichbar sein.

Anmerkung

Der sekundäre Rechner kann mit weniger Ressourcen (Arbeitsspeicher, CPU, Festplattenkapazität) ausgestattet sein, da dieser nur im Notfall, wenn das primäre System ausfällt, zum Einsatz kommt.

Folgende Schritte müssen bei der Einrichtung durchgeführt werden:

- Beenden Sie beide d.3 search Instanzen.
- Legen Sie auf dem Datenvolumen, das den Datenbestand halten soll, einen neuen Ordner an; z.B. C:\d3\corpora\.
- Wenn bereits Dokumentensammlungen bestehen, kopieren Sie diese in den soeben angelegten Ordner.
- Löschen Sie das bestehende Corpora-Verzeichnis auf den beiden d.3 search Rechnern, um im nächsten Schritt die Netzwerkverknüpfung anlegen zu können.
- Nun legen Sie bei beiden d.3 search Rechnern eine Verknüpfung zum Netzwerkpfad an. Dazu wechseln Sie zu einem der beiden d.3 search Rechner und öffnen die Konsole. Verwenden Sie nun den Befehl:

Syntax

```
mklink /d "<LOKALER PFAD>" "<NETZWERKPFAD>"
```

Beispiel

```
mklink /d "C:\d3\d.3 search\corpora\" "\\D3-CORPORA\C$\d3\corpora\"
```

Anmerkung

Der Benutzer, der den Prozess d.3 search ausführen wird, muss Zugriffsberechtigungen auf die Netzwerkfreigabe besitzen.

- Öffnen Sie nun die Einstellungen (**Settings**) des d.3 process managers auf beiden d.3 search Instanzen nacheinander und aktivieren Sie jeweils **Enable failover support**.
- Geben Sie dann jeweils den anderen Rechner an und definieren Sie ein identisches Passwort bei **Remote access password**.

Settings

☒ Enable failover support

Host: [Empty field]

Port: 3480

Remote access user: admin

Remote access password: [Masked password]

[Save] [Cancel]

d.velop AG d.3 process manager 4.3.1 (196) © 2001-2013 | D3-CLIENT-712

- Definieren Sie nun eine sekundäre d.3 search Instanz, indem Sie auf einem der beiden Rechner den d.3 process manager öffnen und dort die Einstellungen der Instanz mit einem Klick auf **d.3 search (Server)** konfigurieren.
- Selektieren Sie dann in der Sektion **Settings** den Punkt **Only execute if the secondary system is down**.

Settings

Group: d.3 search

Caption: d.3 search (server)

Target: C:\d3\d.3 search\ldsearchsvr.exe

Working path: C:\d3\d.3 search\

Start delay: 0,00 sec.

Min Runtime: 0,00 sec.

Execute: RUN

Priority: IDLE

Execution times:

☒ Su	☒ Mo	☒ Tu	☒ We	☒ Th	☒ Fr	☒ Sa	00:00:00	-	00:00:00
☐ Su	☐ Mo	☐ Tu	☐ We	☐ Th	☐ Fr	☐ Sa	00:00:00	-	00:00:00
☐ Su	☐ Mo	☐ Tu	☐ We	☐ Th	☐ Fr	☐ Sa	00:00:00	-	00:00:00

Failover: ☒ Only execute if the secondary system is down.

[Save] [Cancel]

Nachdem Sie diese Einstellungen abgespeichert haben, erscheint ein Monitorsymbol vor dem d.3 search Prozesseintrag.



Auf der primären d.3 search Instanz können Sie den d.3 search Prozess nun starten.

Sie haben nun ein funktionsfähiges Aktiv/Passiv-Cluster aufgebaut.

Nutzungsanweisungen

Clusterschwenk

Ein Clusterschwenk kann manuell vorgenommen werden oder wird automatisch vom d.3 process manager getätigt, sobald der zweite d.3 process manager nicht mehr über das Netzwerk erreichbar ist.

Wenn Sie beabsichtigen, eine d.3 search Instanz upzudaten, nehmen Sie einen manuellen Clusterschwenk vor.

Dazu beenden Sie den d.3 process manager der Instanz, die Sie updaten wollen.

Weitere Informationen zum Update von d.3 search in einem Aktiv/Passiv-Cluster finden Sie im Kapitel [Update](#).

Automatisch wird ein Clusterschwenk vorgenommen, wenn:

- der d.3 process manager nicht über den definierten Port erreichbar ist: weitere Informationen finden Sie im Kapitel [Einrichtung](#)
- der d.3 process manager heruntergefahren wird
- der Rechner der d.3 search Instanz heruntergefahren wird

Ein automatischer Clusterschwenk bedeutet, dass es in irgendeiner Form einen Ausfall gegeben hat.

Wie Sie bei einem Ausfall handeln sollten, finden Sie im Kapitel [Ausfall](#).

Ausfall

Ein Ausfall eines Aktiv/Passiv-Clusters tritt dann auf, wenn die beiden d.3 process manager, in denen die d.3 search server eingetragen sind, nicht mehr miteinander kommunizieren können. Einen Ausfall können Sie anhand von folgenden Merkmalen erkennen:

- ein unerwarteter automatischer Clusterschwenk hat stattgefunden
- der d.3 object monitor meldet, dass ein d.3 process manager nicht mehr erreichbar ist; zur Einrichtung des d.3 process managers in den d.3 object monitor schauen Sie bitte in das Handbuch von d.3 object monitor.
- der entsprechende Rechner ist nicht gestartet
- der entsprechende d.3 process manager ist nicht gestartet

Wenn ein Ausfall eingetreten ist, muss dafür gesorgt werden, dass die d.3 process manager wieder miteinander kommunizieren können. Dazu bearbeiten Sie bitte folgende möglichen Fehlerquellen:

- Der Rechner, auf dem der d.3 process manager läuft, ist nicht gestartet.
- Der d.3 process manager ist auf dem Rechner nicht gestartet. Dieser kann entweder als Dienst oder als Prozess gestartet werden.
- Der Host/Port, der im d.3 process manager unter **Einstellungen** (Settings) eingestellt worden ist, ist von außen nicht erreichbar.

Wenn eine Instanz ausfällt, übernimmt das zweite System alle Funktionalitäten. D.h., der Anwender kann weiterhin Suchen vornehmen und Änderungen der Datenbasis können ebenfalls weiterhin stattfinden.

Wenn die fehlerhafte Instanz wieder vom anderen d.3 process manager erreicht werden kann, wird automatisch ein Clusterschwenk vorgenommen, so dass das sekundäre System nur aktiviert wird, wenn das primäre System nicht erreichbar ist.

Einrichtungsbeispiele

Zwei-Knoten-Cluster

Im Folgenden werden die Konfigurationsbeispiele für ein Zwei-Knoten-Cluster aufgelistet.

Auf beiden Knoten sind die Konfigurationen identisch:

Cluster	☒ Active	☒	
	Group	Clustergroup	
	Port	6134	
Gateway	HandlerCount	6	
	Host	d3-sql-712	
	Port	3400	
Communication	SSLCertFile	server.crt	
	SSLKeyFile	server.key	
	SSLPassword		
General	LogFile		
	LogLevel	9	
	MinRelevance	0	
	OMOMsgDist	10	
	Port	3406	
	WantHints	☒	
	JobQueue	JobThreadCount	100
Settings	MaxJobsForRead	-1	
	MinutesBetweenReadJobs	-1	
	CorporaPath	.\corpora\	
Webinterface	EnginesPath	.\engines\	
	JobsPath	.\jobs\	
	Password		
	Port	6133	
	SSLPort	6135	
Username	admin		
UseSSL	☐		

Dokumentensammlung: d3P

dsearch.ini

General		
 DestroyJobs	☒	
 Engine	shaft	
 JobSynchron	☐	
 ShaftUTLPath	.\\shaft\\64Bit\\	

corpus.ini

General		
 CurrentVersionsOnly	☒	
 UseFileStream	☐	

Search		
 Expand Search	☐	
 Expand SearchMinResult	20	

Shaft_Global		
 ForceNoHints	☐	
 UTF8Support	☒	

SHAFT1		
 Host	localhost	
 Port	3546	
 SHAFTAgentPort	6133	

SHAFT2		
 Host	localhost	
 Port	3547	
 SHAFTAgentPort	6133	

SHAFT3		
 Host	localhost	
 Port	3548	
 SHAFTAgentPort	6133	

shaft.ini

Backup		
 CheckDay	0	
 CheckDB	☒	
 Time	00:00	

Engine			
	ArchiveLogs	☐	
	AutoCorrect	☒	
	AutoDefrag	☒	
	AutoRepair	☒	
	CacheSize	256000	
	CommitCount	200	
	Cursor	64	
	WordRevalue	150	
	Delimitermode	1	
	FirstWildCard	3	
	IcuExcludedDelimiter	-	
	Journal	☒	
	LogLevel	9	
	MaxDocsForHints	100	
	MaxFilesPerWord	1000000	
	MaxWordsForHints	100	
	MinWordLen	2	
	Numeric	☒	
	Orderbyrank	☒	
	WildCardLimit	1000000	
	SemanticExtend	☐	
	StoreText	☒	
	UseMostRecent	☒	
PerfLogs			
	Abstract	5000	
	Delete	5000	
	Insert	5000	
	Search	5000	
	Spellcheck	5000	

Zwei-Knoten-Cluster mit einer ausgelagerten d.3 SHAFT-Engine

Bei einem Zwei-Knoten-Cluster, in dem Sie eine d.3 SHAFT-Engine auslagern wollen, ist die Konfiguration größtenteils gleich.

Um mehr über die Auslagerung einer d.3 SHAFT-Engine zu erfahren, lesen Sie das Kapitel [SHAFT-Agent \(SHAFT-Engine auslagern\)](#).

Bei dem Rechner mit der auszulagernden Engine wird lediglich ein neuer **Host** angegeben, statt bei **Host** den Wert **localhost** zu verwenden.

Dokumentensammlung: d3P

corpus.ini

SHAFT1	Host	d3-sql-712	
	Port	3547	
	SHAFTAgentPort	6133	

1.2.8. Integration im d.ecs monitor

Einleitung

Wenn d.ecs monitor agent auf dem System installiert ist, werden regelmäßige Statusmeldungen an d.ecs monitor gesendet, um den Administrator über Probleme zu informieren.

Folgende Meldungen werden von d.3 search gesendet:

- Start/Stop von d.3 search server
- TCP/IP-Listen-Port konnte nicht initialisiert werden.
- d.3 search Lizenz ist nicht vorhanden
- Fehlermeldungen, die bei der Initialisierung von d.3 search server auftraten und zum Herunterfahren des Prozesses führen.
- Im Clusterbetrieb
 - Das Cluster ist inaktiv.

Für eine bessere Überwachung werden zudem folgende Monitoring-Objekte an d.ecs monitor gesendet:

Aktualisierungsintervall: 5 Minuten

- Anzahl Add-/Update-/Delete-Jobs in der Queue

Aktualisierungsintervall: 60 Minuten

- Ausreichend Arbeitsspeicher pro d.3 search engine (SHAFT)
- Ausreichend Speicherplatz für die automatische Defragmentierung pro d.3 search engine (SHAFT)
- Ausreichend Speicherplatz für die automatische Reperatur pro d.3 search engine (SHAFT)
- Größe der Datenbank pro d.3 search engine (SHAFT)
- Dauer des Backups pro d.3 search engine (SHAFT)
- Fragmentierungsgrad der Datenbank pro d.3 search engine (SHAFT)
- Stunden bis zum letzten erfolgreichen Backup pro d.3 search engine (SHAFT)

Anmerkung

Wenn eine Dokumentensammlung hinzugefügt/geändert/entfernt wird, werden die Monitoring-Objekte die sich auf eine d.3 search engine (SHAFT) beziehen erst nach einem Neustart von d.3 search korrekt zum d.ecs monitor übermittelt.

1.2.9. SSL-Verschlüsselung

Einrichtung

Um alle Verbindungen, die mit d.3 search aufgebaut werden, zu verschlüsseln, gehen Sie wie folgt vor:

Eine verschlüsselte Kommunikation bei direkter Kommunikation zwischen d.3 server und d.3 search ist nicht möglich.

Wenn Sie die SSL-Verschlüsselung wieder deaktivieren wollen, gehen Sie bitte genauso wie bei der Aktivierung vor, aber deaktivieren Sie die Schalter anstatt sie zu aktivieren.

→ Verschlüsseln Sie das Webinterface, indem Sie den Schalter `[webinterface]UseSSL` in der `dsearch.ini` aktivieren.

Weitere Informationen finden Sie im Kapitel [Konfiguration des d.3 search servers](#).

→ Verschlüsseln Sie die Kommunikation zwischen d.3 server und d.3 search über das d.3 gateway.

→ Überprüfen Sie, ob in der Datei <Windowsinstallationsverzeichnis>\dlink.ini der [D3] SecretKey des d.3 gateways gesetzt ist.

→ Kopieren Sie die dlink.ini in das Windows-Verzeichnisses jeder d.3 search-Instanz und starten Sie die Instanzen daraufhin neu.

→ Kopieren Sie die public.key-Datei aus dem Installationsverzeichnis des d.3 gateways, über den der d.3 server kommunizieren soll, in das Installationsverzeichnis des d.3 servers und starten Sie diesen neu. Nun wird die Verbindung zwischen dem d.3 server zur ersten d.3 search-Instanz verschlüsselt.

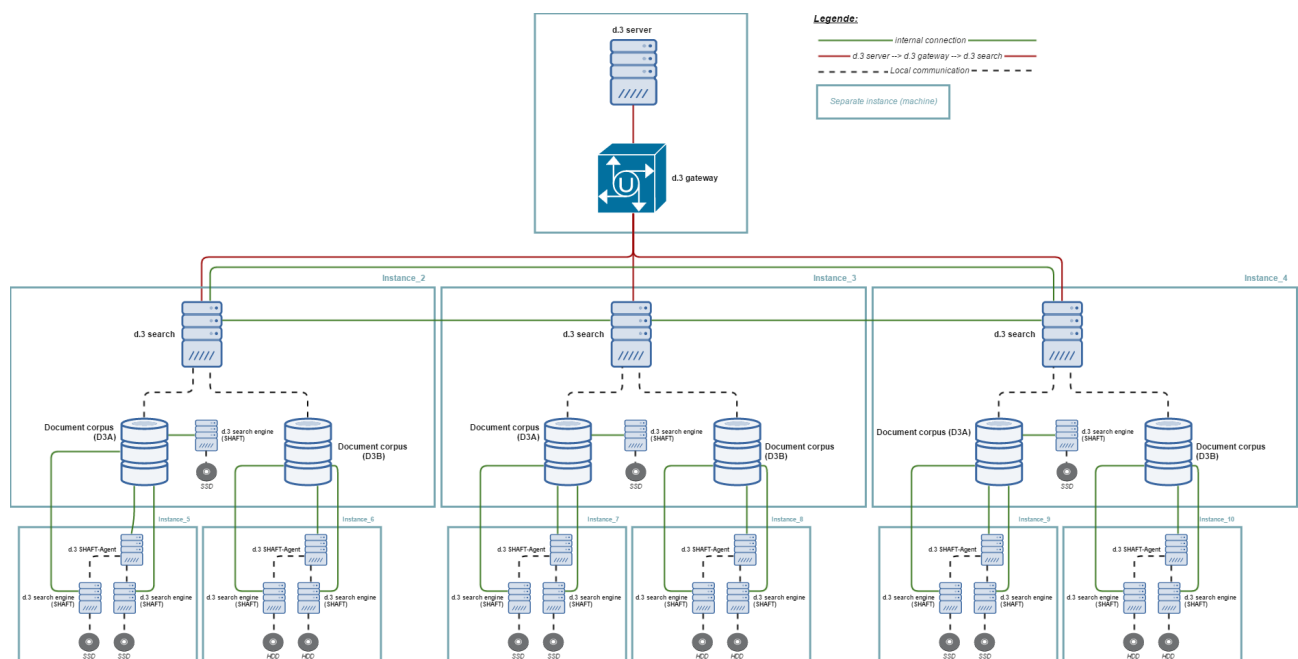
Anmerkung

Eine verschlüsselte Kommunikation bei direkter Kommunikation zwischen d.3 server und d.3 search ist nicht möglich.

Wenn Sie die SSL-Verschlüsselung wieder deaktivieren wollen, gehen Sie bitte genauso wie bei der Aktivierung vor, aber deaktivieren Sie die Schalter anstatt sie zu aktivieren.

Architekturschaubild

Das Schaubild zeigt eine skalierte Installation im Cluster. Alle roten Verbindungen stellen die Verbindungen aus Punkt 2 dar.



1.3. Konfiguration

In diesem Kapitel erfahren Sie mehr über die Konfigurationsmöglichkeiten.

1.3.1. Konfiguration per Weboberfläche

Allgemein

Mit der Version 2.4.0 können Konfigurationseinstellungen über eine web-basierte Oberfläche vorgenommen werden. Zudem bietet die Oberfläche weitere Informationen zum aktuellen Zustand des d.3 search Systems.

Der Aufruf der Administrationsoberfläche erfolgt über `http://host:6133/` oder `https://host:6133`, falls SSL aktiviert wurde. Einen entsprechenden Link finden Sie ebenfalls unter **Start | Programme | d.velop | d.3 search**.

Wird die Oberfläche vom Host aus aufgerufen, auf dem installiert wurde, ist die Eingabe von Benutzername und Passwort nicht notwendig.

Warnung

Nach der Installation von d.3 search ist die Administrationsoberfläche aufrufbar. Ändern Sie Zugangsdaten Username und Password ab und aktivieren Sie SSL.

Warnung

Die Administrationsoberfläche setzt einen Microsoft Internet Explorer in der Version 8, 9, 10 oder 11 voraus.

Ab dem Microsoft Internet Explorer 11 muss TLS 1.2 abgeschaltet werden, wenn eine SSL-Verbindung für das Webinterface eingerichtet wird. TLS 1.0 und 1.1 müssen weiterhin aktiviert bleiben. Die Einstellungen finden Sie im Microsoft Internet Explorer unter: **Internetoptionen | Erweitert**.

Die Sprache der Weboberfläche ist abhängig von der Browsersprache. Das heißt, ist diese Deutsch, wird die deutsche Oberfläche angezeigt, sonst die englische.

Bedienung

In der linken Menüleiste haben Sie die Möglichkeit zwischen den einzelnen Bereichen der Administration zu wechseln. Wenn Sie das Fenster kleiner ziehen, z.B. auf einem Smartphone oder Tablet, öffnen Sie die Menüleiste durch einen Klick auf .

Wurde ein Menüpunkt z.B. **Einstellungen** ausgewählt, erscheint am linken Rand unterhalb des ausgewählten Menüpunktes, eine Auswahl an weiteren Sektionen, die innerhalb des ausgewählten Bereichs konfiguriert werden können.

Anmerkung

Global oder die Einstellungen der bereits angelegten Dokumentensammlungen.

Prozessübersicht

Einstellungen

Global

Dokumentensammlung

Skalierung

Statistik

Funktionen

Cluster

Engines

Support Paket

Assistent

Einstellungen - Global

Cluster

Active

Group

Port

SearchCluster12

6134

Gateway

HandlerCount

Host

Port

3

wathe01

3400

Wenn Sie weitere Informationen zur aktuellen Eingabe benötigen, klicken Sie auf das entsprechende Informationszeichen neben der Eingabe.

Anschließend erscheint ein Hilfetext aus der Dokumentation.

Cluster

Active

Group

Port

SearchCluster12

6134

Standard: SearchCluster
Gibt den Gruppennamen des Clusters an.

Damit weitere Installationen im Netzwerk nicht automatisiert zum Cluster hinzugefügt werden, sollte der Gruppenname vor der Aktivierung der Clusterfunktion angepasst werden.

Folgende Menüpunkte stehen zur Auswahl:

Bezeichnung	Beschreibung
Prozessübersicht	In diesem Bereich finden Sie aktuelle Informationen zum System-Zustand.
Einstellung	In diesem Bereich können Sie Ihr d.3 search konfigurieren.
Skalierung	In diesem Bereich kann die Konfiguration für die skalierten d.3 search engines (SHAFT) vorgenommen werden.
Anmerkung Dieser Bereich steht nur zur Auswahl, wenn eine Skalierung vorgenommen worden ist.	
Statistik	In diesem Bereich können Sie Statistiken seit dem letzten Neustart abrufen.
Funktionen	In diesem Bereich stehen Ihnen verschiedene Funktionen zur Verfügung, mit denen Sie Ihre Dokumentensammlungen testen oder ändern können.

Bezeichnung	Beschreibung
Cluster	In diesem Bereich können Sie den aktuellen Status von Ihrem Cluster einsehen und manuell Instanzen zum Cluster hinzufügen/entfernen.
Engines	In diesem Bereich können Sie Informationen zu den d.3 search engines (SHAFT) einsehen und diese manuell starten/stoppen.
Support Paket	In diesem Bereich können Sie im Support-Fall ein Informationspaket für den d.velop Support erstellen.
Assistent	In diesem Bereich können Sie neue Dokumentensammlungen erstellen.

Prozessübersicht

In diesem Bereich finden Sie aktuelle Informationen zum System-Zustand. Dieser ist unterteilt in **Globale Information** und **Cluster Information**.

Anmerkung

Der Bereich **Cluster Information** ist nur sichtbar, wenn ein Cluster eingerichtet worden ist.

Globale Informationen

Die **Globalen Informationen** sind in vier weitere Bereiche unterteilt: **Status**, **Wartende Jobs**, **d.3 gateway** und **Verzeichnisse**.

Anmerkung

Der d.3 gateway Bereich ist nur sichtbar, wenn die Verbindung zum d.3 gateway eingerichtet ist.

Bereich	Bezeichnung	Beschreibung
Status	Laufzeit	Gibt an, wie lange der letzte Start zurückliegt.
	Verbindungen	Gibt die Anzahl der zur Zeit geöffneten Verbindungen an.
	Anfragen	Gibt die Anzahl aller Anfragen seit dem Start an, die d.3 search entgegengenommen hat.
	Anzahl Dokumentensammlungen	Gibt die Anzahl aller Dokumentensammlungen an.
	Lizenzinformation	Gibt an, welche Lizenzen gültig geladen worden sind.
	Server Information	Gibt den Rechnernamen und den Port an, unter dem d.3 search erreichbar ist.
	CPU-Auslastung	Gibt die CPU-Auslastung des d.3 search server an.
Wartende Jobs	Add-Jobs	Gibt die Anzahl der Add-Jobs in der Queue an.
	Delete-Jobs	Gibt die Anzahl der Delete-Jobs in der Queue an.
	Update-Jobs	Gibt die Anzahl der Update-Jobs in der Queue an.
Job Threads	ID	Gibt die ID des Threads an. Die Threads sind für die Verarbeitung der Job-Queue zuständig.
	Status	Der Status in Form einer grafischen Darstellung zeigt welchen Status der Thread aktuell hat. Folgende Status sind möglich: • pausierend • Warten auf Jobs • <JobNr> <DocID> wird verarbeitet -> <JobFileName>
d.3 gateway	<IP> (<Rechnername>)	Gibt die IP-Adresse und den Rechnernamen an, über die mit dem d.3 gateway kommuniziert wird.
	Abbildung	Die Abbildung zeigt, wie viele Threads aktuell eine Anfrage vom d.3 gateway verarbeiten und wie viele eine Verbindung mit dem d.3 gateway hergestellt haben. Die Summe aus beiden Zahlen ergibt die Gesamtanzahl Threads, welche eine Verbindung mit dem d.3 gateway herstellen, die in den Einstellungen angegeben ist.
Verzeichnisse	Corpora-Verzeichnis	Gibt das Verzeichnis an, unter dem die Dokumentensammlungen abgelegt sind.

Bereich	Bezeichnung	Beschreibung
	Engine-Verzeichnis	Gibt das Verzeichnis an, unter dem die Suchmaschinen-Libraries abgelegt sind.
	Job-Verzeichnis	Gibt das Verzeichnis an, unter dem die Jobs abgelegt werden.

Globale Informationen

Status

Laufzeit	00:01:31
Verbindungen	0
Anfragen	0
Anzahl Dokumentensammlungen	4
Lizenzinformation	d.3 search basic d.3 search Addon: File-Content d.3 search Addon: enterprise (Clustering/Scaling)
Server Information	D3-SQL-712 (3406)
CPU-Auslastung	0.00%

Wartende Jobs

Add-Jobs	0
Delete-Jobs	0
Update-Jobs	0

Job Threads

ID: 0	pausierend
ID: 1	pausierend

d.3 gateway

10.95.0.80 (d3-sql-712)

0 arbeitend

6 anfragend

Verzeichnisse

Corpora-Verzeichnis	C:\d3\d.3 search\corpora\
Engine-Verzeichnis	C:\d3\d.3 search\Jobs\
Job-Verzeichnis	C:\d3\d.3 search\engines\

Cluster Informationen

Unter dem Bereich **Cluster Information** finden Sie folgende Informationen:

Bezeichnung	Beschreibung
Aktiv	Gibt an, ob das Cluster aktiv ist.
Cluster-Name	Gibt den Namen des Clusterverbundes an.
Anzahl Server	Gibt an, wie viele Server im Cluster vorhanden sind.
Status	Gibt an, welchen Status der aktuelle Server im Cluster hat.

Cluster Informationen

Aktiv	true
Cluster-Name	AndreThesker
Anzahl Server	2
Status	Slave

d.3 SHAFT-Agenten

Unter dem Bereich **d.3 SHAFT-Agenten** finden Sie sämtliche Informationen zu allen verwendeten d.3 SHAFT-Agents.

Bezeichnung	Beschreibung
Laufzeit	Gibt an, wie lange der letzte Start zurückliegt.
TCP-Port	Gibt den TCP-Port an, über dem der d.3 SHAFT-Agent erreichbar ist.
Corpora-Verzeichnis	Gibt das Verzeichnis an, in dem die Dokumentensammlungen liegen.
SHAFT-Verzeichnis	Gibt das Verzeichnis an, aus dem die d.3 search engine (SHAFT) verwendet wird.
LogLevel	Gibt das LogLevel des d.3 SHAFT-Agents an.
CPU-Auslastung	Gibt an, die CPU-Auslastung von dem d.3 SHAFT-Agent Prozess an.

d.3 SHAFT-Agenten

DSEARCH1 (10.95.4.244)

Laufzeit	19:00:33
TCP-Port	6133
Corpora-Verzeichnis	C:\d3\d.3 search\corpora
SHAFT-Verzeichnis	C:\d3\d.3 search\shaft\64Bit\
LogLevel	6
CPUAuslastung	0.15%

Zu jedem verwendeten d.3 SHAFT-Agenten werden Informationen über die Dokumentensammlungen aufgelistet, die bei dem d.3 SHAFT-Agenten liegen.

Bezeichnung	Beschreibung
d.3 search engine für: <Dokumentensammlung-ID>	Gibt an, ob die Engine für die Dokumentensammlung aktuell online oder offline ist.
Engine-ID	Gibt die ID der Engine an.
Rechnername	Gibt den Rechnernamen an, der diese Engine verwendet.
Prozessaufrufparameter	Gibt an, mit welchem Aufrufparameter die d.3 search engine (SHAFT) gestartet wird.
Prozesspfad	Gibt an, welcher Prozess für die d.3 search engine (SHAFT) gestartet wird.
CPU-Auslastung	Gibt an, die CPU-Auslastung von der d.3 search engine (SHAFT) an.

d.3 search engine (SHAFT) für: d3p	Online
→ Engine-ID	shaft1
→ Rechnername	d3-sql-712
→ Prozessaufparameter	C:\d3\d.3 search\corpora\d3p\d3-sql-712\shaft1\shaft.ini
→ Prozesspfad	C:\d3\d.3 search\shaft\64Bit\shaft.exe
→ CPUAuslastung	0.00%

Einstellungen

Der Menüpunkt **Einstellungen** unterteilt sich in die zwei Sektionen **Global** und **Dokumentensammlung**.

Im Folgenden wird beschrieben, welche Einstellungen Sie auf den entsprechenden Seiten vornehmen können.

Anmerkung

Wenn sich der aktuelle Rechner in einem Cluster befindet, werden die Einstellungen in globale  und lokale  Einstellungen unterteilt.

Globale Einstellungen haben eine Auswirkung auf alle Instanzen im Cluster, lokale Einstellungen haben lediglich Auswirkungen für die aktuelle Instanz.

Die globalen Einstellungen können nur auf dem **Master** geändert werden.

Global

In dieser Sektion werden die globalen Einstellungen für d.3 search hinterlegt.

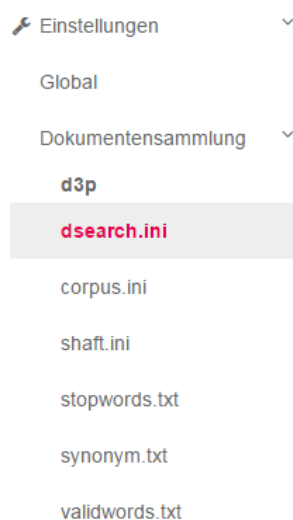
Informationen zu den Parametern entnehmen Sie dem Kapitel [Konfiguration des d.3 search Servers](#).

Dokumentensammlung

In der Menüleiste finden Sie unterhalb dieses Punktes die von Ihnen angelegten Dokumentsammlungen.

Für jede Dokumentensammlung sind die Unterpunkte **dsearch.ini**, **corpus.ini**, **shaft.ini**, **stopwords.txt**, **synonym.txt** und die **validwords.txt** vorhanden.

Wenn der **Delimitermode=3** in der **shaft.ini** gesetzt ist, erscheint ein zusätzlicher Punkt **delimiter.txt**.



Informationen zu den Parametern entnehmen Sie bitte den entsprechenden Kapiteln:

Unterpunkt	Informationen zu den Parametern
dsearch.ini	Konfiguration von Dokumentensammlungen
corpus.ini	Konfiguration von Dokumentensammlungen
shaft.ini	Konfiguration der d.3 search engine
stopwords.txt	Verwaltung der Stoppwörter
synonym.txt	Verwaltung der synonym.txt
validwords.txt	Verwaltung der ValidWords
delimiter.txt	Konfiguration der d.3 search engine

Skalierung

Der Menüpunkt **Skalierung** beinhaltet mehrere Sektionen. Jede Sektion trägt den Namen einer Dokumentensammlung.

Durch einen Klick auf die jeweilige Dokumentensammlung können die Einstellungen bezüglich der Skalierung dieser Dokumentensammlung vorgenommen werden.

⌘ Skalierung

d3q

d3r

Pro d.3 search engine (SHAFT) ist ein Eintrag auf der Seite zu finden. Ein Eintrag besteht aus den folgenden Informationen bzw. Einstellungsmöglichkeiten:

SHAFT1 - dsearcht3

L Path

C:\d3\d.3 search\corpora\d3r\dsearcht1\sr



L Database

corpora\d3r\dsearcht1\shaft1\



Im Titel ist die ID der d.3 search engine (SHAFT) für die aktuelle Dokumentensammlung zu finden und dahinter, getrennt durch einen Bindestrich, der Hostnamen, auf welchem die d.3 search engine (SHAFT) ausgelagert ist.

Die darauf folgenden Einstellungsmöglichkeiten werden wie folgt beschrieben:

Parameter	Beschreibung
Path	Ablageort für das Online-Backup

Parameter	Beschreibung
Database	Relativer Pfad zum Ablageort der Indexdaten

Anmerkung
 Wenn dieser Pfad geändert wird, muss zuvor die jeweilige d.3 search engine (SHAFT) beendet werden. Anschließend kann der Pfad angepasst werden.

 Der letztendliche Umzug der Indexdateien zum neuen Pfad muss manuell erfolgen.

Statistik

Der Menüpunkt **Statistik** unterteilt sich in die zwei Sektionen **Global** und **Dokumentensammlung**.

Die Sektion selbst ist unterteilt in

- **Funktionen**
- **Informationen (Gesamt)**
- **Informationen (pro d.3 search engine (SHAFT))**

Funktionen

Dem Abschnitt **Funktionen** können Sie jeweils Informationen über die aufgerufenen API-Funktionen von d.3 search entnehmen.

Bei einem Neustart von d.3 search werden die Statistiken zurückgesetzt.

Zudem kann ein manueller Reset durch einen Klick auf **Reset Statistik** erfolgen.

Hierbei wird nur die Statistik des aktuellen Corpus zurückgesetzt.

Für jede API-Funktion von d.3 search werden folgende Informationen in der Statistik festgehalten:

Bezeichnung	Beschreibung
Aufruf	Name der aufgerufenen API-Funktion
Geöffnete Verbindungen	Anzahl der insgesamt geöffneten Verbindungen, egal ob Sie erfolgreich oder fehlerhaft waren. Dazu zählt jede geöffnete Verbindung zu der Dokumentensammlung, d.h. jede Verbindung zur d.3 search engine (SHAFT). Bei einer Skalierung wird die Verbindung zu alle d.3 search engines (SHAFT) als eine Verbindung gezählt.
Geschlossene Verbindungen	Anzahl der erfolgreich geschlossenen Verbindungen. Inkl. durchschnittlich benötigte Zeit pro Verbindung. Dazu zählt jede geöffnete Verbindung zu der Dokumentensammlung, d.h. jede Verbindung zur d.3 search engine (SHAFT). Bei einer Skalierung wird die Verbindung zu alle d.3 search engines (SHAFT) als eine Verbindung gezählt.
offene Verbindungen	Anzahl der aktuell offenen Verbindungen. Dazu zählen alle Verbindungen zu aktuell zu einer Dokumentensammlung geöffnet sind. D.h. in der Regel eine geöffnete Verbindung zu einer d.3 search engine (SHAFT).
Auf Fehler gelaufene Verbindungen	Anzahl, der auf Fehler gelaufenen Verbindungen. Jede Versuch eine Verbindung zu der Dokumentensammlung aufzubauen (der d.3 search engine (SHAFT)) wird als neuer Fehler gezählt. Somit kann es sein, dass die Queue von d.3 search eine Vielzahl an Fehler hervorruft, die alle auf einen Verbindungsfehler zurück zu führen sind.

Anmerkung
 Eine Verbindung ist z.B. fehlerhaft, wenn:

- d.3 search engine (SHAFT) nicht erreichbar ist
- d.3 search engine (SHAFT) ReadOnly ist

Global

Unter der Sektion **Global** sind sämtliche API-Funktionen gelistet, die keiner Dokumentensammlung zugeordnet werden können.

Der folgenden Tabelle können Sie entnehmen, welche Bedeutung die gelisteten API-Funktionen haben.

Aufruf	Beschreibung
GetEngine	Listet die von d.3 search unterstützten Suchmaschinen mit deren Fähigkeiten auf.
GetServerStatus	Liefert den aktuellen Status vom d.3 search server zurück.

Statistik - Global

Funktionen

Aufruf	Geöffnete Verbindungen	Geschlossene Verbindungen	Offene Verbindungen	Auf Fehler gelaufene Verbindungen
GetEngine	0	0	0	0
GetServerStatus	0	0	0	0
Summe	0	0	0	0

Reset Statistik (28.05.2015 14:28:48.059)

Dokumentensammlung

Unter der Sektion **Dokumentensammlung** befindet sich für jede konfigurierte Dokumentensammlung eine Statistik.

Die gelisteten API-Funktionen beziehen sich auf die ausgewählte Dokumentensammlung.

Der folgenden Tabelle können Sie entnehmen, welche Bedeutung die gelisteten API-Funktionen haben.

Aufruf	Beschreibung
Add	Fügt ein Dokument zu einer Dokumentensammlung hinzu.
Update	Ändert ein Dokument in einer Dokumentensammlung.
Delete	Entfernt ein Dokument aus einer Dokumentensammlung.
GetCorpora	Listet alle angelegten Dokumentensammlungen und deren Eigenschaften auf.
QueryDocuments	Führt eine Suchanfrage in der Dokumentensammlung durch.
GetSimilarWords	Bestimmt alle Worte, die ähnlich zu einem gegebenen Wort sind.
GetDocument	Gibt eine Liste aller Wörter zurück, die zu einem Dokument in der Dokumentensammlung gespeichert sind. Der Inhalt entspricht, auf Grund von Stopwortlisten und Entfernung von doppelten Wörtern, in den meisten Fällen nicht dem Originaldokument.
GetCorpusProperties	Ermittelt die spezifischen Eigenschaften einer Dokumentensammlung.
SetReadOnlyModus	Versetzt die Dokumentensammlung in den Modus "Nur Lesen".
GetSummary	Rekonstruiert ein TextSummary (Abstract) für ein Dokument für die angegebenen Dokumente. D.h. es werden Sätze zurückgegeben, in denen die ursprüngliche Suche (Query) gefunden worden ist.

Statistik - d3p

Funktionen

Aufruf	Geöffnete Verbindungen	Geschlossene Verbindungen	Offene Verbindungen	Auf Fehler gelaufene Verbindungen
Add	0	0 (0 ms)	0	0
Update	0	0 (0 ms)	0	0
Delete	0	0 (0 ms)	0	0
GetCorpora	0	0 (0 ms)	0	0
QueryDocuments	0	0 (0 ms)	0	0
GetSimilarWords	0	0 (0 ms)	0	0
GetDocument	0	0 (0 ms)	0	0
GetCorpusProperties	0	0 (0 ms)	0	0
SetReadOnlyMode	0	0 (0 ms)	0	0
GetSummary	0	0 (0 ms)	0	0
Summe	0	0 (0 ms)	0	0

Reset Statistik (28.05.2015 14:28:48.059)

Informationen (Gesamt)

Dem Abschnitt **Informationen (Gesamt)** können Sie die folgenden Werte pro Sektion entnehmen.

In der Sektion **Global** werden die jeweiligen Werte über alle d.3 search engines (SHAFT) aller Dokumentensammlungen kumuliert.

Bezeichnung	Beschreibung
Gesamtgröße	Gesamtgröße aller d.3 search engine (SHAFT) - Datenbanken, die zu der ausgewählten Dokumentensammlung gehören
Größe des Suchindex	Größe der reinen d.3 search engine (SHAFT) - Datenbanken, die zu der ausgewählten Dokumentensammlung gehören, ohne StoreText- und nGramm-Dateien
Größe aufgrund der Verwendung von "FirstWildcard" kleiner als 3	Größe der d.3 search engine (SHAFT) - Datenbankdateien, für die Verwendung eines Wildcards an erster oder zweiter Stelle in der Suche Weitere Informationen zu dem Schalter FirstWildcard finden Sie unter Konfiguration der d.3 search engine .
Anmerkung Die Prozentangabe hinter diesem Wert besagt, wie viel Prozent die Größe im Vergleich zu der Größe des Suchindex ausmacht.	

Bezeichnung	Beschreibung
Größe aufgrund der zusätzlichen Speicherung des Volltextes	Größe der d.3 search engine (SHAFT) - Datenbankdateien, für die zusätzliche Speicherung der Volltextinformationen Weitere Informationen zu dem Schalter StoreText finden Sie unter Konfiguration der d.3 search engine .
Anmerkung Die Prozentangabe hinter diesem Wert besagt, wie viel Prozent die Größe im Vergleich zu der Größe des Suchindexes ausmacht.	
Anzahl Dokumente	Anzahl der Dokumente, innerhalb einer Dokumentensammlung
Anzahl Wörter	Anzahl der unterschiedlichen Wörter innerhalb einer Dokumentensammlung

d3p - SHAFT1 (localhost)

Datenbankgröße

Gesamtgröße	0.018 GB
Größe des Suchindexes	0.002 GB
Größe aufgrund der Verwendung von "FirstWildCard" kleiner als 3	0.016 GB
Größe aufgrund der zusätzlichen Speicherung des Volltextes	0 GB

Datenbankinhalt

Anzahl Dokumente	3
Anzahl Wörter	2

Informationen zum Datenbankverzeichnis

Lesemodus	Nein
Datenbankverzeichnis	C:\d3\d.3 search\corpora\d3p\SHAFT1
Speicherkapazität	15.388 GB frei von 44.9 GB
Belegter Speicherplatz	66%

Informationen (pro d.3 search engine (SHAFT))

Unterhalb des Abschnittes **Informationen (pro d.3 search engine (SHAFT))** finden Sie Informationen für die einzelnen d.3 search engines (SHAFT).

Folgende Werte können diesem Abschnitt entnommen werden:

Bezeichnung	Beschreibung
Gesamtgröße	Gesamtgröße der d.3 search engine (SHAFT) - Datenbanken
Größe des Suchindexes	Größe der reinen d.3 search engine (SHAFT) - Datenbanken
Größe aufgrund der Verwendung von "FirstWildCard" kleiner als 3	Größe der d.3 search engine (SHAFT) - Datenbankdateien, für die Verwendung eines Wildcards an erster oder zweiter Stelle in der Suche Weitere Informationen zu dem Schalter FirstWildCard finden Sie unter Konfiguration der d.3 search engine .
Größe aufgrund der zusätzlichen Speicherung des Volltextes	Größe der d.3 search engine (SHAFT) - Datenbankdateien, für die zusätzliche Speicherung der Volltextinformationen Weitere Informationen zu dem Schalter StoreText finden Sie unter Konfiguration der d.3 search engine
Anzahl Dokumente	Anzahl der Dokumente, innerhalb der d.3 search engine (SHAFT)-Datenbank
Anzahl Wörter	Anzahl der unterschiedlichen Wörter innerhalb der d.3 search engine (SHAFT)-Datenbank
Lesemodus	Gibt an, ob sich die d.3 search engine (SHAFT) gerade im Lesemodus befinden. Gründe dafür könnten sein: Backup läuft, Engine wird verschoben.
Datenbankverzeichnis	Gibt den Pfad zum Datenbankverzeichnis der d.3 search engine (SHAFT) an.

Bezeichnung	Beschreibung
Speicherkapazität (Datenbankverzeichnis)	Gibt die freie und gesamte Speicherkapazität des Datenbankverzeichnisses der d.3 search engine (SHAFT) an.
Belegter Speicherplatz (Datenbankverzeichnis)	Stellt grafisch dar, wie viel Prozent der Speicherkapazität des Datenbankverzeichnisses bereits belegt ist.
Sicherung läuft	Gibt an, ob gerade ein Backup durchgeführt wird.
Backupverzeichnis	Gibt den Pfad zum Backupverzeichnis der d.3 search engine (SHAFT) an.
Speicherkapazität (Backupverzeichnis)	Gibt die freie und gesamte Speicherkapazität des Backupverzeichnisses der d.3 search engine (SHAFT) an.
Belegter Speicherplatz (Backupverzeichnis)	Stellt grafisch dar, wie viel Prozent der Speicherkapazität des Backupverzeichnisses bereits belegt ist.
Benötigter Speicherplatz für die automatische Defragmentierung	Gibt an, wie viel Speicherplatz für die automatische Defragmentierung der d.3 search engine (SHAFT) benötigt wird.
	Warnung Ist dieser Wert rot hinterlegt, ist nicht genügend Speicherplatz vorhanden.
Benötigter Speicherplatz für das online Backup	Gibt an, wie viel Speicherplatz für das online Backup der d.3 search engine (SHAFT) benötigt wird.
	Warnung Ist dieser Wert rot hinterlegt, ist nicht genügend Speicherplatz vorhanden.
Benötigter Speicherplatz für die automatische Reparatur	Gibt an, wie viel Speicherplatz für die automatische Reparatur der d.3 search engine (SHAFT) benötigt wird.
	Warnung Ist dieser Wert rot hinterlegt, ist nicht genügend Speicherplatz vorhanden.

Informationen zum Backupverzeichnis

Sicherung läuft	Nein
Backupverzeichnis	C:\d3\d.3 search\corpora\d3p\SHAFT1\backup
Speicherkapazität	15.388 GB frei von 44.9 GB
Belegter Speicherplatz	66%

Benötigter Speicherplatz

Benötigter Speicherplatz für die automatische Defragmentierung	0.5 GB
Benötigter Speicherplatz für das online Backup	0.518 GB
Benötigter Speicherplatz für die automatische Reparatur	0.518 GB

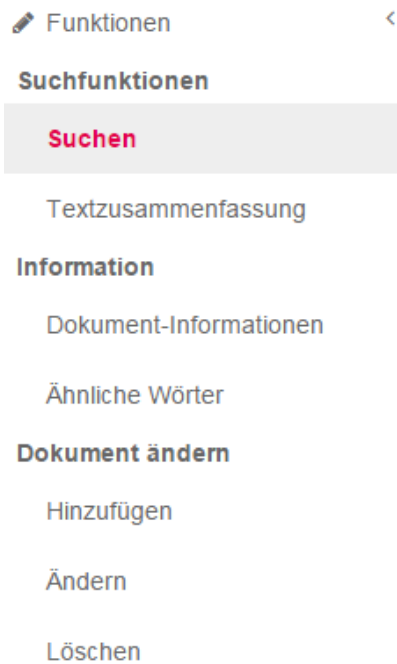
Arbeitsspeicher

Arbeitsspeicherkapazität	0.282 GB frei von 4 GB
Verwendeter Arbeitsspeicher	93%

Funktionen

Der Menüpunkt **Funktionen** ist in drei Sektionen **Suchfunktionen**, **Informationen** und **Dokument ändern** unterteilt.

Im Folgenden werden die einzelnen Funktion kurz vorgestellt und die entsprechenden Parameter erläutert.



Suchfunktionen

In dieser Sektion befinden sich sämtliche Funktionen, die bei einer Suche nach einem Dokument verwendet werden können.

Suchen

Die Funktion **Suchen** bietet Ihnen die Möglichkeit auf der aktuellen Instanz eine Suchabfrage abzusetzen.

Die folgenden Kriterien stehen Ihnen dabei zur Verfügung:

Parameter	Beschreibung
Korpus	Name der zu durchsuchenden Dokumentensammlung
Query	gesuchter Inhalt (Suchanfrage selbst)
QueryCodepage	Gibt die Codepage der Suchanfrage vor.
SimilarDocId	Dokument-Id eines Dokumentes in der Dokumentensammlung; wird dieses Feld gefüllt, werden ähnliche Dokumente zu dem angegebenen Dokument gesucht.
StartRow	Gibt an, ab welcher Zeile die Treffer ausgegeben werden.
MaxRowCount	Gibt an, wie viele Treffer maximal angezeigt werden sollen.
MinRelevance	Gibt die Relevance an, die ein Suchergebnis mindestens erfüllen muss.
wantHints	Gibt an, ob Suchwolken zurück geliefert werden sollen.
wantAbstract	Gibt an, ob ein Abstract (kurze Textzusammenfassung über den Treffer) mitgeliefert werden soll.
NoExtend	Gibt an, ob eine erweiterte Suche an Hand der Suchwolken statt finden soll.
QueryByExample	Ist dieser Haken gesetzt, kann ein Dokument hochgeladen werden. Bei der Suche werden Dokumente gesucht, die dem hoch geladenen Dokument ähnlich sind.
Handle	Handle, welches verwendet werden kann, damit eine Suchanfrage nicht erneut ausgeführt wird, sondern aus dem Cache abgerufen wird

Suchen

Korpus	d3andre
Query	
QueryCodePage	
SimilarDocId	
StartRow	0
MaxRowCount	1000
Minimale Relevanz	0
WantHints	☐
WantAbstract	☐
NoExtend	☐
QueryByExample	☐
Handle	0

Textzusammenfassung

Die Funktion **Textzusammenfassung** bietet Ihnen die Möglichkeit auf der aktuellen Instanz eine Textzusammenfassung (Abstract) eines Dokumentes zu erhalten.

Die folgenden Kriterien stehen Ihnen dabei zur Verfügung:

Parameter	Beschreibung
Korpus	Name der zu durchsuchenden Dokumentensammlung
DocId	Dokumenten-ID des Dokuments, welches zusammengefasst werden soll
AbstractSize	Gibt an, wie viele Zeilenumbrüche maximal zurückgegeben werden sollen.
Context	Gibt an, unter welchen Kontext eine Zusammenfassung erfolgen soll. In den meisten Fällen entspricht das dem gesuchten Wort
HintWords	Liste der Wörter aus der Suchwolke, welche bei der Zusammenfassung berücksichtigt werden sollen

Textzusammenfassung

Korpus	d3p
DocId	
AbstractSize	100
Context	
HintWords	

Senden

Information

In dieser Sektion befinden sich sämtliche Funktionen, die verwendet werden können, um Informationen zu einem Dokument zu bekommen.

Dokument-Informationen

Die Funktion **Dokument-Information** bietet Ihnen die Möglichkeit auf der aktuellen Instanz alle gespeicherten Wörter zu einem Dokument zu erhalten.

Die folgenden Kriterien stehen Ihnen dabei zur Verfügung:

Parameter	Beschreibung
Korpus	Name der zu durchsuchenden Dokumentensammlung
Dokumenten-ID	Dokumenten-ID des Dokuments, dessen gespeicherte Wörter ermittelt werden sollen

Anmerkung
Wenn der Schalter `CurrentVersionsOnly` deaktiviert ist, muss die Dokumenten-ID im Format `<Dokument-ID>.<Dokument-Version>` angegeben werden. Ist der Schalter aktiviert, muss das Format `<Dokument-ID>` verwendet werden.

Dokument-Informationen

Korpus	d3p
Dokumenten-ID	

Senden

Ähnliche Wörter

Die Funktion **Ähnliche Wörter** gibt vergleichbare/ähnliche Wörter zu einem gegebenen Wort zurück.

Parameter	Beschreibung
Korpus	Name der zu durchsuchenden Dokumentensammlung
Wort	das Wort, für das ähnliche Wörter gesucht werden sollen
Minimale Relevanz	Gibt die Relevanz an, die mindestens gegeben sein muss.

Ähnliche Wörter

Korpus

d3andre ▼

Wort

Minimale Relevanz

0

Senden

Dokument ändern

In dieser Sektion befinden sich sämtliche Funktionen, die verwendet werden können, um Dokumente zu ändern, hinzuzufügen und zu löschen.

Hinzufügen

Die Funktion **Hinzufügen** bietet Ihnen die Möglichkeit auf der aktuellen Instanz ein Dokument zum Suchindex hinzuzufügen.

Die folgenden Kriterien stehen Ihnen dabei zur Verfügung:

Parameter	Beschreibung
Korpus	Name der zu durchsuchenden Dokumentensammlung
Dokumenten-ID	ID, die das zu importierende Dokument eindeutig referenziert
Document	das zu importierende Dokument

Hinzufügen

Korpus

d3p ▼

Dokumenten-ID

Document

Datei auswählen

Auswählen

Senden

Ändern

Die Funktion **Ändern** bietet Ihnen die Möglichkeit auf der aktuellen Instanz ein Dokument im Suchindex zu Ändern.

Die folgenden Kriterien stehen Ihnen dabei zur Verfügung:

Parameter	Beschreibung
Korpus	Name der zu durchsuchenden Dokumentensammlung
Dokumenten-ID	ID, die das zu ändernde Dokument eindeutig referenziert
Document	das zu importierende Dokument

Ändern

Korpus

d3p ▼

Dokumenten-ID

Document

File auswählen

Auswählen

Senden

Löschen

Die Funktion **Löschen** bietet Ihnen die Möglichkeit auf der aktuellen Instanz ein Dokument im Suchindex zu Löschen.

Die folgenden Kriterien stehen Ihnen dabei zur Verfügung:

Parameter	Beschreibung
Korpus	Name der zu durchsuchenden Dokumentensammlung
Dokumenten-ID	ID, die das zu löschende Dokument eindeutig referenziert
Synchron	Gibt an, ob das Dokument direkt (Synchron) gelöscht werden oder ob ein Job in der Queue hinzugefügt werden soll (asynchron).

Löschen

Korpus

d3p ▼

Dokumenten-ID

Synchron

☐

Senden

Cluster

Der Menüpunkt **Cluster** erscheint nur, wenn ein Cluster eingerichtet worden ist. Auf dieser Seite können Sie aktuelle Informationen zu Cluster einsehen, Instanzen zum Cluster hinzufügen und diese ggfs. wieder entfernen. Der Bereich Cluster ist unterteilt in **Information**, **Einstellung**, **Cluster-Server** und **Log**.

Information

In diesem Abschnitt stehen wesentliche Informationen zum Cluster, dazu zählen:

Information	Beschreibung
Aktiv	Gibt an, ob das Cluster aktiv ist. Ein Cluster ist nicht aktiv, wenn eine Instanz im Cluster einen Fehler aufweist. Wenn das Cluster nicht aktiv ist, können keine Änderungen am Suchindex vorgenommen werden.
Status	Gibt an, welchen Modus die aktuelle Instanz hat.

Information

Aktiv

true

Status

Slave

Einstellungen

In diesem Abschnitt kann eingestellt werden, ob die Informationen unter **Cluster-Server** stetig aktualisiert werden sollen.

Einstellungen

Automatische Aktualisierung



Cluster-Server

In diesem Abschnitt werden alle Server aufgelistet, die sich im Cluster befinden.

Die Auflistung enthält die folgenden Informationen:

Information	Beschreibung
IP-Adresse	Gibt die IP-Adresse des Servers im aktuellen Netzwerk an
Rechnername	Gibt den Rechnernamen des Servers im aktuellen Netzwerk an.
	Warnung Im Cluster wird ein Server an Hand seines Rechnernamen identifiziert. Eine spätere Änderung des Rechnernamen hat somit Auswirkungen auf das Cluster.
Modus	Gibt den Modus an, den eine Instanz zur Zeit hat (Master/Slave/Unbekannt)
Letzte Meldung	Gibt an, wann das letzte mal eine Kommunikation zwischen der aktuellen Instanz und der ausgewählten Instanz statt gefunden hat.
Information	Gibt zusätzliche Informationen zu einem Server an (z.B. Startet gerade neu)
Fehlermeldungen	Gibt die Fehlermeldungen einer Instanz an. (z.B. d.3 search engine (SHAFT) ist nicht gestartet)

Der Button **Löschen** ist nicht aktiv, wenn die Instanz aktiv ist. Ein Server kann nur aus dem Cluster entfernt werden, wenn die jeweilige Instanz abgeschaltet ist. Siehe dazu [Clustermitglied entfernen](#). Durch einen Klick auf den Button **Hinzufügen** kann eine Instanz manuell zum Cluster hinzugefügt werden. Nach dem Klick erscheint eine zusätzliche Zeile, in der der Rechnername oder die IP-Adresse angegeben werden kann. Für weitere Informationen siehe [Neue d.3 search Instanz für das Cluster einrichten](#)

Log

In diesem Abschnitt werden die Log-Ausgaben des Clusters angezeigt. Die Farben haben dabei folgende Bedeutung:

Farbe	Bedeutung
Grün	Der Request wurde erfolgreich versendet.
Rot	Der Request konnte nicht erfolgreich versendet werden (evtl. Timeout).
Blau	Information (z.B. Eine Instanz hat einen Fehler gemeldet).
Grau	Der Request wurde erfolgreich empfangen.

Log

```
28.01.2015 13:08:36: [3420] [RECV]: dsearchdc1 - IAMALIVEMASTER
28.01.2015 13:08:42: [3928] [RECV]: dsearchdc1 - IAMALIVEMASTER
28.01.2015 13:08:47: [3868] [RECV]: dsearchdc1 - IAMALIVEMASTER
```

Engines

Über den Menüpunkt **Engines** können Sie Informationen zu den d.3 search engines (SHAFT) erlangen. Zudem können einzelne Engines gestartet, gestoppt und neu gestartet werden. Der Menüpunkt enthält für jede eingerichtete Dokumentensammlung einen separaten Punkt.

Im Abschnitt **Globale Funktionen** können Sie alle Engines einer Dokumentensammlung gleichzeitig starten, stoppen oder neu starten. Klicken Sie dazu auf den entsprechenden Button.

Globale Funktionen

Alle starten

Alle stoppen

Alle neustarten

Im Anschluss finden Sie für jede d.3 search engine (SHAFT) einen separaten Abschnitt.

Der Abschnitt enthält folgende Informationen:

Information	Beschreibung
Status	Gibt an, ob der Prozess gestartet, oder gestoppt ist.

Information	Beschreibung
ID	Gibt die ID an, die die Engine innerhalb der Dokumentensammlung hat.
Korpus	Gibt den Korpus für den die Engine läuft.
Rechnername	Gibt den Rechnernamen an, auf dem die Engine läuft.
Dateiname	Gibt den Dateinamen des Prozesses an. Diese Information entfällt, wenn die Engine ausgelagert ist.
Befehlszeile	Gibt die Befehlszeile an, mit der der Prozess aufgerufen wird. Diese Information entfällt, wenn die Engine ausgelagert ist.
Prozess-ID	Gibt die aktuelle Prozess-ID des Prozesses an. Diese Information entfällt, wenn die Engine ausgelagert ist.
CPU-Auslastung	Gibt die CPU-Auslastung für die d.3 search engine (SHAFT) an.
Startzeit	Gibt an, wann der Prozess gestartet worden ist.

SHAFT2

Status	Online
ID	SHAFT2
Korpus	d3p
Rechnername	localhost
Dateiname	C:\d3\d.3 search\shaft\64Bit\shaft.exe
Befehlszeile	C:\d3\d.3 search\corpora\d3p\SHAFT2\shaft.ini
Prozess-ID	4956
CPUAuslastung	0%
Startzeit	16.02.2016 11:40:28.300

Stop

Neu starten

Service

Support Paket

Über den Menüpunkt **Support Paket** können Sie ein Support Paket erstellen lassen.

Die Inhalte des Support Paketes können nach belieben verändert werden. Um das Support Paket zu erstellen, müssen Sie die folgenden Schritte durchgehen.

- Wählen Sie zunächst die entsprechenden Dokumentensammlungen aus.
- Wählen Sie aus, welche Paketgröße das Support Paket haben soll, bzw. welcher Inhalt in dem Support Paket zur Verfügung gestellt werden soll.
- Erzeugen Sie das Support Paket.
- Laden Sie das Support Paket herunter.

Anmerkung

Bei der Erstellung des Support Pakets wird zusätzlicher Festplattenplatz benötigt. Stellen Sie sicher, dass dieser zur Verfügung steht. Ein Support Paket wird wieder gelöscht, sobald es vom Anwender herunter geladen worden ist.

Service Funktionen

Über den Menüpunkt **Service Funktionen** können Sie spezielle Funktionen für einzelne Dokumentensammlungen ausführen.

Folgende Funktionen können pro Dokumentensammlung ausgeführt werden:

Warnung

Durch die Ausführung der Funktionen können Sie den Betrieb der Volltextsuche lahm legen. Führen Sie diese Funktionen also nur aus, wenn Sie sich absolut sicher sind.

Bezeichnung	Beschreibung
Dokumentensammlung umbenennen	Sie können die erstellte Dokumentensammlung mit dieser Funktion umbenennen. Dadurch werden die Ordnerstruktur und die Einstellungen geändert. Verwenden Sie diese Funktion z.B., wenn Sie beim parallelem Neuaufbau den Schalter Dokumentensammlung löschen nicht gesetzt haben.
	Anmerkung Beachten Sie, dass eine Änderung Auswirkungen auf die Zuordnungen Dokumentart zu Dokumentensammlung im d.3 admin haben kann.
Backup manuell starten	Durch diese Funktion können Sie das Backup der d.3 search engines (SHAFT) manuell starten. Führen Sie diese Aktion zur Sicherheit durch, wenn Sie Änderungen an den Einstellungen der d.3 search engine (SHAFT) vornehmen.
	Warnung Während des Backups ist nur ein lesender Zugriff auf die Dokumentensammlung möglich.
Lesemodus aktivieren / deaktivieren	Versetzen Sie die Dokumentensammlung im Lesemodus, damit ein Flush stattfindet und Sie die Datenbankdateien der d.3 search engine (SHAFT) manuell sichern wollen.
Dokumentensammlung entfernen	Das Löschen einer Dokumentensammlung sollte nur erfolgen, wenn Sie die Dokumentensammlung nicht mehr benötigen. Verwenden Sie diese Funktion z.B., wenn Sie beim parallelem Neuaufbau den Schalter Dokumentensammlung löschen nicht gesetzt haben.
	Warnung Dieser Vorgang kann nicht rückgängig gemacht werden.

Service Funktionen

d3p

Dokumentensammlung umbenennen	
Backup manuell starten	
Lesemodus aktivieren	
Dokumentensammlung entfernen	

Assistent

Unter dem Menüpunkt **Assistent** befinden sich mehrere Assistenten zur Verwaltung von d.3 search.

- [Der Assistent zur Anlage einer neuen Dokumentensammlung](#)
- [Der Assistent zur Reindizierung einer Dokumentensammlung](#)

1.3.2. Konfiguration des d.3 search Servers

Globale Einstellungen für den d.3 search Server können in der Datei `dsearch.ini` konfiguriert werden.

Die Datei befindet sich im d.3 search Verzeichnis] (z.B. `D:\d3\d.3 search\`).

Die Einträge, die Sie dort finden, haben folgende Bedeutung:

Sektion	Eintrag	Bedeutung
[General]		
	Port	Standard: 3406 TCP/IP-Port, unter dem der d.3 search Server für den d.3 Server erreichbar ist. Bei einer direkten Kommunikation muss dieser Wert mit dem Serverparameter <code>d3fc_dsearch_server_tcpip_port</code> in der Administration übereinstimmen, damit ein Verbindungsaufbau möglich ist. Wenn eine Kommunikation über den d.3 gateway eingerichtet ist, sollte der Standardwert beibehalten werden.
	wantHints	(1: ja), (0: nein) Standard: 1 Ist dieser Schalter gleich "0", werden keine Worte für eine Suchverfeinerung geliefert. Ist der Schalter gleich "1" werden Worte für die Suchverfeinerung geliefert. Durch das Ausschalten dieser Funktion kann die Performance der Such-Engine gesteigert werden.

Sektion	Eintrag	Bedeutung
	LogLevel	Standard: 6 LogLevel=0: nur Fehler werden im Logviewer ausgegeben LogLevel=6: normale Protokollierung von d.3 search API Calls LogLevel=9: alle Meldungen (inkl. Debug-Ausgaben) LogLevel=10: Die für die Shaft-Engine präparierte Datei (.\$\$\$) bleibt im Jobs-Verzeichnis als .tmp liegen. Warnung Wenn das LogLevel dauerhaft auf 10 gesetzt ist, könnte die Festplattenkapazität nicht mehr ausreichend sein. Anmerkung Ist das d.3 search Server Fenster im Vordergrund, kann der LogLevel über die Tastenkombination <Strg>+<Alt>+<D> vom aktuellen in den Debug-Modus versetzt werden.
	MinRelevance	Standard: 0 Gibt die Relevanz an, die ein Such-Ergebnis mindestens erfüllen muss. Der Wert wird nur verwendet, wenn die Suchanfrage keine Angabe für diesen Wert enthält. Ansonsten wird der Wert der Suchanfrage verwendet.
	LogFile	Gibt den Dateinamen einer Logdatei an, in der zusätzlich geloggt wird. Wenn diese Einstellung nicht gesetzt ist, wird keine separate Logdatei erzeugt. Anmerkung C:\d3\log.txt
	OMOMsgDist	Standard: 10 Gibt die minimale Distanz zwischen zwei OMO-Nachrichten in Sekunden an.
[Settings]		
	JobsPath	Standard:d.3 search Verzeichnis\jobs\ (z.B. D:\d3\d.3 search\jobs\) Gibt das Verzeichnis an, unter dem die Jobs geschrieben werden. Zu den Jobs zählen Add-/Update- und Delete-Jobs, die in jeweiligen Unterverzeichnissen geschrieben werden.
	Corpora-Path	Standard:d.3 search Verzeichnis\corpora\ (z.B. D:\d3\d.3 search\corpora\) Gibt das Verzeichnis an, unter dem die Dokumentensammlungen gespeichert sind.
	Engine-sPath	Standard: d.3 search Verzeichnis\engines\ (z.B. D:\d3\d.3 search\engines\) Gibt das Verzeichnis an, unter dem die Engines gespeichert sind. Der jeweilige Ordner enthält die DLL-Dateien zur Verständigung mit der jeweiligen Suchmaschine.
[Webinterface]		
	Username	Standard: admin Gibt den Benutzernamen an, um sich von einem anderen Rechner aus, in das Webinterface einzuloggen. Beim Logging über den lokalen Rechner findet keine Authentifizierung statt.

Sektion	Eintrag	Bedeutung
	Password	Standard: dvelop Gibt das Passwort an, um sich von einem anderen Rechner aus, in das Webinterface einzuloggen. Beim Logging über den lokalen Rechner findet keine Authentifizierung statt. Anmerkung Das Passwort wird verschlüsselt gespeichert und kann nur über das Webinterface gesetzt werden. Warnung Ändern Sie das Passwort direkt nach der Installation.
	Port	Standard: 6133 Gibt den Port an, über den das Webinterface zu erreichen ist.
	UseSSL	(1: ja), (0: nein) Standard: 0 Gibt an, ob das Webinterface über eine SSL Verbindung erreichbar ist. Wenn dieser Schalter aktiviert wird, kann weiterhin lokal auf dem Rechner über http auf die Webseite zugegriffen werden. Von außerhalb ist eine Verbindung nur noch über https möglich. Im Browser findet dazu automatisch ein Redirect statt. Anmerkung Wurde kein SSL Zertifikat hinterlegt, wird beim nächsten Neustart ein unsigniertes Zertifikat automatisch hinterlegt. Warnung Ab dem Microsoft Internet Explorer 11 muss TLS 1.2 abgeschaltet werden, wenn eine SSL-Verbindung für das Webinterface eingerichtet wird. TLS 1.0 und 1.1 müssen weiterhin aktiviert bleiben. Die Einstellungen finden Sie im Microsoft Internet Explorer unter: Internetoptionen Erweitert.
	SSLPort	Standard: 6135 Gibt den Port an, unter dem die SSL-Verbindung für das Webinterface erfolgen soll. Wenn SSL aktiviert ist, steht der unverschlüsselte Zugang zu dem Webinterface lokal immer noch zur Verfügung. Von einem anderen Rechner aus, leitet der Browser automatisch auf die HTTPS-Verbindung um. Anmerkung Der Schalter greift nur, wenn UseSSL für das Webinterface aktiviert worden ist.
[Communication]		
	SSLCertFile	Gibt den Dateipfad zum SSL-Zertifikat an. Anmerkung Das SSL-Zertifikat wird für die HTTPS-Verbindung zum Webinterface verwendet.
	SSLKeyFile	Gibt den Dateipfad zum SSL-Schlüssel für das definierte SSL-Zertifikat an. Anmerkung Das SSL-Zertifikat wird für die HTTPS-Verbindung zum Webinterface verwendet.
	SSLPassword	Gibt das Passwort des SSL-Zertifikats an. Anmerkung Das SSL-Zertifikat wird für die HTTPS-Verbindung zum Webinterface verwendet.

Sektion	Eintrag	Bedeutung
	D3FCPass-word	Gibt das Passwort für die d3fc-Kommunikation an. Anmerkung Dasselbe Passwort muss im d.3 admin für die Kommunikation mit d.3 search hinterlegt werden.
[Cluster]		
	Active	(1: ja), (0: nein) Standard: 0 Der Schalter gibt an, ob die Clusterfunktionalität aktiviert ist. Anmerkung Wenn die Clusterfunktionalität aktiviert wird, finden sich die d.3 search Instanzen mit der identischen Clustergruppe automatisch über einen UDP-Broadcast zu einem Cluster zusammen. Die jeweiligen Dokumentensammlungen werden dann automatisiert zu den jeweiligen Instanzen übertragen.
	Group	Standard: SearchCluster Gibt den Gruppennamen des Clusters an. Anmerkung Damit weitere Installationen im Netzwerk nicht automatisiert zum Cluster hinzugefügt werden, sollte der Gruppenname vor der Aktivierung der Clusterfunktion angepasst werden Warnung Der Clustername darf keinen Unterstrich (_) enthalten. Ansonsten kann d.3 search nicht starten.
	Port	Standard: 6134 Gibt den Port an, über dem die Clusterkommunikation statt findet. d.3 search baut über diesen Port einen TCP- und einen UDP-Server auf. Der UDP-Server ist für das automatische Hinzufügen neuer Instanzen verantwortlich. Der TCP-Server regelt anschließend die Überwachung und Kommunikation zwischen den Instanzen. Anmerkung Bei der Erstellung eines Clusters müssen alle Instanzen denselben Port eingestellt haben.
[Gateway]		
	Host	Standard: definiert den Hostnamen des d.3 gateways; wenn Sie keinen Hostnamen angeben, ist die Funktionalität deaktiviert Warnung Geben Sie hier möglichst einen Hostnamen an, der auch von anderen Rechnersystemen aus erreichbar ist. Angaben wie z.B. localhost oder 127.0.0.1 sind möglich, sollten aber gemieden werden.
	Port	Standard: 3400 Gibt den Port an, über dem der d.3 Gateway erreichbar ist.

Sektion	Eintrag	Bedeutung
	Handler - Count	Standard: 3 Gibt an, wie viele Verbindungen zum d.3 gateway aufgebaut werden sollen. Anmerkung Wenn die Anfragen vom d.3 gateway nach 30 Sekunden nicht von d.3 search entgegengenommen werden, sollte der HandlerCount erhöht werden.
[JobQueue]		
	JobThread - Count	Standard: 2 Gibt die Anzahl der Threads an, die die Job-Queue abarbeiten. Die Anzahl sollte erhöht werden, wenn die Queue nicht abgearbeitet werden kann und stetig anwächst.
	MaxJobs - ForRead	Standard: -1 Gibt an, wieviele Jobs beim Start maximal aus dem Jobs-Ordner geladen werden sollen. Bei dem Wert -1 werden alle Jobs aus dem Jobs-Ordner geladen, ansonsten die entsprechende Anzahl. Anmerkung Die Verwendung ist sinnvoll, wenn sich so viele Jobs in den Ordner befindet, dass der Start von d.3 search eine längere Zeit in Anspruch nimmt. Durch die Limitierung der Anzahl der Jobs, die beim Start eingelesen werden, kann somit der Startprozess von d.3 search verringert werden und die Volltextsuche steht früher wieder zur Verfügung. Warnung Bei der Verwendung dieses Schalters werden evtl. einige Jobs nicht eingelesen. Um diese Jobs im späteren Verlauf nachzuladen, sollte der Schalter [Job-Queue] MinutesBetweenReadJobs ebenfalls verwendet werden. Ansonsten werden die Jobs erst bei einem erneuten Neustart wieder eingelesen.
	Minutes - Between - ReadJobs	Standard: -1 Der Schalter gibt an, nach wie vielen Minuten der Jobs-Ordner erneut nach Jobs durchsucht werden soll. Um die Job-Queue gering zu halten und unnötige Zugriffe auf der Festplatte zu vermeiden, werden die Jobs erst nachgeladen, wenn die aktuelle Job-Queue keine Einträge mehr hat. Bei dem Wert -1 wird der Jobs-Ordner nicht nachträglich durchsucht. Anmerkung Die Verwendung ist sinnvoll, wenn der Schalter [JobQueue] MaxJobsRead verwendet wird, um den Start von d.3 search zu beschleunigen.

1.3.3. Konfiguration einer Dokumentensammlung

Der d.3 search Server verwaltet eine oder mehrere Dokumentensammlungen.

Eine Dokumentensammlung enthält eine Menge von Textdokumenten (ASCII-Format), die über einen eindeutigen Bezeichner referenziert werden.

Dieser Bezeichner wird vom d.3 server in Form von `DOK_ID.VERSIONS_NUMMER` definiert.

Jede Dokumentensammlung wird von einem Prozess einer der Sammlung zugeordneten Suchmaschine überwacht. Somit werden die Indexdaten einer Suchmaschine nicht vom d.3 search Server, sondern von einem externen Prozess beaufsichtigt. Der d.3 search Server gibt lediglich die Änderungsaktionen (z. B. Einfügen, Ändern oder Löschen eines Dokuments) an diesen externen Prozess weiter.

Da der d.3 search Server die Verbindung zu mehreren Dokumentensammlungen herstellen kann, ist es nicht zwingend erforderlich, dass alle Textinformationen eines Archivs in einer Dokumentensammlung abgelegt werden.

Anlage einer Dokumentensammlung

Der Menüpunkt im Webinterface **Assistent | Anlage** bietet die Möglichkeit eine neue Dokumentensammlung anzulegen.

Für die Anlage einer Dokumentensammlung befolgen Sie einfach die Anweisungen des Wizards.

Anmerkung

Dieser Assistent ersetzt die bisherige Anwendung **d.3 search Assistent**.

Schritt 1: Willkommen

Im ersten Schritt werden Sie darauf hingewiesen, dass Sie, um eine Dokumentensammlung anzulegen, die nachfolgenden Schritte durchgehen müssen.

Im Clusterbetrieb wird eine erfolgreich erstellte Dokumentensammlung automatisch an alle Knoten verteilt.

Durch **Weiter >>** gelangen Sie zur nächsten Seite.

Warnung

Berücksichtigen Sie, dass einige Einstellungen nachträglich nicht mehr ohne Reindexierung geändert werden können. Diese kann je nach Größe der Dokumentensammlung viel Zeit in Anspruch nehmen.

Anmerkung

Eine Dokumentensammlung kann nur erstellt werden, wenn der eingerichtete Cluster aktiv ist.

Assistent zur Anlage einer Dokumentensammlung

Willkommen

Basis

Indexierung

Suche

SHAFT-Einstellung

Backup

Fertigstellen

Schritt 1: Willkommen

Willkommen beim d.3 search Assistenten zur Anlage einer Dokumentensammlung.
Führen Sie die folgenden Schritte durch, um eine neue Dokumentensammlung zu erstellen.

WICHTIG:

Bitte berücksichtigen Sie bei der Konfiguration die d.3 search Dokumentation und bedenken Sie, dass einige Einstellungen später, ohne Reindexierung, nicht geändert werden können.

<< Zurück

Weiter >>

Schritt 2: Basis

Im zweiten Schritt werden die Basiseinstellungen der neuen Dokumentensammlung festgelegt.

Folgende Basiseinstellungen können vorgenommen werden:

Einstellung	Beschreibung
Name der Dokumentensammlung	der Name der neuen Dokumentensammlung; dieser darf nicht länger als 255 Zeichen sein und darf keine Sonderzeichen enthalten Anmerkung Sofern der d.3 server nicht anders konfiguriert ist, werden alle Dokumente zunächst automatisch in die Dokumentensammlung D3"XX" importiert, wobei "XX" für das Archivkürzel steht! Weitere Informationen finden Sie hier: d.3 server Einstellungen für d.3 search.
Spiegelung aktivieren	Wenn diese Option aktiviert ist, muss zunächst in einem untergeordneten Menü der Name einer Dokumentensammlung ausgewählt werden. Bei Schreibvorgängen in der gespiegelten Dokumentensammlung werden alle Änderungen ebenso auf die spiegelnde Dokumentensammlung angewendet. Anmerkung Dies gilt beispielsweise für Import-/ Delete- und Updatevorgänge. Anmerkung Dieser Punkt sollte ausschließlich zur Migration einer Dokumentensammlung verwendet werden. Weitere Informationen zu einer Migration finden Sie hier: Assistent zur Reindizierung einer Dokumentensammlung.
Arbeitsspeicher pro SHAFT-Engine	Die Größe des Hauptspeichers (in KB) für die Datenzwischenspeicherung. Dieser wird bei Transaktionen bzw. Suchanfragen von d.3 search engine (SHAFT) verwendet.
Anzahl Suchcursor	Dieser Wert beschreibt, wie viele Anfragen von d.3 search engine (SHAFT) gleichzeitig bearbeitet werden können. Wird diese Zahl überschritten, müssen weitere Anfragen darauf warten, dass d.3 search engine (SHAFT) einen besetzten Suchcursor freigibt.

Nach erfolgreicher Konfiguration klicken Sie auf **Weiter >>**, um zur nächsten Seite zu gelangen.

Schritt 3: Indexierung

Anmerkung

Beachten Sie, dass bei einer nachträglichen Änderung in einer bestehenden Dokumentensammlung der Suchindex neu aufgebaut werden muss.

Im dritten Schritt geben Sie an, wie die Dokumente in den Suchindex aufgenommen werden sollen.

Die folgenden Einstellungsmöglichkeiten stehen Ihnen dazu zur Verfügung:

Einstellung	Beschreibung
Numerische Werte	Ist diese Einstellung aktiviert, werden numerische Informationen aus den Dokumenten ebenfalls zu dem Index hinzugefügt.

Einstellung	Beschreibung
Anzahl Wortvorkommen	Dieser Wert bestimmt, ab wann ein Wort als "Stoppwort" klassifiziert und damit nicht mehr weiter indiziert wird. Gezählt werden alle Dokumente, in denen das Wort vorkommt. Anmerkung Diese Grenze kann angegeben werden, da z.B. eine Suche nach einem Wort, dass in mehr als 1.000.000 Dokumente vorkommt keinen Sinn macht. Zum einen wäre eine Treffermenge von 1.000.000 eine zu unfeine Suche für den Endanwender und zum anderen nimmt das Zusammenjoinen von Treffermengen jenseits von 1.000.000 Treffer vermehrt Rechenressourcen in Anspruch.
Minimale Wortlänge	Ein Wort muss mindestens so lang sein, wie der hier vorgegebene Wert, um in den Index aufgenommen zu werden.
StoreText	Gibt an, ob die vollständige Speicherung, des übermittelten Dokumentes, im Index aktiviert werden soll. Warnung Wenn Sie diese Funktion deaktivieren stehen Ihnen folgende Funktionen der Volltextsuchmaschine nicht zur Verfügung: • Semantische Suche • Phrasensuche • Bilden von Textzusammenfassungen (Abstracts) • Reindexierung über die SHAFUTL . exe Anmerkung Die vollständige Speicherung des Textinhaltes verursacht eine erhöhte Speicherplatzverwendung.
Nur die neueste Version im Index halten	Ist dieser Schalter gesetzt, wird nur die neueste Version eines Dokumentes im Index gehalten. Falls Sie den Index von alten Versionen der Dokumente bereinigen wollen, müssen Sie den Index durch eine Neuverschlagerung erneut aufbauen. Wird der Schalter zur Laufzeit aktiviert, wird das neue Verhalten nur bei neuen Dokumenten angewendet. Warnung Bei einer bestehenden Dokumentensammlung darf dieser Wert nachträglich nur aktiviert werden. Eine Deaktivierung würde dazu führen, dass Dokumente nicht mehr gefunden werden. Warnung Dieser Schalter setzt den d.3 server Version 7.2.1 voraus!

Nach erfolgreicher Konfiguration klicken Sie auf **Weiter >>**, um zur nächsten Seite zu gelangen.

Schritt 3: Indexierung

Geben Sie an, welche Textinhalte im Suchindex aufgenommen werden sollen.
Beachten Sie, dass bei einer nachträglichen Änderung der Suchindex neu aufgebaut werden muss.

Numerische Werte	☒	
Minimale Wortlänge	2	
Vollständige Speicherung der Textinhalte	☒	
Nur die neueste Version im Index halten	☒	

Schritt 4: Suche

Im vierten Schritt können Sie das Suchverhalten der d.3 search engine (SHAFT) einstellen.

Dazu stehen Ihnen die folgenden Einstellungsmöglichkeiten zur Verfügung:

Einstellung	Beschreibung
Erste Wildcardposition	Gibt die Zeichenposition an, ab der ein Wildcard-Symbol in einem Suchwort frühestens verwendet werden darf.
Autokorrektur	Gibt an, ob bei einer Suchanfrage, die keine Treffer geliefert hat, automatisch nach einem ähnlich geschriebenes Wort gesucht werden soll.
Nach Relevanz sortieren	Gibt an, ob zu jedem Treffer ein Relevanzwert errechnet werden soll und ob die Treffermenge anschließend nach den Relevanzwerten sortiert werden sollen.
Erweiterte Suche	Gibt an, ob die Suche um eine Wildcardsuche erweitert werden soll, wenn die Suche zu wenige Treffer zurück liefert.
Grenze für die erweiterte Wildcardsuche	Gibt an, ab welcher Treffermenge eine erweiterte Wildcardsuche durchgeführt werden soll.

Nach erfolgreicher Konfiguration klicken Sie auf **Weiter >>**, um zu nächsten Seite zu gelangen.

Schritt 5: SHAFT-Einstellung

Im fünften Schritt können Sie für die Dokumentensammlung mehrere d.3 search engines (SHAFT) erzeugen.

Diese können sofort oder später auf separate Rechner ausgelagert werden. Somit können Sie die Dokumentensammlung für eine spätere Skalierung schon an dieser Stelle vorbereiten.

Die Skalierung ermöglichen Sie durch das Setzen des Schalters **Skalierung aktivieren**. Anschließend erscheint ein neues Eingabefeld in dem Sie die Anzahl der zu erstellenden Engines eingeben können.

Anmerkung

Ein Suchindex sollte je nach Rechenkapazität maximal eine Größe von ca. 40 GB annehmen. Dies entspricht in etwa 30 Mio. Dokumenten (abhängig von der Größe der Dokumente).

Für jede d.3 search engine (SHAFT) können Sie die folgenden Einstellungen festlegen:

Einstellung	Beschreibung
Rechnername	Gibt den Hostname des Rechners an, auf dem die d.3 search engine (SHAFT) laufen soll. Warnung Auf dem Rechner muss der d.3 SHAFT-Agent installiert sein. Weitere Informationen finden Sie hier: d.3 SHAFT-Agent
Port	Gibt den TCP/IP Port an, unter dem die d.3 search engine (SHAFT) Anfragen entgegennimmt. Anmerkung Achten Sie auf unterschiedliche Portnummern, falls in Ihrem System mehrere d.3 search engine Prozesse aktiv sind.
Backup Verzeichnis	Gibt den Ablageort für das Online-Backup an. Anmerkung Damit das Backup zum Zeitpunkt der Ausführung die Suchanfragen nicht behindert, sollte dieses Verzeichnis auf einer extra Festplatte abgelegt werden.

Einstellung	Beschreibung
Datenbank Verzeichnis	Gibt das Verzeichnis an, in dem die d.3 search engine (SHAFT) seine Volltextdaten speichert. Dieses Arbeitsverzeichnis enthält später den vollständigen Index.
Anmerkung Als Daumenregel benötigt der unkomprimierte d.3 search engine (SHAFT) Index die 1,4-fache Größe der zu indizierenden Textmenge. Dieser Faktor kann unter 1 gebracht werden, wenn sich das Ablageverzeichnis auf einer NTFS-Partition befindet und Sie die Kompressionseigenschaft unter Microsoft Windows für dieses Verzeichnis setzen.	
Warnung Wenn Sie die Kompression unter Microsoft Windows einschalten, verlangsamt sich die Indizierungsgeschwindigkeit. Es wird daher nicht empfohlen, die Kompression einzuschalten. Die d.velop AG testet Kompressions-Szenarien nicht explizit. Überwacht ein Virens Scanner das Verzeichnis, verlangsamt sich die Indizierungsgeschwindigkeit. Um die bestmögliche Performance zu erzielen, sollte das Verzeichnis daher im Virens Scanner als Ausnahme definiert werden.	

Warnung

Bitte beachten Sie, dass die Anzahl der d.3 search engines (SHAFT) im Nachhinein nicht geändert werden kann. Eine nachträgliche Skalierung ist nur durch eine parallele Nachverschlagerung in eine neue skalierte Dokumentensammlung möglich.

Nach erfolgreicher Konfiguration klicken Sie auf **Weiter >>**, um zur nächsten Seite zu gelangen.

Assistent zur Anlage einer Dokumentensammlung

Willkommen
Basis
Indexierung
Suche
SHAFT-Einstellung
Backup
Fertigstellen

Schritt 5: SHAFT-Einstellung

Sie haben an dieser Stelle die Möglichkeit mehrere SHAFT-Engines für eine Dokumentensammlung zu erzeugen. Diese können sofort oder später auf separaten Rechnern ausgelagert werden, um die Performance zu erhöhen.

Bitte beachten Sie, dass die Anzahl der d.3 search engines (SHAFT) im Nachhinein nicht geändert werden kann. Eine nachträgliche Skalierung ist nur durch eine parallele Nachverschlagerung in eine neue skalierte Dokumentensammlung möglich.

Skalierung aktivieren ☐

SHAFT-Engine 1

Rechnername ⓘ

Port ⓘ

Backup-Verzeichnis ⓘ

Datenbank-Verzeichnis ⓘ

<< Zurück
Weiter >>

Schritt 6: Backup

Im sechsten Schritt können Sie das Backupverhalten der d.3 search engine (SHAFT) konfigurieren.

Einstellung	Beschreibung
Journaling	Gibt an, ob das Schreiben einer Journal-Datei eingeschaltet werden soll. Dieses bewirkt, dass Änderungen an der Datenbank zunächst in ein Journal geschrieben werden, wodurch bei einem plötzlichen Prozessabbruch die Indexdateien in einem konsistenten Zustand gehalten werden können.
Zeit für Backup	Gibt die Uhrzeit an, wann das Backup durchgeführt wird.
Datenbank vor Backup prüfen	Gibt an, ob der Index vor einem Backup geprüft werden soll.
Anmerkung Die Konsistenzprüfung vor einem Backup ist sinnvoll und sollte durchgeführt werden! Schalten Sie diese Option daher nur ab, wenn es aus zeitlichen Gründen nicht vertretbar ist.	
Tag für Backup	Gibt den Tag an, an welchen Tagen das Backup durchgeführt werden soll.

Nach erfolgreicher Konfiguration klicken Sie auf **Weiter >>**, um zu nächsten Seite zu gelangen.

Assistent zur Anlage einer Dokumentensammlung

Willkommen Basis Indexierung Suche SHAFT-Einstellung **Backup** Fertigstellen

Schritt 6: Backup

Ändern Sie die Parameter, um das Backupverhalten der SHAFT-Engine einzustellen.

Journaling	☒	
Online Backup		
▶ Zeit für Backup	00:00	
▶ Datenbank vor Backup prüfen	☒	
▶ Tag für Backup	Jeden Tag	

Schritt 7: Fertigstellen

Im siebten Schritt können Sie die Dokumentensammlung fertigstellen. Dabei wird die soeben konfigurierte Dokumentensammlung erstellt und gestartet.

Wenn ein Cluster vorliegt, wird die Dokumentensammlung an das Cluster verteilt und auf den anderen Instanzen ebenfalls gestartet.

Im Browser Google Chrome haben Sie die Möglichkeit, bereits an dieser Stelle die Einstellungen der Dokumentensammlung im JSON-Format herunterzuladen. Diese Einstellungen können als Sicherung dienen bzw. bei einer Installation dem Setup mitgegeben werden.

Im Browser Microsoft Internet Explorer wird diese Funktion erst nach dem Klick auf **Fertigstellen** angeboten.

Assistent zur Anlage einer Dokumentensammlung

Willkommen Basis Indexierung Suche SHAFT-Einstellung Backup **Fertigstellen**

Schritt 7: Fertigstellen

Klicken Sie auf "Fertigstellen", damit der d.3 search Assistent die gewählte Konfiguration vornimmt.
Einstellungen im JSON-Format herunterladen.

<< Zurück

Fertigstellen

Assistant for the generation of a document corpus

Welcome Basics Indexing Search SHAFT settings Backup **Finish**

Step 7: Finish

Document corpus has been successfully created.

This page is automatically refreshed after 11 seconds. Click on this hyperlink to refresh the page immediately.

[Download settings in JSON-format.](#)

Anmerkung

Wird d.3 search im Clusterbetrieb eingesetzt, schaltet sich der Cluster nach dem Anlegen der neuen Dokumentensammlung kurz inaktiv, um sich zu synchronisieren.

Dies können Sie an dem kleinen roten Warndreieck  vor dem Menüpunkt Cluster sehen.

Nach erfolgreicher Anlage der Dokumentensammlung erscheint der folgende Bestätigungstext. Zudem können Sie an dieser Stelle die Einstellungen im JSON-Format herunterladen.

Diese Einstellungen können als Sicherung dienen, bzw. bei einer Installation dem Setup mitgegeben werden.

Nach fünfzehn Sekunden wird diese Seite automatisch neu geladen, damit die neue Dokumentensammlung unter den Einstellungen zu finden ist.

Semantische Suche

Bei der semantischen Suche werden Korrelationen vom angefragten Suchausdruck zu anderen Dokumenten und Begriffen erzeugt. Die einzelnen, gefundenen Begriffe werden nach d.velop eigenen Analyseverfahren bewertet und dann als sogenannte Hints zurückgegeben.

Anmerkung

Wenn ein Nutzer nach dem Wort "d.velop" sucht, liefert die Suchengine die Wörter "Gescher" und "ECM". Dies resultiert daraus, da diese drei Wörter häufig zusammen in Briefköpfen auftauchen.

Da die Ressourcen eines Rechners endlich und zeitlich beschränkt sind, muss die Analysemenge eingeschränkt werden, über die die Statistik erzeugt wird.

Begrenzt wird die Analyse durch eine maximale Anzahl von Dokumenten ([MaxDocsForHints](#)) aus der Treffermenge. Die Wahl der Dokumente kann durch den Schalter [UseMostRecent](#) beeinflusst werden. Ist dieser Schalter gesetzt, werden zuerst Dokumente verwendet, die eine hohe Relevanz besitzen, ansonsten werden die Dokumente gleichverteilt aus der Ergebnismenge gewählt. Durch die Einstellung [MaxWordsForHints](#) wird bestimmt, wie viele der, für die Dokumente relevantesten, Wörter pro Dokument maximal in Betracht gezogen wird.

Warnung

Je größer die Werte [MaxDocsForHints](#) und [MaxWordsForHints](#) gewählt werden, desto mehr Rechenzeit wird für eine semantische Suche in Anspruch genommen.

Zusätzlich kann durch den Schalter [WordRelValue](#) bestimmt werden, dass Wörter, die in zu vielen Dokumenten vorkommen, herausgefiltert werden. Dies ist begründet dadurch, dass diese Wörter eine hohe Relevanz bekommen, einzig weil sie in vielen verschiedenen Dokumenten vorkommen, jedoch durch ihr häufiges Vorkommen zur Suche keinen Mehrwert beitragen.

Wird nun die Option [SemanticExtend](#) aktiviert, wird bei einer Suche nach Hints zusätzlich die Suche automatisch um die Hint-Begriffe erweitert.

Anmerkung

Wenn bei einer Suche nach Hints gesucht wird und zusätzlich die Option [SemanticExtend](#) aktiviert ist, werden durch die erweiterte Suche in der Ergebnismenge auch Dokumente vorhanden sein, in dem das ursprünglich gesuchte Wort nicht vorhanden ist.

Anmerkung

Bei der Suche nach "d.velop" werden z.B. die Begriffe "Gescher", "Softwareunternehmen", "Schildarpstraße" und "d.forum" als Wortwolke gefunden.

Bei der sematischen Erweiterung der Suche wird nun auch nach den Begriffen aus der Wortwolke gesucht. Dies führt dazu, dass mehr Treffer gefunden werden.

Anmerkung

Die automatisierte semantische Erweiterung ist in den meisten Fällen nur bei einer reinen Recherche-Suche sinnvoll. D.h. ein Anwender will sich über die d.velop AG informieren. Dabei will er zum einen direkte Treffer zur d.velop AG haben, andererseits interessieren ihn auch der Ort und evtl. andere Softwareunternehmen.

Bei einer gezielten Suche ist die automatische semantische Erweiterung eher irreführend für den Anwender.

Warnung

Wenn bei aktivierter semantischen Suche in einer Dokumentart gesucht wird, werden die semantischen Erweiterung nicht innerhalb dieser Dokumentart gesucht.

D.h. bei der Suche nach "d.velop" innerhalb der Dokumentart "Kontakte" werden die Begriffe aus der Wortwolke, z.B. "Softwareunternehmen", auch aus anderen Dokumentarten gesucht.

Das führt dazu, dass der d.3 server viele Treffer nachfiltern muss.

Zudem kann es bei aktivierter automatischen Erweiterung dazu kommen, dass eine Verfeinerung der Suche, z.B. "d.velop Softwareunternehmen Schildarpstraße", mehr Treffer liefert, als die Suche nach dem Wort "d.velop". Dies liegt an der jeweiligen Wortwolke, die zur semantischen Erweiterung verwendet wird.

Konfiguration der d.3 search engine

Beim Start eines d.3 search engine Prozesses (shaft . exe) wird der Dateiname einer Konfigurationsdatei als Startparameter übergeben (in der Regel shaft . ini).

Diese Datei enthält folgende Parameter, die z. T. bei der [Anlage einer Dokumentsammlung](#) abgefragt werden:

Sektion	Parameter	Bedeutung
[Engine]		
Archive-Logs	Archive-Logs	Journale archivieren (1: ja) , (0: nein) Standard: 1 Voraussetzung: Journal=1 Bewirkt, dass jedes Journal im UnterverzeichnisARCHLOGSgespeichert wird, um die Datenbank im Notfall wiederherstellen zu können (SHAFTUTL {INI-Datei} -i). Die Journal-Dateien können Sie dann mit einem Backup-Programm sichern.
	AutoCorrect	Automatische Korrektur unbekannter Suchbegriffe (0: aus), (1: an) Standard: 1 Ist dieAutomatische Korrektur unbekannter Suchbegriffeaktiviert, wird bei Suchanfragen, die zu keinen Treffern führen, automatisch ein ähnlich geschriebenes Wort verwendet und danach gesucht.

Sektion	Parameter	Bedeutung									
	Cache - Size	Speicherplatz für Datenzwischenspeicherung Standard: 256000 Dieser Wert in KB gibt an, wie viel Hauptspeicher für die Datenzwischenspeicherung von Transaktionen bzw. Suchanfragen verwendet werden soll. Je nach dem, ob die Engine in 32 Bit oder 64 Bit betrieben wird, kann der Cache unterschiedlich groß konfiguriert werden: <table border="1"> <thead> <tr> <th></th><th>Shaft 32 Bit</th><th>Shaft 64 Bit</th></tr> </thead> <tbody> <tr> <td>Microsoft Windows 32 Bit</td><td>max. 1,5 GB</td><td>-</td></tr> <tr> <td>Microsoft Windows 64 Bit</td><td>max. 3,5 GB</td><td>max. 16384 GB</td></tr> </tbody> </table> Der Cache sollte nicht größer als Arbeitsspeicher-1GB konfiguriert werden. Bei virtuellen Festplatten sollte der Cache größer oder gleich der Indexgröße betragen. Ist der Cache zu klein dimensioniert und führt dies dazu, dass die Laufzeit einzelner Anfragen länger als fünf Sekunden beträgt, legt die Engine ein PerfLogsVerzeichnis an und legt darin Protokolle über das Laufzeitverhalten des Anfrage an. Darin wird unter anderem ein Prozentwert für die Cachehitratio angegeben. Ist dieser Wert kleiner als 98%, sollte der Cache vergrößert werden.		Shaft 32 Bit	Shaft 64 Bit	Microsoft Windows 32 Bit	max. 1,5 GB	-	Microsoft Windows 64 Bit	max. 3,5 GB	max. 16384 GB
	Shaft 32 Bit	Shaft 64 Bit									
Microsoft Windows 32 Bit	max. 1,5 GB	-									
Microsoft Windows 64 Bit	max. 3,5 GB	max. 16384 GB									
	Commit - Count	Standard: 200 Der Parameter gibt an, nach wie vielen Einfügeoperationen der Index auf der Festplatte aktualisiert wird.									
	Database	Ablage der Indexdaten Pfad zur d.3 search engine Datenbank. Dieser Pfad muss auf einer lokalen Festplatte liegen. Netzwerkpfade werden von der Suchmaschine nicht akzeptiert.									
	Logpath	Pfad zum d.3 search engine Transaktions-Log Um die Performance der d.3 search engine zu optimieren und die Festplatte, auf dem die Datenbank liegt, weniger zu beanspruchen, sollte dieser Pfad auf ein separates Laufwerk geändert werden.									

Sektion	Parameter	Bedeutung
	Delim- termode	Dieser Parameter wird standardmäßig nicht in <code>dieshaft.in</code> eingetragen. Standard: bei alten Indizes: 0, bei neuen Indizes: 1 0: Standardworttrennung, d. h. alle Zeichen außer Buchstaben oder Zahlen werden als Trennzeichen interpretiert 1: <code>;;_,-/\</code> werden nicht als Trennzeichen interpretiert, allerdings am Anfang und am Ende eines Wortes entfernt 2: <code>;;_,-/\</code> werden nicht als Trennzeichen interpretiert, allerdings aus den Worten entfernt 3: Es wird eine Datei mit dem Namen <code>delimter.txt</code> im <code>CorpusVerzeichnis</code> erwartet, welche die Trennzeichen enthält. Diese Datei enthält je Zeile ein Trennzeichen. Es können auch Multibyte Trennzeichen definiert werden (z.B. UTF8). Dazu werden einfach die gewünschten Zeichen in eine Zeile geschrieben. Der Delimitermode als auch <code>delimter.txt</code> darf zur Laufzeit nicht mehr geändert werden. Soll der Delimitermode oder <code>delimter.txt</code> geändert werden, muss der Index neu aufgebaut werden. Eine Beispiel-Datei <code>delimter.txt</code> finden Sie im Ordner <code>Vorlagen</code>. Sofern <code>Stortext</code> aktiviert ist, kann dies mit Hilfe von <code>desshaft.tut</code> und dem Parameter <code>-rid</code> durchgeführt werden. Andernfalls muss der Index vollständig neu gefüllt werden. 4: Es wird der BreakIterator der OpenSource Komponente ICU (International Components for Unicode) verwendet. Dieser Modus ist erforderlich, wenn z.B. asiatische Sprachen verwendet werden sollen, die nur über Wörterbücher getrennt werden können. Warnung Bei diesem Delimitermode muss d.3 search im Unicode Modus betrieben werden, sonst können die asiatischen Sprachen nicht verarbeitet werden. Anmerkung Weitere Informationen zu der Komponente ICU finden Sie unter http://www.icu-project.org und Internationalisierung.
	IcuExc- ludedDe- limiter	Dieser Parameter hat nur in Verbindung mit dem Parameter <code>DelimiterMode=4</code> eine Funktion. Er gibt an, welche Zeichen, die von der ICU Komponente als Trennzeichen interpretiert werden, nicht als Trennzeichen interpretiert werden sollen. Standard: - Hier können hintereinander ohne Trenner die Zeichen angegeben werden, die von der ICU Komponente nicht als Trennzeichen interpretiert werden sollen. Ist der Parameter nicht angegeben, wird das Minus/Bindestrich Zeichen nicht als Trenner interpretiert, so dass Worte die mit Bindestrich verbunden sind, nicht getrennt abgelegt werden. Dies ist vor allem für Produktbezeichnungen wichtig, die häufig Bindestriche enthalten. Wird der Bindestrich nicht ausgeklammert, würde die Engine bei einer Suche nach einem mit Bindestrich verbundenem Wort eine UND Suche mit zwei Worten durchführen und ggf. mehr Treffer finden als gewünscht.

Sektion	Parameter	Bedeutung
	First-WildCard	Frühest mögliche Wildcardposition Standard: 3 Gibt die Zeichenposition an, ab der ein Wildcard-Symbol in einem Suchwort frühestens auftreten darf. Als Wildcard-Zeichen können Sie "?" oder "*" nutzen. Dieser Parameter steht im Standard auf "3". Wird dieser Parameter auf "0" oder "1" gesetzt, erzeugt die Engine einen weiteren nGramm Index (die Dateien nGramm.idx und nGramm.dat). Sofern der Index schon Daten enthält, wird der Index beim Neustart zunächst angelegt und mit den bereits vorhanden Daten gefüllt. Dieser Index verlangsamt den Import und benötigt zusätzlich Platz auf der Festplatte, ermöglicht der Engine aber bei Suchen mit führendem Wildcard gefolgt von mindestens drei Zeichen, auf das Vergleichen aller Worte im Index zu verzichten. Dadurch können derartige Suchanfragen wesentlich schneller beantwortet werden. Wird der Wert nachträglich wieder auf einen Wert größer 1 gesetzt, werden die nGramm Dateien automatisch entfernt und der Festplattenplatz wieder freigegeben.
	Journal	Schreibe Journale (Redo-Log) (1: ja), (0: nein) Standard: 1 Der Wert "1" aktiviert das Schreiben einer Journal-Datei. Die Journal-Datei bewirkt, dass Änderungen an der Datenbank zunächst in ein Journal geschrieben werden. Anhand der Journal-Datei können bei einem plötzlichen Prozessabbruch die Indexdateien in einem konsistenten Zustand gehalten werden.
	LogLevel	Standard: 9 Setzt die Höhe des Loglevels des Transaktions-Protokolls. Bei "0" werden nur Fehler protokolliert. Bei "9" wird alles protokolliert, inklusive der Texte, die beim Einfügen in die Datenbank übergeben wurden. In Verbindung mit dem Onlinebackup kann dieser Parameter problemlos auf "9" eingestellt werden, da das Log automatisch gelöscht wird. Wird kein Onlinebackup verwendet, wächst das Verzeichnis sehr schnell an.
	Cursor	Max. Anzahl der Suchcursor Standard: 64 Der Parameter definiert, wie viele Suchanfragen im Hauptspeicher gehalten werden. Somit beschreibt dieser Wert, wie viele Anfragen von d.3 search engine gleichzeitig bearbeitet werden können. Es muss ein Wert > "0" eingetragen werden.
	MaxDocs-ForHints	Standard: 100 Gibt an, wie viele Dokumente einer Treffermenge maximal für die semantische Analyse herangezogen werden. Der ParameterMaxDocsForHintswird explizit auf "0" gesetzt, wenn keine Semantik (Semantische Erweiterungnicht aktiviert) gewünscht ist.
	Max-Words-ForHints	Standard: 100 Gibt an, wie viele Worte pro Dokument aus der Treffermenge maximal für die semantische Analyse benutzt werden. Verwendet werden dabei jeweils die für das Dokument relevantesten Werte.
	Numeric	Numerische Werte (inkl. Datumsangaben) erfassen (1: ja), (0: nein) Standard: 1 Bewirkt, dass numerische Informationen aus dem Dokument ebenfalls in der Datenbank indiziert werden.

Sektion	Parameter	Bedeutung
	Orderby-rank	Sortierung von Suchergebnissen nach Relevanz (1: an), (0: aus) Standard: 1 Aktivieren Sie diesen Parameter, wird für jedes Dokument ein Relevanzwert zur verwendeten Suchanfrage berechnet. Anhand dieses Relevanzwertes wird die Treffermenge anschließend sortiert, so dass das Dokument mit der höchsten Relevanz als erstes ausgeliefert wird.
	Port	d.search TCP/IP Port Standard: 3541 TCP/IP Port unter der die d.3 search engine Anfragen entgegennimmt. Es ist darauf zu achten, dass alle lokal laufenden Prozesse einen anderen Port benutzen.
	SemanticExtend	Suchbegriffe vor der Suche erweitern (1: an), (0: aus) Standard: 0 Aktivieren Sie diesen Parameter, fließt die Wortwolke in die Suche mit ein. D. h., die Begriffe der Wortwolke werden dem Anwender nicht nur angezeigt, sondern es wird direkt danach gesucht, wodurch die Ergebnismenge größer wird. Die Worte aus der Wolke bekommen dabei je nach Abstand zur eigentlichen Suche ein geringeres Gewicht für die Bestimmung des Rankings. Weitere Informationen finden Sie auch hier: Semantische Suche
	Store-Text	Vollständige Speicherung der Textinhalte (1: an), (0: aus) Standard: 1 StoreText=1 aktiviert die vollständige Speicherung des Textes, vom übermittelten Dokument, im Index Warnung Wenn Sie diese Funktion deaktivieren, stehen Ihnen folgende Funktionen der Volltextsuchmaschine nicht zur Verfügung: • Phrasensuche • Bilden von Textzusammenfassungen (Abstracts) • Reindexierung über die SHAFTUTL . exe Anmerkung Die vollständige Speicherung des Textinhaltes verursacht eine erhöhte Speicherplatzverwendung.
	UseMost-Recent	Semantik auf Basis der Relevanz Regelt die Auswahl der Dokumente aus der Treffermenge, die für die semantische Analyse ausgewählt werden. 1: Verwendet zuerst die Dokumente mit höchster Relevanz zur semantischen Analyse 0: Ermittelt anhand einer Gleichverteilung die Dokumente Standard: 1
	WordRel-value	Standard: 150 Relativer Relevanzwert eines Wortes, der nicht unterschritten werden darf. Wird dieser Wert für ein Wort überschritten, wird es aus den Suchvorschlägen entfernt und durch ein anderes ersetzt. Der Wert sollte auf "150" eingestellt werden, wenn der Parameter MaxWordsForHints auf "100" eingestellt ist.

Sektion	Parameter	Bedeutung
	MinWordLen	Standard: 2 Gibt an, wie lang ein Wort mindestens sein muss, um in den Index aufgenommen zu werden. Um die Begrenzung aufzuheben, kann dieser Parameter auf "0" gesetzt werden.
	AutoRepair	Standard: 0 1: Wird bei einem Backup ein Fehler im Index festgestellt, wird dieser automatisch berichtigt. Dies ist nur möglich, wenn ein Transaktions-Log vorhanden ist und ein Onlinebackup vorliegt. Warnung Bei der Behebung der Fehler im Index wird eine zusätzliche Festplattenkapazität in der Größe des aktuellen Indexes benötigt. Die Festplattenkapazität muss auf der Festplatte zur Verfügung stehen, auf der der Index abgelegt ist.
	WildcardLimit	Standard: 1000000 Gibt an, nach wie vielen Treffern die Engine beendet. Das heißt, sobald die Engine 1.000.000 Treffer gefunden hat, hört diese auf zu suchen, auch wenn es evtl. noch relevantere Treffer gibt.
	AutoDefrag	(1: an), (0: aus) Standard: 1 Dieser Schalter gibt an, ob bei einem Backup der Index automatisch defragmentiert werden soll. Die Defragmentierung wird nur durchgeführt, wenn die Fragmentierung des Index größer als 15% ist. Je nach Größe des Index kann eine Defragmentierung einige Zeit in Anspruch nehmen. Warnung Bei der Defragmentierung wird eine zusätzliche Festplattenkapazität in der Größe der Dateien link.dat und words.dat benötigt. Die Festplattenkapazität muss auf der Festplatte zur Verfügung stehen, auf der der Index abgelegt ist.
[PerfLogs]		
	Abstract	Hinweise zur Dokumentensammlung mit der d.3 search engine Standard: 5000 Zeit in Millisekunden als Grenzwert für das Ausführen der Operation Abstract. Wird der Grenzwert überschritten, wird eine Log-Datei im Verzeichnis <code>... \d3\d.3 search\corpora\[CorpusId]\PerfLogs</code> angelegt.
	Delete	Hinweise zur Dokumentensammlung mit der d.3 search engine Standard: 5000 Zeit in Millisekunden als Grenzwert für das Ausführen der Operation Delete. Wird der Grenzwert überschritten, wird eine Log-Datei im Verzeichnis <code>... \d3\d.3 search\corpora\[CorpusId]\PerfLogs</code> angelegt.
	Search	Hinweise zur Dokumentensammlung mit der d.3 search engine Standard: 5000 Zeit in Millisekunden als Grenzwert für das Ausführen der Operation Search. Wird der Grenzwert überschritten, wird eine Log-Datei im Verzeichnis <code>... \d3\d.3 search\corpora\[CorpusId]\PerfLogs</code> angelegt.
	Insert	Hinweise zur Dokumentensammlung mit der d.3 search engine Standard: 5000 Zeit in Millisekunden als Grenzwert für das Ausführen der Operation Insert. Wird der Grenzwert überschritten, wird eine Log-Datei im Verzeichnis <code>... \d3\d.3 search\corpora\[CorpusId]\PerfLogs</code> angelegt.

Sektion	Parameter	Bedeutung
	Spell-check	Hinweise zur Dokumentensammlung mit der d.3 search engine Standard: 5000 Zeit in Millisekunden als Grenzwert für das Ausführen der Operation Spellcheck. Wird der Grenzwert überschritten, wird eine Log-Datei im Verzeichnis ... \d3\d.3 search\corpora\[CorpusId]\PerfLogs angelegt.
[Backup]		
	CheckDay	Prüfung erfolgt ... Standard: 0 Gibt an, an welchem Tag das Backup geprüft werden soll. 0: jeden Tag 1: Sonntag, 2: Montag, usw. Anmerkung Die Archlogs und die Translogs werden nur nach einem erfolgreichen Check gelöscht, um die Datenbank wieder herstellen zu können, wenn das Backup nicht konsistent war. Anmerkung Während des Backups können Suchanfragen weiterhin verarbeitet werden. Änderungen sind während der Zeit nicht möglich.
	CheckDB	Die Datenbank wird vor der Sicherung überprüft Standard: 1 Gibt an, ob der Index überhaupt vor einem Backup geprüft werden soll. Anmerkung Die Überprüfung versichert, dass das Backup anschließend konsistent ist. Eine Abschaltung dieser Funktion führt zwar zu einem schnelleren Backup, versichert Ihnen aber nicht, dass das Backup in sich konsistent ist.
	Path	Ablage in: Ablageort für das Online-Backup Anmerkung Damit die Festplatte, auf dem die Datenbank gespeichert ist, nicht unnötig beim Backup belastet wird, empfiehlt es sich das Backup-Verzeichnis auf einer separaten Festplatte anzugeben. Warnung Wenn dieser Wert nicht definiert wird, findet kein Backup statt. Ohne Backup stehen die folgenden Funktionen nicht zur Verfügung: • Automatische Defragmentierung • Wiederherstellung einer Datenbank • Regelmäßige Überprüfung der Datenbank Zudem muss bei einer defekten Datenbank eine Reindizierung über den d.3 server durchgeführt werden.

Sektion	Parameter	Bedeutung
	Time	Uhrzeit Standard: 00:00 Zeitpunkt, wann das Online-Backup erfolgen soll Anmerkung Wählen Sie am besten einen Zeitpunkt, an dem möglichst wenige Mitarbeiter mit dem d.3 System arbeiten.

Anmerkung

Ein Beispiel einer shaft.ini:

```
[Engine]
Database=C:\d3\d.3 search\corpora\D3P
Port=3541
Journal=1
MaxFilesPerWord=10000000
ArchiveLogs=1
Numeric=1
CacheSize=256000
PhysicalCache=1
MaxCursor=10
WordRelValue=150
MaxWordsForHints=100
FirstWildCard=3
AutoCorrect=1
StoreText=0
Orderbyrank=1
MaxDocsForHints=0
SemanticExtents=0
UseMostRecent=0
LogLevel=9
[Backup]
Path=C:\d3\d.3 search\corpora\D3P\backup
Time=00:00
CheckDB=1
CheckDay=0
```

Anmerkung

Ein Auszug aus einer stopwords.txt Datei (Sie finden eine solche Datei im Unterverzeichnis corpora \ d.3 search engine Name).

```
im
aber
als
am
an
auch
auf
aus
bei
bin
bis
bist
da
```

Verwaltung der stopwords.txt

Die stopwords.txt können Sie über das Webinterface der d.3 search Instanz unter dem Menüpunkt **Einstellungen | Dokumentensammlung | <Corpus-ID> | stopwords.txt** editieren.

Über das Webinterface wird die stopwords.txt in das Datenbankverzeichnis der jeweiligen d.3 search engines (SHAFT) abgelegt.

Jede Zeile steht für ein Stopwort. Erst nach einem Neustart der d.3 search engine (SHAFT) greifen die Änderungen.

Der Eintrag eines Wortes in der Stopwortliste hat folgende Auswirkungen:

- Das Wort wird bei neuen Dokumenten nicht mit in den Suchindex aufgenommen
- Das Wort wird bei einer Suche aus der Suchanfrage entfernt. Dadurch greift das Stopwort auch bei bereits importierten Dokumenten.

Anmerkung**Beispiel: Ignorieren von Artikeln**

Bei einer Suchanfrage sollen Artikel, wie z.B. "der, die das" nicht berücksichtigt werden, da die Artikel bei der Suche zu vernachlässigen sind

Stopworte:

- der, die, das

Suchanfrage:

- "Verwaltung der Stopworte"

Ergebnis:

- Alle Dokumente in denen die Worte "Verwaltung" und "Stopworte" vorkommen.
- Unter den Treffern kann es Dokumente OHNE "der" geben

Gute Erweiterungen für die Stopwortliste:

- Worte, die für eine Suche nicht aussagekräftig sind, da Sie in jedem Dokument vorkommen.
- der eigene Firmennamen, der in jedem Dokument vorkommt.

Warnung

Der Inhalt der stopwords.txt **MUSS** in Kleinschrift erweitert werden. Das gilt auch für Umlaute.

Anmerkung

Sofern die Engine im UTF-8 Modus läuft, muss die Datei beim Speichern UTF-8 kodiert abgelegt werden.

Ob eine d.3 search Dokumentensammlung im UTF-8 Modus läuft, können Sie der Datei corpus.ini entnehmen.

Verwaltung der synonym.txt

Es ist möglich, einen Katalog mit Synonymen zu pflegen. Ziel ist es, die Suchfunktionalität zu verbessern, indem zu den Sucheingaben des Benutzers weitere Begriffe in die Suche aufgenommen werden, die in der synonym.txt gepflegt werden.

Anmerkung

Synonym.txt:

Auto PKW Wagen

Sucht der Benutzer nach dem Wort Auto, wird die Suche zusätzlich um die Wörter PKW und Wagen erweitert.

Anmerkung

Wenn ein Synonym-Katalog gepflegt wird, findet keine automatische Korrektur der Suchanfrage statt.

Die `synonym.txt` können Sie über das Webinterface der d.3 search Instanz unter dem Menüpunkt **Einstellungen | Dokumentensammlung | <Corpus-ID> | synonym.txt** editieren.

Pro Begriff wird eine Zeile verwendet. Es gibt zwei Möglichkeiten Einträge vorzunehmen:

1. Der Eintrag in dem Katalog sieht wie folgt aus: **Auto Pkw Wagen**
In diesem Fall wird nach allen drei Wörtern gesucht, egal welches Wort der Benutzer eingeben hat.
2. Der Eintrag in dem Katalog sieht wie folgt aus: **Auto = Pkw Wagen**
In diesem Fall wird nur nach allen drei Worten gesucht, wenn der Benutzer das Wort Auto eingeben hat.

Um die Einstellungen wirksam zu machen, muss die entsprechende Engine neu gestartet werden.

Verwaltung der `validwords.txt`

Warnung

Wird dieser Katalog angepasst, gelten die Änderungen nur für neuverschlagwortete Dokumente. Um die Änderungen auch rückwirkend wirksam zu machen, muss der Index neu verschlagwortet werden.

Weitere Informationen dazu finden Sie im Kapitel [Reindexierung einer Dokumentensammlung](#).

Es ist möglich, einen Katalog mit `ValidWords` zu pflegen. Diese stehen für Begriffe, welche bei der Textanalyse von der Suchmaschine nicht getrennt werden sollen, also ohne Beachtung der normalen Trennzeichen in der Suche gefunden werden können.

Anmerkung

`validwords.txt`:

`d.velop!AG`

`d-velop.de`

Der Begriff "d.velop!AG" wird nun nicht mehr in drei Worte geteilt (d und velop und AG) abgespeichert, sondern ist der Suchmaschine als d.velop!AG bekannt!

Die `validwords.txt` können Sie über das Webinterface der d.3 search Instanz unter dem Menüpunkt **Einstellungen | Dokumentensammlung | <Corpus-ID> | validwords.txt** editieren. Pro Begriff wird eine Zeile verwendet.

Um die Einstellungen wirksam zu machen, muss die entsprechende Engine neu gestartet werden.

Anmerkung

Wenn die `validwords.txt` beim Start der Engine nicht gefunden wird, wird die Erkennung von `ValidWords` aus Performanzgründen deaktiviert. Sie finden dazu auch einen entsprechenden Eintrag im d.3 log.

Erweiterte Konfiguration einer Dokumentensammlung

Für jede Dokumentensammlung legt d.3 search ein eigenes Verzeichnis an, in dem Konfigurations- und Indexdaten von der Suchmaschine abgelegt werden.

Dieses Verzeichnis lautet im Allgemeinen `d.3 search Verzeichnis\corpora\[CorpusId]` (z. B. `D:\d3\d.3 search\corpora\d3p`).

Die Hauptkonfigurationsdatei für den d.3 search Server ist in diesem Verzeichnis die Datei `dsearch.ini`.

Die Einträge, die Sie dort finden, haben folgende Bedeutung:

Sektion	Parameter	Bedeutung
[Mirrors]		
	CorpusId	Im Standard ist dieser Parameter nicht vorhanden. Wenn eine Dokumentensammlungs-ID bei diesem Parameter gesetzt wird, wird die Dokumentensammlung mit der aktuellen gespiegelt.
		Anmerkung Es werden nur neue Dokumente gespiegelt, eine Synchronisierung mit den Altdaten findet nicht statt.

Die Konfigurationsdatei für die Dokumentensammlung ist in diesem Verzeichnis die Datei `corpus.ini`.

Die Einträge, die Sie dort finden, haben folgende Bedeutung:

Sektion	Parameter	Bedeutung
[Shaft_Global]		
	UTF8Support	(1: ja), (0: nein) Standard: 1 Dieser Wert gibt an, ob die Dokumentensammlung die Zeichenkodierung UTF8 unterstützt.
		Anmerkung Wenn Ihre Dokumente im d.3 UTF8 kodiert sind, sollten Sie diesen Schalter aktivieren.
	ForceNoHints	(1: ja), (0: nein) Standard: 0 Durch diesen Schalter wird die Erstellung von Wortwolken für diese Dokumentensammlung bei einer Suche deaktiviert.
[ShaftX]		
	Host	Standard: localhost Hostname des Rechners, auf dem die d.3 search engine (SHAFT) läuft.
	Port	Standard: 3541 Gibt den Port an, auf dem die d.3 search engine (SHAFT) reagiert.
	SHAFTAgentPort	Standard: 6133 Gibt den Port an, über den der d.3 SHAFT-Agent erreichbar ist.
		Anmerkung Diese Einstellung wird nur benötigt, wenn die d.3 search engine (SHAFT) auf einem anderen Rechner ausgelagert wird.
[Search]		

Sektion	Parameter	Bedeutung
	Expand-Search	(1: ja), (0: nein) Standard: 0 Dieser Schalter gibt an, ob die Suche um eine Wildcardsuche erweitert werden soll, wenn die Suche zu wenige Treffer zurückliefert. Dabei wird hinter jedem Wort ein * angehängt, um die Treffermenge zu erhöhen. Anmerkung Wenn eine automatische Wildcardsuche durchgeführt wird, findet kein Abgleich mit dem Synonymkatalog statt.
	Expand-SearchMin-Result	Standard: 20 Gibt an, ab welcher Treffermenge eine erweiterte Wildcardsuche durchgeführt werden soll. Wenn der Wert 0 angegeben wird, findet bei jeder Suche eine Wildcardsuche statt. Bedenken Sie, dass eine Wildcardsuche zeitaufwändiger ist und nur wenn nötig ausgeführt werden sollte.
[General]		
	Current-Versions-Only	(1: ja), (0: nein) Standard: 1 Dieser Schalter gibt an, ob nur eine Version eines Dokumentes im Index gehalten werden soll oder mehrere Versionen. Im Standard wird ein Dokument im Index im Format <Dokumenten_ID> . <Versions_ID> abgespeichert. Wenn der Schalter aktiviert wird (auf 1 gesetzt wird), werden die Dokumente im Index im Format <Dokumenten_ID>, also ohne Versionsangabe, abgespeichert. Es ist möglich, den Schalter zur Laufzeit umzuschalten, jedoch wird das Verhalten nur bei neuen Dokumenten angewendet. Warnung Der Schalter darf nur von 0 auf 1 während der Laufzeit gestellt werden. Eine Umschaltung in die andere Richtung würde dazu führen, dass Dokumente nicht mehr gefunden werden. Falls Sie den Index um die alten Versionen der Dokumente bereinigen wollen, müssen Sie den Index durch eine Neuverschlagwortung erneut aufbauen. Anmerkung Damit die Einstellung aktiviert wird, müssen alle d.3 server-Prozesse manuell neu gestartet werden. Warnung Der Schalter setzt die Version 7.2.1 des d.3 servers voraus.
	UseFile-Stream	(1: ja), (0: nein) Standard: 0 Dieser Schalter gibt an, ob die importierten Dateien über einen FileStream gelesen werden sollen. Dies hat den Vorteil, dass nicht die komplette Datei in den Arbeitsspeicher geladen wird. Durch das Setzen dieses Schalters wird jedoch der Import verlangsamt. Es wird empfohlen, den Schalter nur zu setzen, wenn der Arbeitsspeicher beim Lesen größerer Dateien nicht ausreicht.

Löschen einer Dokumentensammlung

Warnung

Eine Dokumentensammlung zu löschen bedeutet, diese unwiderbringlich, inklusive eingestelltem Backupverzeichnis, zu entfernen.

Bei großen Dokumentensammlung kann ein Neuaufbau über den d.3 server bis zu mehrere Wochen in Anspruch nehmen.

Bitte gehen Sie äußerst bedacht mit diesem Befehl um!

Warnung

Befindet sich d.3 search im Clusterbetrieb, wird dieser Befehl automatisch an alle Clus-termitglieder weitergeleitet. Dies führt zum Löschen der Dokumentensammlung auf jedem Knoten des Clusterverbundes!

Das Löschen einer Dokumentensammlung kann mit einem Aufruf über das Webinterface gestartet werden. Rufen Sie dazu zunächst das Webinterface einer d.3 search Instanz auf, z.B.: <http://localhost:6133> oder rufen Sie die Verknüpfung d.3 search Webinterface aus dem d.3 search Startmenü auf.

Ergänzen Sie dann die vorhandene URL um den Befehl `/deleteCorpora?Corpora=<Name der Dokumentensammlung>`.

Anmerkung

Eine mögliche URL zum Löschen einer Dokumentensammlung wäre:

`http://localhost:6133/deleteCorpora?Corpora=d3A`

Umbenennen einer Dokumentensammlung

Warnung

Eine Dokumentensammlung umzubenennen kann dazu führen, dass die Zuordnung Dokumentart zu Dokumentensammlung in der d.3 Administration nicht mehr korrekt ist.

Bitte gehen Sie äußerst vorsichtig mit diesem Befehl um!

Warnung

Befindet sich d.3 search im Clusterbetrieb, wird dieser Befehl automatisch an alle Clus-termitglieder weitergeleitet. Dies führt zum Umbenennen der Dokumentensammlung auf jedem Knoten des Clusterverbundes!

Das Umbenennen einer Dokumentensammlung kann mit einem Aufruf über das Webinterface gestartet werden. Rufen Sie dazu zunächst das Webinterface einer d.3 search Instanz auf, z.B.: <http://localhost:6133> oder rufen Sie die Verknüpfung d.3 search Webinterface aus dem d.3 search Startmenü auf.

Ergänzen Sie dann die vorhandene URL um den Befehl `/renamecorpus?OldCorpusname=<Name der bestehenden Dokumentensammlung>&?NewCorpusname=<Name der neuen Dokumentensammlung>`.

Anmerkung

Eine mögliche URL zum Löschen einer Dokumentensammlung wäre:

<http://localhost:6133/renamecorpus?OldCorpusName=d30ld&NewCorpus-Name=d3New>

Konfiguration des d.3 Systems

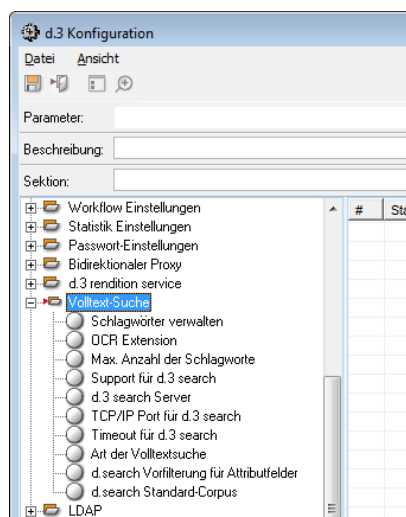
Bitte gehen Sie in folgenden Schritten zur Konfiguration des d.3 Systems vor:

1. [Konfiguration in der d.3 Administration unter d.3 config](#)
2. [Konfiguration in der d.3 Administration unter d.3 search](#)

Konfiguration in der d.3 Administration unter d.3 config

Zur Anbindung von d.3 search an d.3 müssen verschiedene Parameter in der d.3 Konfiguration über die d.3 Administration konfiguriert werden.

Diese Parameter finden Sie in der Sektion **Volltext-Suche**.



Dazu zählen:

Schlagwörter verwalten (KEYWORD_OPTION)

Standardwert: 0

Die folgenden Optionen stehen für die Schlagwortverwaltung zur Verfügung:

0	es werden keine Wörter beim Import in das Schlagwortverzeichnis übernommen
1	nur die Attribute des Dokumentes werden in das Schlagwortverzeichnis übernommen
2	die Attribute des Dokumentes sowie der Dokumenten-Inhalt (OCR-Datei) werden in das Schlagwortverzeichnis übernommen
3	die Attribute des Dokumentes sowie der Inhalt der Notiz-Datei werden in das Schlagwortverzeichnis übernommen
4	die Attribute, der Dokumenten-Inhalt (OCR-Datei) sowie der Inhalt der Notiz-Datei werden in das Schlagwortverzeichnis übernommen

Das Schlagwortverzeichnis wird für die Volltextsuche genutzt.

Anmerkung

Wenn d.3 search eingesetzt wird, muss dieser Wert mindestens auf "1" stehen.

OCR Extension (ocr_ext)

Standardwert: OCR

Endung, an der d.3 die Schlagwort-Datei erkennt.

Beim Batch-Import wird nur dann eine Verschlagwortung durchgeführt, wenn eine Datei mit der angegebenen Endung existiert und die Konfigurationsvariable KEYWORD_OPTION auf "2" steht.

Die Schlagwort-Datei sollte eine reine ASCII-Datei ohne Sonderzeichen sein.

Dokumente, die mit Textverarbeitungsprogrammen erstellt wurden, z.B. .doc-Dateien mit Winword, sind nicht geeignet. Sie müssen erst konvertiert werden, das heisst, die Formatierungs-Information muss entfernt werden.

Max. Anzahl der Schlagworte (OCR_MAX_WORDS)

Standardwert: 100

Anzahl der Worte, die maximal aus jeder Schlagwortdatei in d.3 übernommen werden.

Diese Anzahl gilt für alle Dokumentarten, sofern unter ocr_dokuart und ocr_dokuart_max_woerter nicht etwas anderes für die jeweilige Dokumentart eingestellt ist.

Support für d.3 search (DSEARCH_SUPPORT)

Standardwert: Nein

Einstellung der Nutzung des optionalen Moduls d.3 search (externe Volltextsuchmaschinen) für d.3.

Für den Import muss der Parameter KEYWORD_OPTION mindestens auf "1" stehen.

d.3 search Server (d3fc_dsearch_server_name)

Standardwert: localhost

Name des Rechners, auf dem d.3 search installiert ist.

Über d.3 search können externe Volltextengines eingebunden werden.

TCP/IP Port für d.3 search (d3fc_dsearch_server_tcpip_port)

Standardwert: 3406

TCP/IP-Port des Rechners, auf dem d.3 search installiert ist.

Timeout für d.3 search (d3fc_dsearch_server_timeout)

Standardwert: 30

Maximale Zeit in Sekunden, die d.3 auf das Ergebnis wartet, wenn es einen Auftrag an d.3 search geschickt hat.

Über dieses Timeout werden Staus in den d.3 Prozessen verhindert, falls die Verarbeitung bei d.3 search unerwartet lang dauert.

Art der Volltextsuche (DSEARCH_PLUS_D3)

Standardwert: Nein

Da die Suche über Wörterbuch-Tabellen nicht mehr unterstützt wird, ist es hier nicht sinnvoll einen anderen Wert als den Standardwert einzutragen.

In diesem Fall erfolgt die Suche nur über d.3 search.

d.search Vorfilterung für Attributfelder (DSEARCH_DIRECT_DOCFIELD_SEARCH)

Standardwert: Ja

Aktiviert die direkte d.3 search-Suche für die Dokument-Attributfelder.

Bei kombinierter Volltext- und Attributfeld-Suche verringert dies die Treffermenge von d.3 search und beschleunigt somit die Nachfilterung durch den d.3 Server.

d.search Standard-Corpus (DSEARCH_DEFAULT_CORPUS)

Standardwert: -

Wird verwendet, wenn kein Corpus für die Dokumentart explizit angegeben ist; bei Leerstring wird dieser Wert vom d.3 gateway Server-Kürzel abgeleitet (d3[d3fc_server_id]).

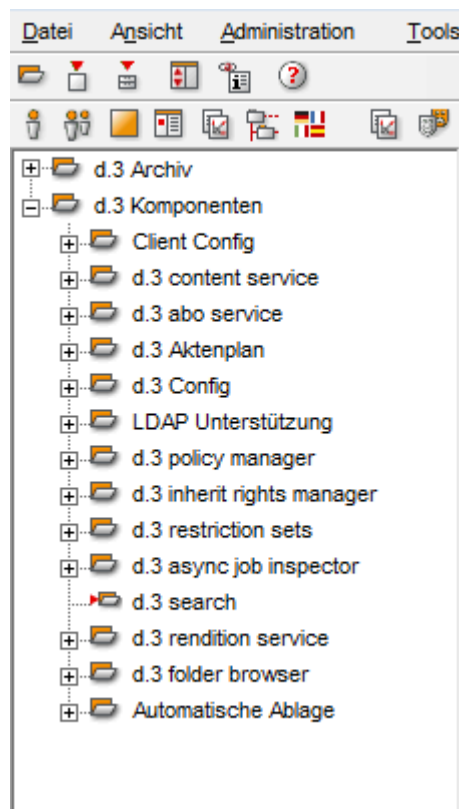
Weitere Informationen finden Sie in der Dokumentation zur d.3 Administration.

Weitere [Einstellungen für die d.3 search Unterstützung](#) finden Sie in dieser Dokumentation.

Konfiguration in der d.3 Administration unter d.3 search

Wenn die d.3 search [Unterstützung für ein d.3 Archiv](#) aktiviert wurde (Parameter DSEARCH_SUPPORT im **d.3 Config** Modul), haben Sie über das **d.3 search** Modul der d.3 Administration die Möglichkeit, folgende Einstellungen aufzurufen:

- den d.3 search server Status
- die Dokumentartenzuordnung
- die Eigenschaften der Dokumentsammlungen
- Rufen Sie in der d.3 Administration unter **d.3 Komponenten** den Punkt **d.3 search** per Doppelklick auf, um zur d.3 search Administration zu gelangen.



Der Dialog der d.3 search Administration ist gegliedert in die Registerkarten [Serverstatus](#), [Dokumentarten](#) und [Dokumentsammlungen](#).

Registerkarte Serverstatus

Auf der Registerkarte **Serverstatus** sehen Sie die aktuellen Eigenschaften des d.3 search servers. Diese Ansicht wird im Sekundentakt automatisch aktualisiert.

Da jede Aktualisierung eine Anfrage aus Sicht des d.3 search Servers darstellt, erhöht sich die Zahl der verarbeiteten Anfragen ebenfalls im Sekundentakt um eins.



Die Zeile **d.search Server** gibt Ihnen an, welcher d.3 search server das d.3 System zur Ablage der Volltextinformation verarbeitet.

Die Zahl in **Aktuelle Verbindungen** gibt an, wie viele Anfragen zu diesem Zeitpunkt vom d.3 search verarbeitet werden (Aktualisierungs- oder Suchanfragen).

Die Zahl in **Verarbeitete Anfragen** wird seit Start des d.3 search server Prozesses berechnet.

Die Information zum **letzten Server Start** bzw. der Laufzeit beschreibt, wie lange der d.3 search server Prozess aktiv war.

Mit der Angabe der verarbeiteten Anfragen und der Laufzeit lässt sich die Auslastung des Systems ablesen.

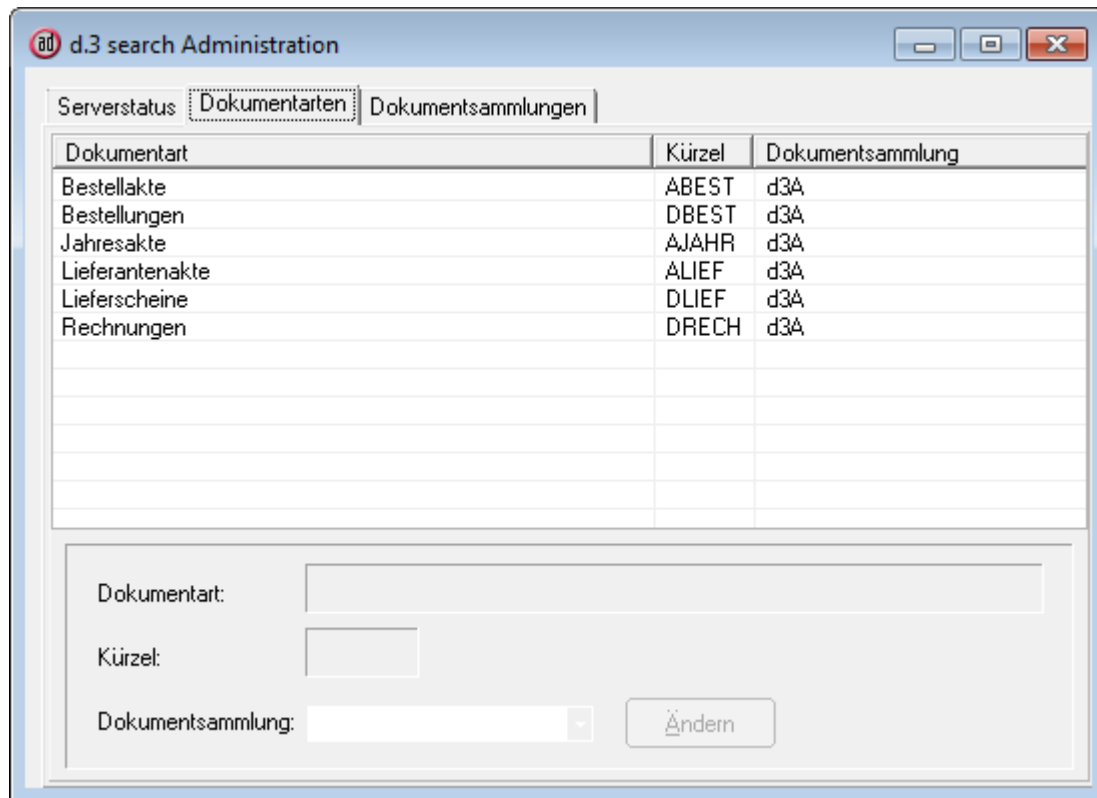
Die Punkte **Offene Einfüge-Operationen**, **Offene Entfernen-Operationen** und **Offene Aktualisierungen** geben an, wie viele Änderungen an den Dokumentsammlungen (Neues Dokument einfügen, Dokument ändern bzw. Dokument löschen) noch nicht durchgeführt wurden. Größere Werte für diese Anzahlen signalisieren, dass entweder eine Suchmaschine nicht erreichbar ist oder d.3 search bei einem Massendatenimport bei der Abarbeitung der Aufträge nicht mit dem d.3 hostimp mithält.

Registerkarte Dokumentarten

Hier können Sie die Zuordnung zwischen Dokumenten einer Dokumentart und der d.3 search Dokumentsammlung verändern.

In der Übersicht sehen Sie zu allen angelegten Dokumentarten den "Langtext", sowie das Dokumentarkürzel und den Namen der zugeordneten Dokumentsammlung.

Die Standarddokumentsammlung, die der d.3 server für neue Dokumentarten vorgibt, ermittelt sich aus dem Präfix "d3" und dem Archivkürzel als Suffix. D. h. für ein Archiv "DE" würde als Standarddokumentsammlung "d3DE" angesprochen.



- Wählen Sie die Dokumentart aus, um die Zuordnung einer Dokumentart zu einer Dokumentsammlung zu ändern.

Im unteren Fensterbereich wird Ihnen die gewählte Dokumentart sowie die zugehörige Dokumentsammlung angezeigt.

- Bestimmen Sie nun für diese Dokumentart, ob sie nicht indexiert werden soll, bzw. welche Dokumentsammlung zur Aufnahme der Schlagwortinformation dieser Dokumentart dient.
- Klicken Sie auf **Ändern**, um die Änderungen an den d.3 server zu senden.

Anmerkung

Diese Änderung bezieht sich auf alle nachfolgend importierten oder aktualisierten Dokumente. Für die bisherigen Dokumente dieser Dokumentart sollte eine Migration der Volltextdaten von der alten Dokumentsammlung in die neue über ein d.3 Datenbankskript angestrebt werden.

Nutzung verschiedener Dokumentsammlungen

Prinzipiell ist es ausreichend eine Dokumentsammlung pro d.3 Archiv anzulegen und alle Schlagworte in diese ablegen zu lassen. Es gibt aber einige Umstände für die es sinnvoll ist verschiedene Dokumentsammlung zu erzeugen. Grundsätzlich sollte man zu Beginn der Konfiguration darüber Gedanken machen, ob ein oder mehrere Dokumentsammlungen erzeugt werden sollen. Wie viele Dokumente werden zukünftig erwartet? Eignen sich die Dokumente für eine Volltextsuche?

Gründe für Einsatz von mehreren Dokumentsammlungen:

- Die Dokumentsammlung wird zu groß
Bei einer Anzahl von 50 Millionen Dokumenten und kleiner Datei-Größe oder 30 Millionen und kleiner Datei-Größen der Dokumente empfiehlt es sich eine Trennung in einzelne Dokumentsammlungen vorzunehmen.

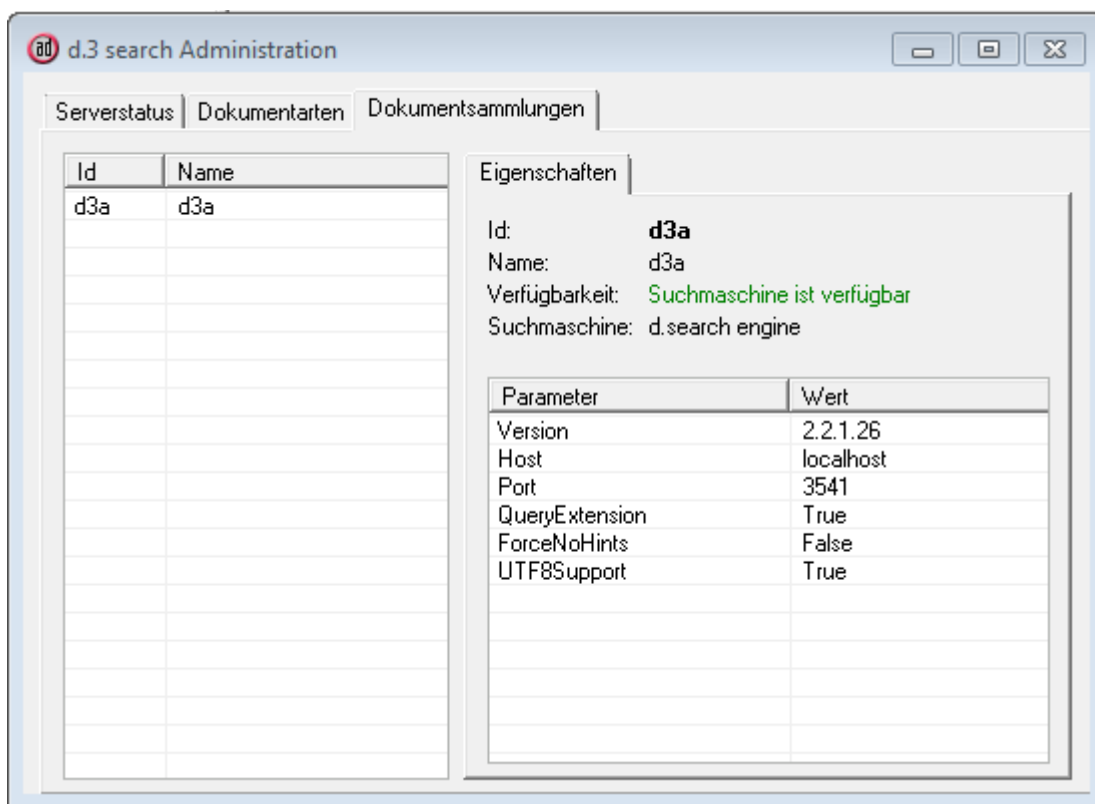
- Einsatz von Wortwolken
 - Für manche Dokumentarten sollen keine Wortwolken genutzt werden, da sie Inhalte enthalten, die nicht für jeden gedacht sind. Es sollen dem Benutzer somit auch keine Vorschläge im Client dazu angezeigt werden. Zu allen anderen Dokumenten wiederum sollen schon Vorschläge gebildet werden.
 - Für manche Dokumentarten macht es keinen Sinn eine semantische Unterstützung anzubieten, da sich aus dem Inhalt keine sinnvollen Wortwolken bilden lassen.

Registerkarte Dokumentsammlungen

Die Registerkarte **Dokumentsammlungen** der d.3 search Administration gibt Ihnen einen Überblick der auf dem d.3 search Server verfügbaren Dokumentsammlungen.

- Wählen Sie in der Liste eine Dokumentsammlung aus, werden Ihnen in der rechten Fensterhälfte weitere Eigenschaften dieser Dokumentsammlung angezeigt.

Die Anzeige enthält ebenfalls die Information, ob auf eine Dokumentsammlung aktiv zugegriffen werden kann oder nicht (Eigenschaft: **Verfügbarkeit**).



1.4. Technischer Hintergrund zum d.3 search Server

Hier finden Sie eine technische Beschreibung des d.3 search Servers, die Ihnen die Konfigurationsmöglichkeiten des d.3 Servers detailliert beschreibt.

1.4.1. Verzeichnisstruktur

Das Standardprogrammverzeichnis des d.3 search servers ist `..\d3\d.3 search`. Dort befinden sich einige Unterverzeichnisse für temporäre Arbeitsdateien oder Konfigurationsdateien.

Anmerkung

Hier eine mögliche Liste der Unterverzeichnisse:

```

corpora\
corpora\d3p\
corpora\d3p\idbackup
d3index
d3index\d.3 server\
d3index\d.3 server\ID\
d3index\d.3 server\ID\dtiff
d3index\d.3 server\ID\Filter\Break
d3index\d.3 server\ID\Filter\Finish
d3index\d.3 server\ID\Filter\Work
d3index\d.3 server\ID\Process
Database
engines\
shaft
shaft\64Bit
jobs\
jobs\corpusID\add
jobs\corpusID\update
jobs\corpusID\delete

```

Verzeichnis corpora

Sie finden in diesem Unterverzeichnis Konfigurations- und Indexdaten für einzelne Dokumentsammlungen.

Für jede Dokumentsammlung wird vom d.3 search server ein eigenes Unterverzeichnis angelegt. In diesem Unterverzeichnis der jeweiligen Dokumentensammlung befinden sich folgende Dateien und Ordner:

Datei / Ordner	Bedeutung
SHAFT1	In diesem Ordner ist bei Standardeinstellungen der Suchindex und die Konfiguration der d.3 search engine (SHAFT) zu finden. Anmerkung Bei einer Skalierung wird pro d.3 search engine (SHAFT) ein weiterer Ordner mit der Benennung SHAFT2, SHAFT3 usw. daneben liegen. Anmerkung Wenn die d.3 search engine (SHAFT) auf einem anderen Rechner ausgelagert ist, ist der Ordner ohne Inhalt.
corpus.ini	Die Konfigurationsdatei beschreibt die Dokumentensammlung. Weitere Informationen finden Sie hier: Erweiterte Konfiguration der Dokumentensammlung .
dsearch.ini	Die Konfigurationsdatei enthält Informationen, wie d.3 search mit der Dokumentensammlung umzugehen hat. Weitere Informationen finden Sie hier: Erweiterte Konfiguration der Dokumentensammlung .

Verzeichnis d3index

Alle Programmverzeichnisse und -dateien von der d.3 search Nachverschlagnwortung werden bei der Installation im Unterverzeichnis d3index innerhalb des d.3 search-Programmverzeichnisses abgelegt.

Verzeichnis engines

Diese Verzeichnis enthält die Kopplungsmodule zu den einzelnen Suchmaschinen. Seit der d.3 search Version 2.2 liegt die SHAFT-Anbindung (shaft.dll) unter diesem Verzeichnis.

Verzeichnis shaft

Enthält die Programmdateien der SHAFT Engine d.3 search engine in 32 Bit. Im Ordner 64Bit befindet sich die 64 Bit Version der Engine.

Verzeichnis jobs

Unterhalb des Job-Verzeichnisses werden die Änderungsaufträge für die Dokumentsammlungen zwischengespeichert. D. h. der d.3 search server nimmt z. B. beim Neuimport eines Dokumentes die Textdaten vom d.3 Server zu einem Dokument entgegen und quittiert den Eingang. Bis die Textdaten zur Suchmaschine übertragen wurden, verbleiben die Job-Daten in dem jobs-Verzeichnis.

1.4.2. Die d.3 search engine

Das Produkt d.3 search wird zusammen mit der Volltext-Suchmaschine d.3 search engine ausgeliefert. Diese Suchmaschine arbeitet intern mit der sogenannten SHAFT-Technologie aus dem Hause d.velop, daher wird auch oft von der SHAFT-Suchmaschine gesprochen.

Die Suchmaschine besteht aus den Dateien SHAFT . EXE und SHAFTUTL . EXE.

Hinweise zum Update der d.3 search engine

Warnung

Wird von einer d.3 search Version kleiner 1.2.4 aktualisiert, ist eine Re-Indexierung (sogenanntes „Rebuild“) der Datenbank zwingend erforderlich.

Wird dieses nicht durchgeführt, wird die Datenbank bei einem Testlauf als "defekt" deklariert und die Suchmaschine kann nicht mehr gestartet werden.

Die Re-Indexierung kann allerdings dann immer noch durchgeführt werden.

Anmerkung

Die SHAFT Engine ab Version 1.4 liegt in 32 Bit und 64 Bit vor. Wird die Engine wie gewohnt aus dem Shaft Ordner gestartet, wird die 32 Bit Version ausgeführt. Um auf die 64 Bit Version umzustellen, muss lediglich der Pfad auf . . \SHAFT\64Bit angepasst werden.

Der Index muss dazu nicht geändert oder angepasst werden. Es kann jederzeit zwischen 32 und 64 Bit gewechselt werden.

Hinweise zur Dokumentensammlung mit der d.3 search engine

Um eine Suchmaschine nutzen zu können, muss im Kontext von d.3 search eine sogenannte Dokumentensammlung (oft wird auch der Begriff „Corpus“ synonym benutzt) angelegt werden und eine Instanz (also ein laufender Prozess der SHAFT . EXE) vorhanden und über das TCP/IP-Protokoll ansprechbar sein.

Dies alles wird mit dem d.3 search Assistenten zur [Anlage einer Dokumentsammlung](#) erledigt.

Der Benutzer hat bei der Benutzung des Assistenten die Wahl, welche Suchmaschine die neue Dokumentsammlung verwalten soll.

Im Fall der d.3 search engine kann danach gewählt werden, ob eine bereits bestehende Instanz der Suchmaschine genutzt oder ob eigens für die neue Dokumentensammlung auch eine neue Suchmaschinen-Instanz erzeugt werden soll.

Anmerkung

Soll Letzteres geschehen, ist es empfehlenswert darauf zu achten, die Datenbank zur Ablage der Indexdaten der Suchmaschinen-Instanz auf einem lokalen oder äquivalenten Laufwerk anzulegen.

Warnung

Beachten Sie, dass auf einem PC mehrere Instanzen der Suchmaschine laufen können, die jede ihre eigene Volltext-Datenbank verwalten. Da das TCP/IP Protokoll für die Kommunikation zwischen d.3 search und der d.3 search engine genutzt wird, ist lediglich darauf zu achten, dass für jede Suchmaschinen-Instanz ein anderer Port benutzt wird.

Warnung

Die Suchmaschinen-Datenbank sollte niemals auf einem NAS-Dateisystem oder ähnlichem (wie zum Beispiel bei NetApps) erstellt werden, da dies die Verfügbarkeit der Datenbank einschränken und damit zu Inkonsistenzen der Daten führen kann, die wiederum schließlich zu einer korrupten Datenbank führen können.

Zur Erklärung: Datenbanken von Suchmaschinen müssen meist eine hohe Performanz für das Antwortzeitverhalten aufweisen und sollten daher grundsätzlich hochverfügbar sein und keinen externen Caching-Mechanismen unterliegen. Diese können zum einen die Performanz, zum anderen unter Umständen auch die Funktionsfähigkeit erheblich beeinträchtigen.

Performance

Datenbanken von Suchmaschinen müssen eine hohe Performanz für das Antwortzeitverhalten aufweisen und sollten daher grundsätzlich hochverfügbar sein und keinen externen Caching-Mechanismen unterliegen. Diese können zum einen die Performanz, zum anderen unter Umständen auch die Funktionsfähigkeit erheblich beeinträchtigen.

Benötigt die d.3 search engine für die Ausführung einer Operation (Insert, Delete, Abstract, Search, Spell Check) einen längeren Zeitraum als der definierte Grenzwert (siehe [Konfiguration der d.3 search engine](#)), wird automatisch eine Log-Datei in das Verzeichnis ...\\d3\\d.3 search\\corpora\\[CorpusId]\\PerfLogs geschrieben. In der Log-Datei sind Details zu diesem einen Call hinterlegt.

Warnung

Da die Log-Dateien nicht automatisch gelöscht werden, muss das Verzeichnis überwacht werden.

```

2012-12-11-10-25-05-879-5084.txt - Editor
Datei Bearbeiten Format Ansicht ?
Global informations:
Cache Hit:64926086 Miss:4381026 Ratio:93%
Read access: Time:5286,437676s Bytes:21011829355

This operation:
Search+Hints Start:0 Maxhits:1000 Expression: searchterm 7452581818666
Run time:00:00:07.172
Cache Hit:1215784 Miss:15283 Ratio:98%
Disk access: Time:4,779042s Bytes:62592194

Read operations:
F:\d.3 search corpora\d3D\link.dat, Pos:867172, Bytes:4032, Sec:0,000006
F:\d.3 search corpora\d3D\link.dat, Pos:3959693, Bytes:4032, Sec:0,000003
F:\d.3 search corpora\d3D\link.dat, Pos:996345, Bytes:4032, Sec:0,000002
F:\d.3 search corpora\d3D\link.dat, Pos:4009740, Bytes:4032, Sec:0,000002
F:\d.3 search corpora\d3D\link.dat, Pos:4047642, Bytes:4032, Sec:0,000002
F:\d.3 search corpora\d3D\link.dat, Pos:4057741, Bytes:4032, Sec:0,000002
F:\d.3 search corpora\d3D\link.dat, Pos:4110393, Bytes:4032, Sec:0,000003
F:\d.3 search corpora\d3D\Sound.dat, Pos:199347, Bytes:4096, Sec:0,000007
F:\d.3 search corpora\d3D\Sound.dat, Pos:199049, Bytes:4096, Sec:0,000003
F:\d.3 search corpora\d3D\Sound.dat, Pos:199018, Bytes:4096, Sec:0,000003
F:\d.3 search corpora\d3D\Sound.dat, Pos:198983, Bytes:4096, Sec:0,000004
F:\d.3 search corpora\d3D\Sound.dat, Pos:198961, Bytes:4096, Sec:0,000003
F:\d.3 search corpora\d3D\Sound.dat, Pos:187296, Bytes:4096, Sec:0,000003
F:\d.3 search corpora\d3D\words.dat, Pos:27996009, Bytes:4095, Sec:0,000004
F:\d.3 search corpora\d3D\Sound.dat, Pos:187254, Bytes:4096, Sec:0,000004
F:\d.3 search corpora\d3D\words.dat, Pos:27990229, Bytes:4095, Sec:0,000003
F:\d.3 search corpora\d3D\words.dat, Pos:27990797, Bytes:4095, Sec:0,000003
F:\d.3 search corpora\d3D\words.dat, Pos:27991804, Bytes:4095, Sec:0,000003
F:\d.3 search corpora\d3D\words.dat, Pos:27992079, Bytes:4095, Sec:0,000002
F:\d.3 search corpora\d3D\Sound.dat, Pos:187207, Bytes:4096, Sec:0,000003
F:\d.3 search corpora\d3D\words.dat, Pos:27980520, Bytes:4095, Sec:0,000004
F:\d.3 search corpora\d3D\words.dat, Pos:27981435, Bytes:4095, Sec:0,000002
F:\d.3 search corpora\d3D\words.dat, Pos:27981671, Bytes:4095, Sec:0,000003
F:\d.3 search corpora\d3D\words.dat, Pos:27982023, Bytes:4095, Sec:0,000002
F:\d.3 search corpora\d3D\words.dat, Pos:27982887, Bytes:4095, Sec:0,000003
F:\d.3 search corpora\d3D\words.dat, Pos:27983348, Bytes:4095, Sec:0,000003

```

Beispiel einer Log-Datei für einen Such-Call

Die Log-Datei ist in drei Bereiche aufgeteilt:

- **Global Information:** Die Informationen in dieser Sektion beziehen sich auf den Zeitraum letzter Start der d.3 search engine bis zum Ausführen der jeweiligen Operation im Log.

Zeile Cache	
Hit	gibt an, wie viele Daten-Zugriffe aus dem Cache bedient werden konnten
Miss	gibt an, wie viele Daten-Zugriffe nicht vom Cache bedient werden konnten, sondern über einen lesenden Zugriff auf der Festplatte ermittelt wurden
Ratio	gibt an, wie viel Prozent der Anfragen aus dem Cache bedient werden konnten
Zeile Read access	
Time	gibt an, wie viel Zeit mit dem Lesen von Daten von der Festplatte insgesamt verbraucht wurde
Bytes	gibt an, wie viele Daten bislang von der Festplatte gelesen wurden

- **This Operation:** Die Informationen beziehen sich auf die protokollierte Operation.

Zeile Search Hints	
Start	gibt an, ab welcher Position in den Treffern gelesen wird
Maxhits	gibt an, wie viele Treffer ab der Position maximal gelesen werden sollen
Expression	Suchterm, der vom Benutzer in der Oberfläche eingegeben wurde

<u>Zeile Run time</u>	gibt die benötigte Zeit an, um das Ergebnis zu ermitteln
<u>Zeile Cache</u>	
Hit	gibt an, wie viele Lesezugriffe auf den Cache erfolgen
Miss	gibt an, wie viele Lesezugriffe nicht vom Cache bedient werden konnten, sondern über einen lesenden Zugriff auf der Festplatte ermittelt wurden
Ratio	gibt an, wie viel Prozent der Anfragen aus dem Cache bedient werden konnten
<u>Zeile disk access</u>	
Time	gibt die benötigte Zeit für das Lesen von der Festplatte an
Bytes	gibt die Größe der gelesenen Daten an

- **Read operations:** In diesem Bereich werden Details zum Lesezugriff auf die jeweilige Datei protokolliert.

Ist der Wert Ratio aus den globalen Informationen kleiner 85%, kann dies ein Hinweis dafür sein, dass der d.3 search engine mehr RAM für den Cache zugewiesen werden sollte. Ist der Ratio Wert größer 85%, ist die Ursache für eine lange Dauer der Operation eher bei den Zugriff-Zeiten auf die Festplatte zu suchen.

Für detaillierte Analysen der Log-Dateien wenden Sie sich bitte an Ihren d.velop Partner oder an den d.velop Support.

Die Suchmaschinen-Datenbank

Zur Datenbank einer SHAFT-Instanz gehören die folgenden grundlegenden Dateien:

```

DATABASE.DAT
FILES.DAT
FILES.IDX
FILES.WRD
LINK.DAT
SOUND.DAT
SOUND.IDX
WORDS.DAT
WORDS.IDX
SHAFT.INI

```

Hinzu kommt je nach Konfiguration noch die Datei JOURNAL.DAT.

Für ein vollständiges Backup werden die oben genannten Dateien bis auf die Dateien SHAFT.INI und JOURNAL.DAT benötigt.

Wenn das automatische tägliche Backup der SHAFT- Suchmaschine eingeschaltet wurde, werden täglich zu einer bestimmten Uhrzeit genau die notwendigen Dateien in dem konfigurierbaren Ordner abgelegt.

Start der d.3 search engine

Die Instanzen der d.3 search engine laufen gewöhnlich in Form eines Prozesses der SHAFT.EXE im d.3 process manager.

Ihre Bezeichnung lautet im Allgemeinen **d.3 search – engine (d3?)**.

Gestartet wird ein solcher Prozess immer unter Angabe einer Konfigurationsdatei mit dem Namen SHAFT.INI (siehe Aufruf im d.3 process manager).

Anmerkung

```
shaft c:\d3\d.3 search\corpora\d3x\shaft.ini
```


Mit dieser Datei werden sowohl Konfigurations-Parameter für den SHAFT-Prozess, als auch für die Verwaltung seiner Datenbank festgelegt.

Sollte der angegebene Port bereits belegt oder die Datenbank als defekt markiert sein, startet die Suchmaschine dennoch, die Datenbank kann allerdings nicht verwendet werden.

Beenden Sie in diesem Fall die Suchmaschine und überprüfen Sie das Transaktionsprotokoll (`trans.log`) im Transaktionsprotokoll-Verzeichnis `TLOG`.

Insbesondere während des laufenden Betriebes ist zu beachten, dass der/die Prozess(e) der Suchmaschine immer ordnungsgemäß beendet („heruntergefahren“) werden, damit auch diesbezüglich Probleme mit den zugehörigen Datenbanken vermieden werden können. Dies ist unter anderen auch bei Nutzung externer Backup-Mechanismen von essentieller Bedeutung.

Sicherungs-Mechanismen

Um sicherzustellen, dass die Datenbank der Suchmaschine nicht durch Software- oder Hardwarefehler zerstört wird, enthält sie diverse Mechanismen, um Fehler zu erkennen und zu beheben.

Die Behebung der Fehler ist teilweise automatisch möglich, bei schwerwiegenden Problemen wird die Datenbank allerdings als "defekt" markiert und es muss mit Hilfe des `SHAFTUTL` manuell eingegriffen werden.

Online-Sicherung

Nach Möglichkeit sollte die Datenbank nicht gesichert werden, während die Suchmaschine läuft, da Änderungen an der Datenbank dazu führen können, dass eine inkonsistente Datenbank gesichert wird.

Daher sollte die Suchmaschine entweder zur Sicherung angehalten oder die Online Backup Funktion verwendet werden.

Die Online Backup Funktion wird aktiviert, indem der Zielpfad und der Zeitpunkt des Backups in der Datenbank Start-Konfiguration angegeben wird.

Dies bewirkt, dass zum angegebenen Zeitpunkt die Suchmaschine auf "nur Lesen" umschaltet und die Datenbankstruktur auf Konsistenz prüft. Sofern diese Prüfung fehlerfrei verläuft, wird die Datenbank in das angegebene Verzeichnis kopiert. Ist die Sicherung erfolgreich, wird das Archivprotokoll und Transaktionsprotokoll anschließend gelöscht.

Somit enthalten beide Logs nur Informationen, die ab der Sicherung relevant sind. Backupsoftware kann somit nach diesem Backup die Datenbank aus dem angegebenen Zielverzeichnis jederzeit sichern. Zusätzlich ist es ratsam, die Archivprotokolle direkt nach der Erstellung mit einer Backupsoftware zu sichern, sofern diese ein solches Feature anbietet.

Journal

Dieses ist der Grundmechanismus, welcher über den Konfigurations-Parameter `Journal` aktiviert wird.

Bei dieser Funktion werden alle Änderungen, die in die Datenbank geschrieben werden sollen, zuvor in die `journal.dat` geschrieben. Erst wenn diese Datei vollständig geschrieben wurde, werden die Änderungen in die Datenbank geschrieben.

Sollte die Hardware ausfallen bzw. das Programm terminiert werden, merkt es beim Start, dass es beim Schreiben in die Datenbank unterbrochen wurde und wiederholt diese Schreibzugriffe. So wird verhindert, dass ein Stromausfall oder ein Bluescreen, etc. die Datenbank zerstören.

Transaktions-Protokoll

Beim Transaktions-Protokoll handelt es sich um eine Textdatei `trans.log`, die je nach `LogLevel` mehr oder weniger Informationen zu Transaktionen enthält, die auf der Datenbank ausgeführt wurden.

Die Datei befindet sich im Unterverzeichnis `TLOG` des Verzeichnisses einer Dokumentsammlung.

Bei einem LogLevel von "9" werden zusätzlich die Inhalte der Dateien die indiziert wurden, in einer Unterverzeichnisstruktur abgelegt. Mit Hilfe der SHAFTUTL können diese Transaktionen zu einem späteren Zeitpunkt noch mal durchgeführt werden.

Wurde ein Fehler beim Schreiben nicht sofort erkannt und enthalten daher auch die Archiv-Protokolle Fehler, können alle logischen Änderungsoperationen auf der letzten Sicherung der Datenbank wiederholt werden.

Liegt keine Sicherung vor und wurde das Transaktionsprotokoll nie gelöscht oder abgeschaltet, kann die Datenbank auch vollständig mit Hilfe dieses Logs wieder hergestellt werden.

Sofern Archivprotokolle und Transaktionsprotokolle vorliegen, sollten die Transaktionsprotokolle eingespielt werden, da diese sicherer zum Ziel führen. Sind allerdings sehr viele Transaktionen wiederherzustellen, sollten zunächst die Archivprotokolle verwendet werden, da diese schneller verarbeitet werden können.

Archiv-Protokoll

Hierbei werden die Journale nicht überschrieben, sondern in einem Verzeichnis gesichert.

Somit enthält dieses Verzeichnis ein vollständiges Änderungsprotokoll der Datenbank.

Mit Hilfe einer Sicherung und aller danach erstellten Protokoll- Dateien kann daher der aktuelle Zustand der Datenbank wieder hergestellt werden.

Um Protokoll-Dateien einer Sicherung zuzuordnen, werden diese unter einem Dateinamen abgelegt, der den Zeitpunkt ihrer Sicherung enthält ({Datum}- {Uhrzeit}-{Millisekunden}).

Die Protokolle werden mit dem SHAFTUTL wieder eingespielt. Das Programm spielt diese Dateien dabei automatisch in der richtigen Reihenfolge ein. Liegt keine Sicherung der Datenbank vor, kann sie auch vollständig wieder aufgebaut werden, sofern noch alle Archivprotokolle vorliegen.

Wartung mit SHAFTUTL.EXE

Zusammen mit der Suchmaschine wird die Datei SHAFTUTL . EXE ausgeliefert.

Hierbei handelt es sich um ein Werkzeug, mit dem die Datenbank der Suchmaschine sowohl untersucht als auch repariert bzw. wiederhergestellt werden kann.

Dieses Werkzeug ist von der Kommandozeile zu benutzen. Bei der Benutzung ist immer die Angabe einer Konfigurations-Datei (* . INI) als erster Kommandozeilen-Parameter nötig. Hierbei handelt es sich um die Datei SHAFT . INI der jeweiligen Suchmaschinen-Datenbank. Sie befindet sich immer im jeweiligen Corpus-Verzeichnis von d.3 search.

Anmerkung

```
SHAFTUTL {Pfad+Name der entsprechenden SHAFT.INI}
<ac:macro ac:name="beispiel-doku"
><ac:parameter ac:name="atlassian-macro-output-
type">BLOCK</ac:parameter><ac:rich-text-body><ac:macro
ac:name="code"><ac:plain-text-body><![CDATA[SHAFTUTL "C:\d3\d.3
search\corpora\d3p\shaft.ini"
```

Danach erfolgt die Angabe weiterer Parameter, die im Folgenden nun erläutert werden. Wird das Programm ohne Parameter gestartet, gibt es eine Übersicht über die Funktionen aus.

```
C:\d3\d.3 search\shaft\64Bit>shaftutl "c:\d3\d.3 search\corpora\d3p\SHAFT1\shaft.ini"
SHAFT tool 3.0.0 (372)
Usage: <DB-Start.cfg> <option> [optionparam]
Options
-c Check DB structure
-i Import archivelogs
  optionparam = Importpath
-is Import synonym file
  optionparam = filename
-r Rebuild
-ri Reindex <Storetext must be enabled!>
-rt Restore and redo transaction log
-d Defrag
-f file dump
-w Word dump
-ws Stopword dump by MaxFilesPerWord
-t redo transaction log
```

Updates für die d.3 search engine

Je nach Situation kann es vorkommen, dass die Suchmaschine d.3 search engine (SHAFT) aktualisiert werden muss.

Nach diesem Update kann es sein, dass Sie aufgefordert werden, den Index der Suchmaschinen-Datenbanken sämtlicher Dokumentsammlungen mit dem Programm ShaftUtl.EXE neu aufzubauen.

Dazu führen Sie den folgenden Befehl aus:

Anmerkung

```
SHAFTUTL {Pfad+Name der entsprechenden SHAFT.INI} -r
<ac:macro ac:name="beispiel-doku"
><ac:parameter ac:name="atlassian-macro-output-
type">BLOCK</ac:parameter><ac:rich-text-body><ac:macro
ac:name="code"><ac:plain-text-body><![CDATA[SHAFTUTL
..\corpora\d3p\shaft.ini -r
```

Anmerkung

```
SHAFTUTL "C:\d3\d.3 search\corpora\d3p\shaft.ini" -r
```

Sollte dieses nicht geschehen, kann die Datenbank nach einem Update als "defekt" erkannt werden und hierdurch die d.3 search-Jobverarbeitung stehenbleiben. Die Gefahr des Datenverlustes besteht jedoch nicht.

Allgemeine Parameter für das Programm ShaftUtl.EXE

Reindex	-ri
dies funktioniert nur, wenn der Text im Index gespeichert ist; in diesem Fall wird der Index vollständig neu aufgebaut.	
Restore	-rt
nimmt dem Administrator das Zurückkopieren mit anschließendem Transaktionseinspielen ab, d.h., die Datenbank wird aus dem Backup-Verzeichnis zurückgespielt und anschließend die protokollierten Transaktionen wiederholt. Es verringern sich die Fehlerquellen, so dass ggf. ein Backup zurückgespielt wird und anschließend versehentlich ein altes Journal dieses Backup wieder zerstört.	
Defrag	-d
Defragmentiert den Index, so dass Treffermengen schneller gelesen werden können	

Datenbank prüfen

Um die Konsistenz der Datenbank zu prüfen wird das SHAFTUTL mit dem Schalter -c gestartet.

Diese Prüfung ist bei laufender Suchmaschine möglich, allerdings kann es dann passieren, dass fälschlicherweise ein Fehler gefunden wird. Dies liegt daran, dass die Suchmaschine Änderungen nicht sofort in

die Datenbank schreibt und bei der Prüfung ggf. Daten geprüft wurden, die nach einer Änderung durch die Suchmaschine nicht mehr richtig erscheinen.

Um ein wirklich sicheres Ergebnis zu erhalten, sollte die Suchmaschine daher nicht laufen. Sofern ein Fehler in der Datenbank gefunden wird, wird diese als "defekt" markiert und die Suchmaschine kann erst wieder gestartet werden, nachdem eine wiederholte Prüfung keinen Fehler aufdeckt, also z.B. nach einer Reparatur.

Archiv-Protokolle wieder einspielen

Mit dem Schalter `-i` können die Archiv-Protokolle wieder eingespielt werden.

Dazu wird nach diesem Schalter der Pfad angegeben, in dem die Protokoll-Dateien gespeichert sind.

Anmerkung

```
shaftutl c:\d3\d.3 search\corpora\d3x\shaft.ini -i  
c:\backup\archlogs
```

Um Archiv-Protokolle einzuspielen, muss die Suchmaschine angehalten sein.

Sofern die Suchmaschine läuft, bricht SHAFTUTL mit einem Fehler ab. Sollte keine Sicherung der Datenbank vorhanden sein, ist es ratsam vor dem Einspielen von Archiv-Protokollen eine Sicherung zu erstellen.

Transaktions-Protokoll wiederholen

Wird SHAFTUTL mit dem Schalter `-t` gestartet, werden alle aufgezeichneten Einfüge- und Löschoperationen auf der Datenbank nochmal ausgeführt. Diese entspricht also einem „Neuimport“ aller im Transaktionsprotokoll aufgezeichneten Dateien.

Vor dem Einspielen eines Transaktionsprotokolls sollte eine Sicherung der Datenbank erstellt werden bzw. vorhanden sein, da die Journalfunktion beim Einspielen der Transaktionen deaktiviert ist.

Datenbank wiederherstellen

Warnung

Diesen Schritt sollten Sie nur ausführen, wenn es nach einem Update erforderlich ist ODER die obigen Verfahren nicht verwendet werden können.

Sollten keine Sicherungen zur Verfügung stehen oder wurde eine neue Suchmaschine eingespielt, die eine andere Datenstruktur verwendet, kann mit dem Schalter `-r` eine Re-Indexierung („Rebuild“) gestartet werden.

Dies bewirkt, dass die Struktur ggf. auf eine neue Version gehoben wird und alle Verknüpfungen neu geschrieben werden. Bevor ein Rebuild gestartet wird, sollte auch hier ein Backup der Datenbank erstellt werden bzw. vorhanden sein.

Sollte der Rebuild- Prozess abbrechen, kann er zwar in der Regel erneut gestartet und durchgeführt werden, dies kann aber nicht garantiert werden.

Alle Wörter der Datenbank ausgeben

Mit dem Schalter `-w` werden alle Wörter, die in der Datenbank gespeichert sind, auf der Konsole ausgegeben. Mit dem Umleitungssymbol `>` kann diese Ausgabe in eine Datei gelenkt werden.

Alle Wörter der Datenbank ausgeben, die zu Stoppwörtern geworden sind

Mit dem Schalter `-ws` werden alle Wörter, die in der Datenbank gespeichert sind und durch den Wert **MaxFilesPerWord** zu Stoppwörtern geworden sind, auf der Konsole ausgegeben. Mit dem Umleitungssymbol `>` kann diese Ausgabe in eine Datei gelenkt werden.

Alle Dateinamen der Datenbank ausgeben

Mit dem Schalter `-f` werden wie bei den Worten alle Dateinamen ausgegeben, die in der Datenbank gespeichert sind.

Re-Indizieren der Datenbank

Mit dem Schalter `-r i` wird eine Re-Indizierung der Datenbank ausgelöst.

Anmerkung

Dies funktioniert nur, wenn der Text im Index gespeichert ist. In diesem Fall wird der Index vollständig neu aufgebaut. Dies bedeutet: alle Texte werden extrahiert und vollständig neu in einen leeren Index eingefügt.

Wiederherstellen von Transaktionen

Mit dem Schalter `-r t` leiten Sie ein Zurückkopieren mit anschließendem Transaktionseinspielen ein.

Das heißt, die Datenbank wird aus dem Backupverzeichnis zurückgespielt und anschließend die protokollierten Transaktionen wiederholt.

So verringern sich die Fehlerquellen, so dass ggf. ein Backup zurückgespielt wird und anschließend versehentlich ein altes Journal dieses Backup wieder zerstört.

Defragmentieren der Datenbank

Starten Sie SHAFTUTL mit dem Parameter `-d`, leiten Sie eine Defragmentierung der Datenbank ein.

Auf diese Weise wird der Index restrukturiert, so dass Treffermengen schneller gelesen werden können.

Warnung

Bei der Defragmentierung wird eine zusätzliche Festplattenkapazität in der Größe der Dateien `link.dat` und `words.dat` benötigt. Die Festplattenkapazität muss auf der Festplatte zur Verfügung stehen, auf der der Index abgelegt ist.

1.5. Wartung

In diesem Kapitel erhalten Sie weitere Informationen zur Wartung.

1.5.1. Hinweise zur Optimierung von d.3 search

Es kann sinnvoll sein, d.3 search auf einem eigenen Server zu installieren, da das Befüllen der Datenbanken der Dokumentensammlungen ein sehr schreibintensiver Prozess ist.

Dieser Server sollte dann zwei physikalische Datenträger haben, dass die Schreibaktivitäten möglichst optimal verteilt werden können. Des Weiteren sollte beim Start der Engine-Prozesse die errechnete und protokollierte Plattengeschwindigkeit (Diskspeed ~61440 KB/s) beachtet werden. Diese sollte für eine produktiv genutzte Dokumentensammlung nicht unter 50.000 KB/s liegen.

Ferner ist eine schnelle Netzwerk-Anbindung erforderlich, um die Performance zu steigern. Weiterhin kann ein Mehrprozessor-Kern unterstützen.

Dabei sollten

- das Programmverzeichnis der Applikation (`. \d.3\d.3 search\`) und
- das Jobs-Verzeichniss des d.3 search Servers (`. \d.3\d.3 search\jobs\`) und
- das Konfigurationsverzeichnis der Dokumentensammlungen (`. \d.3\d.3 search\corpora\`) und
- das Transaktionslog-Verzeichnis (`. \d.3\d.3 search\{Name der Dokumentensammlung}\t log\`)

auf einer physikalischen Partition liegen.

Die Datenbank-Dateien der Dokumentensammlung (shaft.ini Parameter Database unterhalb der Sektion [Engine]) sollten mit dem ggf. aktivierten ARCHLOGS-Verzeichnis (shaft.ini Parameter ArchiveLogs=1 unterhalb der Sektion [Engine]) auf einer anderen physikalischen Partition abgelegt werden.

Das Backup-Verzeichnis der Dokumentensammlungen (shaft.ini Parameter Path unterhalb der Sektion [backup]) sollte auf einen langzeitstabilen Datenspeicher abgelegt und kann via UNC-Pfadangabe konfiguriert werden.

Die Zuweisung des Transaktionslog-Verzeichnisses und des Archlog-Verzeichnisses erfolgt für die Dokumentensammlungen in der shaft.ini unterhalb der Sektion [Engine] über die Parameter Log-path={Pfadangabe zum tlog-Ordner} und ArchivePath={Pfadangabe zum ARCHLOGS-Ordner}, die Sie zusätzlich hinzufügen müssen.

Die Zuweisung des Jobs-Verzeichnis für den d.3 search Server wie auch die Vorgabe des Konfigurationsverzeichnisses muss im jeweiligen Programmverzeichnis über die ebenfalls manuell zu erstellende Konfigurationsdatei dsearch.ini mit folgendem Inhalt erfolgen:

Anmerkung

```
[Settings]
JobsPath={Pfadangabe zum JOBS-Ordner}
CorporaPath={Pfadangabe zum Corpora-Ordner}
```

Anmerkung

Wenn Sie nur das Database-Verzeichnis und das entsprechende Archlogs-Verzeichnis der Dokumentensammlungen auf eine andere physikalische Partition verschieben, brauchen Sie lediglich nur die Anpassung in der jeweiligen shaft.ini vornehmen.

1.5.2. Hinweise zur Optimierung von d.3 search auf einem RAID5-System

Warnung

Wenn Sie den d.3 search Server auf einem RAID5-System betreiben wollen, beachten Sie folgendes:

- RAID5 ist von Natur aus eher langsam und
- bei falscher Konfiguration des RAID5-Systems leidet die Performance des d.3 search Servers.

Die Plattenanzahl bei RAID5-Systemen muss so gewählt werden, dass 1024 / 2048 bzw. 4096 Byte so verteilt werden, dass sie mit der NTFS-Clustergröße korrespondieren. Dies ist der Fall, wenn drei, fünf oder neun Platten zu einem RAID-5 mit einer Paritätsplatte geschaltet sind. Dazu müssen die Platten mit 512 Byte Blöcken arbeiten.

Anmerkung

Moderne Platten haben oft größere Sektoren als 512 Byte.

Anmerkung

Hier ein Beispiel, wie eine RAID5-Konfiguration NICHT aussehen sollte:

Es werden sechs Festplatten zu einen RAID5 zusammengefasst.

Demnach gibt es fünf Datenplatten und eine Paritätsplatte, welche die Prüfsummen zur Wiederherstellung im Fehlerfall speichert.

Nun bilden alle Sektoren über die sechs Festplatten hinweg eine Einheit.

Da jeder Sektor 512 Byte groß ist und fünf Platten dieser Einheit Daten aufnehmen, müssen immer $5 \times 512 = 2.560$ Byte zusammenhängend geschrieben, bzw. geändert werden, da für diese 2.560 Byte der Paritätsblock der sechsten Platte berechnet werden muss.

NTFS verwendet allerdings immer Clustergrößen von 4, 8, 16, 32 oder 64 KB (in der Regel 4 KB).

Bei der Verwendung einer 4 KB-Clustergröße und einem 2.560 Byte Block bleiben bei jedem Schreibzugriff ~ 1.536 Byte über, die in einem 2.560 Byte Block geschrieben werden: $4 \text{ KB} \times 1.024 = 4.096$, $2.560 - 4.096 = -1.536$.

Daher muss der RAID Controller vor dem Schreiben zunächst den Block lesen, die 1.536 Byte einfügen, die neue Parität berechnen und geändert zurückschreiben. Bei wahlfreien Schreibzugriffen, z.B. von Microsoft SQL Server, SHAFT und kleinen Dateien führt dieses Verhalten dazu, dass die Schreibgeschwindigkeit bis auf 25% der möglichen Schreibgeschwindigkeit sinkt.

Lösung: Richten Sie einen RAID5-Controller mit fünf Platten ein. In diesem Fall wäre der Block $4 \times 512 = 2.048$ Byte groß. Die Leseoperation des Controller entfällt, da ein 4 KB Cluster genau zwei Blöcke ergibt.

1.5.3. Hinweise zur Wiederherstellung einer Dokumentensammlung

Es gibt verschiedene Möglichkeiten zur Wiederherstellung einer Dokumentensammlung bzw. eines Index.

Diese sind abhängig von den zuvor gewählten Backup-Strategie (online oder offline) und welche Log-Typen (ArchivLogs, TLogs, Journal.dat) aktiviert wurden. Des Weiteren sind sie abhängig vom gewählten LogLevel.

Szenario 1	
Online-Backup	ja
ArchivLogs	ja
Tlogs	ja
Journal.dat	ja
ToDo	• Engine herunterfahren • Online-Backup ins Corpus-Verzeichnis kopieren (vorhandene Dateien überschreiben) • Journal.dat liegt ebenfalls im Corpus Verzeichnis • Archivlogs zurückspielen (shaftutl.exe ".\shaft.ini" - i ".\archlogs") • Engine starten

Szenario 1	
Alternative	• Engine herunterfahren • Journal.dat aus dem Corpus-Verzeichnis löschen, falls noch vorhanden • Sicherstellen, dass sich im Ordner tlog mehrere Ordner mit OCR-Informationen befinden (LogLevel >= 9) Methode 1: • Online-Backup in das Corpus-Verzeichnis kopieren (vorhandene Dateien überschreiben) • TransaktionsLogs zurückspielen (shaftutl.exe ".\shaft.ini" -t) Methode 2: • TransaktionsLogs zurückspielen (shaftutl.exe ".\shaft.ini" -rt) => Datenbankdateien werden automatisiert überschrieben • Engine starten
Unterschied der Alternative	• Fehler in den Archive-Protokollen werden ebenfalls wiederhergestellt • Die Wiederherstellung durch die Archiv-Protokolle geht schneller • OCR-Informationen müssen bei dem Wiederherstellen durch die Transaktionslogs doppelt gespeichert werden (LogLevel >= 9)

Szenario 2	
Online-Backup	ja
Archivlogs	ja
Tlogs	ja
Journal.dat	nein
ToDo	• Engine herunterfahren • Online-Backup in das Corpus-Verzeichnis kopieren (vorhandene Dateien überschreiben) • Archivlogs zurückspielen (shaftutl.exe ".\shaft.ini" -i ".\archlogs") • Engine starten • Dokumente ab dem letzten ArchLog nachverschlagnworten
Alternative	• Engine herunterfahren • Journal.dat aus dem Corpus-Verzeichnis löschen, falls noch vorhanden • Sicherstellen, dass sich im Ordner tlog mehrere Ordner mit OCR-Informationen befinden (LogLevel >= 9) Methode1: • Online-Backup in das Corpus-Verzeichnis kopieren (vorhandene Dateien überschreiben) • TransaktionsLogs zurückspielen (shaftutl.exe ".\shaft.ini" -t)
Unterschied der Alternative	• Fehler in den Archive-Protokollen werden ebenfalls wiederhergestellt • Das Wiederherstellen durch die Archive-Protokolle geht schneller • OCR-Informationen müssen bei dem Wiederherstellen durch die Transaktionslogs doppelt gespeichert werden (LogLevel >= 9)

Szenario 3	
Online-Backup	ja
Archivlogs	ja
Tlogs	nein
Journal.dat	ja
ToDo	• Engine herunterfahren • Online-Backup in das Corpus-Verzeichnis kopieren (vorhandene Dateien überschreiben) • Journal.dat liegt ebenfalls im Corpus-Verzeichnis • Archivlogs zurückspielen (shaftutl.exe ".\shaft.ini" -i ".\archlogs") • Engine starten

Szenario 4	
Online-Backup	ja
Archivlogs	ja
Tlogs	nein
Journal.dat	nein

Szenario 4	
ToDo	• Engine herunterfahren • Online-Backup in das Corpus-Verzeichnis kopieren (vorhandene Dateien überschreiben) • Archivlogs zurückspielen (shaftutl.exe ".\shaft.ini" - i".\archlogs") • Engine starten • Dokumente ab dem letzten ArchLog nachverschlagnworten
Alternative	Nachverschlagnwortung wurde nicht getestet

Szenario 5	
Online-Backup	ja
Archivlogs	nein
Tlogs	ja
Journal.dat	ja
ToDo	• Engine herunterladen • Journal.dat aus dem Corpus-Verzeichnis löschen, falls noch vorhanden • Sicherstellen, dass sich im Ordner tlog mehrere Ordner mit OCR-Informationen befinden (LogLevel>= 9). Methode1: • Online-Backup in das Corpus-Verzeichnis kopieren (vorhandene Dateien überschreiben) • TransaktionsLogs zurückspielen (shaftutl.exe ".\shaft.ini" -t) Methode2: • TransaktionsLogs zurückspielen (shaftutl.exe ".\shaft.ini" -rt) => Datenbankdateien werden automatisiert überschrieben • Engine starten
Alternative	anschließend mehr Dokumente als zuvor

Szenario 6	
Online-Backup	ja
Archivlogs	nein
Tlogs	ja
Journal.dat	nein
ToDo	• Engine herunterfahren • Journal.dat aus dem Corpus-Verzeichnis löschen, falls noch vorhanden • Sicherstellen, dass im Ordner tlog mehrere Ordner mit OCR-Informationen befinden (LogLevel>= 9) Methode1: • Online-Backup in das Corpus-Verzeichnis kopieren (vorhandene Dateien überschreiben) • TransaktionsLogs zurückspielen (shaftutl.exe ".\shaft.ini" -t) Methode2: • TransaktionsLogs zurückspielen (shaftutl.exe ".\shaft.ini" -rt) => Datenbankdateien werden automatisiert überschrieben • Engine starten
Alternative	anschließend mehr Dokumente als zuvor

Szenario 7	
Online-Backup	ja
Archivlogs	nein
Tlogs	nein
Journal.dat	ja
ToDo	• Engine herunterfahren • Online-Backup in das Corpus-Verzeichnis kopieren (vorhandene Dateien überschreiben) • Konsistenz des Online-Backups überprüfen (shaftutl.exe ".\shaft.ini" -c) • Engine starten • Dokumente ab dem letzten Online-Backup nachverschlagnworten

Szenario 8	
Online-Backup	ja
Archivlogs	nein
Tlogs	nein
Journal.dat	nein
ToDo	• Engine herunterfahren • Online-Backup in das Corpus-Verzeichnis kopieren (vorhandene Dateien überschreiben) • Konsistenz des Online-Backups überprüfen (shaftutl.exe ".\shaft.ini" -c) • Engine starten • Dokumente ab dem letzten Online-Backup nachverschlagnworten

Szenario 9	
Online-Backup	nein
Archivlogs	X
Tlogs	X
Journal.dat	X
ToDo	• Engine starten • Alle Dokumente nachverschlagnworten

1.5.4. Reindexierung einer Dokumentensammlung

Im Laufe der Verwendung einer Dokumentensammlung kann es dazu kommen, dass Einstellungen angepasst werden, welche den Aufbau der Dokumentensammlung beeinflussen. Um die Änderungen der Einstellungen für die gesamte, bereits bestehende Dokumentensammlung zu übernehmen muss der Suchindex, die Datenbank, reindexiert werden.

Gründe für das Anpassen der Einstellungen könnten sein:

- der Delimitermodus wird zur Internationalisierung angepasst
- der Katalog ValidWords wird firmenspezifisch angepasst
- Storetext soll aktiviert werden

Anmerkung

Bei den jeweiligen Schaltern wird darauf hingewiesen, dass eine Reindexierung durchgeführt werden muss.

Im Folgenden werden die beiden Möglichkeiten einer Reindexierung beschrieben.

Assistent zur Reindizierung einer Dokumentensammlung

Der Assistent zur Reindizierung einer Dokumentensammlung bietet Ihnen zwei Funktionen:

- **Paralleler Neuaufbau (PN)** einer Dokumentensammlung in eine andere Dokumentensammlung. Diese Funktion steht Ihnen nur zu Verfügung, wenn Sie eine Spiegelung bei der Anlage einer Dokumentensammlung angegeben haben.
- **Nachverschlagnworten (N)** einer Dokumentensammlung. Um fehlende Dokumente vom d.3 System abzufragen, können Sie erneut alle OCR-Informationen vom d.3 System für eine Dokumentensammlung abfragen.

Warnung

Dieser Assistent kann erst ab der d.3 Server-Version 7.2.2.7 verwendet werden. Bei älteren Versionen steht Ihnen diese Funktion nicht zur Verfügung.

Anmerkung

Bitte beachten Sie, dass bei einer Reindizierung immer nur die aktuelle Version eines Dokumentes reindiziert wird. Ältere Versionen werden nicht berücksichtigt.

Reindizierung einrichten

Im Folgenden wird beschrieben, welche Einstellungen Sie für die Reindizierung vornehmen können.

Allgemein

Allgemein

Dokumentensammlung	Migrieren von d3p (M) ▼ in d3ps	
Anzahl d.3 server-Prozesse	1	
Zeitraum der Reindizierung	Vom 01.02.2015  bis zum 20.02.2015 	
Anzahl der Dokumente pro Call	500	
OCR-Datei generieren	☐	
Dokumentensammlung löschen	☐	

Einstellung	Beschreibung
Dokumenten-sammlung	Wählen Sie die entsprechende Dokumentensammlung aus, für welche eine Reindizierung stattfinden soll. Ein (PN) kennzeichnet, dass ein paralleler Neuaufbau von dieser Dokumentensammlung in eine andere vorgenommen wird. Es wurde dementsprechend eine Spiegelung für diese Dokumentensammlung eingerichtet. Ein (N) kennzeichnet, dass eine Nachverschlagwortung der Dokumentensammlung stattfindet. Das heißt, dass alle Dokumente erneut vom d.3 System abgerufen und neu verschlagwortet werden.
Anzahl d.3 server-Prozesse	Geben Sie an wie viele d.3 server-Prozesse, zur selben Zeit, maximal für die Reindizierung verwendet werden dürfen. Warnung Die angegebene Anzahl an d.3 server-Prozessen steht Ihnen für das Produkktivsystem nicht mehr zur Verfügung. Berücksichtigen Sie dies und richten Sie ggfs. für den Zeitraum der Reindizierung zusätzliche d.3 server-Prozesse ein.
Zeitraum der Reindizierung	Wenn Sie nicht alle Dokumente in den Dokumentensammlungen neu reindizieren wollen, können Sie den Zeitraum an dieser Stelle festlegen. Anmerkung Als Zeit wird die letzte Änderungszeit der Attribute eines Dokumentes verwendet.
Anzahl der Dokumente pro Call	Mit diesem Parameter können Sie angeben, wie viele Dokumente auf einmal vom d.3 server abgefragt werden sollen. Eine hohe Anzahl sorgt dafür, dass das Netzwerk entlastet wird. Dafür werden dann Verbindungen entsprechend länger offen gehalten.
OCR-Datei generieren	Wenn dieser Schalter aktiviert wird, werden OCR-Dateien erzeugt, falls noch keine OCR-Informationen für ein Dokument vorhanden sind. Anmerkung Wenn diese Auswahl nicht angehakt wird und es Dokumente ohne OCR-Informationen gibt, werden diese später als Fehler gelistet.

Einstellung	Beschreibung
Dokumenten-sammlung löschen	Dieser Schalter gibt an, ob die Dokumentensammlung nach erfolgreichem parallelen Neuaufbau gelöscht werden soll. Anmerkung Bei einem parallelen Neuaufbau von d3P nach d3PS bedeutet dies folgendes: • d3P wird gelöscht • d3PS wird nach d3P umbenannt Dies erspart Ihnen eine nachträgliche manuelle Konfiguration im d.3 admin für die Dokumentensammlung. Anmerkung Bei einer Nachverschlagwortung steht Ihnen diese Option nicht zur Verfügung.

d.3 Benutzerdaten

Einstellung	Beschreibung
d.3 Benutzername	Geben Sie hier den d.3 Benutzer an, der über lesende Rechte für sämtliche Dokumentarten in der Dokumentensammlung verfügt. Zudem muss der Benutzer über Administrationsrechte verfügen. Anmerkung Eine Verfeinerung der zu reindizierenden Dokumente kann über das Rechtesystem des angegebenen Benutzers erfolgen. Anmerkung Die Administrationsrechte werden benötigt, um die Job-Queue im d.3 auszulesen. Durch die Abfrage der Job-Queue soll verhindert werden, dass die Queue durch die Nachverschlagwortung mit TIF001-Jobs und TIF002-Jobs voll läuft.
d.3 Passwort	Geben Sie das d.3 Passwort entsprechend zu dem gewählten d.3 Benutzer an.

d.3 Serverdaten

d.3 Serverdaten

Rechnername	localhost	
d.3-Server-ID	P	
Port	3400	

Einstellung	Beschreibung
Rechnername	Geben Sie den Rechnernamen an, auf dem der d.3 gateway eingerichtet ist.
d.3 Serverkürzel	Geben Sie hier das d.3 Serverkürzel von Ihrem d.3 System an, welches Sie reindizieren wollen.
Port	Geben Sie den Port an, über den der d.3 gateway erreichbar ist.

Laufzeiten

In diesem Abschnitt können Sie bis zu drei Laufzeiten einstellen. Wählen Sie dazu erst die Tage aus, an dem die Laufzeit gelten soll. Anschließend geben Sie die Start- und Endzeit an.

Um die Belastung des Produktivsystems so gering wie möglich zu halten empfiehlt es sich, den parallelen Neuaufbau nicht während der Hauptarbeitszeit laufen zu lassen.

Warnung

Die Backupzeiten der d.3 search engine (SHAFT) sollten berücksichtigt werden, um unnötige Fehler zu vermeiden.

Anmerkung

Um eine Laufzeit über 00:00 hinweg einzurichten (z.B. von 22:00 - 05:00) gehen Sie bitte wie folgt vor:

1. **Zeit 1** von 22:00 - 23:59
2. **Zeit 2** von 00:00 - 05:00

Wichtige Hinweise

Bevor Sie eine Reindizierung starten, sollten Sie folgendes überprüfen:

1. Stellen Sie sicher, dass der verwendete Benutzer Rechte aus sämtliche Dokumente hat, die Sie Migrieren wollen. Das die Reindizierung auf Basis einer d.3-Suche basiert sind die Rechte des Benutzer entscheidend.
2. Die Reindizierung kann die Job-Queue von d.3 search unter Umständen zum voll laufen bringen. Stellen Sie daher sicher, dass Sie mindestens so viele **Job-Threads** einrichten, wie Sie **d.3 server Prozesse** für die Reindizierung verwenden. Bedenken Sie, dass das Produktive Anfragen auch weiterhin über die **Job-Threads** verarbeitet werden. Daher sollte die **Jobs-Thread** Anzahl immer größer als die Anzahl der **d.3 server Prozesse** für die Reindizierung sein.
3. Wenn Sie d.3 search über den d.3 gateway angebunden haben, legen Sie den **HandlerCount**, für die d.3 gateway-Verbindungen, auf die selbe Anzahl fest, wie die Anzahl der **Job-Threads**.

Reindizierung überwachen

Dieser Übersicht können Sie den aktuellen Status der Reindizierung entnehmen.

Status

Status

Status	Reindizierung läuft
Startzeit	23.4.2015 14:10
Vorraussichtlich fertig in	00:00:25
Bearbeitete Dokumente	17
Erstellte OCR-Jobs	3
Fehlerhafte Dokumente	Geamtanzahl: 0
Warnungen für Dokumente	Geamtanzahl: 0
Zeitlich erstes Dokument	18.2.2015 12:47
Zeitlich letztes Dokument	23.4.2015 23:59
Zeitlich aktuelles Dokument	23.4.2015 23:59
Aktuelle Dokumentart	Bestellung (3 von 40)
Fortschritt aktuelle Dokumentart	100%
Gesamtfortschritt	8%

Eigenschaft	Beschreibung
Status	Diese Eigenschaft beschreibt den aktuellen Status der Reindizierung. Folgende Status können auftreten: • Reindizierung läuft Die Reindizierung wird aktuell durchgeführt • Programmtechnische Pause Die Reindizierung befindet sich nicht innerhalb der definierten Laufzeiten • Manuelle Pause Der Benutzer hat die Reindizierung manuell angehalten • Reindizierung wird pausiert... Die Reindizierung wird aktuell angehalten, sobald die Möglichkeit besteht • Reindizierung auf Grund von Fehlern pausiert. Die Reindizierung wurde gestoppt, weil Fehler aufgetreten sind • Reindizierung erfolgreich beendet Die Reindizierung ist erfolgreich beendet • Reindizierung mit Fehlern beendet Die Reindizierung ist erfolgreich beendet, jedoch sind fehlerhafte Dokumente vorhanden

Eigenschaft	Beschreibung
Startzeit	Gibt an, wann die Reindizierung gestartet worden ist. Anmerkung Alle Dokumente bis zu diesem Datum werden reindiziert. Neuere Dokumente werden nicht reindiziert, da diese über den normalen Weg in die Dokumentensammlung aufgenommen werden.
Voraussichtlich fertig in	Gibt an, wie lange es noch dauert, bis die Migration durchgelaufen ist. Anmerkung Bei der Angabe handelt es sich um einen Schätzwert auf Basis des Gesamtfortschritts im Verhältnis zu der bereits abgelaufenen Zeit.
Bearbeitete Dokumente	Gibt an, wie viele Dokumente bereits bearbeitet worden sind.
Erstellte OCR-Jobs	Gibt an, wie viele OCR-Jobs bereits erstellt werden mussten. Anmerkung Beachten Sie, dass OCR-Jobs asynchron vom d.3 async verarbeitet werden.
Fehlerhafte Dokumente	Gibt an, wie viele Dokumente fehlerhaft waren. Zusätzlich liefert es Informationen darüber, welcher Fehlercode wie oft vorgekommen ist.
Warnungen für Dokumente	Gibt an, für wie viele Dokumente eine Warnung vorliegt. Eine Warnung für ein Dokument liegt vor, wenn keine OCR-Informationen vorliegen und keine OCR-Dateien automatisch bei der Reindizierung erstellt werden sollten.
Zeitlich erstes Dokument	Gibt an, welches Dokument das zeitlich erste in der aktuellen Dokumentart ist. Mit Hilfe dieser Informationen wird der Fortschritt für die Dokumentart erstellt.
Zeitlich letztes Dokument	Gibt an, welches Dokument das zeitlich letzte in der aktuellen Dokumentart ist. Mit Hilfe dieser Informationen wird der Fortschritt für die Dokumentart erstellt.
Zeitlich aktuelles Dokument	Gibt an, welches Dokument das zeitlich aktuelle in der aktuellen Dokumentart ist. Mit Hilfe dieser Informationen wird der Fortschritt für die Dokumentart erstellt.
Aktuelle Dokumentart	Gibt an, welche Dokumentart aktuell reindiziert wird.
Fortschritt aktuelle Dokumentart	Gibt den Fortschritt für die aktuelle Dokumentart an. Der Fortschritt besagt, dass X-Prozent der Dokumente aus der aktuellen Dokumentart bearbeitet worden sind. Anmerkung Wenn ein OCR-Job für ein Dokument erstellt worden ist, gilt das Dokument als bearbeitet, auch wenn der OCR-Job asynchron verarbeitet wird.
Gesamtfortschritt	Gibt den Gesamtfortschritt für die Reindizierung an. Der Fortschritt besagt, dass X-Prozent der Dokumente bearbeitet worden sind. Anmerkung Wenn ein OCR-Job für ein Dokument erstellt worden ist, gilt das Dokument als bearbeitet, auch wenn der OCR-Job asynchron verarbeitet wird.

Einstellungen

In diesem Bereich werden die Einstellungen angezeigt, die zuvor für diese Reindizierung eingestellt worden sind. Die Beschreibung für die Einstellungen können Sie aus dem vorherigen Kapitel entnehmen.

Einstellungen

Dokumentensammlung

Migrieren von **d3p** in **d3ps**

Anzahl d.3 server-Prozesse

1

Zeitraum der Reindizierung

Vom 01.02.2015

bis zum

20.02.2015



Anzahl der Dokumente pro Call

500

OCR-Datei generieren



Dokumentensammlung löschen



d.3 Benutzername

dvelop

Rechnername

10.95.1.166

d.3-Server-ID

P

Port

3400

	Mo	Di	Mi	Do	Fr	Sa	So	Startzeit	Endzeit
Zeit 1	☒	☒	☒	☒	☒	☒	☒	00:00	23:59
Zeit 2	☐	☐	☐	☐	☐	☐	☐	00:00	23:59
Zeit 3	☐	☐	☐	☐	☐	☐	☐	00:00	23:59

Ergebnisprotokoll

Nach einer erfolgreichen Reindizierung ändert sich der Reindizierungsstatus entsprechend auf **Reindizierung erfolgreich beendet**. Wenn Fehler bei der Reindizierung aufgetreten sind, die einzelne Dokumente betreffen, werden diese unterhalb der Sektion **Einstellungen** im Abschnitt **Fehlerliste** in Form von **Dokumenten-ID** und **Fehlercode** aufgelistet. Die Bedeutung der Fehlercodes entnehmen Sie bitte der d.3 server-Dokumentation.

Anmerkung

Der Fehlercode -23 bedeutet, dass keine OCR-Information zu diesem Dokument vorliegt. Da der Schalter **OCR-Datei generieren** nicht gesetzt worden ist, erscheint dieses Dokument als fehlerhaft.

Fehlerliste

Dokumenten-ID	Fehlercode
P000000001	-23
P000000002	-23
P000000153	-23
P000000154	-23

Reindexierung über die shaftutl.exe

Warnung

Grundvoraussetzung hierfür ist, dass Sie den Schalter StoreText bei der entsprechenden Dokumentensammlung seit Erzeugung der Datenbank aktiviert haben.

Die Reindexierung über die Verwendung von StoreText bietet den Vorteil, dass die Dokumente nicht über den d.3 server angefordert werden müssen. Beachten Sie jedoch, dass die Dokumentensammlung für die Dauer der Neuindexierung nicht verfügbar ist.

Warnung

Je nach Größe der Dokumentensammlung kann eine Reindexierung einige Zeit (Stunden/Tage) in Anspruch nehmen.

Ist für Sie der Ausfall der Dokumentensammlung für die Dauer der Reindexierung keine Option, lesen Sie bitte das Kapitel [.Reindizierung über den d.3 Server v3.0.0.](#)

Um die Reindexierung durchzuführen gehen Sie wie folgt vor:

- Gehen Sie in das Webinterface der entsprechenden d.3 search Instanz.
- Wählen Sie nun in der Menüleiste den Punkt **Engines** aus und im Untermenü die **Corpus-ID** der Engine, die reindexiert werden soll.
- Drücken Sie anschließend auf den Button **Stop** in der Sektion **Funktionen**.
- Öffnen Sie eine Kommandozeile (CMD-Konsole cmd . exe) und navigieren Sie in das Verzeichnis der d.3 search Installation.

Anmerkung

```
C:\Users\Administrator> cd C:\d3\d.3 search
```

- Geben Sie folgenden Befehl ein: `shaft\shaftutl.exe corpora\<CorpusID>\shaft.ini -ri`

Anmerkung

```
C:\d3\d.3 search> shaft\shaftutl.exe corpora\d3p\shaft.ini -ri
```

- Warten Sie, bis shaftutl.exe erfolgreich beendet ist.
- Gehen Sie nun wieder in das Webinterface von d.3 search und starten Sie die entsprechende Engine (**Engines** | **<Corpus-ID>** und dann unter **Funktionen** den Button **Start**).

Anmerkung

Falls Sie sich im Clustermodus befinden, führen Sie die einzelnen Schritte Eins bis Vier für jeden einzelnen Knoten durch.

Reindexierung über den d.3 Server

Wenn Sie nicht die Möglichkeit haben eine Reindexierung mittels der `shaftutil.exe` vorzunehmen können Sie die Reindexierung über den d.3 server verwenden.

Beachten Sie bitte, dass die Reindexierung insgesamt mehr Zeit in Anspruch nehmen wird.

Anmerkung

Verwenden Sie, wenn möglich, den Assistenten zur Reindizierung im Webinterface.

Informationen zu den Assistenten finden Sie hier: [Assistent zur Migration einer Dokumentensammlung](#).

Reindexierung

Für eine Neuindizierung über den d.3 server gehen Sie bitte wie folgt vor:

- Gehen Sie in das Webinterface der entsprechenden d.3 search Instanz.
- Wählen Sie nun in der Menüleiste den Punkt **Engines** | **<Corpus-ID>** der Engine, die reindexiert werden soll.
- Drücken Sie anschließend auf den Button **Funktionen** | **Stop**.
- Navigieren Sie zum Verzeichnis der Dokumentensammlung (im Standard: `C:\d3\d.3 search\corpora\<Corpus-ID>`).

Anmerkung

Verzeichnis für den Corpus d3p: `C:\d3\d.3 search\corpora\d3p\`

- Löschen Sie in diesem Verzeichnis alle Dateien und Ordner außer:
 - *.ini Dateien
 - *.txt Dateien
- Gehen Sie in das Webinterface der entsprechenden d.3 search Instanz.
- Wählen Sie nun in der Menüleiste den Punkt **Engines** | **<Corpus-ID>** der Engine, die reindexiert werden soll.
- Drücken Sie anschließend auf den Button **Funktionen** | **Start**.
- Navigieren Sie zu dem Installationsverzeichnis von d.3 search.
- Besorgen Sie sich das Skript zum Nachverschlagworten der Dokumente (`reindex_fulltext_corpus.jpl` oder `dokumente_nachverschlagworten.jpl`) und kopieren Sie dieses in das `ext_jpl`-Verzeichnis des d.3 server.

Anmerkung

Das Skript können Sie beim d.velop Support anfordern.

Anmerkung

Kopieren von `C:\%Pfad_zum_Script%\dokumente_nachverschlagworten.jpl` nach `C:\d3\d3server.prg\ext_jpl`

- Passen Sie das JPL-Skript vor der Benutzung entsprechend an.
- Dazu navigieren Sie in die Zeile 131 und geben hier das Kürzel der Dokumentensammlung an, wenn Sie nicht die Einstellungen vom d.3 server verwenden wollen.

```
## !! Die Dokumentensammlung braucht nicht mehr angepasst werden! Es
werden die Einstellungen des d.3 server verwendet! Geben Sie die ID als
5. Parameter an, wenn Sie Dokumente in eine andere Dokumentensammlung
hinzufügen wollen.
error = dsearch_add_document(dok_dok_id, "", 1, "", "", 1)
```

Zusätzlich können Sie ab Zeile 100 den Select-Befehl zur Ermittlung der DokumentIds anpassen.

```
#####
# Select zur Ermittlung der DokumentIds
# Kann bei Bedarf angepasst werden.
#####

DBMS DECLARE CURSOR work_cursor FOR \
  SELECT doku_id \
  FROM phys_datei :lock_hint \
  WHERE (dokuart <> '$LOG$') \
  :DOKU_ID_TERM \
  ORDER BY doku_nr DESC

#####
```

- Öffnen Sie das entsprechende d.3 server Interface des Archivs, aus dem Dokumente in ihre Datensammlung importiert werden sollen
- Öffnen Sie unter **Programs** das kopierte JPL-Skript `dokumente_nachverschlagworten.jpl` aus dem `ext_jpl`-Verzeichnis
- Warten Sie, bis alle Dokumente vernachschlagwortet sind.

Anmerkung

Befinden Sie sich im Clustermodus, muss nach dem Schritt Vier das Cluster aktiv sein. Überprüfen Sie dies im Webinterface über den Menüpunkt **Cluster**.

Reindizierung der Notizdateien

Die Notizdateien werden mit bei dem Skript `dokumente_nachverschlagworten.jpl` nicht berücksichtigt. Dazu können Sie das Skript `Notizdateien_nachverschlagworten.jpl` verwenden.

Für eine Neuindexierung der Notizdateien über den d.3 server gehen Sie bitte wie folgt vor:

- Navigieren Sie zu dem Installationsverzeichnis von d.3 search.
- Besorgen Sie sich das Skript zum Nachverschlagworten der Notizdateien (`Notizdateien_nachverschlagworten.jpl`) und kopieren Sie dieses in das `ext_jpl`-Verzeichnis des d.3 Servers.

Anmerkung

Das Skript können Sie beim d.velop Support anfordern.

Anmerkung

Kopieren von `C:\%Pfad_zum_Skript%\Notizdateien_nachverschlagworten.jpl` nach `C:\d3\d3server.prg\ext_jpl`

- Passen Sie das JPL-Skript vor der Benutzung entsprechend an.
- Dazu navigieren Sie in die Zeile 38 und geben hier das Kürzel der Dokumentensammlung an, wenn Sie nicht die Einstellungen vom d.3 server verwenden wollen.

```
// Konfiguration
vars DSEARCH_CORPUS = "d3P";
```

Zusätzlich können Sie ab Zeile 62 den Select-Befehl zur Ermittlung der DokumentIDs anpassen.

```
// Doku-IDs aller Notizdateien ermitteln
DBMS DECLARE CURSOR work_cursor FOR \
    SELECT doku_id \
    FROM dependent_documents \
    WHERE UPPER(doc_extension) = 'NOT' \
    ORDER BY doku_id DESC
```

- Öffnen Sie das entsprechende d.3 server Interface des Archivs, aus dem Dokumente in ihre Datensammlung importiert werden sollen.
- Öffnen Sie unter **Programs** das kopierte JPL-Skript Notizdateien_nachverschlagworten.jpl aus dem ext_jpl-Verzeichnis.
- Warten Sie, bis alle Dokumente vernachschlagwortet sind.

Anmerkung

Befinden Sie sich im Clustermodus, muss das Cluster aktiv sein. Überprüfen Sie dies im Webinterface über den Menüpunkt **Cluster**.

OCR-Dateien im d.3 server erzeugen

Wenn in Ihrem d.3-System keine OCR-Dateien vorhanden sind, da Sie in der Vergangenheit keine OCR-Dateien beim Import erzeugt haben, können Sie diese durch das Skript nach_vertiffen.jpl erstellen lassen.

Anmerkung

Verwenden Sie, wenn möglich, den Assistenten zur Reindizierung im Webinterface. Informationen zu den Assistenten finden Sie hier [Assistent zur Migration einer Dokumentensammlung](#).

Für die Erstellung der OCR-Dateien über den d.3 server gehen Sie bitte wie folgt vor:

- Navigieren Sie zu dem Installationsverzeichnis von d.3 search.
- Besorgen Sie sich das Skript zum nachträglichen vertiffen von Dokumenten (nach_vertiffen.jpl) und kopieren Sie dieses in das ext_jpl-Verzeichnis des d.3 Servers.

Anmerkung

Kopieren von C:\%Pfad_zum_Skript%\nach_vertiffen.jpl nach
C:\d3\d3server.prg\ext_jpl

- Passen Sie das JPL-Skript vor der Benutzung entsprechend an.
- Dazu navigieren Sie in die Zeile 101 und geben hier die Tabelle an, die die jeweiligen Dokument-IDs enthält. Standardmäßig wird die Tabelle phys_datei angegeben, die alle Dokumente enthält.

```
// Tabelle mit den DokumentIDs in der Spalte doku_id
vars table_dokuid = "phys_datei"
```

Zusätzlich können Sie ab Zeile 103 den Benutzer angeben, unter dessen Namen die Klassen ermittelt werden und das OCR-Datei erstellen angestoßen wird. Wenn der Eintrag leer bleibt, wird der Besitzer des Dokumentes verwendet oder der Bearbeiter für den Status "Bearbeitung"

```
// User, der für das Vertiffen "benutzt" werden soll
vars user_script = ""
```

- Öffnen Sie das entsprechende d.3 server Interface des Archivs, aus dem Dokumente in ihre Datensammlung importiert werden sollen.
- Öffnen Sie unter **Programs** das kopierte JPL-Script `nach_vertiffen.jpl` aus dem `ext_jpl`-Verzeichnis.
- Warten Sie, bis für alle Dokumente eine OCR-Datei erstellt worden ist.

Anmerkung

Befinden Sie sich im Clustermodus, muss das Cluster aktiv sein. Überprüfen Sie dies im Webinterface über den Menüpunkt **Cluster**.

1.6. Fehler und Probleme

Falls beim Betrieb der d.3 search engine Probleme auftreten,

- öffnen Sie bitte den Log-Viewer auf Ihrem d.3 Server (**d.velop | d.3 gateway | Log-Viewer**) und überprüfen das Log-Viewer Protokoll.

Hier werden alle fehlerhaften Vorgänge im Zusammenhang mit dem d.3 System protokolliert.

Alle Protokollmeldungen von d.3 search werden mit dem Modul ID DSEARCH ausgegeben.

File Edit View Options Help				
Refresh	Critical Errors	Errors	Alerts	Information
			Start/Stop	Debug
Date	Source	ID	Message	
18.05 15:04:13,121	DSEARCH	00000E68	CALL 42: 410100 - d3_admin 049 - DSEARCHGetCorpora	
18.05 15:04:13,121	DSEARCH	00000E68	Number: 1	
18.05 15:04:13,121	DSEARCH	00000E68	ReturnCaption: SUCCESS	
18.05 15:04:13,121	DSEARCH	00000E68	Header->retcode: 0	
18.05 15:04:13,121	DSEARCH	00000E68	Function terminated normaly.	
18.05 15:04:13,121	DSEARCH	00000E68	#####	
18.05 15:04:13,137	DSEARCH	00000E68	ISM_CLOSE	
18.05 15:04:13,731	DSEARCH	00000E68	Connecting 10.5.0.119	
18.05 15:04:13,731	DSEARCH	00000E68	#####	
18.05 15:04:13,731	DSEARCH	00000E68	CALL 43: 410100 - d3_admin 049 - DSEARCHGetServerStatus	
18.05 15:04:13,731	DSEARCH	00000E68	ReturnCaption: SUCCESS	
18.05 15:04:13,731	DSEARCH	00000E68	Header->retcode: 0	
18.05 15:04:13,731	DSEARCH	00000E68	Function terminated normaly.	
18.05 15:04:13,731	DSEARCH	00000E68	#####	

1.6.1. d.3 search

Aktiv/Aktiv-Cluster ist inaktiv

Ist Ihr Aktiv/Aktiv-Cluster inaktiv, überprüfen Sie folgende Einstellungen, die den fehlerfreien Betrieb ihres Clusters stören können:

- Überprüfen Sie, ob alle gewünschten d.3 search engine (SHAFT) -Prozesse laufen. Dies können Sie im Webinterface, standardmäßig aufrufbar unter der Adresse <http://localhost:6133/>, unter dem Menüpunkt **Engines | <CorpusId>** überprüfen:

Einstellungen

Status	Online
Korpus	d3p
Rechnername	localhost
Dateiname	C:\d3\d.3 search\shaft\64Bit\shaft.exe
Befehlszeile	C:\d3\d.3 search\corpora\d3p\shaft.ini
Prozess-ID	5452
Startzeit	20.08.2014 15:23:32.281

- Überprüfen Sie, ob Ihr d.3 search server Prozess noch im d.3 process manager eingetragen und aktiv ist:

d.3 search						
Execute selected Stop selected Restart selected						
Process	Process ID	Memory	Virtual Memory	Starttime	Starts	
d.3 search (server)	5788 (0x0000169C)	11736 K	10692 K	20.08.2014 14:59:51	3	

- Überprüfen Sie, ob die Clustermitglieder kommunizieren, indem Sie im d.3 search Webinterface, standardmäßig aufrufbar unter der Adresse <http://localhost:6133/>, das **Log** überprüfen. Dieses finden Sie unter dem Menüpunkt **Cluster**. Hier sollten Sie, bei erfolgreicher Verbindung zwischen den Mitgliedern im Cluster, Einträge in folgender Form finden:

```
21.08.2014 11:34:31: [SEOK]: [REDACTED] - IAMALIVEMASTER
21.08.2014 11:34:38: [SEOK]: [REDACTED] - IAMALIVEMASTER
```

d.3 search wird gestartet und beendet sich sofort wieder

d.3 search beendet sich automatisch, wenn der konfigurierte TCP/IP Port (3406 bzw. die Vorgabe in der `dsearch.ini`) bereits von einem anderen Serverprozess belegt ist.

Hierzu finden Sie auch entsprechende Protokollmeldungen im Log-Viewer.

d.3 search ist langsam

Im folgenden werden verschiedene Lösungsansätze aufgelistet, die eine Performancesteigerung von d.3 search ermöglichen.

Shaft-Engine mehr Arbeitsspeicher zuweisen

Um die Festplattenzugriffe so niedrig wie nur eben möglich zu halten, sollte der zur Verfügung stehende Arbeitsspeicher komplett genutzt werden.

Sie können der Shaft-Engine mehr Arbeitsspeicher zuweisen, in dem Sie die `CacheSize` in der `shaft.ini` erhöhen.

Festplattenzugriff optimieren

Um festzustellen, ob ein optimaler Festplattenzugriff existiert, können Sie wie folgt vorgehen:

1. Kontrollieren Sie, ob die Shaft-Engine ein PerfLogs-Verzeichnis im Corpus-Verzeichnis angelegt hat. Ist dies der Fall überprüfen Sie, ob die Lesegeschwindigkeit außergewöhnlich langsam ist. Dazu entnehmen Sie die entsprechenden Werte aus der Perflogs-Datei ([Hinweise zur Dokumentensammlung mit der d.3 search engine](#)) und nehmen folgende Rechnung vor:

$$(\text{Bytes}/1024/1024)/\text{Time} = x \text{ MB/s.}$$
 Dieser Wert sollte optimaler Weise immer über 30 MB/s liegen. Ist dies nicht der Fall überprüfen Sie bitte den Festplattenzugriff. Achten Sie dabei auf die unten angegebenen Hinweise.
2. Beim Start der Shaft-Engine wird ein Speed-Test vorgenommen. Dieser Wert sollte ebenfalls optimaler Weise immer über 30 MB/s liegen.

Da oftmals aus Ressourcenmangel nicht der ganze Index im Arbeitsspeicher gehalten werden kann ist der größte Schwachpunkt bei einer Suche der Zugriff auf die Festplatte.

Kontrollieren Sie daher, ob noch andere Programme auf die physikalische Festplatte zugreifen, auf der die Shaft-Datenbank (Index) hinterlegt ist.

Optimal wäre eine physikalische Festplatte, die ausschließlich nur von der Shaft-Engine verwendet wird.

Windows Cache-Verwaltung

Überprüfen Sie, ob die Seitenfehler pro Sekunde enorm ansteigen, wenn Suchanfragen gesendet werden.

Wenn dies der Fall ist, lagert Windows viele Daten auf die Festplatte aus. Da der Zugriff auf die Festplatte wesentlich langsamer ist als der auf dem Arbeitsspeicher, kommt es hier zu langen Wartezeiten. Dieses Phänomen wird auch Swapping genannt.

Um dieses zu verhindern bzw. zu Minimieren muss der Shaft-Engine mehr Arbeitsspeicher zur Verfügung gestellt werden, der nicht gewappt wird.

Dies kann durch das Tool SetSystemFileCacheSize eingestellt werden. Sie können das Tool kostenlos aus dem Internet herunterladen.

Führen Sie das Tool bitte mit folgenden Aufruf aus:

`SetSystemFileCacheSize.exe MinCache MaxCache`

wobei:

`MaxCache` = (Physicalischer Cache, der der Shaft-Engine zugewiesen wurde) - (1024 bis 2048 MB)

`MinCache` = 8 MB

Anschließend sollten die Seitenfehler pro Sekunde bei der Ausführung einer Suchanfrage nicht mehr so enorm steigen und somit kein Swapping mehr statt finden.

Aktiv/Aktiv-Cluster durch fehlerhaften Knoten inaktiv - Aufstauen nicht verarbeitete Jobs vermeiden

Wenn Sie d.3 search in einem Clusterbetrieb führen und das Cluster nun durch einen Fehler inaktiv wird, stauen sich nicht verarbeitete Jobs auf.

In manchen Fällen ist es aber nicht sinnvoll, dass diese tatsächlich gestoppt und nachträglich in großer Menge nachgepflegt werden. Hier empfiehlt sich, den fehlerhaften Knoten temporär aus dem Cluster zu entfernen und ihn nachträglich sauber wieder einzufügen.

Gehen Sie dafür wie folgt vor:

1. Die d.3 search Instanz auf dem defekten Knoten muss heruntergefahren werden. Nehmen Sie dann den fehlerhaften Knoten aus dem Cluster, indem Sie den Knoten im Webinterface des Masters über den Button **Löschen** entfernen.

Cluster-Server

IP-Adresse	Rechnername	Modus	Letzte Meldung	Information	Fehlermeldung
192.168.1.10	server1	Unbekannt	2023-01-01 10:00:00		
192.168.1.11	server2	Master	2023-01-01 10:00:00		

2. Deinstallieren Sie danach d.3 search von dem entfernten Rechner. Sorgen Sie dafür, dass der Knoten danach keinerlei Spuren der alten d.3 search Installation aufweist.
3. Wenn der Knoten wieder einsatzbereit ist wird d.3 search auf diesem neu installiert. Fügen Sie dann den Knoten auf normale Weise wieder zum Cluster hinzu (siehe Kapitel [Neue d.3 search Instanz für das Cluster einrichten](#)).

Warnung

Führen Sie Schritt 2 nicht richtig durch, wird die schon bestehende Dokumentensammlung auf dem ehemals defekten Knoten beim erneuten Integrieren in das Cluster nicht geupdatet.

Durch die asynchronen Zustände der Knoten ist das Cluster damit inkonsistent.

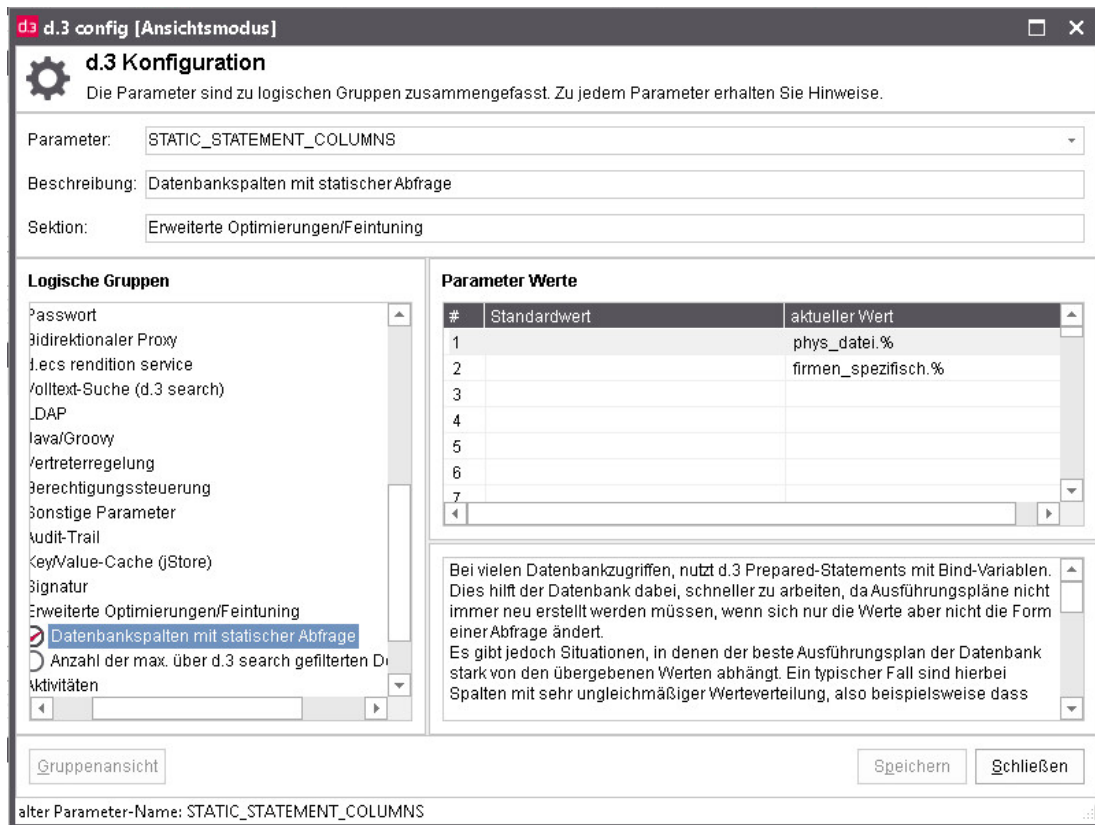
Langsame Reindizierung über den Assistenten

Unter gewissen Umständen kann es sein, dass die Suche bei der Reindizierung über den Assistenten unnötig lange dauert.

Wenn dies der Fall ist, kann der folgende Parameter gesetzt werden, um die Suche in der Datenbank zu beschleunigen.

- Wechseln Sie dazu im d.3 admin zur Konfiguration **d.3 config**.
- Wählen Sie dort den Punkt **Erweiterte Optimierungen/Feintuning** aus und dann den Punkt **Datenbankspalten mit statischer Abfrage**. Dort tragen Sie den Wert: **firmen_spezifisch.%** ein.

Die Änderung führt dazu, dass die Datenbank den SQL-Befehl besser analysieren kann und somit den Index der Datumswerte verwendet.



1.6.2. d.3 search engine

Checkliste und Fragenkatalog für den Support

Fragen, die bei Problemen mit der d.3 search engine gestellt werden sollten:

- Wie sieht die SHAFT . INI der betreffenden Dokumentsammlungen aus?
- Wurden Journal, Backup und Logging eingeschaltet?
- Laufen die SHAFT-Prozesse durch oder werden Sie von Zeit zu Zeit heruntergefahren?
- Hart oder korrekt? Wann, wie spät?
- Seit wann tritt das Problem auf?
- Datum / Uhrzeit definierbar?
- Wie äußern sich die Probleme?
- Prozess-Verhalten, Fehlermeldungen, Speicherplatz-Verhalten
- Wie groß ist das Dokument-Volumen, das von d.3 search verwaltet werden soll?
- Wieviele / Welche Dokumentarten werden verwaltet?
- Von welcher Suchmaschine?
- Wieviele Dokumente müssten eigentlich in einer Dokumentsammlung verwaltet werden? (Per SQL-Abfragen auf der d.3 Datenbank feststellbar)
- Wie sieht die Job-Verarbeitung von d.3 search aus?

- Welche Versionen haben
- SHAFT . EXE
- SHAFTUTL . EXE
- DSEARCHSVR . EXE
- SHAFT . DLL
- evtl. TextUpdate . exe
- Wie ist d.3 search in der d.3 Administration konfiguriert? (Screenshots, Beschreibungen)
- d.3 Parameter SUCHWORT_OPTION (KEYWORD_OPTION)
- DUMMY-Dokumentart(en)
- Dokumentsammlungen-Dokumentarten-Zuordnungen
- d.3 Parameter DSEARCH_SUPPORT
- Art der Volltextsuche (d.3 search + d.3 Wörterbuchtabellen?)

Anfragen an die d.3 search engine dauern lang

Die Verarbeitungsgeschwindigkeit der d.3 search engine hängt stark von dem verfügbaren Hauptspeicher ab.

Der Parameter CacheSize in der shaft . ini im Konfigurationsverzeichnis der Dokumentsammlung sollte erhöht werden.

Auch der Parameter PhysicalCache sollte aktiviert sein ("1"), falls auf dem Server genügend Hauptspeicher frei ist.

Reparatur einer SHAFT-Datenbank

Reparatur bei vorhandener Sicherung und vorhandenem Transaktions-Protokoll

Annahmen

Betrachtet wird die Dokumentsammlung "d3x", Corpus-Verzeichnis also auch \d3x (Unterordner im Ordner corpora).

Der SHAFT-Prozess ist ausgeschaltet; SHAFTUTL -C liefert: "Corrupt Database".

Es existiert ein Backup im lt. SHAFT . INI konfigurierten Backupordner.

Seit diesem letzten Backup (Dateidatum) sind alle Transaktionen im TLOG Ordner mitprotokolliert worden (die ersten Einträge in der trans . Log-Datei stammen aus dem Zeitraum direkt nach dem letztem Backup -> nahtloser Übergang).

Sicherungsphase

- d.3 search engine (d3x) - Prozess (SHAFT) als ausgeschaltet verifizieren
- d.3 search Server - Prozess als ausgeschaltet verifizieren
- Corpus-Verzeichnis sichern (also z. B. nur "d3p", zunächst ohne Unterordner)
- Unterverzeichnis ARCHLOGS sichern
- Unterverzeichnis TLOG sichern

Wiederherstellungsphase

- Datenbank-Dateien (alle bis auf .INI-Dateien) aus dem Corpusverzeichnis d3x löschen.
- Inhalt des Backup-Verzeichnisses in das Verzeichnis der Dokumentsammlung (z. B. d3x) zurückholen.
- Zurückgeholtes Backup mit SHAFTUTL -c prüfen (Die Sicherung bleibt im konfigurierten Backup Ordner).

Anmerkung

Wie Sie die Datei SHAFTUTL . EXE aufrufen, erfahren Sie [hier](#).

- SHAFT-Prozess per DOS-Kommando starten.
- Warten, bis dieser komplett hochgefahren ist (Fenster in Vordergrund und wiederhergestellt: „Database opened“).
- SHAFT-Prozess wieder anhalten (SHAFT- Fenster schließen).
- Mit SHAFTUTL -t Transaktionen nachholen.
- Mit SHAFTUTL -c die Datenbank mit den nachgeholten Transaktionen prüfen (und ggf. sichern).

Prüfungsphase

- SHAFT-Prozess wieder per DOS-Kommando starten und warten bis dieser komplett hochgefahren ist („database opened“, Speicherauslastung mindestens lt. Cachegröße in SHAFT . INI).
- Wenn OK, diesen wieder ausschalten und im d.3 process manager starten.
- Warten, bis der Prozess im d.3 process manager hochgefahren ist (Prozess läuft zunächst, danach nicht mehr, siehe grafische Kurve im Web-Interface des d.3 process manager).
- Speicherauslastung prüfen.
- d.3 search Server wieder starten.
- Verbindung von d.3 search zur Dokumentsammlung in der d.3 Administration prüfen.
- Jobverarbeitung von d.3 search prüfen.

Reparatur bei vorhandener Sicherung und Archiv-Protokoll ohne Transaktions-Protokoll

Es werden dieselben Schritte wie bei einer Reparatur bei vorhandenem Transaktionsprotokoll durchgeführt, jedoch wird in der Wiederherstellungsphase NICHT der Parameter -t, sondern -i mit dem SHAFTUTL eingesetzt.

Re-Indexierung einer SHAFT-Datenbank

Sollte die Datenbank der SHAFT-Suchmaschine als korrupt erkannt werden und die Alternativen unter **Reparatur einer SHAFT-Datenbank** nicht zur Verfügung stehen, bzw. nicht funktionieren, besteht noch die Möglichkeit der Re-Indexierung der Datenbank.

Dies wird mit dem SHAFTUTL-Parameter -r durchgeführt und stellt die letzte Option dar, um eine korrupte Datenbank zu reparieren.

Die Re-Indexierung kann zudem auch dann sinnvoll sein, wenn viele Dokumente aus der Dokumentsammlung gelöscht wurden. Dies wird bei einer Überprüfung mit dem SHAFTUTL und dem Parameter -c ebenfalls analysiert. Die Re-Indexierung wirkt dann wie eine Art Defragmentierung und verringert die Größe der Datenbank.

Anmerkung

Wie Sie die Datei SHAFTUTL . EXE aufrufen, erfahren Sie [hier](#).

Suchmaschine ist nicht erreichbar

Problem

Im d.3 smart explorer können keine Volltext-Suchanfragen durchgeführt werden.

Diese werden mit entsprechenden Fehlermeldungen abgebrochen:



Symptome

Die Volltext-Informationen von (neuen) Dokumenten werden zwar an den d.3 search Server geliefert, jedoch erhöht dieser lediglich die Zahl seiner zu verarbeitenden Jobs, kann die Jobs jedoch nicht an die Suchmaschine weitergeben.

Im d.3 logviewer erscheinen Meldungen wie z.B.:

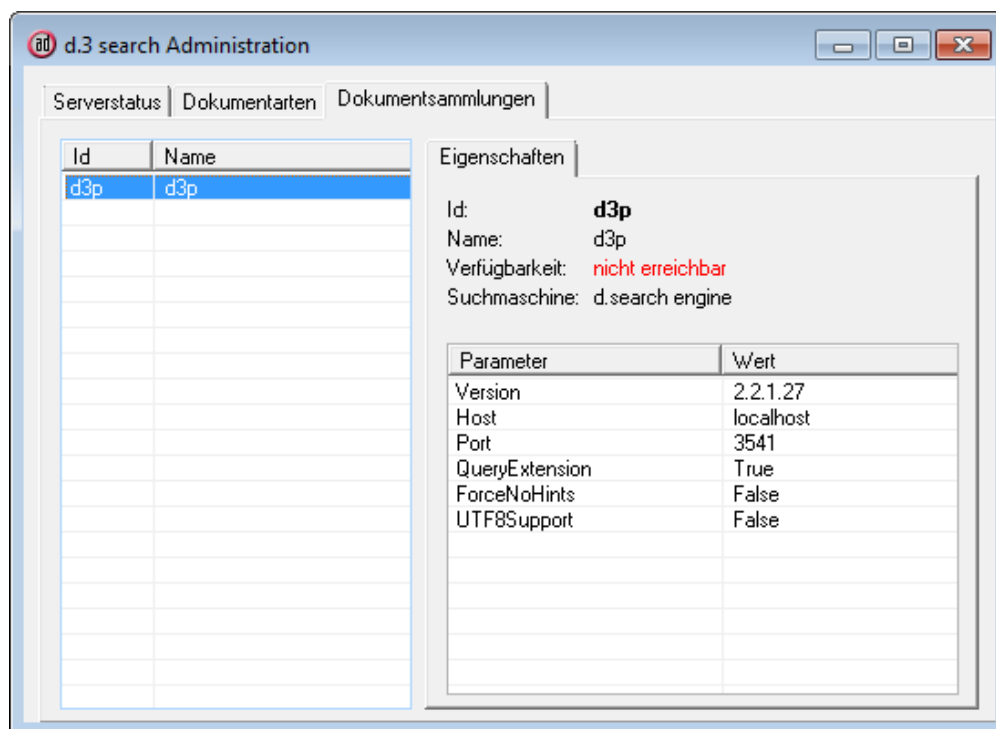
"Exception occurred."

"Context: Shaft - connect"

"Type: EldConnectException"

"Message: Socket-Fehler # 10061#0D#0AVerbindung abgelehnt."

In der d.3 Administration kann bei Wahl des Konfigurations-Moduls d.3 search und der Registerkarte **Dokumentsammlungen** beim Klick auf die entsprechende Dokumentsammlung keine Verbindung mit der Instanz der Suchmaschine aufgebaut werden:



Maßnahmen

Prüfen Sie, ob der betreffende d.3 search engine Prozess vollständig gestartet ist.

Dies kann festgestellt werden, indem die Speicherlast des jeweiligen Prozesses überprüft wird. Zu überprüfen ist auch, ob sogenannte „Geisterprozesse“ laufen, die den zu startenden Prozess blockieren.

Des Weiteren ist zu prüfen, ob die TCP/IP-Verbindung zwischen d.3 search und der Instanz der d.3 search engine aufgebaut werden kann. Hierbei müssen die Ports in CORPUS.INI und SHAFT.INI übereinstimmen und frei bzw. verfügbar sein.

32bit-Version der Shaft Engine

Fehler:

Ist eine 32bit Version der Shaft-Engine unter einem 64bit Betriebssystem mit einer Speicherzuweisung von mehr als 1 GB (CacheSize > 1048576) im Einsatz, kann diese beim Start des d.3 search engine Prozesses (SHAFT) die Error-Meldung im d.3 Logfile

"Open Error (Database or Listenport)"

protokollieren und dann wird der Prozess neu durchgestartet.

Auch eine Prüfung der Dokumentensammlung mit dem Shaftutl-Tool schlägt mit der Meldung

"WARNING! Database not closed!"

Try to close y/[n]?"

fehl und diese über die Eingabe "y" zu schließen, ist nicht möglich.

"Unable to close the database!"

Ursache:

Die 32bit Version der Shaft-Engine kann maximal mit einer Speicherzuordnung von 1 GB (CacheSize=1048576) arbeiten.

Erst ab der d.3 search Version 2.2.2.0 wird eine Shaft-Engine der Version 1.4 ausgeliefert, die auch eine 64-Bit Version enthält, die mit höheren Speicherzuweisungen arbeiten kann.

Lösung:

Bitte prüfen Sie in der zugehörigen Konfigurationsdatei `shaft.ini` der d.3 search Dokumentensammlung den zugewiesenen CacheSize-Wert und korrigieren Sie dieses, wenn notwendig. Dieser wird in KB angegeben.

Eine weitere Alternative ist ein Update der d.3 search Version auf eine Version, die eine 64-Bit Version der Shaft-Engine enthält.

1.7. Weitere Informationsquellen und Impressum

Wenn Sie Ihre Kenntnisse rund um die d.velop-Software vertiefen möchten, besuchen Sie die digitale Lernplattform der d.velop academy unter <https://dvelopacademy.keelelearning.de/>.

Mithilfe der E-Learning-Module können Sie sich in Ihrem eigenen Tempo weiterführende Kenntnisse und Fachkompetenz aneignen. Zahlreiche E-Learning-Module stehen Ihnen ohne vorherige Anmeldung frei zugänglich zur Verfügung.

Besuchen Sie unsere Knowledge Base im d.velop service portal. In der Knowledge Base finden Sie die neusten Lösungen, Antworten auf häufig gestellte Fragen und How To-Themen für spezielle Aufgaben. Sie finden die Knowledge Base unter folgender Adresse: <https://kb.d-velop.de/>

Das zentrale Impressum finden Sie unter <https://www.d-velop.de/impressum>.