



HULBEE ENTERPRISE SEARCH

Installation und Setup-Handbuch

Modified: 20-Sep-2016

Version: 2.2.27

Inhaltsverzeichnis

1	ARCHITEKTUR	3
1.1	HULBEE ENTERPRISE SEARCH.....	3
1.2	SWISSCOWS COMPANY SEARCH.....	4
2	ERSTINSTALLATION	5
2.1	VOLLSTÄNDIGE INSTALLATION	5
2.1.1	Indexstore	5
2.1.2	Verarbeitungsserver	6
2.1.2.1	Hardware-Voraussetzungen für HES Processing Server	6
2.1.2.2	Software-Voraussetzungen für HES Processing Server	6
2.1.2.3	IFilters setup.....	7
2.1.2.4	Authentifizierungseinstellungen	7
2.1.2.5	HES Auspacken.....	7
2.1.2.6	IndexUtil und benutzerdefinierte Indexstruktur	7
2.1.2.7	Ersteinrichtung.....	8
2.1.2.8	Lizenz einrichten	13
2.2	SWISSCOWS COMPANY SEARCH.....	13
2.2.1	Einrichtung einer Domain	13
2.2.2	Dateispeicher	14
3	ERSTDURCHLAUF	14
3.1	INDEX REINIGUNG UND MODIFIZIERUNG	16
4	EINRICHTUNG DER ADMIN SEITE	16
4.1	WIEDERHERSTELLUNGSOPTIONEN	16
4.2	ERSTE SCHRITTE DER EINSTELLUNGEN	17
4.3	REGISTRIERUNG DER CONNECTOREN.....	17
4.4	HINZUFÜGEN VON SPEICHERPLÄTZEN.....	17
4.5	ADMINISTRATIONSBEREICH	18
4.5.1	Dashboard	19
4.5.2	Anschlüsse	20
4.5.3	Nutzer	20
4.5.4	Protokolle.....	21
4.5.5	Unverarbeitet.....	21
4.6	EINSTELLUNGEN.....	22
4.6.1	Authorization	22
4.6.2	Verarbeitung.....	24
4.6.3	Löschen	27
4.6.4	Logging	28
4.6.5	Importieren / Exportieren	29
4.6.6	Vorschau	29
4.7	DATA CLOUD.....	30
4.7.1	Benutzerdefinierte Suchanfragen	30
4.7.2	Stoppwörter	31
5	CONNECTORS	31
5.1	ALLGEMEINE EINSTELLUNGEN.....	31
5.1.1	Einstellungen der Connector über Konfiguration.....	32

5.1.2	Einstellungen des Connectors im Admin-Bereich.....	33
5.1.3	Manuelles Connector-Setup.....	37
5.1.4	Connector Registrierung Aufheben.....	37
5.2	DATEISYSTEM CONNECTOR	38
5.2.1	Einstellungen des Dateisystem-Connector.....	38
5.3	MS-EXCHANGE CONNECTOR.....	39
5.3.1	MS-Exchange Connector Einstellungen	39
5.4	WEB-CONNECTOR.....	39
5.4.1	Einstellungen des Web-Connectors	39
5.4.2	Feineinstellungen des Web-Connectors.....	42
5.4.3	Indexierung eines Teils einer Web-Seite deaktivieren.....	43
5.4.4	Kanonische Links.....	43
5.4.5	X-HES-Users headers.....	43
6	HILFREICHE LINKS	45
7	BEKANNTE PROBLEME.....	45

1 Architektur

1.1 Hulbee Enterprise Search

Hulbee Enterprise Search (HES)¹ ist für Nutzer von MS Active Directory vorgesehen. Die aktuelle Version ist für die Suche in Dateispeichern unter Berücksichtigung der Rechte von Nutzern an den einen oder anderen Dateien vorgesehen. Für eine Einbeziehung von Dateien in die Suche muss der Nutzer über Rechte, jedoch mindestens über Leserechte verfügen.

Seit der Version 2.1 sind neben dem Enterprise-Dateisystem auch anderen Arten von Ressourcen „aus der Box“ verfügbar, z.B. Web-Ressourcen, MS Exchange-Ressourcen usw. Sie werden mit Hilfe des Connectors an das System angeschlossen. Zudem wird ein API für die Entwicklung individueller Connectoren angeboten. (Alle).

Die Quellen, die für die Suche verfügbar sind und einzelne Speicher innerhalb jeder Quelle, werden zu dem System hinzugefügt und durch den Administrator des Unternehmens bezeichnet.

Es werden folgende Dateiformate unterstützt:

Dateityp	Dateierweiterung	Extrahieren von Text	Extrahieren von Meta-Tags	Extrahieren von eingebetteten Dateien
Textdateien	txt, rtf, doc/dot, odt, wri, sxw	✓		
	docx/docm/dotx	✓	✓	
Verarbeitungsdateien	pdf	✓	✓	
	xps	✓		

¹ Einige der Module haben Swisscows oder SES im Dateinamen oder Texten. Es ist der alte Name des Projekts und ist ein Synonym für Hulbee Enterprise Search oder HES.

Hypertextdateien	html, htm, xml	✓		
	mht, shtml	✓	✓	
Tabellen	xsl, xslt, xls, ods, csv	✓		
	xlsx	✓	✓	
Präsentationen	pptx	✓	✓	
	ppt, pps, odp	✓		
Grafische Dateien	bmp, jpg/jpeg, png, jfif, tif, tiff, jpe		✓	
E-mail	msg, eml	✓	✓	✓
Archive	zip, rar, 7zip			✓
Media	avi, mp3, mp4, wav, m4a, wma, wmv,ogg, flac, mkv, ape, mpc		✓	
Source Code und Scripting	cs, vb, js, csproj, h, c, cpp, vbs,vcproj, vbproj, pl, sql, bat, cmd	✓		
	css	✓	✓	

Als Nutzeroberfläche kann jeder beliebige Browser (Mozilla Firefox, Chrome u.a.) mit offenem Link zur Suchseite im Intranet verwendet werden. In ihm werden die Suche selbst und Änderungen des Nutzerprofils vorgenommen. Auch Administratoren können unter Nutzung der Administratorenleiste viele Einstellungen vornehmen.

Dokumente, die als Dateien im Dateisystem (Netzwerk-Ressourcen) zur Verfügung stehen, versucht HES genau wie Dateien mit einem assoziierten Programm (z.B. Microsoft Word oder Acrobat Reader) zu öffnen, lädt diese aber nicht über den Browser. Um dies zu ermöglichen, muss auf dem Computer des Benutzers das Modul Desktop Manager ausgeführt sein. Details werden im Benutzerhandbuch beschrieben.

Zum Öffnen der gefundenen Dateien ist es erforderlich, dass der Nutzer mit seinem echten Account in einem lokalen Netz mit Dateispeicher arbeitet. Das Öffnen der Dateien erfolgt mithilfe der HES Desktop Manager Utility, die auf dem Computer des Kunden installiert sein muss. Der Nutzer muss über die erforderlichen Rechte für die Installation von Software verfügen oder sich hierfür an seinen Administrator wenden.

HES-Software-Komplex besteht aus zwei Hauptteilen:

- Indexstore (Linux-Server mit installiertem und konfiguriertem Elasticsearch).
- Processing server / Verarbeitungsserver (Windows-Server mit allen anderen Komponenten).

Bei kleinen Filestores, könnten diese Server mit den einzelnen Windows-Server-Maschinen kombiniert werden. Grosse Filestores können eine Anpassung von Elasticsearch Cluster mit einer Anzahl von Servern erfordern.

1.2 Swisscows Company Search

Die genannte Konfiguration stellt einen Appliance Server dar, der sämtliche Komponenten von Hulbee Enterprise Search enthält. Aufgrund dessen, dass hier ein vollwertiger Windows Server 2012 R2 Standard enthalten ist, kann die genannte Lösung, neben einer gesonderten Suche, auch zur

Erweiterung von Diensten auf der Grundlage der MS Active Directory² und zur Speicherung von Daten in kleinen Strukturen³, die hierfür noch gleichrangige Netze nutzen, dienen.

Die Swisscows Company Search enthält Releases, die sich sowohl durch die Hardwareleistung als durch die Begrenzung der Nutzerzahlen unterscheiden.

2 Erstinstallation

2.1 Vollständige Installation

2.1.1 Indexstore

Als Indexspeicher wird Elasticsearch v.2.3.* verwenden. Führen Sie die Installation entsprechend der auf der Internetseite des Herstellers enthaltenen Anleitung (6 - Hilfreiche Links) durch. Elasticsearch kann sowohl auf einem Computer mit Komponenten eines Processing Server als auch auf einem Einzelcomputer installiert werden. Bei der Installation von Elasticsearch auf einem Einzelcomputer kann ein anderes Betriebssystem ausser Windows, jedoch mit JRE (GNU/Linux, Solaris, etc), installiert werden.

Bei besonders hoher Auslastung wird die Nutzung eines Clusters aus mehreren Elasticsearch empfohlen.

Nach der Installation von Elasticsearch entsprechend der Anleitung sind zusätzliche Einstellungen in der Konfiguration von Elasticsearch (Datei elasticsearch.yml) erforderlich. Es wäre von Vorteil, dem Server Elasticsearch einen Cluster-Namen zuzuordnen, der sich vom Standard (cluster-Schlüssel.name) unterscheidet. Dies ist notwendig, um das Risiko einer Beschädigung oder Kompromittierung des Indexes bei der Bereitstellung von anderen Instanzen von Elasticsearch auf dem lokalen Netzwerk zu minimieren.

Achtung! Der Index kann vertrauliche Daten enthalten. Um Lecks solcher Daten zu verhindern, deaktivieren Sie bitte alle TCP/IP-Verbindungen für alle Komponenten ausser Application Server.

Die Leistung von Elasticsearch wird durch die Leistung des Datenträgersubsystems, sowie die Menge an Arbeitsspeicher, dem verfügbaren Puffer unter Elasticsearch und dem Dateicache beeinflusst. Die optimale Menge an Speicherpuffer für Elasticsearch ist abhängig von vielen Parametern, einschliesslich der Art der Daten, aber als Startwert für HES.

Dafür kann man diesen Algorithmus empfehlen: Es sind 25% RAM für den Puffer von Elasticsearch notwendig wenn die Indexstore Installation auf einem separaten Rechner abgelegt ist. Bei der Installation auf einem Rechner mit Processing Server wären 20% RAM notwendig. Aber diese Speichermenge sollte nicht weniger als 3GB und nicht mehr 31GB betragen.

In Linux-Systemen wird die Menge an Arbeitsspeicher, die für Elasticsearch zur Verfügung steht, als ES_HEAP_SIZE eingegeben. Bei der Installation unter Windows als Service, kann die Grösse des Speichers über "Elasticsearch Eigenschaften" GUI eingestellt werden. Es steht unter dem Kommando "service.bat manager" über der Befehlszeile zur Verfügung.

² Links zur Einführung in Active Directory: siehe Abschnitt 6 (Hilfreiche Links).

³ Denken Sie daran, dass ein backup installiert sein muss und regelmässig durchzuführen ist. Allgemeine Informationen über Sicherheitskopien: Siehe Abschnitt 6 (Hilfreiche Links).

2.1.2 Verarbeitungsserver

2.1.2.1 Hardware-Voraussetzungen für HES Processing Server

Component	Minimum	Recommended
Processor Cores	4	>=8
Memory	16 GB	64 GB
Hard disks and available storage space	256 GB	512 GB
Network adapter speed (to filestorage and indexstorage)	1 Gb/s	>=10 Gb/s

2.1.2.2 Software-Voraussetzungen für HES Processing Server

Installieren Sie bitte Windows Server 2012 R2 Standard mit den neuesten Updates und den folgenden Komponenten:

NetFx4ServerFeatures	IIS-RequestFiltering	IIS-ISAPIExtensions	IIS-WebServerManagementTools
NetFx4	IIS-StaticContent	IIS-ISAPIFilter	
NetFx4Extended-ASPNET45	IIS-DefaultDocument	IIS-ASPNET45	IIS-ManagementConsole
IIS-WebServerRole	IIS-DirectoryBrowsing	IIS-HealthAndDiagnostics	WCF-Services45
IIS-WebServer	IIS-HttpErrors	IIS-HttpLogging	WCF-TCP-PortSharing45
IIS-CommonHttpFeatures	IIS-ApplicationDevelopment	IIS-Performance	
IIS-Security	IIS-NetFxExtensibility45	IIS-HttpCompressionStatic	IIS-WebSockets
		IIS-WindowsAuthentication	

Vor der Installation können Sie folgenden Befehl ausführen:

```
Dism /Online /Enable-Feature /FeatureName:NetFx4ServerFeatures  
/FeatureName:NetFx4 /FeatureName:NetFx4Extended-ASPNET45  
/FeatureName:IIS-WebServerRole /FeatureName:IIS-WebServer  
/FeatureName:IIS-CommonHttpFeatures /FeatureName:IIS-Security  
/featurename:IIS-WebSockets /FeatureName:IIS-RequestFiltering  
/FeatureName:IIS-StaticContent /FeatureName:IIS-DefaultDocument  
/FeatureName:IIS-DirectoryBrowsing /FeatureName:IIS-HttpErrors  
/FeatureName:IIS-ApplicationDevelopment /FeatureName:IIS-  
NetFxExtensibility45 /FeatureName:IIS-ISAPIExtensions /FeatureName:IIS-  
ISAPIFilter /FeatureName:IIS-ASPNET45 /FeatureName:IIS-  
HealthAndDiagnostics /FeatureName:IIS-HttpLogging /FeatureName:IIS-  
Performance /FeatureName:IIS-HttpCompressionStatic /FeatureName:IIS-  
WebServerManagementTools /FeatureName:IIS-ManagementConsole  
/FeatureName:WCF-Services45 /FeatureName:WCF-TCP-PortSharing45  
/FeatureName:IIS-WindowsAuthentication /All
```

Auch diese Handlung kann unter Nutzung des Systemapplets "Turn Windows features on or off" vorgenommen werden.

2.1.2.3 IFilters setup

Bitte installieren Sie folgende IFilter, um einen besseren Text-Auszug für Word- und PDF-Dateien zu bekommen:

- MS Office: <http://www.microsoft.com/en-US/download/details.aspx?id=17062> mit Service pack <http://support.microsoft.com/kb/2687447>. Installieren Sie bitte 64-Bit Version.
- PDF iFilter 64 11.0.01: <http://www.adobe.com/support/downloads/detail.jsp?ftpID=5542>.

2.1.2.4 Authentifizierungseinstellungen

Der HES Installer installiert die WEB-Einheit (Benutzeroberfläche und Admin-Panel) als Anwendung (Application) auf die Standard IIS-Website (IIS Default Web Site). Der Administrator kann die Einstellungen der Anwendung ändern, indem er die Standardtools von IIS und Windows verwendet.

Die automatische Anmeldung, die der schnellen Realisierung der Authentifizierung ohne Eingabe von Login und Passwort des aktuellen Windows-Benutzers dient, erfordert die Aktivierung einer entsprechenden IIS-Einstellungen: in dem IIS-Konsole Fenster wählen Sie die Website mit HES (in der Regel die Standard-Website). In dem Bereich Features wählen Sie bitte „Authentication“. Hier ist es notwendig, folgende Punkte zu aktivieren: „Anonymous Authentication“ und „Windows Authentication“. Für „Windows Authentication“ wählen Sie unter „Action“> „Enable“> „Providers...“ den entsprechenden „Negotiate“ Anbieter.

2.1.2.5 HES Auspacken

Das HES Anwendungspaket sieht so aus: **HES.2.2.XX.XXXXXX.zip** (XX ist Ihre spezifische Version). Es enthält folgende Komponenten:

- Connectors
 - Hes.Connectors.Exchange
 - Hes.Connectors.FileSystem
 - Hes.Connectors.Web
- Helpers
 - ConfigTransformationHelper
- Services
 - Hes.Services.ConnectorManager
 - Hes.Services.IndexCleaner
- Utilities
 - HESCoreMock
 - IndexUtil
- Web
 - Ses.Web
- Ses.Setup.*

Entpacken Sie es bitte in demselben Ordner, zum Beispiel C:\HES (d.h. C:\HES\Web\, C:\HES\Services\, C:\HES\Utilities\, usw).

2.1.2.6 IndexUtil und benutzerdefinierte Indexstruktur

Im Verlauf der weiteren Installation erstellt das Installationsprogramm einen Index mit den Standardeinstellungen, die auch weiterhin in den meisten Fällen verwendet werden sollten. Es gibt jedoch Fälle, wo es notwendig ist, nicht ganz der üblichen Konfiguration zu folgen. Dies kann mit Hilfe des Dienstprogramms IndexUtil (aus der Distribution HES) erstellt werden. Das Installationsprogramm kann beim Schritt zum Einrichten des zu verwendenden Indexes selbst feststellen, ob der Index bereits

existiert und bietet eine Auswahl an, ob Sie einen neuen erstellen möchten oder einen bereits vorhandenen verwenden möchten.

HES verwendet 3 Indexer:

Titel index	Inhalt	Standardeinstellungen
{HES}	Text-Inhalte, nach denen gesucht wird.	Shards = 4 Replicas = 1
{HES}_settings	Einstellungen	Shards = shards in HES index Replicas = 1
{HES}_diagnostics	Kernel-Logs HES	Shards = shards in HES index Replicas = 1

In diesem Fall ist {HES} der Name des Indexes, der angegebene Parameter dient für das Dienstprogramm IndexUtil oder während der Installation. Das ist derselbe Name, der im Feld Index Name im Schritt 2 der Installation eingegeben wird. Die Benennung der unterstützenden Indizes wird mit Hilfe von Suffixen konstruiert, die an diesen Namen angefügt werden. Um Informationen über Parameter des Tools zu erhalten, starten Sie ihn mit folgendem Schlüssel:

```
IndexUtil.exe /?
```

Bestimmt die Anzahl der shards & replicas abweichend vom Standard. Es macht in dem Fall Sinn, wenn Elasticsearch Cluster mit der Anzahl der nodes nicht gleich 2 oder 4 verwendet wird. Empfehlungen zur Auswahl dieser Optionen finden Sie auf der Website der Entwickler Elasticsearch (siehe 6 – Hilfreiche Links).

Achtung! Wenn Elasticsearch mehrere nodes beinhaltet, dann vermeiden Sie bitte Änderungen an irgendwelchen Einstellungen im Adminbereich während des Neustarts eines Servers.

2.1.2.7 Ersteinrichtung

Führen Sie bitte das Dienstprogramm ses.setup.exe im Stammordner der entpackten distributive aus.

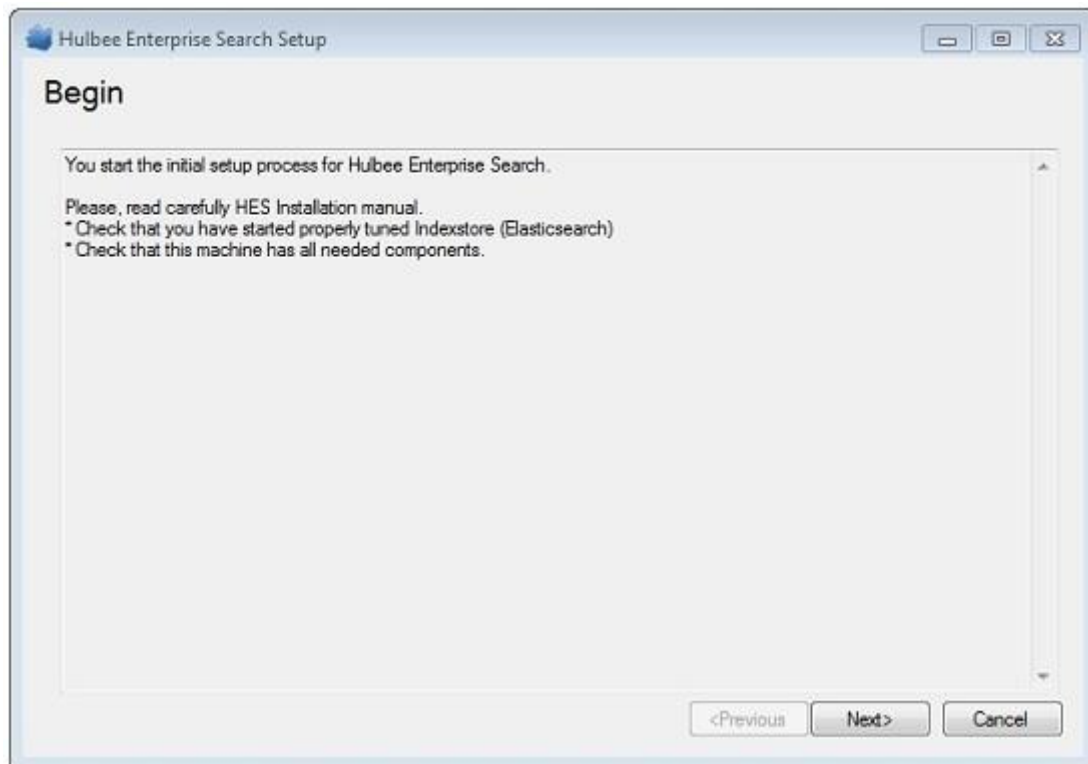


Abb. 1 HES-Installationsassistent – Beginn der Installation.

Klicken Sie bitte auf „Next“, um zum Schritt 1 zu kommen.

1. Schritt 1.

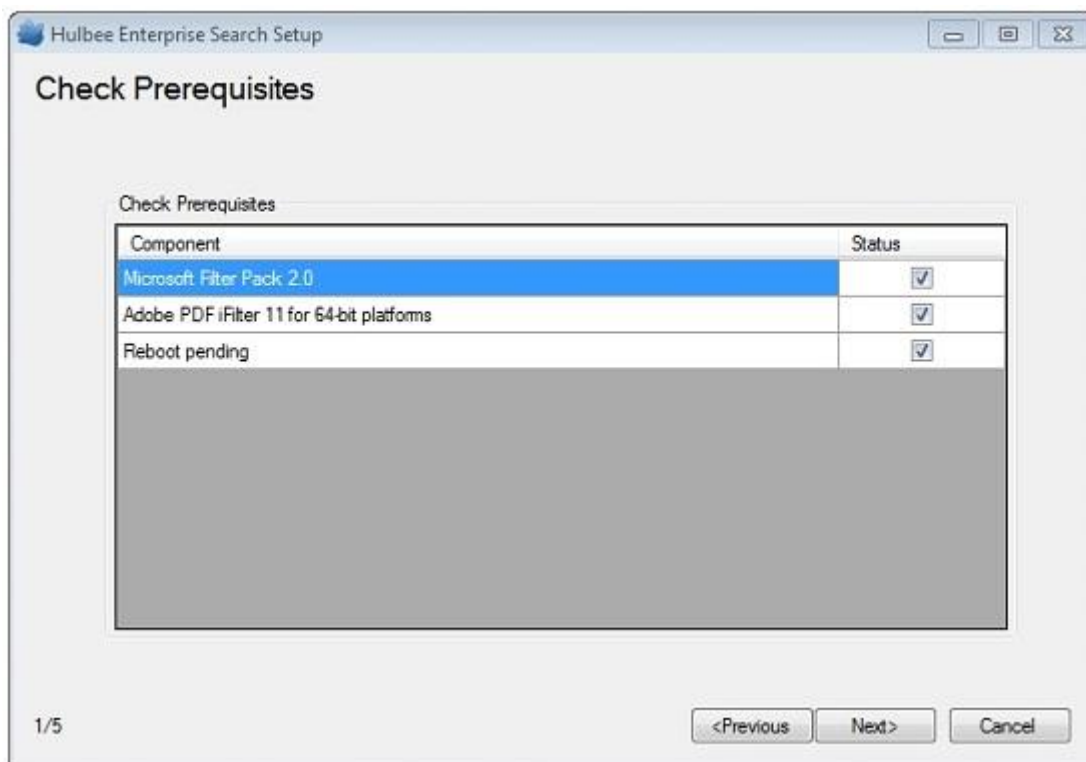
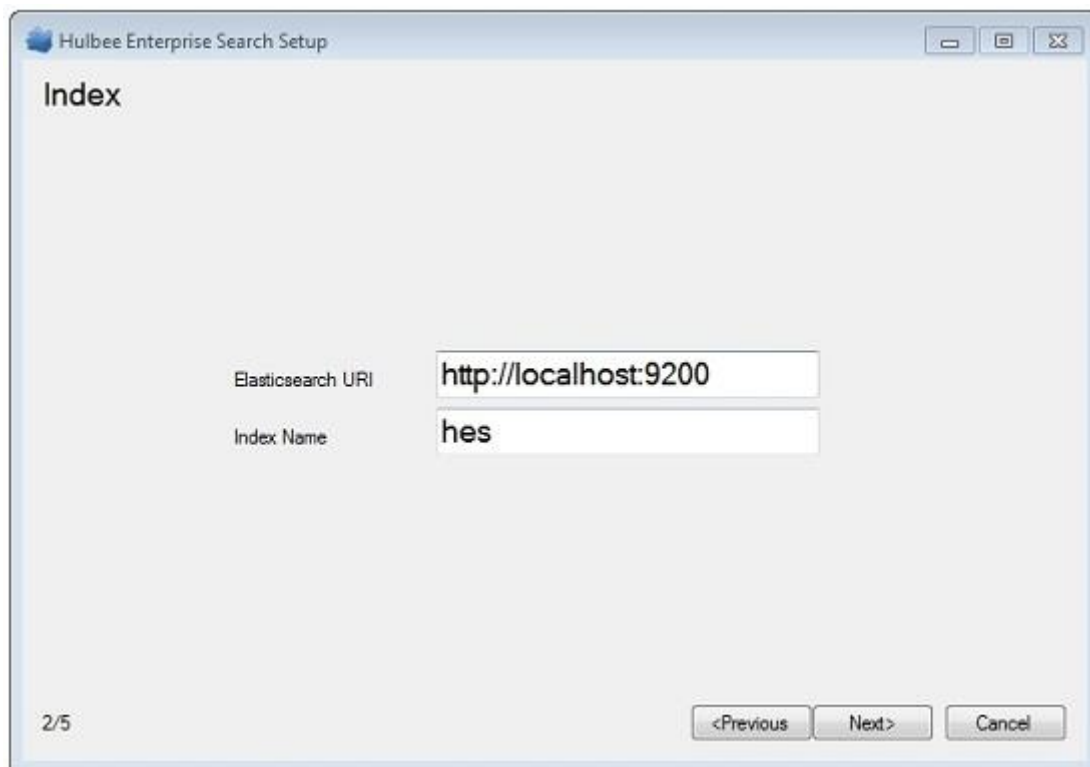


Abb. 2 Schritt 1 der HES-Installationsassistent.

Bei der Prüfung der optionalen Voraussetzungen werden die fehlenden Komponenten angezeigt. Eine Fortsetzung der Installation ist möglich, jedoch kann durch das Fehlen des IFilters die Qualität beim Extrahieren von Text aus verschiedenen Formaten (MS Office und PDF) beeinträchtigt werden.

2. Schritt 2.

Geben Sie bitte die gültige URL zu dem Elasticsearch Server und den Indexnamen ein (Standardwert ist „hes“). URL sollte http-Protokoll-Präfix, die richtige IP-Adresse oder Domain-Namen und Port enthalten.



The screenshot shows a Windows-style dialog box titled "Hulbee Enterprise Search Setup". The main area is titled "Index". It contains two labeled text input fields. The first is labeled "Elasticsearch URI" and contains the text "http://localhost:9200". The second is labeled "Index Name" and contains the text "hes". At the bottom left, it says "2/5". At the bottom right, there are three buttons: "<Previous", "Next>", and "Cancel".

Abb. 3 Schritt 2 der HES-Installationsassistent.

3. Schritt 3.

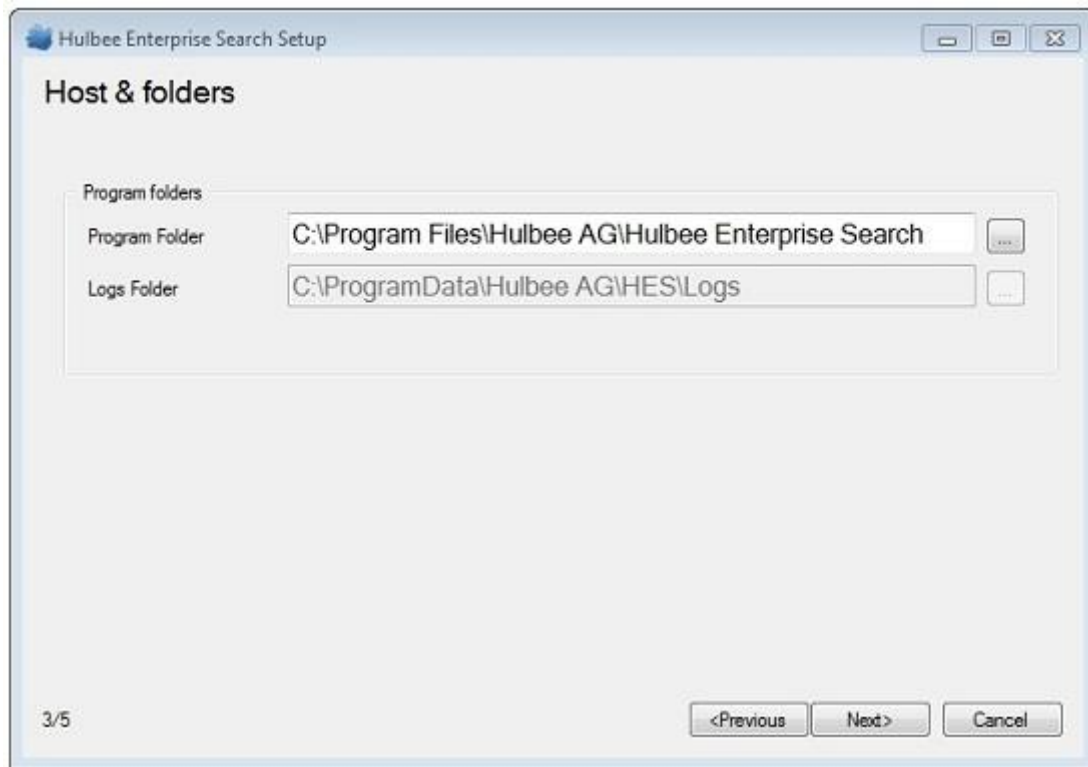


Abb. 4 Schritt 3 der HES-Installationsassistent.

„**Program Folder**“ und „**Logs Folder**“ können mit den Standardwerten belassen werden.

4. Schritt 4.

In diesem Schritt wird das Passwort des HES Administrators eingegeben. Dieser Benutzer ist ein ganz besonderer – es ist kein Active Directory-Benutzer oder eines aktuellen Gerätes, dieser ist notwendig damit die Anmeldung auf der noch nicht konfigurierten HES Software ermöglicht wird. Nur dieser hat Zugriff auf den Admin-Bereich von HES und kann jede administrative Einstellung, wie z.B. die Verbindung zu Active Directory, vornehmen.

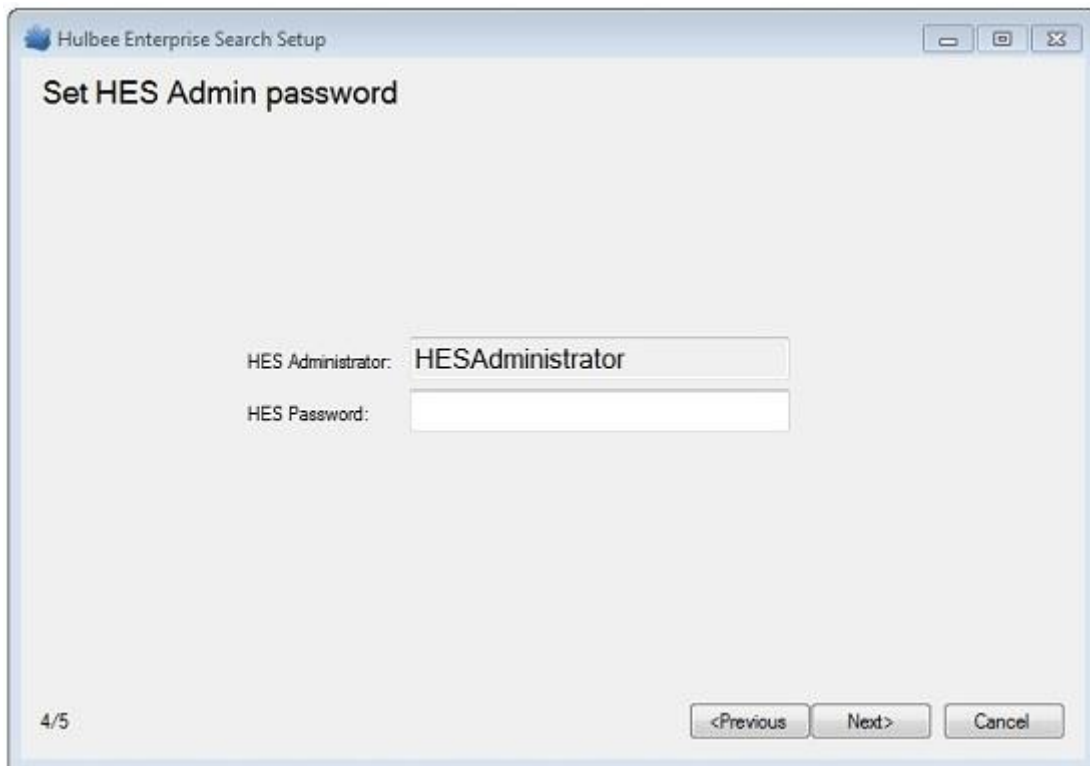


Abb. 5 Schritt 4 der HES-Installationsassistent.

5. Schritt 5.

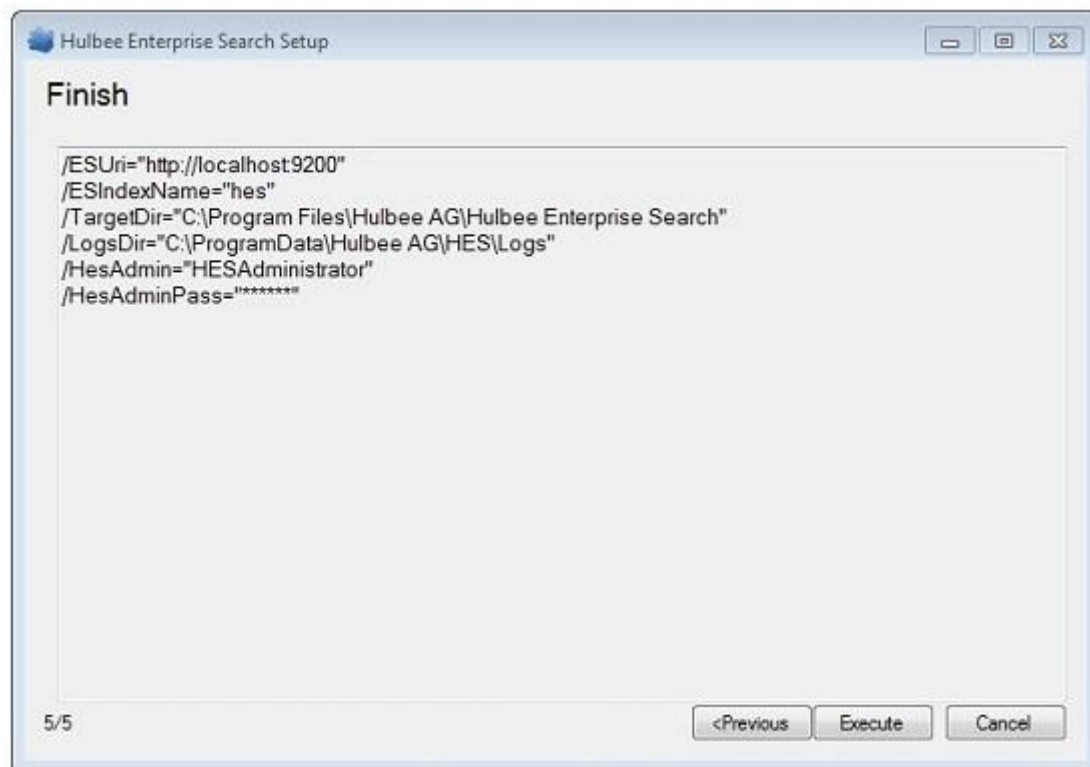


Abb. 6 Schritt 5 der HES-Installationsassistent.

Sie können die ausgewählten Optionen sehen und das Setup mit der Schaltfläche „Execute“ ausführen.

6. Letzter Schritt.

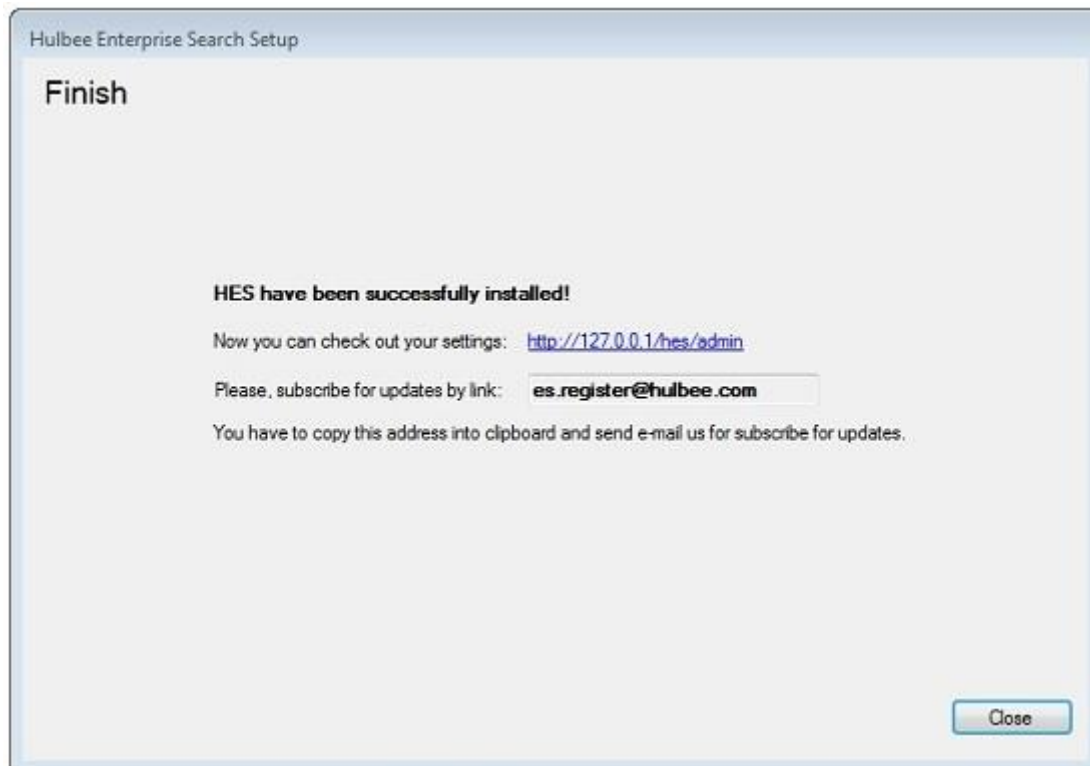


Abb. 7 HES Setup – Setup beendet.

Nach dem erfolgreichen Setup ist es wichtig, den angegebenen Link zu öffnen (Abb. 7), mit demselben Login und Passwort, die in Schritt 4 der Installation für die Ausführung der grundlegenden Einstellungen (Abschnitt „Authorization“ (Kapitel 4.6.1)) und der Connectoren Schnittstelle (Kapitel 4.3) verwendet wurden.

2.1.2.8 Lizenz einrichten

Zusätzlich zu dem Produkt wird eine persönliche Lizenzdatei zur Verfügung gestellt. Wenn die Datei nicht zur Verfügung gestellt wurde, können Sie diese vom Kundenservice support@hulbee.com anfordern. Für die Platzierung der Lizenzdatei wird der folgende Ordner empfohlen: C:\Program Files\Hulbee AG\Hulbee Enterprise Search\Web\Ses.Web\bin (gleicher Ort wo die binären HES site modules installiert sind). Ist die Datei nicht vorhanden, werden Benutzer aus dem Active Directory des Unternehmens für die Verwendung des Produkts nicht zugelassen.

2.2 Swisscows Company Search

Diese Konfiguration kommt mit allen vorinstallierten Komponenten, notwendig für die Arbeit von HES und mit dem installierten aber nicht angepassten HES.

2.2.1 Einrichtung einer Domain

Planen Sie vor allem die Nutzung des bereitzustellenden Appliance Servers. In Abhängigkeit hiervon ist folgende Nutzung möglich:

- Einbindung des Swisscows Company Search Server in eine bestehende Domain.
- Einrichtung einer neuen Domain auf der Grundlage des Dienstes Microsoft Active Directory im Windows Server 2012 R2.
- Verwendung des Impersonalisationsmodus ohne die Einführung von Rechnern in einer Active Directory-Domain.

Einführende Informationen über den Dienst Active Directory: Links in Abschnitt 6 (Hilfreiche Links).

Während der Bereitstellung kann sich der Administrator unter Nutzung des vorinstallierten Bedienerkontos „Administrator“ mit dem Passwort „Admin123“ auf dem Server autorisieren lassen.

Achtung! Achten Sie darauf, das Administrator-Passwort vor der eigentlichen Nutzung des Servers zu ändern.

Verwenden Sie das Standard-Verfahren, um das Windows-Administrator-Passwort zu ändern.

Unternehmen Sie die folgenden Schritte, um das HES Administrator-Passwort zu ändern:

1. Öffnen Sie die Konfigurationsdatei C:\Program Files\Hulbee AG\Hulbee Enterprise Search\Web\Ses.Web\Web.config
2. Passwort ändern bei <add key="SuperUser.Password" value="Admin123" />
3. Datei speichern.

2.2.2 Dateispeicher

Die übliche Grösse des Suchindex beträgt bis zu 10% der binären Grösse eines Dokuments (Richtwert, da diese Grösse im Wesentlichen von der tatsächlichen Datenauswahl abhängig ist). Der Appliance Server kann in diesem Zusammenhang auch als Netzwerkspeicher dienen.

Wenn z.B. vorauszusehen ist, dass der Dateiumfang in den nächsten Jahren nicht mehr als 1 TB beträgt, wenn jedoch der Festplattenspeicher 2 TB beträgt, kann nicht nur die Suche, sondern auch das Speichern von Dokumenten auf dem Appliance Server vollkommen sicher erfolgen. In diesem Fall wird die Belastung des Netzwerks während einer intensiven Indexierung reduziert.

Ist ein höherer Dateiumfang vorauszusehen, sind die Dateien unbedingt auf einem Server zu speichern, der mit einem für diesen Umfang ausreichenden Festplattenspeicher ausgestattet ist, oder es sind zusätzliche Festplatten im Appliance Server zu installieren.

Achtung! Der Dateispeicher muss über ein lokales Netzwerk zugänglich sein und, ebenso wie die übrigen Teile von HES, den Dienst Active Directory nutzen.

3 Erstdurchlauf

Die Adresse, wo die installierte HES Software abgelegt ist, hängt von den administrativen Einstellungen für das lokale Netzwerk ab und an welcher Website in der IIS-Anwendung HES angehängt ist.

Wenn das Öffnen von einem lokalen Rechner durchgeführt wird, so ist es in der Regel:

<http://127.0.0.1/hes> oder <http://localhost/hes>.

Von anderen Rechnern im lokalen Ethernet-Netzwerk ist HES entweder durch die IP des Rechners oder über den Namen im lokalen Netzwerk erreichbar:

<http://hes-server/hes>.

Nach der Anmeldung füllen Sie das Formular zur Autorisierung aus:

Abb. 8 Formular zur Autorisierung.

Achtung! Um HES zu verwalten, müssen Sie sich mit dem gleichen Login/Passwort des Administrators anmelden, die bei der Installation in Schritt 4 angegeben wurden, oder mit denen die im Nachhinein geändert wurden (Kapitel 2.2.1).

Nach der Anmeldung erscheint die Startseite:

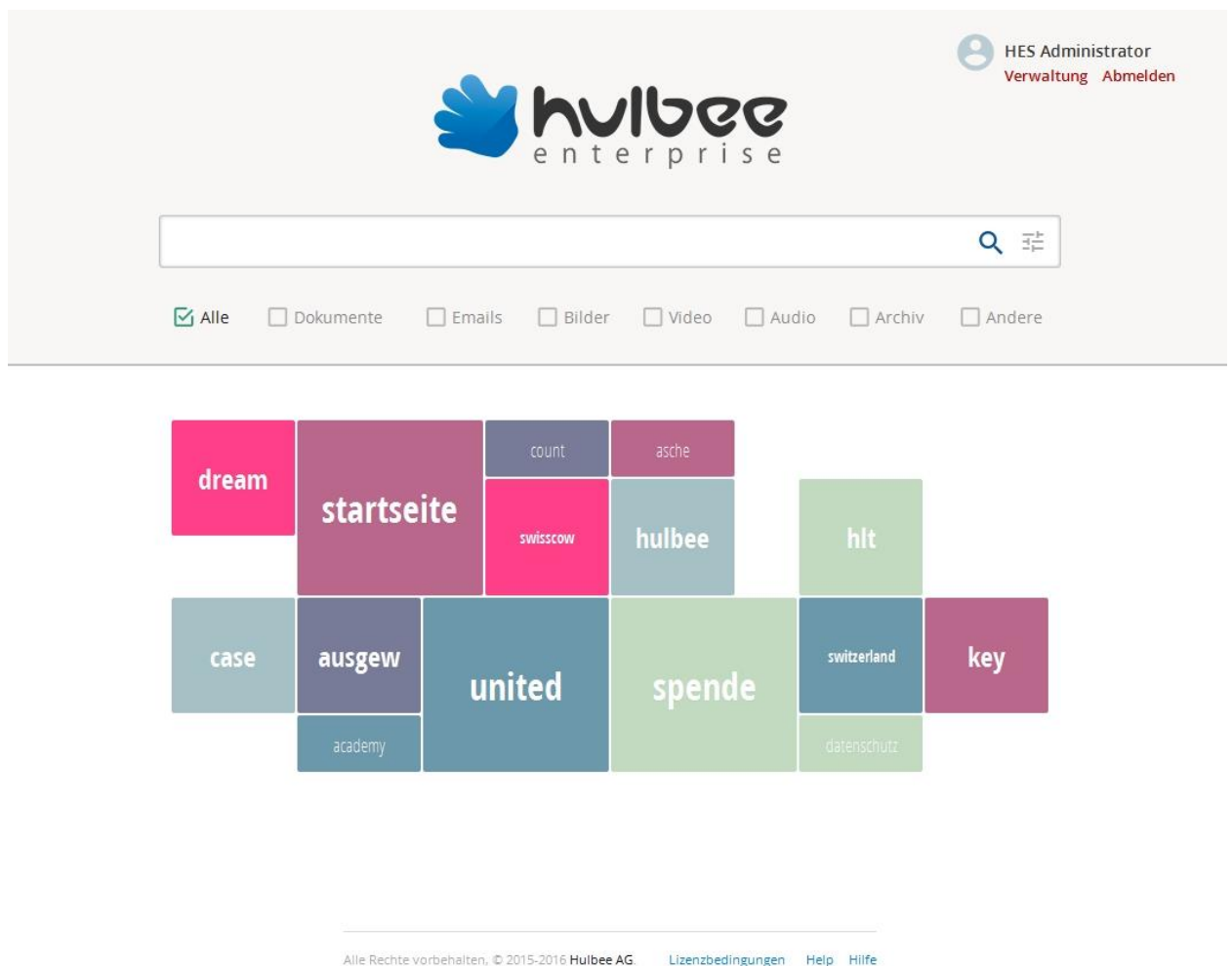


Abb. 9 HES-Startseite.

DataCloud kann auf der Startseite fehlen, solange der Indexer leer ist. Das ist ein normaler Zustand des Systems.

Um in den Administrationsbereich zu gelangen, klicken Sie bitte auf „Administration“.

Achtung! Nach der Installation von HES im Installationsordner, nämlich im Web\Ses.Web, finden Sie die Datei Downloads\Hes.Desktop.Manager.Setup.msi. Diese Dienstprogramm Distribution wird im Benutzerhandbuch (Abschnitt 7.2) beschrieben. Wenn Ihre Sicherheitsrichtlinien verhindern, dass der Benutzer auf seinem Rechner beliebige Software installieren kann, so fügen Sie das Zertifikat aus dieser Distribution der Liste der zugelassenen Installationen hinzu.

3.1 Index Reinigung und Modifizierung

Wenn der Bedarf besteht die Index-Struktur zu modifizieren oder diese zu reinigen (was mit dem Befehl Neuerstellung oder Index löschen funktioniert), kann das Tool IndexUtil verwendet werden. Mit Hilfe des IndexUtil kann der Index neu erstellt werden, falls dieser beschädigt wurde.

Achtung! Vorher sollten alle Services und Connectoren von HES im System-Applet Services beendet werden.

Die Benutzung vom Dienstprogramm IndexUtil für diese Optionen ist die gleiche wie im Abschnitt 2.1.2.6 beschrieben wird. Nach Neuerstellung wird der Index (einschliesslich der Einstellungen) gelöscht und das System wird in den gleichen Zustand gebracht, wie es nach der Installation war (mit Ausnahme der Einstellungen, die in den Konfigurationsdateien abgespeichert wurden). Dementsprechend sollten Sie wieder Einstellungen vornehmen, die im Abschnitt 4 beschrieben werden.

Nach der Neuerstellung des Indexes sollten die HES Dienste neu gestartet werden.

4 Einrichtung der Admin Seite

Die übrigen Setup-Schritte können im Admin-Panel durchgeführt werden. Bitte folgen Sie dem nächsten Link:

<http://enterprise-search.company.com/hes/admin>

Geben Sie anstelle von **enterprise-search.company.com** den Link ein, der für den Installationsschritt 3 genutzt wurde oder folgen Sie dem „Administration“ Link auf der HES-Startseite.

4.1 Wiederherstellungsoptionen

Wenn Sie bereits ein abgestimmtes Exemplar von HES installiert haben, können Sie die zuvor gesicherten Einstellungen wiederherstellen. Bitte, wählen Sie „Exportieren / Importieren“ (Kapitel 4.6.5) und:

1. Klicken Sie auf „Browse...“ und wählen Sie die Sicherungsdatei aus.
2. Klicken Sie auf „Importieren“.

Dieser Vorgang ermöglicht die Wiederherstellung globaler und persönlicher Benutzer-Einstellungen. Das heisst, dass es sich um die Daten handelt, die zu sehen sind und mithilfe der Administratorleiste und mithilfe des Userboards bearbeitet werden. Der Suchindex ist nicht Bestandteil des Einrichtens – er wird mithilfe des Hes.Services.ConnectorManager erneut aktualisiert.

4.2 Erste Schritte der Einstellungen

Bei der Ersteinrichtung werden folgende Schritte empfohlen:

1. Verbindung zum Domain – Autorisierungs-Kontext (Kapitel 4.6.1). Geben Sie diese Einstellungen ein, bevor weitere Schritte durchgeführt werden!
2. Hinzufügen der Connectoren (Kapitel 4.3)
3. Hinzufügen von Speicherplätzen (Kapitel 4.4).

4.3 Registrierung der Connectoren

Bei der ersten Anmeldung sind in der Liste keine Connectoren vorhanden. HES beinhaltet mehrere Typen von Connectoren: für die Verbindung mit dem Dateisystem des Unternehmens, mit den Web-Ressourcen und so weiter. Es gibt die Möglichkeit, eigene individuelle Connectoren unter der Verwendung des HES-Connectors API (6 - Hilfreiche Links) hinzuzufügen.

Für die Verbindung der Connectoren führen Sie folgende Schritte aus:

1. Im Admin-Bereich auf der Seite „Anschlüsse“ geben Sie die Adresse des Connectors ein. Die Adresse von Standard - Connectoren können Sie der Beschreibung des entsprechenden Connectors (5.2 - 5.4) entnehmen. Der Port, auf dem die Connectoren registriert sind, muss bei dem Entwickler angefragt werden.
2. Klicken Sie auf die Schaltfläche „Anmelden“.

Es ist nicht notwendig, mehrere Connectoren des gleichen Typs bei Standard Lieferung zu registrieren.

Achtung! Überprüfen Sie, ob der Connector - Dienst während der Herstellung der Verbindung gestartet wurde.

4.4 Hinzufügen von Speicherplätzen

Sobald zu der HES Software notwendige Connectoren hinzugefügt worden sind, müssen weitere Speicher, wo die Dokumente ablegt sind, hinzugefügt werden. Speicher ist ein Platz wo Dokumente desselben Typs abgespeichert sind. Speicher können sein: Netzwerk-Dateifreigabe, Webseite, CRM System, ERP, Mail-Server, etc. Ein und derselbe Connector kann mit mehreren Speichern des jeweiligen Typs verbunden. Dabei können sich die Dokumente mit demselben uri nur in einem einzigen Speicherplatz des Connectors befinden.

Um Speicher hinzuzufügen, wählen Sie „Details“ im Bereich „Anschlüsse“ aus.

Anschlüsse

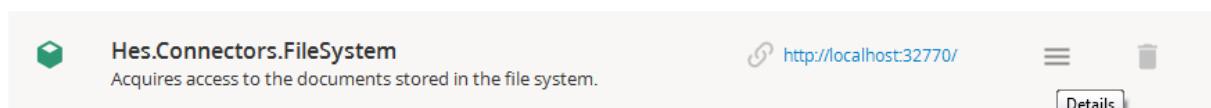


Abb. 10 Bereich „Anschlüsse“, Schaltfläche „Details“.

Als Nächstes wählen Sie „Settings“ aus, wo auch Vorgänge wie hinzufügen, löschen, konfigurieren von Speicherplätzen durchgeführt werden können. Ausführlicher werden diese Vorgänge in den Ziffern 5.2 - 5.4 beschrieben.

4.5 Administrationsbereich

Die Administratorleiste besteht aus einem Navigations- und einem Arbeitsbereich. Der Arbeitsbereich enthält Elemente für die Bearbeitung von Einstellungen und für die Anzeige von verschiedenen Informationen über den Betrieb des Systems.

The screenshot displays the Hulbee Administrator interface. At the top, the Hulbee logo is on the left, a search bar in the center, and the user 'HES Administrator' with 'Verwaltung' and 'Abmelden' links on the right. The left sidebar contains a navigation menu with sections: 'Dashboard' (containing Dashboard, Anschlüsse, Nutzer, Protokolle with a red '23' badge, and Unverarbeitet), 'Einstellungen' (containing Authorization, Verarbeitung, Löschen, Logging, and Importieren / Exporti...), and 'DataCloud' (containing Benutzerdefinierte S... and Stoppwörter). The main content area is titled 'Administrationsbereich' and features three columns of widgets. The first column, 'INDEX', shows 'Indexierte Dokumente' (1'296), 'Größe der Dokumente' (2.87 GB), and 'Index Größe' (88.61 MB). The second column, 'LIZENZ', shows 'Lizenz' (TypeM), 'Max. Benutzer' (20), and 'Erfasste Benutzer' (0), with a 'BENUTZER VERWALTEN >' link. The third column, 'ANSCHLÜSSE', shows 'Registriert' (2) and 'Aktiv' (2), with an 'ANSCHLÜSSE VERWALTEN >' link. Below these is a 'DIENSTE' (Services) section showing 'Hes.Services.Connector...' and 'Hes.Services.IndexCleaner' both in 'Running' status.

Abb. 11 Administratorsite. Allgemeine Ansicht.

In den Screenshots werden lediglich die Bereiche dargestellt, die beim Wechsel zwischen den unterschiedlichen Links der Administratorenleiste geändert werden.

4.5.1 Dashboard

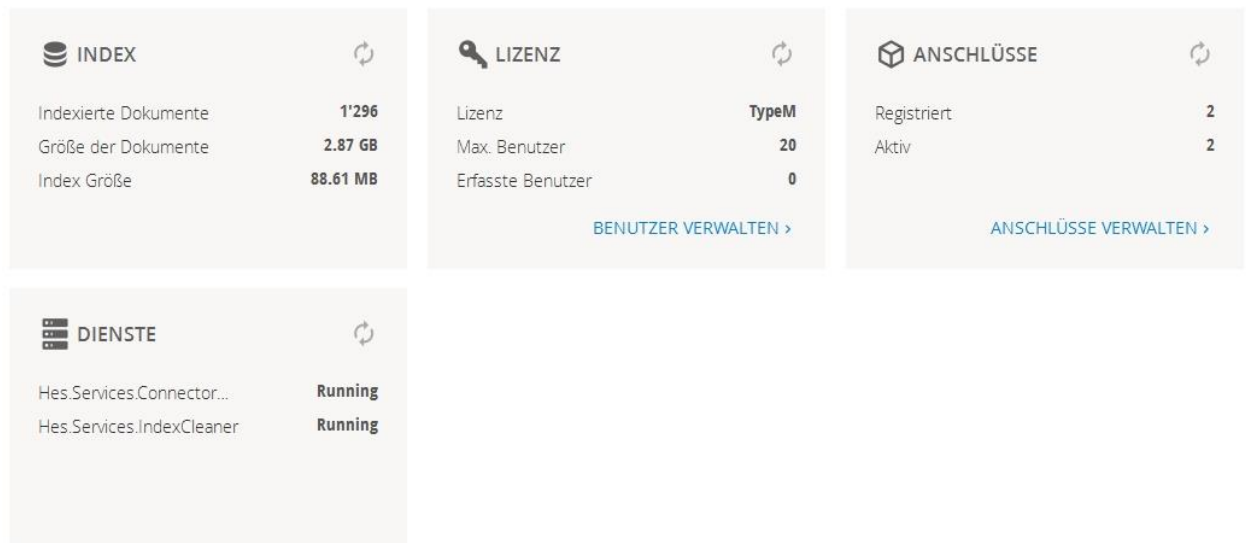










Abb. 12 Dashboard.

Auf dem Dashboard sind verschiedene Statistiken dargestellt:

- INDEXED DOCUMENTS. Anzahl der indexierten Dokumente. Diese Zahl kann nicht mit der Gesamtzahl der im Speicher vorhandenen Dateien übereinstimmen, da ein Teil der Dokumente aufgrund verschiedener Filter (Grösse, Erweiterungen) nicht bearbeitet werden kann. Gleichzeitig kann in den Archiven und Mitteilungen mehr als eine Datei enthalten sein.
- DOCUMENTS SIZE. Gesamtgrösse der im Index vorhandenen Dokumente. Hier ist die Summe der Binärgrössen gemeint. Auch dieser Wert weicht bei dem von der Datei benötigten Speicherplatz ebenfalls aus den vorgenannten Gründen ab.
- INDEX SIZE. Indexgrösse. Elasticsearch auf der Festplatte.
- LICENSE {id}. In diesem Widget ist der Lizenztyp angegeben. Anzahl der Nutzer, die den Server bereits genutzt haben und Anzahl der potenziellen Nutzer.
- CONNECTORS. Informationen zu Connectoren die im System registriert sind.
- Servicewidgets. Mithilfe dieser Widgets kann man sich vom einwandfreien Crawling und vom einwandfreien Indexcleaning überzeugen. Sie ermöglichen ausserdem einen Neustart nach Änderung der Einstellungen.

4.5.2 Anschlüsse

Anschlüsse

	Hes.Connectors.FileSystem Acquires access to the documents stored in the file system.	 http://localhost:32770/		
	Hes.Connectors.Web Acquires access to the documents stored in the World Wide Web.	 http://localhost:32769/		

Anschluss registrieren

Endpoint

Anmelden 

Abb. 13 Anschlüsse Seite.

„Anschlüsse“ zeigt Informationen zu Connectoren die im System registriert sind. Enthält ebenfalls den Button „Details“, wo Sie zu den Einstellungen des entsprechenden Connectors gelangen können.

Die Arbeit mit Connectoren erfolgt über http-Protokoll. Das Feld „**Endpoint**“ dient der Eingabe der Server-url eines bestimmten Connectors, die der Konfigurationsdatei entnommen werden kann.

Vorhandene Connectoren können entfernt oder neue registriert werden (Kapitel 4.3).

4.5.3 Nutzer

< > Seite 1 von 1

Nutzer



	BENUTZERNAME ▲	STATUS	
---	----------------	--------	---

Abb. 14 Nutzer Seite.

In der Benutzerseite kann man sehen, wer als Nutzer im System bereits autorisiert ist. Hier kann für einen Nutzer die HES-Suche HES auch verweigert werden. In diesem Fall bleibt der betreffende Nutzer bei der Prüfung der Richtigkeit der Anzahl der Lizenznutzer unberücksichtigt. Bei der Meldung „Zugriff verweigert“ bleibt dem Nutzer und Administrator nur der Zugang zur Administratorenleiste.

4.5.4 Protokolle

Seite 1 von 10 Kritisch, Fehler, Warnung, Informatio... Hes.Services.ConnectorManager, Hes....			
ZEITSTEMPEL	VERANSTALTUNG	MELDUNG	PROZESS
Heute, 12:22:07	Stop	Finished processing of the feed `archives` (Hes.Connectors.FileSystem).	Hes.Services.ConnectorManager
Heute, 12:21:57	Stop	Finished processing of the feed `smoke` (Hes.Connectors.FileSystem).	Hes.Services.ConnectorManager
Heute, 12:21:54	Stop	Finished processing of the feed `sw` (Hes.Connectors.Web).	Hes.Services.ConnectorManager
Heute, 12:21:54	Stop	Finished processing of the feed `hulbee` (Hes.Connectors.Web).	Hes.Services.ConnectorManager
Heute, 12:21:50	Stop	Finished processing of the feed `size` (Hes.Connectors.FileSystem).	Hes.Services.ConnectorManager
Heute, 12:21:48	Start	Started processing of the feed `sw` (Hes.Connectors.Web).	Hes.Services.ConnectorManager

Abb. 15 Protokolle Seite.

Diese Seite zeigt Fehlermeldungen, Warnmeldungen und andere Informationen. Sie können Ereignistyp oder Dienstleistungen durch Ankreuzen entsprechender Elemente in Dropdown-Liste Ereignisse filtern.

4.5.5 Unverarbeitet

Seite 1 von 2					
Unverarbeitete Dokumente					
<i>Die Liste der Dokumente, die aufgrund einiger Fehler nicht verarbeitet wurden. Sie einige können Elemente abziehen, um zu versuchen, sie erneut zu verarbeiten oder einen Zeitplan für die Verarbeitung mit bestimmten Grenzen für die äußerst wichtige Dokumente, zu erstellen.</i>					
<input type="checkbox"/>	URI	STATUS	FEHLER		
<input type="checkbox"/>	file://storage/files/QA/TEST-DATA/Archives/gz/audio/wav.GZ	Aborted	File is to large to be processed.		
<input type="checkbox"/>	file://storage/files/QA/TEST-DATA/Archives/tar/TEST-DATA-Smoke.tar	Aborted	File is to large to be processed.		
<input type="checkbox"/>	file://storage/files/QA/TEST-DATA/Archives/tar/audio/wav.TAR	Aborted	File is to large to be processed.		

Abb. 16 Unverarbeitet Seite.

Die Liste der Dokumente, die aufgrund einiger Fehler nicht verarbeitet wurden.

Verarbeitung von Dateien ist mit dem Wert „Watchdog timeout“ (siehe 4.6.1) begrenzt. Wenn auf Grund der Zeitüberschreitung das Dokument in die Liste der unbearbeiteten Dateien geraten ist, kann die Verarbeitung für äußerst wichtige Dokumente - mit erweiterter Grenzzeit (maximal 30 Minuten) geplant werden. Auch auf der Seite der "unverarbeitet" kann durch einige Elemente versucht werden, sie erneut zu verarbeiten.

4.6 Einstellungen

4.6.1 Authorization

Berechtigungskontext

Die Art des Speichers für den Hauptkontext (Server oder Domain), gegen die alle Anfragen durchgeführt werden.

Domain (Active Directory Domain Services store) ▼

Domainname

Der Name des Domains oder des Servers.

IhreFirma.com

NetBIOS Domainname

NetBIOS Domainname (standardmäßig, die linke Markierung im DNS-Domainnamen bis zu den ersten 15 Bytes)

Container

Der Container auf dem Speicher als Lösung des Kontext zu verwenden.

CN=Benutzer, DC=IhreFirma, DC=com

Benutzername

Der Benutzername wird für eine Verbindung mit dem Store verwendet.

Passwort

Das Passwort wird für eine Verbindung mit dem Store verwendet.

Zugriffsregeln

Gewährt oder verweigert den Zugriff auf HES-Anwendung für bestimmte Benutzer und Gruppen.

 IDENTITÄT	ZUGANG	
 Benutzer oder Gruppe SAMAccountName, z.B. DOMAIN\	Zulassen	▼

Einstellungen speichern



Abb. 17 Authorization Seite.

Hier sollten Daten für den Anschluss von HES an Active Directory eingefügt werden. Weiter wird dieser Server für die Differenzierung der Nutzerdaten und deren Berechtigungen für den Zugriff auf Dokumente verwendet.

Die „Domain“ Option muss als **„Berechtigungskontext“** ausgewählt werden. Berechtigungskontext ist der Server oder die Domäne, an dem alle Operationen durchgeführt werden. Es ist der Behälter, der als

Basis für diese Operationen verwendet wird, und die Anmeldeinformationen, die verwendet werden, um die Operationen durchzuführen.

Auch steht die „Machine“ Option als Testmodus zur Verfügung. Bei der Auswahl der „Machine“ Autorisation erfolgt ein Abruf an die lokale Datenbank der Benutzerkonten des Computers, auf dem sich HES befindet. Dies kann in kleineren Netzwerken, die Active Directory nicht nutzen oder zu Test-Zwecken verwendet werden. Normalerweise wird die Option nicht in herkömmlichen Szenarien verwendet.

Wenn der Rechner, auf dem der Processing Server installiert ist, nicht in die Active Directory-Domain eingefügt ist, kann es zu Problemen bei der Ermittlung der IP-Adresse eines Domain Controllers führen, wenn dieser in symbolischer Form eingetragen wird. Um dieses Problem zu lösen, gehen Sie bitte wie folgt vor:

1. Wählen Sie bitte Open Network und Sharing Center > Change adapter settings.
2. Im Menü Einstellungen öffnen Sie das Fenster Internet Protocol Version 4 (TCP/IPv4).
3. Als „Preferred DNS server“ fügen Sie die IP-Adresse des Servers ein, auf dem der DNS-Dienst der Ziel-Domain bereitgestellt wird.

Als nächstes sollten die Felder der Autorisierung ausgefüllt werden:

- **„Domain name“:** Domain-Namen-Verzeichnisdienst
- **„NetBIOS Domain Name“:** NetBIOS-Namen-Verzeichnisdienst.
- **„Container“:** Deren Daten werden in den Feldern „Username“ und „Password“ angezeigt. Dieses Feld ist optional. Der Standardwert für dieses Feld ist ein leerer String. Dieses Feld kann in dem Fall leer sein, wenn es notwendig ist, den Zugriff auf gewisse Gruppen/Nutzer zu beschränken oder wenn es mehrere Kataloge in der Gesamtstruktur des Active Directory gibt. Beispiel des Pfades: „CN=Users,DC=hes,DC=hulbee,DC=com“.
- **„Access Rules“:** In diesem Bereich können die Benutzer (Gruppen) exakt festgelegt werden, denen der Zugang zu HES gewährt oder verweigert wird. Wenn das Feld leer bleibt, gilt der Zugang für alle Benutzer der Domain.

In die Felder **„Username“** und **„Password“** können nicht nur die Daten des Administrators eingetragen werden, sondern auch beliebige Benutzer aus dem Active Directory, die über die entsprechenden Zugriffsrechte verfügen.

4.6.2 Verarbeitung

Anzahl der Parallelarbeiten

Die maximale Anzahl der gleichzeitigen Operationen, die von der Verarbeitungsmaschine ausgeführt werden können.

Watchdog timeout

Wartezeit für Ergebnisse vom Worker, bevor die Verarbeitungsaufgabe ausgeschaltet wird. Der Wert muss im Format "HH:MM:SS" sein, z.B. 00:05:00 - fünf Minuten.

Leerlauf

Wartezeit für die Verarbeitungsaufgabe vom Anschluss, bevor der Worker ausgeschaltet wird. Der Wert muss im Format "HH:MM:SS" sein, z.B. 00:05:00 - fünf Minuten.

Der maximal zu verarbeiteten Grenzwert

Die Dokumente, deren Grösse in Byte der angegebene Wert überschreitet, werden nicht bearbeitet.

Die maximale extrahierte Größe

Die maximale Grösse der Inhalt in Bytes, die aus dem Dokument extrahiert werden kann.

Die maximale Größe des Anhangs

Die maximale Grösse der Dokumentanhänge in Bytes.

Abb. 18 Verarbeitung Seite.

Wählen Sie "Verarbeitung" und füllen Sie folgende Felder aus:

- **"Concurrent workers count"**: Um eine bessere Verarbeitungsgeschwindigkeit zu erhalten, können Sie hier eine Zahl eingeben, bis zur Anzahl der CPU-Kerne. Bei der Konfiguration mit Elasticsearch, was sich auf einem anderen Server befindet, ist es empfehlenswert, die Zahl der Arbeitnehmer in Übereinstimmung mit der Anzahl der Kerne im Rechner mit Elasticsearch zuzuweisen. Aber diese Zahl sollte nicht größer sein als die Anzahl der Kerne in dem Verarbeitungsserver. Für die Konfiguration von Swisscows Company Search und in den Fällen, wenn sich alle Module auf einem Rechner befinden, ist es empfehlenswert, die Anzahl der Beschäftigten bis zur Hälfte der Anzahl der CPU-Kerne einzustellen.
- **"Watchdog timeout"**: Zeit um auf die Ergebnisse vom Arbeiter zu warten, bevor die Verarbeitungsaufgabe abgebrochen wird. Stellen Sie einen Zeit-Basiswert ein (er sollte für die meisten Dokumente des Speichers ausreichen, z.B. 15...30 Sekunden). Er entspricht der Dauer des ersten Zeitabschnitts für die Verarbeitung des Dokuments. Es kann durchaus ein wenig Zeit in Anspruch nehmen bis der erste Indizierungszyklus alle Dokumente schnell verarbeitet hat. Wenn ein Dokument nicht in der angegebenen Zeit verarbeitet werden konnte, erfolgt eine gezwungene Unterbrechung und das nächste Dokument wird bearbeitet. Im nächsten Durchgang wird bei Dokumenten, deren Verarbeitung durch ein Timeout beendet wurde, die

Zeit automatisch um das Dreifache vergrößert. Beim dritten Versuch – um das Zwölfwache. Das ermöglicht, große Dateien (Archive, große Dokumente) zu verarbeiten, die nicht während der ersten beiden Zyklen bearbeitet wurden. Weitere Versuche werden nicht unternommen. Die unbearbeiteten Dokumente können in einer Vorschau auf der Seite „Unprocessed“ (Kapitel 4.5.5) angesehen werden.

- **“Idle timeout”**: Zeit um auf die Aufgaben aus dem Connector-Manager zu warten, bevor der Arbeiter abgemeldet wird.
- **“Maximum document size”**: Dokumente, deren Größe den angegebenen Wert in Byte überschreitet, werden nicht bearbeitet. Metadaten werden aus diesen Dokumenten extrahiert, doch diese werden in die Liste der unbearbeiteten Dateien hinzugefügt (siehe 4.5.5).
- **“Maximum extracting content size”**: maximale Größe von Text in Symbolen, der weiterverarbeitet und indiziert wird. Textextraktion aus dem Dokument ist gemeint, und nicht die binäre Dateigröße. Der restliche Text wird abgeschnitten.
- **“Maximum attachments size”**: maximale Größe der angehängten Dateien. Ist diese Größe erreicht, endet die Verarbeitung der Archiv- oder Emailnachricht.

Filter anhand der Dateiendungen

Extraktionsmethode des Inhaltes, die von der Dokumenterweiterung abhängig ist.

<input type="checkbox"/>	DATEIENDUNGEN	FILTER		
<input type="checkbox"/>	pdf	ifilter dann nativev	▼	
<input type="checkbox"/>	docx;doc;docm	ifilter dann nativev	▼	
<input type="checkbox"/>	ppt; pptx	ifilter dann nativev	▼	
<input type="checkbox"/>	xls;xlsx	ifilter dann nativev	▼	
<input type="checkbox"/>	one	ifilter dann nativev	▼	
<input type="checkbox"/>	csv	nur native	▼	
+	Geben Sie eine Dateierweiterung ein	ifilter dann nativev	▼	

Die maximal zu verarbeiteten Grenzwerten anhand der Dateiendungen

Grenze der Dokumenten Größe je nach Erweiterung.

<input type="checkbox"/>	DATEIENDUNGEN	GRENZWERT	
<input type="checkbox"/>	avi;wmv;mp4;mkv	8589934592	
+	Geben Sie eine Dateierweiterung ein	107374182400	

Einstellungen speichern

Abb. 19 Verarbeitung Seite, Abschluss.

- **“Extraction method by extension”**: Auswahl des Konverters, abhängig von der Erweiterung. In der Regel ist iFilter langsamer, aber gültig für Office-Dokumente und PDF-Dokumente.
- **“Processing limits by extensions”**: einige Dateitypen sind groß und fallen unter Größenbeschränkung. Aber ihre Verarbeitung ist sehr einfach, weil nur ein kleiner Teil der Daten verarbeitet wird. Es bezieht sich in erster Linie auf Mediendateien "avi, wmv, mp4, mkv". Zu diesem Zweck können Sie einzelne Einschränkungen für diese Gruppe von Dateien festlegen. Sie können sie in einer Zeile eingeben, und Erweiterungen durch “;” Symbol trennen.

Nachdem diese Optionen geändert wurden, klicken Sie auf „Einstellungen speichern“.

Achtung! Vor dem Hinzufügen von Speicherplätzen sollten die „Verarbeitung“ – Einstellungen konfiguriert werden, da für deren Anwendung z.T. eine neue Indexierung des Speichers benötigt wird.

4.6.3 Löschen

Losgröße

Die Anzahl der Dokumente, die pro eine Iteration verarbeitet werden.

Iteration Zeitüberschreitung

Die Zeitüberschreitung zwischen den Iterationen (in Format "hh:mm:ss" 00:05:00 - fünf Minuten).

Einstellungen speichern ✓

Abb. 20 Löschen Seite.

Bitte wählen Sie "Löschen" und passen Sie die folgenden Felder an:

- **"Batch size"**: Die ungefähre Anzahl der Dokumente, die in einem Schritt gescannt werden. Der wahre Wert jedes Abschnitts kann abhängig von der Anzahl der Dokumente in dem Index variieren. Es wird der Wert 1000 empfohlen.
- **"Iteration timeout"**: Pause bei der Verarbeitung der einzelnen Dateiteile. In den meisten Fällen beträgt diese Pause 10 Sekunden. Hierbei handelt es sich um eine angemessene Zeitspanne. Geht es jedoch um die Verarbeitung von Millionen oder zehn Millionen Einheiten, kann diese Zeitspanne verringert werden, um eine schnellere Löschung von Dokumenten, die aus dem Index gelöscht wurden (oder deren Speicherpfad geändert wurde), zu erreichen (jedoch wird die Belastung des Processing Servers und des Indexstores erhöht).

Nach Änderung dieser Optionen, klicken Sie bitte auf „Einstellungen speichern“.

4.6.4 Logging

Aufzeichnungstiefe

Information

Entfernt automatisch Protokolle älter als

Woche

Einstellungen speichern ✓

Event types

Critical, Error, Warning, Information, Verbose, Start, Stop

Vor

06.06.2016

Löschen

Abb. 21 Logging Seite.

Die Seite Logging ermöglicht folgende Einstellungen:

- **“Trace level”**: Hier öffnet sich eine Liste, in der eingestellt werden kann, welche Ereignisse im Systemjournal aufgezeichnet werden sollen.
- **“Automatically remove logs older than”**: Hier kann die Anzahl der gespeicherten HES-Logs festgelegt werden. Das Cleaning erfolgt sowohl im IndexStore (diese werden in den Administratoreinstellungen angegeben), als auch in Form von Textdateien auf einem Computer mit Processing Server (üblicher Speicherort: c:\hes\logs). Im Laufe der Zeit kann hier ein grosses Volumen angehäuft werden. Deshalb wird empfohlen, die genannte Option nicht abzuschalten, ausser in bestimmten Fällen. Das Cleaning der Logs erfolgt durch den IndexCleaner. Wird der Speicherzeitraum geändert, ist ein Neustart des Programms erforderlich.

Sie können auch einige Ereignisse („**Event Types**“) aus dem Indexstore manuell entfernen. Dazu müssen Sie die Arten von Veranstaltungen wählen, die entfernt werden sollen und das Datum („**Before date**“), bis zu welchem sie entfernt werden sollen. Klicken Sie dann auf „Delete“.

4.6.5 Importieren / Exportieren

Exportieren

Sicherung der allgemeinen und persönlichen Einstellungen für die aktuelle Installation.



Importieren

Hier können Sie die allgemeinen und persönlichen Einstellungen in dem Zustand wiederherstellen, den sie zu einem früheren Zeitpunkt hatten. Dafür nutzen Sie bitte die Datei, die mit der Exportfunktion erstellt wurde.

No file selected.



Abb. 22 Importieren / Exportieren Seite.

Die Importierer / Exportierer Seite enthält Befehle, die eine Speicherung der HES-Einstellungen in einer Datei und die Wiederherstellung derselben aus der Datei heraus ermöglichen. Die HES-Einstellungen können sowohl als Administratoreinstellungen als auch als Einstellungen, die vom Nutzer im persönlichen Bereich eingegeben werden, gespeichert werden. Die genannten Befehle können bei Abstürzen oder bei der Installation von Updates angewendet werden.

Beim Importieren von Einstellungen müssen Connectoren manuell registriert werden.

4.6.6 Vorschau

Auf dieser Seite können Einstellungen für die korrekte Anzeige der Vorschau von bestimmten Arten von Dokumenten (HTML, PDF, MS Word und so weiter) eingegeben werden.

Unter „Preview“ können folgende Einstellungen angepasst werden:

- **“Max. document size”:** Die maximale Grösse des Dokuments, für die die Vorschau noch verfügbar ist. Wenn die Grösse des Dokuments die vorgegebene Grösse überschreitet, wird Register weiterhin zur Verfügung gestellt, jedoch das Dokument dort nicht angezeigt. Der Benutzer erhält eine entsprechende Meldung. Die Eingrenzung gilt für alle Dateiformate, aber hauptsächlich für grössere Textdokumente, die nicht als einzelne Abschnitte sondern als gesamte Datei hochgeladen werden.
- **“Per-domain preview settings”:** Im Feld DOMAIN wird die Basisadresse (base url) (Protokoll, Domain, Optional - Pfad) für Seiten eingefügt, wo die Einstellungen angewendet werden sollen. Wenn die Seite zu keiner angegebenen base-url passt, wird die Regel angewendet, dass sie mit der Bezeichnung „default“ vermerkt wird.

Die Spalte ALLOW SCRIPTS ist dafür da, um Verbot oder Erlaubnis von Skripten auszuführen und gilt nur für html-Dokumente. Um verschiedene Angriffe zu vermeiden, sollten nur Skripte zugelassen werden, die selbst überprüft worden sind.

Verfügbare Dateien die dem Benutzer über das Protokoll file:// (file Storage im LAN) zur Verfügung stehen, werden in jedem Fall über Proxy angezeigt, da dem Benutzer standardgemäss aus Sicherheitsgründen verboten ist mit dem Browser im Dateisystem zu arbeiten.

Proxy muss in folgenden Fällen verwendet werden:

- Befinden sich im Depot (file:// Protokoll).
- Wenn die Schnittstelle HES über das https-Protokoll funktioniert und alle oder ein Teil der Ressourcen, die angezeigt werden sollen über http:// oder file:// - Protokolle verfügbar sind. Da die Rendern Vorschau im Browser des Benutzers angezeigt wird, sind solche Ressourcen aus Sicherheitsgründen (wie im vorherigen Absatz) nicht verfügbar.

In anderen Fällen (z. B. wenn HES über das http-Protokoll verfügbar ist und die Dokumente nicht aus der Dateiablage angezeigt werden müssen) ist es sinnvoll, direkt ohne Proxy zu arbeiten. Unter anderem gibt es die Möglichkeit der Vorschau von Websites mit Autorisierung. Dafür fügen Sie die URL Ihrer Website in das Feld „type url address...“ ein und nehmen das Häkchen in der Spalte USE PROXY raus. Bei der Anzeige geben Sie die Daten in ein Formular zur Autorisierung (falls erforderlich) ein.

4.7 Datacloud

Diese Einstellungen verhalten sich analog zu den Einstellungen im persönlichen Bereich des Nutzers, jedoch beziehen sie sich nicht auf einen einzelnen HES-Nutzer, sondern auf alle gleichzeitig.

4.7.1 Benutzerdefinierte Suchanfragen

BENUTZERDEFINIERTER SUCHANFRAGE	
*	Geben Sie eine neue Suchanfrage ein...

Abb. 23 Benutzerdefinierte Suchanfragen Seite.

Die Einstellung „Benutzerdefinierte Suchanfragen“ ermöglicht das Ergänzen von Schlüsselwörtern, die danach in der DataCloud auf der Startseite aller Nutzer dargestellt werden und eine schnellere Eingabe von üblichen Suchanfragen ermöglichen.

Die Nutzeranfragen können erweitert, bearbeitet oder gelöscht werden. Es ist auch eine Navigation auf der Anfragenliste, speziell mit einzelnen Arbeitsblättern (wenn die Liste mehr als eine Seite umfasst), möglich.

4.7.2 Stoppwörter

< >		Seite 1 von 140	Q	Zeigen: 20
Stoppwörter				
<input type="checkbox"/>	STOPPWORT ▲			
<input type="checkbox"/>	5th			
<input type="checkbox"/>	able			
<input type="checkbox"/>	about			
*	Neues Stoppwort eingeben ..			

Abb. 24 Stoppwörter Seite.

Stoppwörter sind die Wörter, die häufig in den Dokumenten vorkommen, aber keinerlei weitere Informationen für die Ergebnisse geben. Wenn diese Wörter vorhanden sind, kann es einen negativen Einfluss auf die Relevanz der Abfrage haben.

Mit „Stop Words“ Registerkarte können Sie Wörter angeben, die in der weiteren Bearbeitung des Antrags nicht teilnehmen. Sie erscheinen auch nicht in der Datacloud auf der Suchergebnisseite, z. B. der Name des Unternehmens des Nutzers, der praktisch in jedem Dokument angegeben sein kann, der aber diesbezüglich für die Präzisierung der Anfrage keinen Nutzen bringt.

Stoppwörter die vom Administrator hinzugefügt worden sind, werden bei der Abfrage erst nach der Aktualisierung der Indexstruktur (kann durch den Klick auf den Button "Synchronize" gestartet werden) ignoriert. Während der Synchronisierung kann keine Suche durchgeführt werden. HES Benutzer kann diese Einstellung nicht durchführen.

Die Arbeit mit der Stoppwörter-Liste erfolgt parallel zur Arbeit mit der Anfragenliste in der Registerkarte „Benutzerdefinierte Suchanfragen“.

5 Connectors

5.1 Allgemeine Einstellungen

Connectoren im HES werden als Windows-Dienst ausgeführt, die mithilfe des Netzwerkinstallations-Assistenten installiert werden. Diese sind standardmässig in folgendem Ordner installiert C:\Program Files\Hulbee AG\Hulbee Enterprise Search\Connectors\...

Connectoren können auf zwei Arten konfiguriert werden. Die meisten Einstellungen sind über den HES Admin-Bereich verfügbar. Der jeweilige Teil der Konfiguration (wie Port einbinden, die Authentifizierung, etc.), wird in Ihrer Konfigurationsdatei eingefügt, unter dem Namen <Titel_ausführbare_Datei>.config.

5.1.1 Einstellungen der Connector über Konfiguration.

Die grundsätzlichen Möglichkeiten der Konfiguration, die nützlich sein können bei den Einstellungen:

<add key="Service.EndPoint" value="http://+:32770" /> – das Protokoll und Port, die verwendet werden bei der Schnittstelle für die Realisierung des HES Connectors API.

Der Connector kann über http und über https-Protokoll arbeiten. Es kann vielleicht notwendig sein das bestehende Protokoll und den Port mit diesem Befehl zu reservieren (wird im Administrator-Modus gestartet):

```
netsh http add urlacl url=http://+:32769/ user=\Everyone
```

Führen Sie es manuell aus, wenn Sie Port oder Protokoll des Connectors ändern möchten.

Weitere hilfreiche Befehle:

```
netsh http add urlacl url=https://+:32769/ user=\Everyone - ähnlich wie für  
das https-Protokoll  
netsh http show urlacl - reservierte URL anzeigen  
netsh http delete urlacl url=http://+:32769/ - reservierung der angegebenen  
Adresse löschen
```

Für https-Protokoll muss ein SSL-Zertifikat installiert werden. Dies kann ein vollwertiges Zertifikat sein, das das Unternehmen bei einem autorisierten Anbieter kauft. Es kann auch ein selbst signiertes Zertifikat verwendet werden. Für die Erstellung eines selbst signierten Zertifikats müssen folgende Schritte durchgeführt werden:

1. Zertifikat erstellen. Geben Sie in der Eingabeaufforderung PowerShell, das durch Administratorrechte gestartet wird, folgendes ein:

```
New-SelfSignedCertificate -DnsName localhost -CertStoreLocation  
Cert:\LocalMachine\My
```

Als Antwort wird der Hash des Zertifikats angezeigt. Beispiel:

3214979BE7BD608A426404537FCDB90103E157DB

2. Umwandlung des Zertifikats in vertrauenswürdigen Status. Damit ein Zertifikat vertrauenswürdig ist, installieren Sie es in Cert:\LocalMachine\Root. Führen Sie die Befehle in PowerShell aus:

```
$cert = (get-item Cert:\LocalMachine\My\*)  
$store = (get-item Cert:\LocalMachine\Root\  
$store.Open("ReadWrite")  
$store.Add($cert)  
$store.Close()
```

bei * – Hash des Zertifikats, erstellt im 1.Schritt.

3. Installation des Zertifikats auf dem gleichen Port, auf dem der Connector abgelegt ist. Für die Installation eines vertrauenswürdigen Zertifikats in PowerShell folgenden Befehl ausführen:

```
netsh http add sslcert ipport=0.0.0.0:32769 certhash=*appid='**'
```

bei * – Hash des Zertifikats, erstellt im 1.Schritt, ** – jeder gültige GUID (für deren Generierung kann genutzt werden, Beispiel <https://www.guidgenerator.com/>).

Authentifizierung kann unter Verwendung des Bereichs „authentication“ konfiguriert werden (die entsprechenden Daten müssen auch in die Connector-Einstellungen auf der Registerkarte „GENERAL“ eingetragen werden):

```
<authentication>
  <!-- Basic authentication section:

  <basic enabled="true or false" username="allowed user name"
    password="allowed user password" />

  The following code example demonstrates how to allow access to user with
    name "foo" and password "bar".

  <basic enabled="true" username="foo" password="bar" />
  -->
  <basic enabled="true" username="admin" password="pass"/>

  <!-- Windows authentication section:

  <windows enabled="true or false">
  <allow users="comma-separated list of users" roles="comma-separated list of
    roles" />
  <deny users="comma-separated list of users" roles="comma-separated list of
    roles" />
  </windows>

  The following code example demonstrates how to allow access to all members
    of the Admins role
  and deny access to all other user accounts.

  <windows enabled="true">
  <allow users="DOMAIN\Administrators" />
  <deny users="*" />
  </windows>
  -->
  <windows enabled="false" />
</authentication>
```

5.1.2 Einstellungen des Connectors im Admin-Bereich.

Im Admin-Bereich unter dem Punkt „Connector“ befindet sich die Schaltfläche „Details“, über die Sie zu den Registerkarten (die Einstellungen für Connectoren) gehen können.

Auf der Registerkarte „ALLGEMEIN“ sind folgende Einstellungen:

ALLGEMEIN

EINSTELLUNGEN

FEEDS

STATISTIKEN

DIAGNOSEN

☒ **Datenverarbeitung aktivieren**

Wenn es angekreuzt ist, werden die durch den Anschluss zur Verfügung gestellten Dokumente im Index hinzugefügt/aktualisiert. Andernfalls wird es nur vom IndexCleaner behandelt werden.

Anzeigename

Der Name, der für die Anschlüsse in Filtern, Metadaten usw., angezeigt wird.

FileSystem Connector

Polling timeout

The number of milliseconds to wait before the traversing request times out.

100000

^

v

Authentifizierungsverfahren

Die Authentifizierungsmethode, die bei der Durchführung von Anfragen am Anschluss verwendet wird.

None

v

Einstellungen speichern

✓

Abb. 25 Connectors. ALLGEMEIN.

- **“Enable data processing”**: Im eingeschalteten Zustand bedeutet, dass die Dokumente in einem Indexer indiziert, aktualisiert und hinzugefügt werden. Wenn die Aktivierung aufgehoben ist, werden die Dokumente nur gescannt um vom IndexCleaner aus dem Indexer entfernt zu werden, aber ein aktiver Durchgang des Speichers wird nicht durchgeführt. Die gleiche Einstellung ist auch auf der Registerkarte „FEEDS“ für jedes einzelne Storage verfügbar.
- **“Display Name”**: Name des Connectors, der vom Administrator festgelegt wird. Dieser wird auf der Seite des Benutzers im Bereich Filter und auf der Seite der erweiterten Suche angezeigt.
- **“Polling timeout”(milliseconds)**: Die Wartezeit auf die Antwort vom Connector. Für einige Connectoren, die eine lange Zeit für den Datendurchlauf brauchen, kann es sein, dass die Frist für die Bearbeitung der Datenmenge nicht ausreicht. Falls diese Zeit überschritten wird, so wird der Speicher erneut durchlaufen und die Suchmaschine zeigt dann eine entsprechende Warnung im Bereich „Logs“ an. Damit der nächste Durchlauf fortgesetzt werden kann, muss die Wartezeit erhöht werden.
- **“Authentication method”**: Wenn der Connector im Authentifizierungs-Modus konfiguriert ist (einstellbar in der Konfigurationsdatei „Connector“ – siehe 5.1.1), müssen hier die Einstellungen für die Authentifizierung eingestellt werden, damit HES Zugriff auf den Connector hat.

Bei der Registerkarte „EINSTELLUNGEN“ werden Einstellungen angezeigt (Kapitel 5.2.1, 5.3.1, 5.4.1), die bei den jeweiligen Arten von Connectoren unterschiedlich sind und in der jeweiligen Dokumentation beschrieben werden.

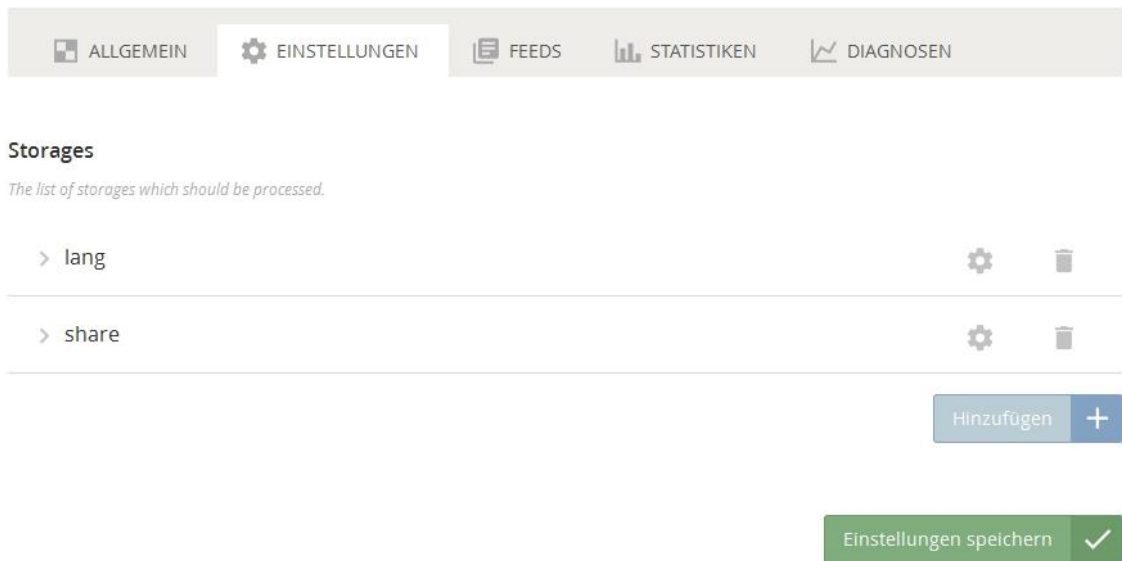


Abb. 26 Connectors. EINSTELLUNGEN.

Liste der Einstellungen auf der Registerkarte „FEEDS“ wird erst angezeigt, wenn einer der Punkte „Common“ oder „Users“ ausgewählt wird.

Abb. 27 Connectors. FEEDS.

Hier sind folgende Einstellungen vom Durchgang des Speichers dargestellt:

- **“Enable feed data processing”**: Wenn aktiviert, werden die vom Feed gelieferten Dokumente dem Index hinzugefügt/aktualisiert; ansonsten nur vom IndexCleaner behandelt.
- **“Display Name”**: Name, der für den Feed in Filtern, Metadaten usw. angezeigt wird
- **“Polling batch size”**: Anzahl der Dokumente pro Durchlauf
- **“Polling interval”**: Minimales Timeout zwischen Anfragen für Durchlauf
- **“Polling session timeout”**: Der minimale Zeitabstand nach einem vollständigen Durchlauf, bis ein neuer Durchlauf starten kann. Wenn Sie wissen, dass ein Feed nur selten aktualisiert wird, ist es sinnvoll, „session timeout“ zu erhöhen. Auf diese Weise kann die Anzahl der zeitgleich stattfindenden Durchläufe der Suchmaschine (verringern die Belastung auf die Suchmaschine) reduziert werden. Der Standardwert beträgt 15 Minuten.

- **“Web requests interval”**: minimales Timeout zwischen Web Anfragen um http(s)/ftp Inhalte zu erlangen. Mit dieser Einstellung kann das timeout erhöht werden, um die Belastung für die Webseite, die das System durchläuft, zu reduzieren.

Als empfohlene Einstellung des Polling interval, wird der Wert vorgeschlagen, der nicht kleiner ist als (Polling batch size*Web requests interval).

Registerkarte „STATISTIKEN“ enthält Kombination von „Namen“ - „Wert“. Diese Kombination wird vom Entwickler des Connectors eingegeben

Registerkarte „DIAGNOSEN“ zeigt Fehlermeldungen an.

5.1.3 Manuelles Connector-Setup

Connectoren repräsentieren die Windows-Dienste, die ebenfalls zu den REST-Services gehören. Standardgemäss werden alle Standard-Connectoren zusammen mit dem HES-Kern installiert und gestartet. Wenn Sie aus irgendeinem Grund einen Connector auf einen anderen Rechner übertragen möchten oder Sie möchten eine weitere Instanz eines Connectors auf dem aktuellen Rechner installieren, können Sie dies wie folgt ausführen:

1. Kopieren Sie den Ordner mit dem Connector an die gewünschte Stelle.
2. Reservieren Sie für den Connector einen freien Port und Protokoll (siehe 5.1.1).
3. Ordnen Sie es der Konfigurationsdatei Connector zu.
4. Installieren Sie den Connector als Dienst im Namen des Benutzers Network Service mit einem freien Service-Namen. Die Installation wird durch die gleiche ausführbare Connector-Datei mit erweiterten Optionen ausgeführt. Die Syntax der erweiterten Optionen erhalten Sie, indem Sie den Service mit dem Parameter „help“ starten.

Das oben beschriebene Verfahren eignet sich nur für Standard-Connectoren, die mit HES mitgeliefert werden. Connectoren die von externen Entwicklern entwickelt werden, werden so installiert und konfiguriert wie es in deren Anweisungen zu diesen Connectoren beschrieben wird.

5.1.4 Connector Registrierung Aufheben

Wenn ein Connector nicht benutzt wird, kann dieser aus der Liste der registrierten entfernt werden. Dies erfolgt durch Klick auf den Button „Papierkorb“ neben dem jeweiligen Header (siehe „Connectors“ Administrations-Bereich).

Anschlüsse

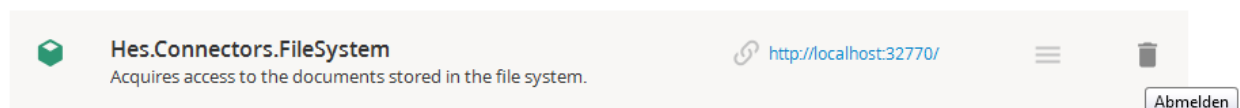


Abb. 28. Connector Bereich, „Unregister“.

Nach dem Ausloggen des Connectors werden alle Dokumente, die zu ihm gehören, aus dem Indexer gelöscht.

Es muss aber berücksichtigt werden, dass ein ausgeloggtter Connector trotzdem weiter als Service funktioniert. Dies kann dazu führen, dass einige Connectorarten (WebConnector) weiterhin den Speicher automatisch umgehen. Um die nutzlose Belastung zu verhindern, sollte der Connector

entweder vor dem Ausloggen aus den Speichereinstellungen entfernt werden, oder einfach der entsprechende Windows-Dienst deaktiviert werden (so wird ein automatischer Start beim Neustart verboten). Im letzteren Fall ist zu berücksichtigen, dass vor einer erneuten Registrierung des Connectors, dieser wieder aktiviert werden muss.

Beim Löschen eines Connectors aus der Liste der registrierten und nach erneuter Aktivierung werden die Speichereinstellungen auf der Registerkarte „SETTINGS“ gespeichert, aber die Einstellungen zur Verarbeitung von RSS-Feeds (Registerkarte „FEEDS“) kehren zu den Standardeinstellungen zurück.

5.2 Dateisystem Connector

Dient für den Zugriff auf Dokumente, die sich im lokalen Dateisystem befinden. Bei der Installation ist der Link standardmässig über <http://localhost:32770/> verfügbar.

5.2.1 Einstellungen des Dateisystem-Connector.

Beim Hinzufügen neuer Archive müssen deren Einstellungen festgelegt werden. Die Einstellungen für den Connector werden verfügbar, sobald diese in einer der Speicherlisten geöffnet werden. Zu den Einstellungen gehören:

- **“Storage name”**: Name des Speicherorts. Der Namen des Speichers. Wenn dieser geändert wird, so beginnt der Durchlauf des Speichers erneut.
- **“Scan location”**: Pfad auf dem lokalen Rechner oder in dem lokalen Netzwerk, das gescannt werden soll.
- **“Scan/watch files wildcard”**: das Suchmuster mit dem die Namen der Dateien am gescannten Ort übereinstimmen müssen. Im Fall einer Übereinstimmung, werden nur Dokumente indiziert, die den angegebenen Mustern entsprechen. Sie können nur die Erweiterung oder nur Dateinamen (Teil des Dateinamens) für die Suche eingeben. Beispiel: ***.tmp** oder **doc*.* (doc*)** oder **document.***. Eingeben von mehreren Wildcard zur gleichen Zeit durch Komma ohne Leerzeichen.
- **“Ignore patterns”**: Das Muster, um Dateien oder Verzeichnisse am gescannten Ort zu überspringen (leer lassen, um keine Dateien oder Verzeichnisse zu überspringen). Beispiel: ***.tmp** –Dateien mit folgender Erweiterung ignorieren oder ***/secure_folder/*** – ignoriert Dateien, deren Pfade dieser Ordner enthält. Jedes einzelne Muster zum Ignorieren wird in eine Zeile geschrieben, mit einem Komma.
- **“Access credentials”**: Sie können die Identitätsanmeldeinformationen angeben, die verwendet werden vom Connector der Zugriff auf das storage hatte. Während die Berechtigungen der Benutzer basierend auf den Einstellungen für den Zugriff auf das Dateisystem definiert werden. Wenn der Connector keine Verbindung zum Speicher herstellen konnte, wird eine entsprechende Meldung im Diagnose Connector aufgezeichnet.
- **“Domain”, “Username”, “Password”**: in diese Felder werden Daten der Benutzer eingegeben, die den Zugriff auf das Projektarchiv besitzen.
- **“Feed traversing timeout (milliseconds)”**: die maximale Zeit, die der Connector beim Sammeln von Dokumenten zur Verarbeitung verbrauchen kann. Diese Sammlung optimiert den Netzwerkverkehr im Falle einer grossen Anzahl von leeren Ordnern und Ordnern mit 1-2 Dateien im Datei-Repository. Wir empfehlen, den Wert um ein Vielfaches kleiner als die typische Timeouts im Netzwerk (5-10 Sekunden) zu halten. Verzeichnisse mit vielen Dateien haben diese Einschränkung nicht, so werden beim Zugriff alle vorhandenen Dateien zur Verfügung gestellt.

Der Wert 0 bedeutet, dass die Antwort sofort gebildet wird, unabhängig von der Anzahl der Dateien die im ersten Verzeichnis gefunden werden.

5.3 MS-Exchange Connector

Dieser Connector ermöglicht der Suchmaschine, den Inhalt der E-Mails von MS Exchange zu indizieren. Im Abschnitt dieses Connectors sind keine separaten Depots vorhanden, da jede Instanz von MS Exchange das Arbeiten mit Mailboxen aller Nutzer in der Domain ermöglicht auch wenn Sie physisch auf einem anderen Server MS Exchange liegen.

Die Anforderungen an den Exchange-Connector-Dienst: Darauf muss Remote Powershell (bisher nur die Basic-Authentifizierung) verfügbar sein.

Wenn die Installation standardgemäss verwendet wird, ist Exchange-Connector unter <http://localhost:32771/> verfügbar.

5.3.1 MS-Exchange Connector Einstellungen

Die Einstellungen für den MS Exchange-Connector enthalten:

- **“Exchange Server url”**: Url des Exchange Server.
- **“Access credentials”**: Sie können die Anmeldeinformationen zur Identität angeben, die verwendet wurden, um auf den MS Exchange Server zuzugreifen.
- **“Domain”, “Username”, “Password”**: In diese Felder werden die Daten des Nutzers eingegeben, der über Lesezugriff für jedes MS Exchange-Postfach verfügt. Dieser Benutzer muss eine eigene Mailbox auf MS Exchange haben, um Personifikationsrechte über Postfächer zu haben. Ausserdem muss der Benutzer zu der Gruppe View-Only Organization Management (oder Organization Management) gehören, um auf die Liste anderer Nutzer zugreifen zu können.
- **“Container”**: Er ermöglicht, die Benutzer aus Active Directory einzuschränken, deren Mail in den Suchindex gelangen soll. Beispiel so eines Pfads zu dem Container: “CN=Users,DC=hes,DC=hulbee,DC=com”. Es können weitere, mehrere solcher Pfade eingegeben werden, doch jeder Pfad benötigt eine neue Zeile. Wenn die Liste leer ist, werden alle Mails von allen Domain-Benutzern indiziert.
- **“Request timeout”(milliseconds)**: Die Wartezeit des MS Exchange Server um Anfrage zum Durchlauf zu erhalten. Falls diese Zeit überschritten wird, wird ein Durchlauf erneut begonnen und die Suchmaschine zeigt dann eine entsprechende Warnung im Bereich "Logs". Damit der nächste Durchlauf fortgesetzt werden kann, muss die Wartezeit erhöht werden.

5.4 Web-connector

Dies ist ein Crawler, der die Seiten im lokalen und auch globalen Netzwerk durchläuft. Web-connector verfügt über umfangreiche Einstellungen des Crawlers. Alle Dokumente, die unter diesem Crawler gefunden werden, haben die gleichen Einstellungen der Sichtbarkeit nach Benutzern innerhalb des Speichers. Die Zugriffsrechte für einzelne Benutzergruppen werden in den entsprechenden Einstellungen („Access list mode“) definiert.

Bei der Installation ist er standardmässig über den Link <http://localhost:32769/> verfügbar.

5.4.1 Einstellungen des Web-Connectors

Die Einstellungen für den Connector werden verfügbar, sobald diese in einer der Optionen in der Speicherliste geöffnet werden. Zu den Einstellungen für Web-Connector gehören folgenden Punkte:

- **“Storage name”**: Name des Speicherorts.
- **“Url”**: url der Startseite von Internet-Ressourcen, die gecrawlt werden sollen. Diese Seite muss Links beinhalten, die zu anderen Seiten der Website führt.
- **“Authentication type”**: legt fest, welche Anmeldeinformation benötigt wird für das Crawlen der Website. Wird verwendet, wenn die Webseite Authentifizierung beinhaltet.
- **“Login”/“Password”**: Login und Passwort für den ausgewählten Authentifizierungstyp. Es muss ein Konto verwendet werden, das alle Seiten und Dokumente anzeigt.
- **“Maximum of concurrent threads”**: die maximale Anzahl gleichzeitiger CPU-Threads, die der Connector für die Verarbeitung einer Ressource verwenden kann.
- **“Maximum pages to crawl”**: die maximale Anzahl der Seiten, die der Crawler crawlen darf. Wenn diese Grenze erreicht ist, so hält der Crawler an und weiter erfolgt nur die Überprüfung der Meta-Informationen bei bereits gefundenen Seiten (Dokumenten).
- **“Maximum pages to crawl per domain”**: wenn eine der Ressourcen einen Verweis auf die externe Domain hat, so können Grenzen für das Crawling der Domain gesetzt werden. Dies macht Sinn, denn bei aktiviertem „Enable external page crawling“ und „Enabled external links crawling“, erfolgt die Verarbeitung von Seiten und Dokumenten, nicht nur in der Domain, auf der angegebenen Homepage.
- **“Max page size (bytes)”**: die maximale Seitengröße für Crawling, sobald diese Anzahl überschritten wird, werden die Links auf der Seite nicht durchlaufen.
- **“Crawl external pages”**: Bestimmt ob die Links auf externen Ressourcen in den Indexer abgespeichert werden.
- **“Crawl external pages links”**: Bestimmt ob die Links mit externen Ressourcen abgespeichert werden sollen, die auf den Indexer übergegangen sind; diese Einstellung ist nur sinnvoll, wenn die vorherige Einstellung aktiviert ist.
- **“Use canonical links”**: Im aktivierten Zustand versucht der crawler, einen link-Tag mit dem Parameter rel="canonical" in dem Abschnitt „head“ auf der Seite zu finden. Wenn die Suche ein Ergebnis ergab, so wird dieser Link als Parameter „href“ link-Tag gespeichert, anstatt dem ursprünglichen Link. Sonst sucht das Meta-Tag mit dem Parameter property="og:url" und speichert den Wert des Parameters „content“ ab. Falls keine Suchergebnisse gefunden werden konnten, bleibt der Link bestehen, der für die Anmeldung auf der Seite genutzt worden ist.
- **“Http request timeout (seconds)”**: Frist für den Umgang mit der Ressource.
- **“Follow redirects”**: die Anzahl der zulässigen http-Weiterleitungen.
- **“Maximum auto redirects”**: Wenn ein Link mit Weiterleitung vorhanden ist, dann wird bei ausgeschalteter Konfiguration der Übergang nicht vorgenommen, der Link selbst wird nicht hinzugefügt.
- **“Enable Cookies”**: Bestimmt ob beim Crawlen durch Webseite-Links Cookies gesendet oder nicht gesendet werden sollen.
- **“Enabled SSL certificate validation”**: bei aktivierter Option wird die Bearbeitung der Seiten nicht erlaubt, von denen das Zertifikat nicht zugänglich ist. Sinnvoll bei der Verarbeitung von https-Ressourcen.
- **“Maximum crawl depth”**: die maximale Tiefe von Links auf externe Ressourcen.
- **“Maximum retry count”**: gibt die Anzahl der Wiederholungsversuche für die Bearbeitung der Datei an, wenn die Datei in der vorgegebenen Zeit nicht bearbeitet werden konnte.
- **“Minimum retry delay (milliseconds)”**: Wartezeit zwischen den Versuchen zum erneuten Zugriff auf die Datei.
- **“Respect robots.txt”**: Bestimmt ob die Regeln aus robots.txt berücksichtigt werden sollen.

- **“Respect meta robots”**: Bestimmt ob der no follow in meta-tags auf der Seite berücksichtigt werden soll.
- **“Respect X-Robots-Tag header”**: Bestimmt ob der no follow in http Überschriften berücksichtigt werden soll.
- **“Respect anchor rel='nofollow'”**: Sollen no follow in Attribute Links (rel="nofollow") berücksichtigt werden.
- **“Ignore robots.txt if root disallowed”**: wenn diese Option aktiviert ist, wird robots.txt ignoriert.
- **“Robots.txt user agent”**: ermöglicht die Angabe von Robots agent, wenn die Regeln von robots.txt ausgewählt sind. Wenn die Option nicht ausgewählt wurde, gelten die allgemeinen Regeln.
- **“Maximum robots.txt crawl delay (seconds)”**: Häufigkeit der Umfrage auf Änderungen der robots.txt Datei.
- **“Minimum crawl delay per domain (milliseconds)”**: die minimale Verzögerung zwischen den Abfragen der Seiten von einer Domain. Die Einstellung ist notwendig, um keine zu hohe Belastung der Webseite durch den Crawler zu produzieren. Wenn zum Beispiel 500 msec. angegeben werden, so würde dies bedeuten, dass pro Sekunde nicht mehr als 2 Seiten von der Web-site angefordert werden.
- **“Additional headers”**: zusätzliche http-Header für die Webseite, werden im Format key: Wert angegeben. Jeder einzelne Titel wird in einer neuen Zeile geschrieben. Dies wird in seltenen Fällen benötigt.
- **“Min generations before deleting”**: die Anzahl der Versuche ein Dokument zur Überprüfung herunterzuladen, das in der Datenbank von Crawling abgelegt ist. Bevor dieses aus der Liste der Dokumente von der verfügbaren Website gelöscht werden kann.
- **“Pause after cycle (seconds)”**: die Verzögerung zwischen den Zeitabschnitten des Durchlaufens der Ressource.
- **“Access list mode”**: Verfahren zur Bestimmung der Zugriffsrechte. Folgende Optionen sind verfügbar:
 - “Public” – Zugriff für alle HES Benutzer erlauben.
 - “Fixed Access List” – Zugriff für Benutzer/Benutzergruppen im Sinne des “Access list allowed”, mit Ausnahme der Verbotenen in dem Feld “Access list denied”.
 - “X-HES-Users only” – Auswahl von Benutzer/Gruppen von header fields “X-HES-Users-Allowed” und “X-HES-Users-Denied”. Detaillierte Infos über X-HES-Users siehe 5.4.5. Einstellungen “Access list allowed\denied” werden ignoriert.
 - “Prefer X-HES-Users” – Wenn es im response-header die Eigenschaften X-HES-Users gibt, so werden gerade diese die Berechtigungen der Benutzer bestimmen, ansonsten – Einstellungen “Access list allowed\denied”. Wenn die Eigenschaften X-HES-Users fehlen, und “Access list allowed\denied” nicht konfiguriert ist, wird die Ressource nur für Systemadministrator HES verfügbar.
 - “X-HES-Users and Access List” – berücksichtigt werden die Zugriffsrechte der Benutzer aus beiden Listen gleichzeitig (die in X-HES-Users und “Access list allowed\denied”).
- **“Access list allowed”**: Die Liste der Benutzer und Gruppen, die Zugriff auf die Dokumente aus dem Speicher haben. Einem Teil kann den Zugang verweigert werden mit Hilfe der Einstellungen aus dem nächsten Absatz.

- **“Access list denied”**: die Liste der Benutzer und Gruppen, denen der Zugriff auf die Ressource verweigert wird, auch wenn diese in der Liste aus dem vorherigen Punkt vorhanden sind. D.h., die Liste der Verweigerung hat Vorrang vor der Liste der Berechtigungen.
- **“Active Directory credentials”**: Active Directory wird vom Connector zur Umwandlung von symbolischen Namen von Benutzern und Gruppen in SID verwendet. Es ist notwendig die Felder (“Domain”, “Container”, “User name”, “Password”) mit den Benutzerdaten auszufüllen, die über die Berechtigung zum Lesen der Liste der Benutzer in Active Directory verfügen. Benutzernamen im Feld „User name“ sollte nach dem Domain-Namen folgen. Diese Rubrik muss nicht ausgefüllt werden, wenn alle Dokumente den Zugriffsmodus „Public“, haben oder die Webseite gibt in X-HES-Users Header nur die Benutzer SID (X-HES-Users-Denied:S-1-5-21-3255245507-3551417498-1381599987-1131) wieder und nicht ihre symbolischen Namen (X-HES-Users-Allowed:HES\Domänen-Benutzer;HES\j.doe).
- **“Taboo rules”**: eine Liste von regulären Befehlen. Wenn der Link mindestens unter eine Regel fällt, so wird er nicht indexiert und kann später entfernt werden. Jeder einzelne des regulären Befehls wird in einer separaten Zeile eingetragen. Dies ermöglicht dem Crawler, weniger wichtige Bereiche der Website zu umgehen (oder jene, wo er hängenbleiben kann, indem der Crawler eine unendliche Anzahl von Seiten durchgeht). Eine weitere Anwendung der Einstellungen ist das Partitionieren einer Website auf mehrere Speicher. Beispiele für reguläre Ausdrücke finden Sie hier: [https://msdn.microsoft.com/en-us/library/az24scfc\(v=vs.110\).aspx](https://msdn.microsoft.com/en-us/library/az24scfc(v=vs.110).aspx). Beispiel: Teil der Webseite, die folgenden Wert enthält: actpos=2, müssen in einen anderen Speicher mit anderen Zugriffseinstellungen eingesetzt werden. Dafür wird in Taboo rules im primären Speicher diese Regel eingegeben „actpos=2“. Dies lässt das Eindringen von solchen Seiten auf den primären Speicher nicht zu. Für diesen ganz besonderen Teil der Website müssen Speicher mit seiner Homepage und eigener Regel hinzugefügt werden „.*actpos=(?!2).*“. Wenn eine solche Option fehlt, wird dies ausgelöst.

Achtung! Bei den Einstellungen, die zu einer „Engstelle“ bei der Anzahl der Ergebnisse führen können (reduzierter Wert im Feld), ist es notwendig, dass der Web-Connector den Speicher von neuem durchläuft. Damit der Speicher nicht neu erstellt werden muss, sollte der Name im Feld **„Storage name“** geändert werden.

Zu den Bereichen, wo es möglicherweise ähnliche Einstellungen gibt, gehören: „Crawl depth“, „Maximum pages to crawl“, „Maximum page size“ usw, sowie die Aktivierung der „Use canonical links“.

5.4.2 Feineinstellungen des Web-Connectors

Einige Web-Connector-Einstellungen (allgemein für alle Archive) sind über die Konfigurationsdatei verfügbar. Sie befinden sich im Bereich „abot“.

Im Bereich „abot“ können die folgenden Eigenschaften nützlich sein:

- `maxPageSizeInBytes` – wenn die Größe der Download-Seite überschritten wird, dann wird „abot“ sie nicht herunterladen und sucht nach den Links. Aber sie kommt immer noch in den Index.
- `downloadableContentTypes` – Zunächst werden die Titel der Seiten heruntergeladen und deren Content-Type überprüft. Wenn sie übereinstimmen, wird die Website für die weitere Suche von den Links auf der neuen Seite (die Arbeit des Crawlers) heruntergeladen.

5.4.3 Indexierung eines Teils einer Web-Seite deaktivieren

Eine Web-Seite enthält oft viele Informationen, die sich auf allen Seiten der Webseite wiederholen und die bei der Suche nutzlos sind. Beispiele für solche Bereiche können sein: Header-Dateien, Fusszeilen, Menüs, verschiedene Navigations-Elemente. Um diese Bereiche aus dem Prozess der Weiterverarbeitung auszuschliessen, kann der Eigentümer der Website mit den folgenden Tags in der HTML-Seite die Kommentare hinzufügen:

```
Ordinal text for processing
<!--allowindexing:off-->
This text is not searchable
<!--allowindexing:on-->
Ordinal text again
```

Ebenso werden ähnliche Tags unterstützt, die von GSA verwendet werden:

- <!--googleoff: all-->
- <!--googleoff: index-->
- <!--googleon: all-->
- <!--googleon: index-->

Dies gilt für alle html-Dokumente, die im Index abgelegt werden. Jedoch wird im Falle der Nutzung des Web-connectors normalerweise die eigene Website indexiert (Crawling). Deshalb kann der Inhaber der HES-Kopie zu diesem Zweck die entsprechenden Tags auf der eigenen Website einfügen.

5.4.4 Kanonische Links

Oft ist es so, dass ein und dieselbe Seite über mehrere URI verfügbar ist. Diese können sich zum Beispiel durch ein Parameter, der anzeigt, von wo Sie auf die Website gelangt sind, unterscheiden. Dies ist sinnvoll für die Analyse von Übergängen, aber nicht wichtig bei der Suche nach Inhalten. Durch die verschiedenen Links wird sie vom Crawler als viele Seiten mit dem gleichen Inhalt wahrgenommen.

In solchen Fällen ist es sinnvoll, den Mechanismus der kanonischen Links zu verwenden, der die Indizierung der Suchmaschine auf mehreren Kopien derselben Seite verhindert. Dazu sollte in den Einstellungen des Web-Connectors die Option „Use canonical links“ aktiviert sein und auf der Website folgender Seiten-Header Tag verwendet werden:

```
<link rel="canonical" href="http://example.com/">
```

Wenn mehrere Seiten den gleichen kanonischen Link beinhalten, wird nur eine in den Index gelangen (Es bedeutet, dass sie inhaltlich identisch sind oder unwesentlich abweichen). Der Übergang von den HES-Suchergebnissen erfolgt einfach durch den Link, der im href-Attribut angezeigt wird.

5.4.5 X-HES-Users headers

Das HTTP-Protokoll ermöglicht das Festlegen von Einschränkungen beim Zugriff auf bestimmte Web-Ressourcen mit Hilfe von Mechanismen für die Autorisierung. Allerdings gibt HES keine Informationen über die Liste der Benutzer, die den Zugriff auf die einen oder anderen Dokumenten haben. Wenn bei der Anzeige der Website individuelle Rechte für die einzelnen Dokumente benötigt werden, können Sie den Mechanismus custom-HTTP-headers nutzen, die mit “X-HES-Users” beginnen.

Es gibt zwei solcher Überschriften: "X-HES-Users-Allowed" und "X-HES-Users-Denied". Diese Überschriften müssen zu der HTTP-Antwort site's backend hinzugefügt werden.

Die Benutzer/Benutzergruppen für X-HES-Users Überschriften können im SID-Format oder symbolischen Formaten eingetragen werden:

- SID (z. B., S-1-5-21-3255245507-3551417498-1381599987-1124),
- domain\user_name,
- user_name.

Mehrere Einträge innerhalb einer X-HES-Users Überschrift werden durch ";" getrennt. Eine X-HES-Users Überschrift kann möglicherweise Namen in verschiedenen Formaten enthalten.

Response Headers:

```
Accept-Ranges:bytes
Content-Encoding:gzip
Content-Length:132
Content-Type:text/plain
Date:Fri, 02 Sep 2016 07:18:34 GMT
ETag:"0f4993e99e9d11:0"
Last-Modified:Fri, 29 Jul 2016 13:00:56 GMT
Server:Microsoft-IIS/8.5
Vary:Accept-Encoding
X-HES-Users-Allowed:S-1-5-21-3255245507-3551417498-1381599987-
1124;HES\Domänen-Benutzer;HES\j.doe;j.doe
X-HES-Users-Denied:S-1-5-21-3255245507-3551417498-1381599987-1131
X-Powered-By:ASP.NET
```

Für das richtige Abrufen der Überschriften müssen Sie Folgendes beachten:

Damit diese Überschriften bearbeitet werden, wird der Modus "Access list mode" gewählt, der die Arbeit mit X-HES-Users Überschriften erlaubt. Falls die Namen der X-headers im symbolischen Format eingetragen sind, müssen Einstellungen der Felder "Active Directory credentials" des Web-Connectors eingetragen werden, damit eine Verbindung zum Active Directory entstehen kann. Bitte beachten Sie, dass das Feld "User name" nach dem Domain-Namen folgen sollte.

Im Falle der korrekt ausgefüllten Felder mit symbolischer Eintragung, werden diese in SID-Format umgewandelt. Wenn Überschriften nicht vom System erkannt wurden, werden diese nicht gespeichert, doch es werden Informationen auf der Registerkarte "DIAGNOSTICS" angezeigt. Die Gründe, warum X-HES-Users Überschriften nicht erkannt wurden, könnten folgende sein:

- Fehler bei der Zusammenstellung der Header selbst (z. B. Eintrag im SID-Format nicht überprüft, d.h. wenn ein Fehler unterlaufen ist, wird dieser nirgends angezeigt).
- Header enthält den Benutzernamen eines anderen Domains oder Subdomains.
- Einstellungen der Felder "Active Directory credentials" falsch ausgefüllt. Bitte beachten Sie, dass die Änderungen an diesen Einstellungen erst nach nochmaligem Durchlauf des Speichers wirksam werden.

6 Hilfreiche Links

Active directory:

- <https://technet.microsoft.com/en-us/library/dn283324.aspx> - Active Directory Services Übersicht.
- <https://technet.microsoft.com/en-us/library/hh472160.aspx> - Bereitstellen von Active Directory- Domain Services (AD DS) in Ihrem Unternehmen.
- <https://technet.microsoft.com/en-us/library/jj574166.aspx> – Installieren eines neuen Windows Server 2012 Active Directory Forest (Level 200)

Organisation des Backup Prozesses:

- <https://technet.microsoft.com/en-US/library/dn390929.aspx> - Windows Server Backup und Storage Pools
- https://en.wikipedia.org/wiki/List_of_backup_software – Liste von Backup Software (unabhängige Anbieter, Open Source).

Elasticsearch

- <https://www.elastic.co/downloads/past-releases> - Download Seite für Elasticsearch 2.3.*
- <https://www.elastic.co/guide/index.html> - Dokumentation.

7 Bekannte Probleme

- HES Deinstallieren. Im Falle einer erfolglosen Deinstallation von HES (HES Dienste laufen immer noch im Dienste-Applet, im Programme und Funktionen Applet ist immer noch eine Verknüpfung vorhanden), versuchen Sie eine manuelle Deinstallation:
 1. Stoppen Sie die Dienste Hes.Services.IndexCleaner, Hes.Connectors.Exchange, Hes.Connectors.FileSystem, Hes.Connectors.Web und Hes.Services.ConnectorManager mithilfe des „Services“ Applets.
 2. Starten Sie die Befehlskonsole (cmd.exe) als Administrator.
 3. Löschen Sie Dienste mit den folgenden Befehlen:
 - `sc delete Hes.Services.IndexCleaner`
 - `sc delete Hes.Connectors.Exchange`
 - `sc delete Hes.Connectors.FileSystem`
 - `sc delete Hes.Connectors.Web`
 - `sc delete Hes.Services.ConnectorManager`
 4. Wenn die Dienste noch unter „Services“ sichtbar sind, starten Sie den Server neu.
 5. Entfernen der Anwendung „hes“, der Standard -Webseite (IIS) und des Anwendungspools „hes“, wenn es notwendig ist.

6. Löschen Sie den Ordner mit dem installierten HES. Es ist c:\Programme\Hulbee AG\Hulbee Enterprise Search\Standardordner.
 7. Öffnen Sie „Programme und Funktionen“, und löschen Sie die „Hulbee Enterprise Search“ Verknüpfung. Klicken Sie auf „Deinstallieren“ und die Option „Verknüpfung löschen“ wird vorgeschlagen werden.
- Wenn es notwendig ist, den Indexer HES zu entfernen, dann können Sie diesen durch Elasticsearch oder mit einem Tool von Utilities\IndexUtil (Hilfe ist beim Start der Anwendung mit der Option „help“ verfügbar) entfernen. Im normalen Fall sollte das Löschen des Indexes nicht erforderlich sein. Beim Update fragt das Installationsprogramm, ob der Index neu erstellt werden soll oder der vorhandene verwendet werden kann.