



# HULBEE ENTERPRISE SEARCH

Installation und Setup-Handbuch

**Erstellt:** 12-Mai-2017

**Version:** 3.0

# Inhaltsverzeichnis

<b>VERWENDETE BEGRIFFE.....</b>	<b>3</b>
<b>1 ARCHITEKTUR.....</b>	<b>4</b>
1.1 HULBEE ENTERPRISE SEARCH.....	4
1.2 HESBOX.....	5
<b>2 ERSTINSTALLATION .....</b>	<b>6</b>
2.1 VORBEREITEN DER INSTALLATION .....	6
2.1.1 Indexstore .....	6
2.1.2 Hardware-Voraussetzungen für HES Processing Server.....	6
2.1.3 Software-Voraussetzungen für HES Processing Server .....	7
2.1.4 IFilters setup.....	7
2.1.5 Apache Tika converter .....	8
2.1.6 Authentifizierungseinstellungen .....	8
2.2 INBETRIEBNAHME.....	8
2.3 NACH DER INSTALLATION .....	17
2.3.1 Lizenz einrichten.....	17
2.3.2 Desktop Manager Installation und Einstellungen .....	17
2.4 OPTIMIERUNGEN (TWEAKS).....	18
2.4.1 IndexUtil and benutzerdefinierte Indexstruktur.....	18
2.5 HESBOX.....	18
2.5.1 Einrichtung einer Domain .....	18
2.5.2 Dateispeicher .....	19
<b>3 ERSTDURCHLAUF.....</b>	<b>19</b>
3.1 INDEX REINIGUNG UND MODIFIZIERUNG .....	21
<b>4 EINRICHTUNG DER ADMIN SEITE.....</b>	<b>22</b>
4.1 WIEDERHERSTELLUNGSOPTIONEN.....	22
4.2 ERSTE SCHRITTE DER EINSTELLUNGEN.....	22
4.3 REGISTRIERUNG DER CONNECTOREN .....	22
4.4 HINZUFÜGEN VON SPEICHERPLÄTZEN .....	23
4.5 ADMINISTRATIONSBEREICH .....	23
4.5.1 Dashboard.....	24
4.5.2 Anschlüsse.....	25
4.5.3 Nutzer .....	26
4.5.4 Ereignisprotokoll .....	26
4.5.5 Unverarbeitet.....	27
4.6 EINSTELLUNGEN.....	28
4.6.1 Authorization .....	28
4.6.2 Verarbeitung .....	31
4.6.3 Löschen .....	33
4.6.4 Protokollierung .....	34
4.6.5 Appearance.....	35
4.6.6 Local öffnen.....	35
4.6.7 Vorschau .....	36
4.6.8 Importieren / Exportieren .....	38

4.7	DATA CLOUD.....	38
4.7.1	Benutzerdefinierte Suchanfragen .....	38
4.7.2	Stoppwörter .....	39
4.7.3	Blockwörter.....	39
5	CONNECTOREN.....	39
5.1	ALLGEMEINE EINSTELLUNGEN .....	39
5.1.1	Einstellungen der Connectoren über Konfiguration .....	40
5.1.2	Einstellungen des Connectors im Admin-Bereich .....	41
5.1.3	Manuelles Connector-Setup.....	45
5.1.4	Connector Registrierung Aufheben .....	46
5.2	DATEISYSTEM CONNECTOR .....	46
5.2.1	Einstellungen des Dateisystem-Connectors .....	46
5.3	MS-EXCHANGE CONNECTOR.....	47
5.3.1	MS-Exchange Connector Einstellungen.....	48
5.4	WEB-CONNECTOR.....	48
5.4.1	Einstellungen des Web-Connectors.....	49
5.4.2	Feineinstellungen des Web-Connectors .....	52
5.4.3	Indexierung eines Teils einer Web-Seite deaktivieren.....	52
5.4.4	Kanonische Links .....	53
5.4.5	X-HES-Users headers.....	53
6	HILFREICHE LINKS .....	54
7	BEKANNTE PROBLEME.....	55

## Verwendete Begriffe

**Index** - im Zusammenhang mit HES eine Datenbank, die Dokumente und Zusatzinformationen enthält. Diese Informationen ermöglichen ein schnelles Auffinden des gewünschten Dokumentes anhand der eingegebenen Suchbegriffe. Sie wird auch zum Speichern von HES-Einstellungen verwendet.

**Storage** - bezeichnet einen Speicher mit Dokumenten wie Dateisystem, Mail-Dienst, "Netzwerk-Festplatte", "FTP-Server", "einige Websites", etc.

**Feed** - Liste der Metadaten für Dokumente aus dem Speicher, die verarbeitet werden sollen. Ein Feed repräsentiert einen Speicher.

**Connector** – ein Plugin, das HES erlaubt, ein Dokument zu suchen, das in verschiedenen benutzerdefinierten Repositories (Enterprise-Dateisystem, Web-Ressourcen, Mail-Service) gespeichert ist, damit es weiter verarbeitet werden kann. Ein Connector ist eine REST-Dienstanwendung, die Methoden (Endpunkt) bereitstellt, die der Connector Manager aufruft, um Dokumente aus bestimmten Datenquellen zu erwerben.

**Crawler** - Teil des Web Connectors, der die Seiten im lokalen und globalen Netzwerk durchsucht. Crawler hat seine eigenen Einstellungen und Zeitraum des Crawlings, die beim Hinzufügen von Storage installiert werden. Das Abmelden des Connectors beendet den Crawler nicht.

# 1 Architektur

## 1.1 Hulbee Enterprise Search

Hulbee Enterprise Search (HES)<sup>1</sup> 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.

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	✓		
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			✓

<sup>1</sup> 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.

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 oder Internet Explorer) 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 HESbox

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 Directory<sup>2</sup> und zur Speicherung von Daten in kleinen Strukturen<sup>3</sup>, die hierfür noch gleichrangige Netze nutzen, dienen.

Die HESbox enthält Releases, die sich sowohl durch die Hardwareleistung als durch die Begrenzung der Nutzerzahlen unterscheiden.

<sup>2</sup> Links zur Einführung in Active Directory: siehe Abschnitt 6 (Hilfreiche Links).

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

## 2 Erstinstallation

### 2.1 Vorbereiten der Installation

#### 2.1.1 Indexstore

Als Indexspeicher wird Elasticsearch v.5.X.X verwendet. 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 Arbeitsspeicher für Elasticsearch ist abhängig von vielen Aspekten, einschliesslich der Art der Daten, aber auch 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.

#### 2.1.2 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

### 2.1.3 Software-Voraussetzungen für HES Processing Server

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

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

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-WindowsAuthentication
/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 /All
```

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

### 2.1.4 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>.

### 2.1.5 Apache Tika converter

Es besteht die Möglichkeit, den Tika-Konverter anstelle von Standard-Iveonik-Konvertern zu verwenden, um Metadaten-Endtext aus verschiedenen Dateien zu extrahieren. Die entsprechende Einstellung wird im Abschnitt Processing (siehe 4.6.2) im Admin-Panel eingerichtet. Dieses Toolkit wird ausführlich auf der offiziellen Tika Apache Seite beschrieben.

### 2.1.6 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:

1. Wählen Sie in dem IIS-Konsole Fenster die Website mit HES (in der Regel die Standard-Website).
2. Wählen Sie bitte in dem Bereich Features „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.

## 2.2 Inbetriebnahme

Zur Installation von HES starten Sie HES Installer HES.3.0.XXX.XXXXX.msi aus dem Paket (XXX ist die Nummer Ihrer spezifischen Version).



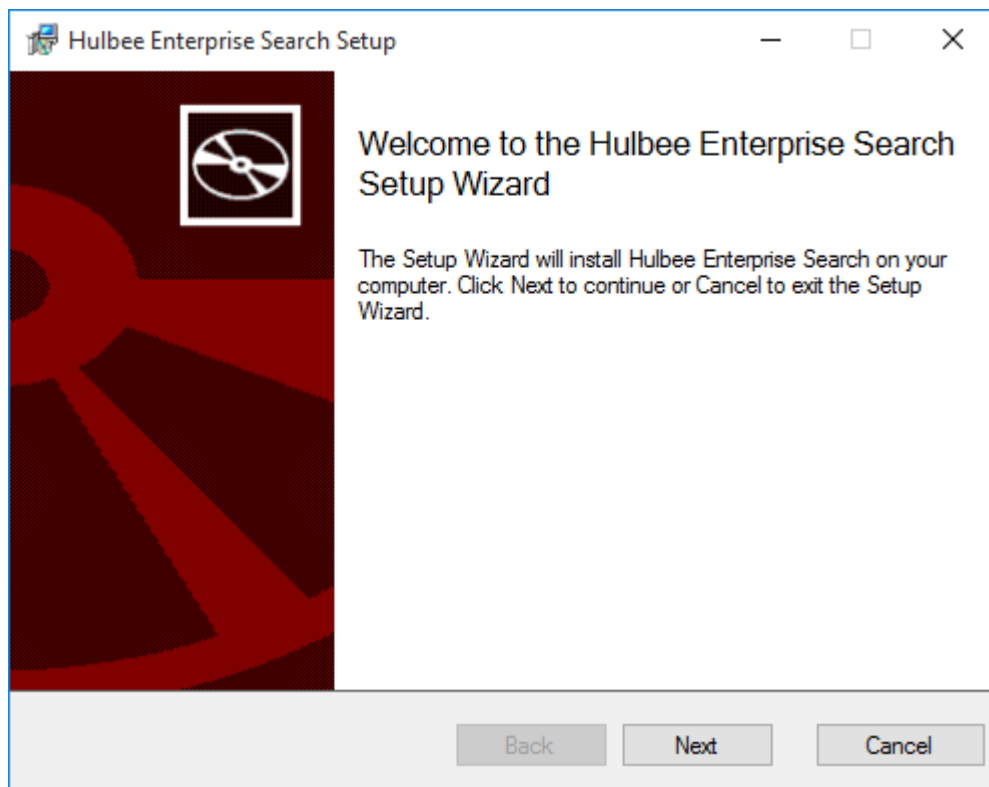


Abb. 1 HES-Installationsassistent – Beginn der Installation.

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

### 1. Schritt 1.



Abb. 2 Schritt 1 der HES-Installationsassistent.

## 2. Schritt 2.

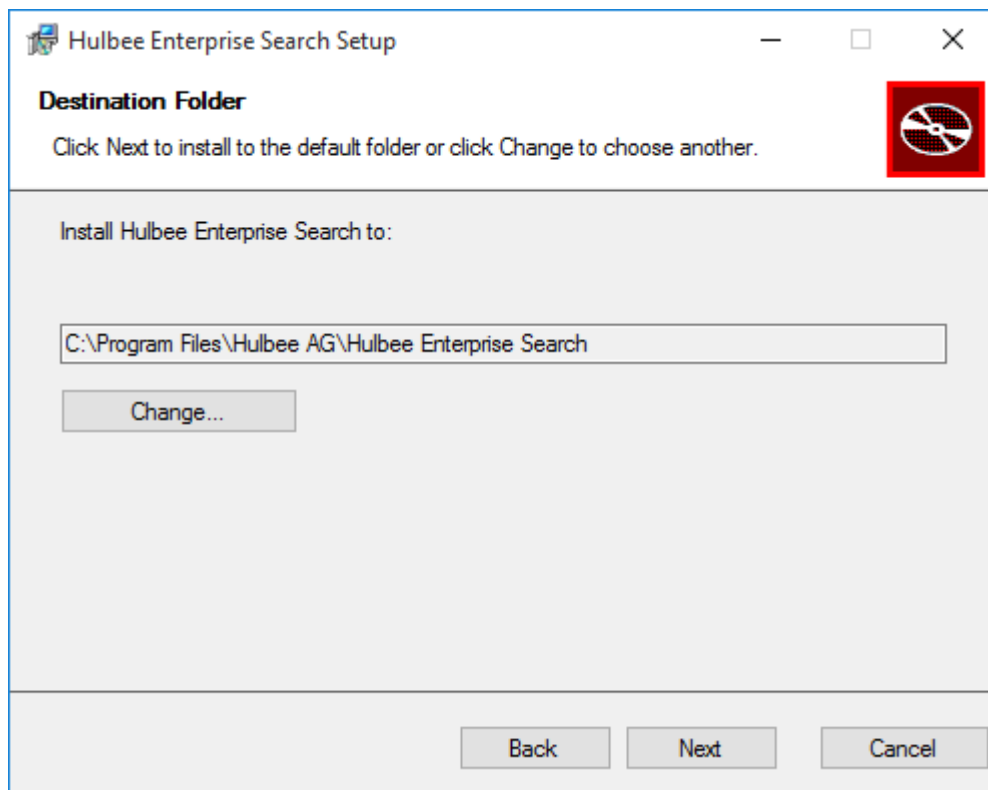


Abb. 3 Schritt 2 der HES-Installationsassistent.

Installation Folder können mit den Standardwerten belassen werden.

## 3. Schritt 3.

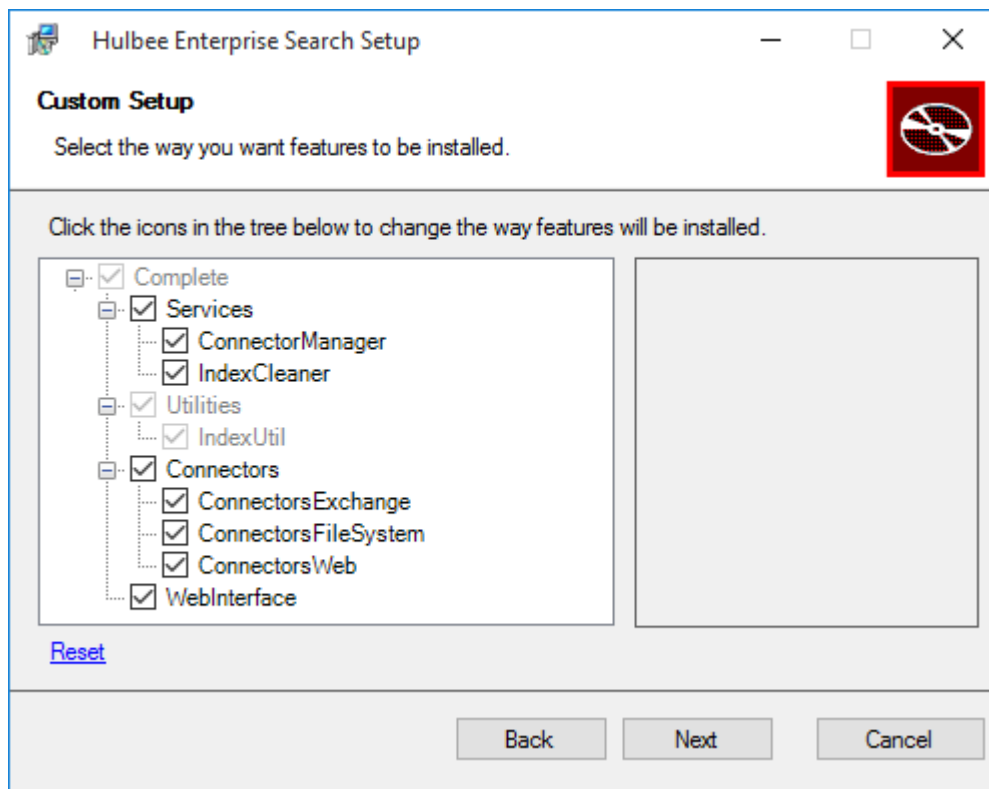


Abb. 4 Schritt 3 der HES-Installationsassistent.

Es ist möglich, das Standard-Set des Features zu verwenden oder spezifisch auszuwählen. Um wieder auf die Standardeinstellung zurückzukommen, klicken Sie auf „Reset“. Falls erforderlich, können deaktivierte Komponenten später installiert werden, bei erneuter Ausführung des Installationsprogrammes oder durch Control Panel\All Control Panel Items\Programs and Features.

#### 4. Schritt 4.

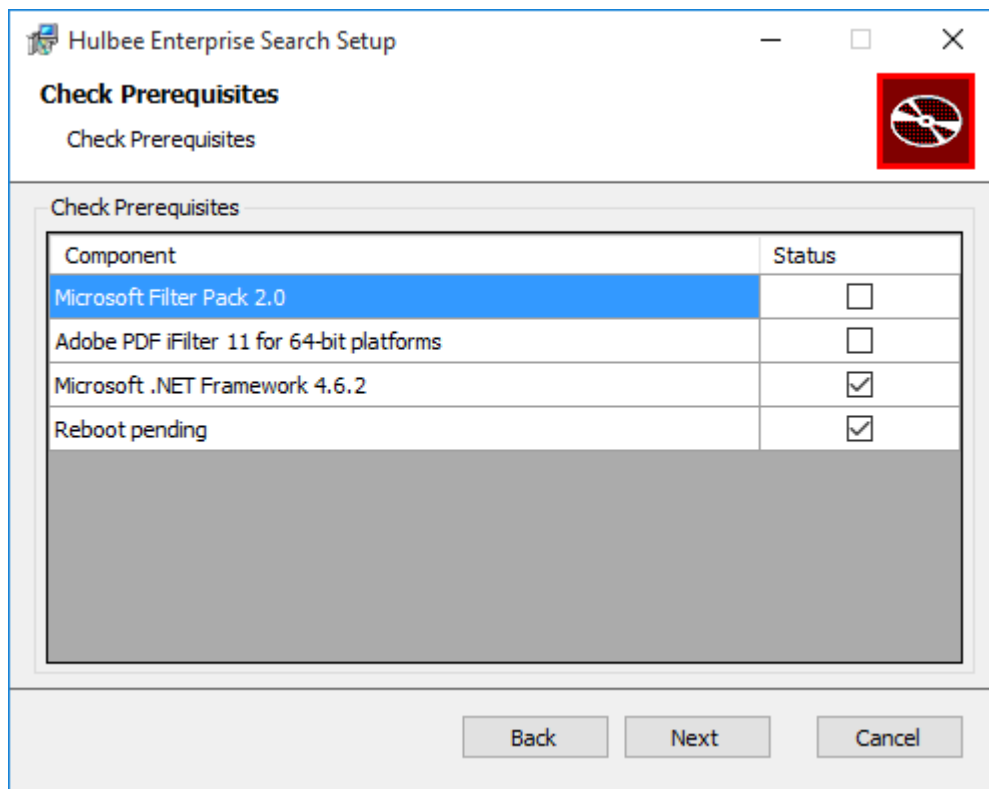


Abb. 5 Schritt 2 der HES-Installationsassistent.

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.

## 5. Schritt 5.

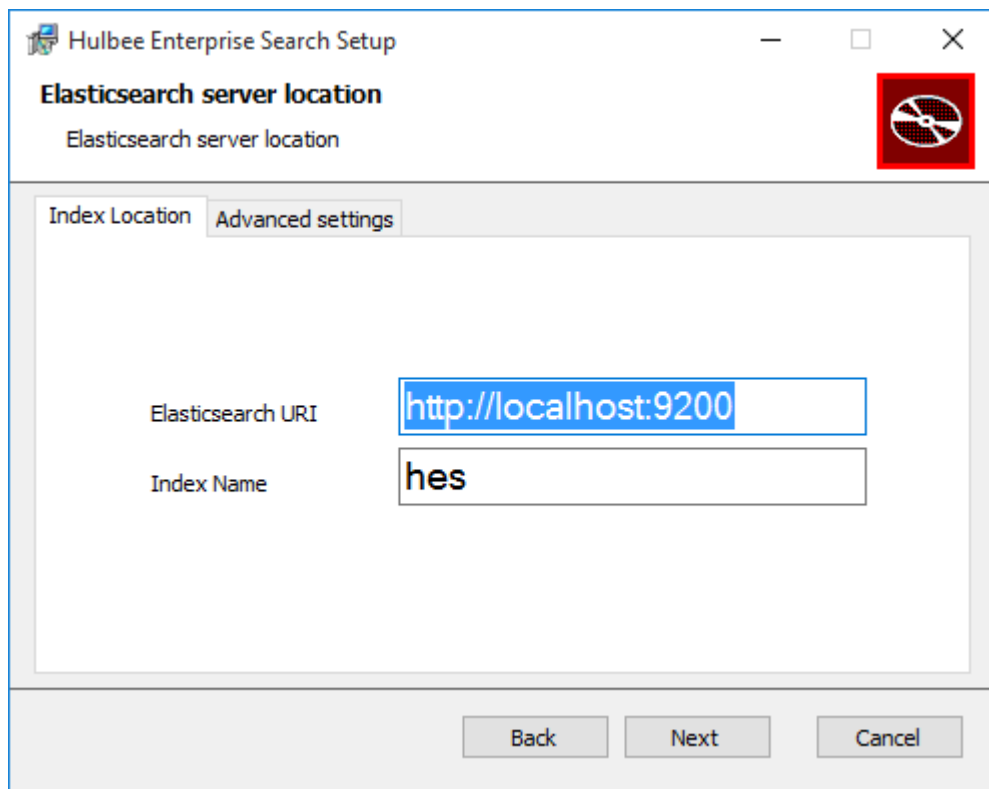


Abb. 6 Schritt 2 der HES-Installationsassistent.

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.

## 6. Schritt 6.

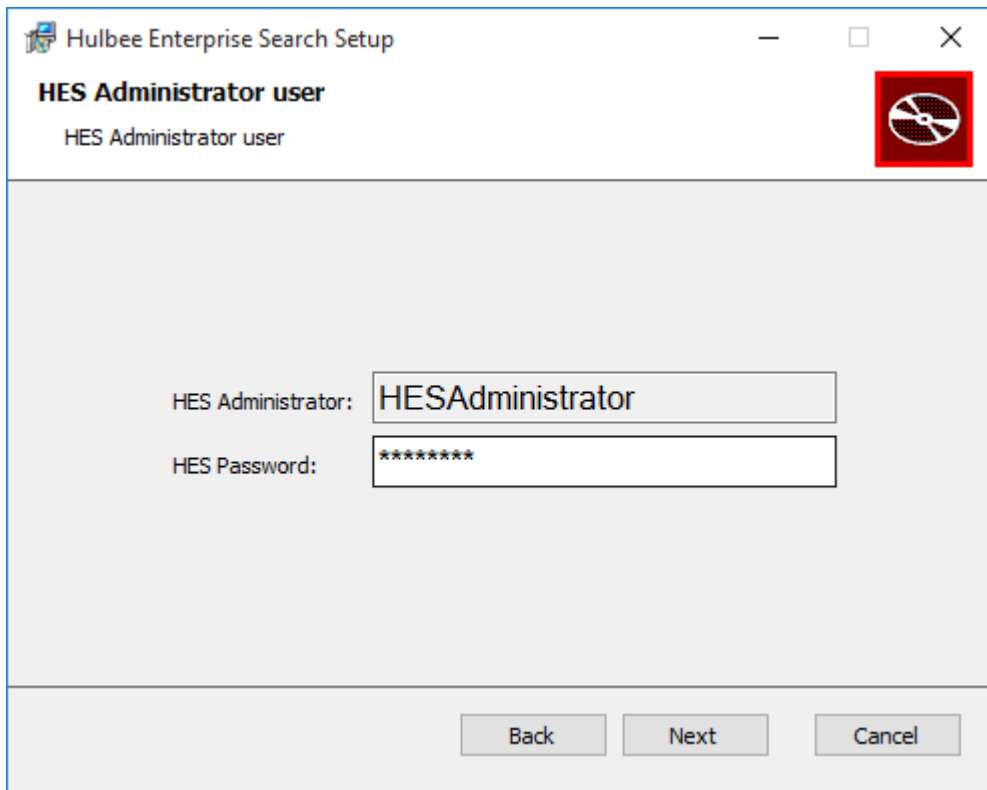


Abb. 7 Schritt 2 der HES-Installationsassistent.

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.

## 7. Schritt 7.

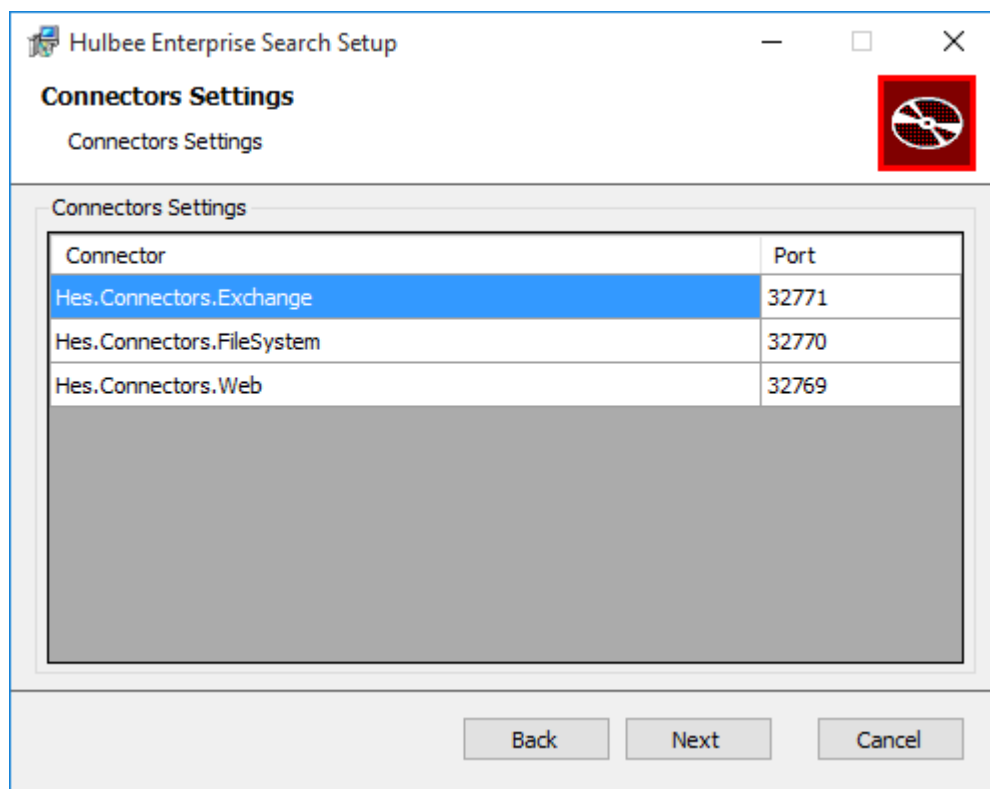


Abb. 8 Schritt 2 der HES-Installationsassistent.

Geben Sie einen anderen Port ein, wenn ein Port besetzt ist.

## 8. Schritt 8.

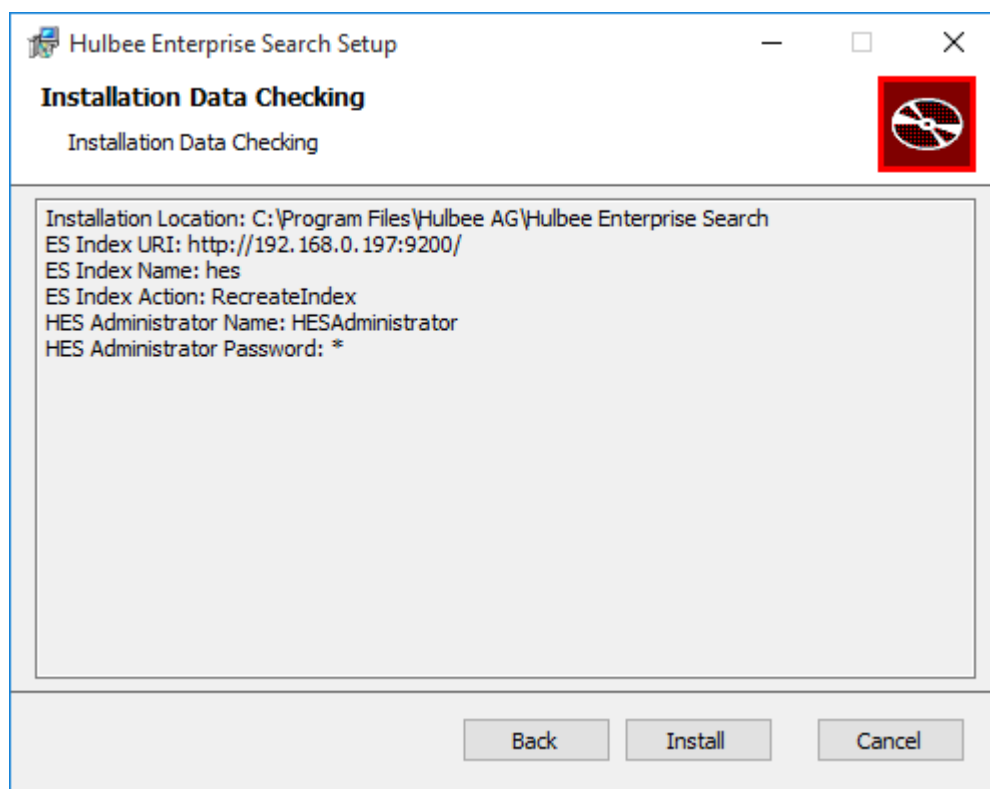


Abb. 9 Schritt 2 der HES-Installationsassistent.

Bericht über die Installation finden Sie unter diesem Link "View Log". Sehen Sie mögliche Probleme, wenn das Setup verweigert wurde.

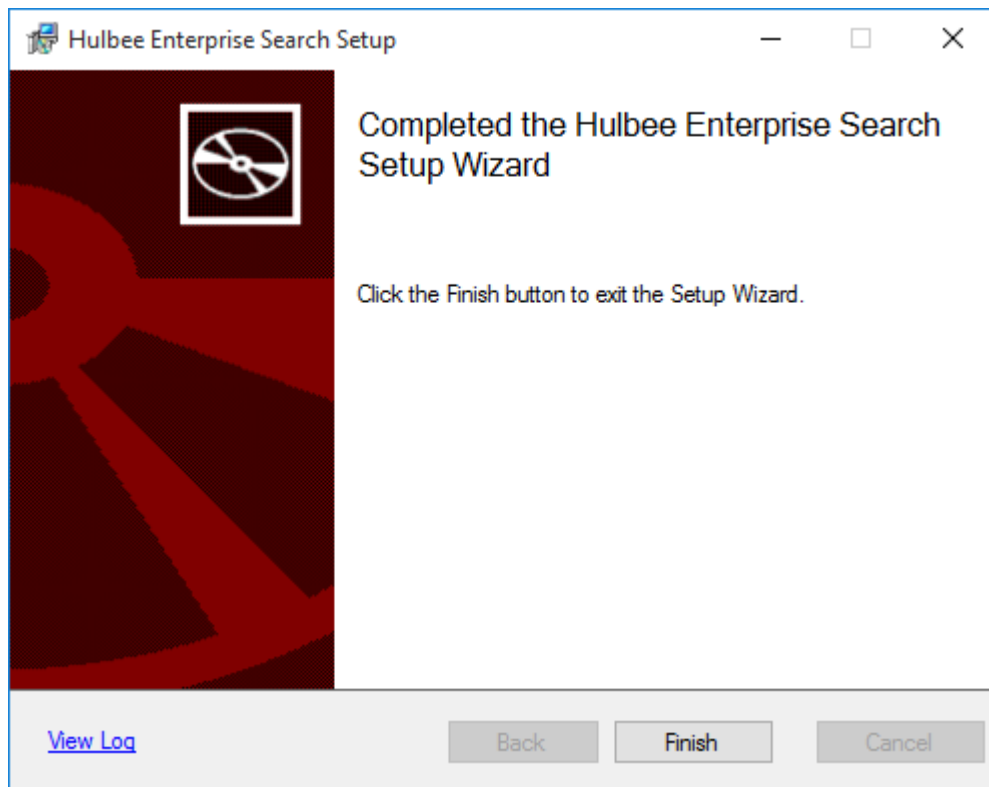


Abb. 10 HES Setup – Setup beendet.

Nach erfolgreichem Setup ist das HES Applications Pack in einem Verzeichnis erreichbar, das in Schritt 2 der Installation angegeben wurde. Es enthält die folgenden Komponenten:

- Connectors
  - Hes.Connectors.Exchange
  - Hes.Connectors.FileSystem
  - Hes.Connectors.Web
- Services
  - Hes.Services.ConnectorManager
  - Hes.Services.IndexCleaner
- Utilities
  - IndexUtil
- Web
  - Ses.Web

Im Lieferumfang von HES enthalten ist auch das Dienstprogramm Desktop Manager zum Öffnen von Dateien aus dem Dateisystem des Unternehmens. Weitere Informationen zur Installation siehe Abschnitt 2.3.2.

Für den ersten Durchlauf (siehe 3) benutzen Sie den gleichen Namen und das gleiche Passwort, das Sie in Schritt 6 der Installation benutzt haben, um die Grundkonfiguration durchzuführen (Abschnitt „Autorisierung“ (siehe 4.6.1) und registrieren Sie die Connectoren (siehe 4.3).



## 2.3 Nach der Installation

### 2.3.1 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](mailto: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.

### 2.3.2 Desktop Manager Installation und Einstellungen

Utility Desktop Manager wird nur verwendet, um Dateien aus dem Dateispeicher des Unternehmens oder dem Ordner mit den Dateien zu öffnen, wenn das Öffnen der Datei nicht möglich ist. Er bietet die Möglichkeit, mit der Original-Datei zu arbeiten und nicht mit der Kopie, die über den Browser heruntergeladen wurde. Dieses Dienstprogramm ist im Benutzerhandbuch (Abschnitt 7.2) beschrieben.

Nach der Installation von HES finden Sie die Datei Hes.Desktop.Manager.Setup.msi im Installationsordner, namentlich Web\Ses.Web\Downloads\.

Desktop Manager sollte auf dem Client-Computer installiert und gestartet werden. Es gibt zwei Installationsszenarien: individuell für jeden Benutzer und für alle Benutzer des Client-Computers. Die zweite Installationsoption wird nur dann ausgeführt, wenn Sie über Administratorrechte verfügen. Nur ein Szenario kann auf derselben Maschine gleichzeitig durchgeführt werden. Nach der Installation wird das Dienstprogramm zum Autostart hinzugefügt.

Der Benutzer kann den Desktop Manager selbst über die Website HES installieren. Der Vorschlag zur Installation erscheint beim Öffnen einer Datei aus der Dateiablage des Unternehmens, wenn die aktuelle Version des Dienstprogramms noch nicht ausgeführt wurde. Während der Installation wird der Benutzer aufgefordert, die HES URL einzugeben.

Um die Installation der Domain-Benutzer zu vereinfachen, installieren Sie zentral auf allen Computern der Domain folgenden Registrierungsschlüssel:

HKEY\_LOCAL\_MACHINE\SOFTWARE\Hulbee AG\HES

der Zeichenfolgenwert „URL“, in dem die URL zur Installation von HES angegeben wird. Wenn das Installationsprogramm des Desktop Manager solch einen Folgenwert findet, so überspringt er den Einstellungsdialog HES URL und verwendet diesen Wert aus der Registrierung.

Im Kontextmenü des Desktop Manager können Einstellungen im Menüpunkt „Options“ korrigiert werden. Standardmässig ist „prefer HES URL from registry“ aktiviert. Wenn es nötig ist, deaktivieren Sie diese Einstellung und fügen HES URL manuell ein. Bei der Installationsoption „for all users“ im Menü „Options“ kann das Dienstprogramm nur als Administrator ausgeführt werden.

**Achtung!** Führen Sie alle erforderlichen Einstellungen durch, um falsche Desktop Manager Einstellungen auf den einzelnen Computern der Benutzer zu verhindern.

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.

## 2.4 Optimierungen (Tweaks)

Nach der Installation von HES stehen weitere zusätzliche Einstellungen zur Verfügung.

### 2.4.1 IndexUtil and benutzerdefinierte Indexstruktur

Bei der nachfolgenden Installation erstellt das Installationsprogramm einen Index mit Standardeinstellungen, die in den meisten Fällen verwendet werden sollen. Allerdings gibt es Fälle, in denen Sie nicht die übliche Konfiguration erstellen müssen. Dann können wir ihn vorher mit einem Utility IndexUtil (Verteilung in HES) erstellen. Wenn es darum geht welcher Index verwendet werden soll, wird das Installationsprogramm feststellen, dass der Index bereits erstellt wurde und es wird eine Wahl angeboten werden, ihn neu zu erstellen oder zu verwenden, was bereits erstellt wurde.

HES verwendet 2 Verzeichnisse:

Indexname	Inhalt	Standardparameter
{HES}	Textinhalt für die Suche	Shards = 4 Kopien = 1
{HES}.service	Einstellungen und interne Daten	Shards = shards in HES index Kopien = 1

In diesem Fall ist {HES} der Indexname, der vom Dienstprogramm IndexUtil als Parameter oder im Installationsvorgang angegeben wird. Dieser Name wird im Feld „Indexname“ in Schritt 5 der Installation eingetragen (siehe 2.2). Der Name des Unterverzeichnisses wird gebildet, indem dem Namen ein Suffix hinzugefügt wird. Um Informationen über das Einstellungsprogramm und die unterstützten Befehle zu erhalten, starten Sie es mit dem folgenden Schlüssel:

```
IndexUtil.exe /?
```

**Achtung!** Verwenden Sie dieses Dienstprogramm mit Vorsicht, es kann die Integrität des Suchindex beschädigen.

**Achtung!** Wenn Elasticsearch mehrere Nodes hat, vermeiden Sie die Änderung von Einstellungen im Admin-Panel während des Neustarts eines Servers.

## 2.5 HESbox

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

### 2.5.1 Einrichtung einer Domain

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

- Einbindung des HESbox 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.

### 2.5.2 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.

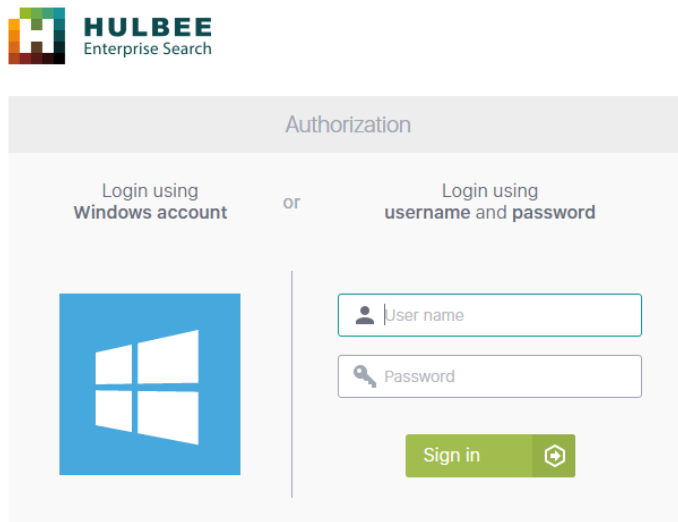
Wenn das Öffnen von einem lokalen Rechner durchgeführt wird, so ist es in der Regel:

<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:



The screenshot shows the 'Authorization' form for HULBEE Enterprise Search. At the top left is the HULBEE logo with the text 'Enterprise Search' below it. The form has a light gray header with the word 'Authorization' in the center. Below the header, there are two login options separated by the word 'or'. The first option is 'Login using Windows account' with a large blue Windows logo icon below it. The second option is 'Login using username and password' with two input fields: 'User name' (with a person icon) and 'Password' (with a key icon). Below these fields is a green 'Sign in' button with a right-pointing arrow icon.

**Abb. 11** 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 (Abschnitt 2.5.1).

Standardmässig sind die Admin-Berechtigungen die folgenden:

Benutzername: *HESAdministrator*

Passwort: *Admin123*

Nach der Anmeldung erscheint die Startseite:



**Abb. 12 HES-Startseite.**

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

Um auf den Admin-Bereich zuzugreifen, folgen Sie dem Link „Administrationsbereich“ (das „Cog“-Symbol in der Nähe des Administratorprofils).

### 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!** Vor der Arbeit am Index, beenden Sie bitte alle HES Dienste und Connectoren im Dienst-Applet.

Die Benutzung vom Dienstprogramm IndexUtil für diese Optionen ist die gleiche wie im Abschnitt 2.4.1 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“ (Abschnitt 4.6.8) 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 (Abschnitt 4.6.1). Geben Sie diese Einstellungen ein, bevor weitere Schritte durchgeführt werden!
2. Hinzufügen der Connectoren (Abschnitt 4.3).
3. Hinzufügen von Speicherplätzen (Abschnitt 4.4).

### 4.3 Registrierung der Connectoren

Standardmässig werden alle Connectoren aus dem HES-Paket während des Installationsvorgangs registriert. Nach dem ersten Durchlauf (siehe) sind sie in der Liste der Connectoren verfügbar.

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 „Connector“ 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. Der gleiche Connector kann mit einer Vielzahl von Speichern des jeweiligen Typs verbunden werden (mehrere Netzwerkspeicher oder mehrere lokale Standorte mit den Dokumenten). Darüber hinaus können die Dokumente mit dem gleichen uri nur in einem Connector-Speicher sein.

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

##### Anschlüsse

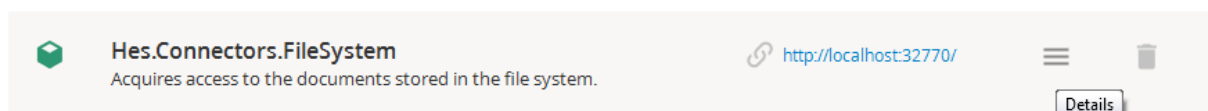


Abb. 13 Bereich „Anschlüsse“, Schaltfläche „Details“.

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

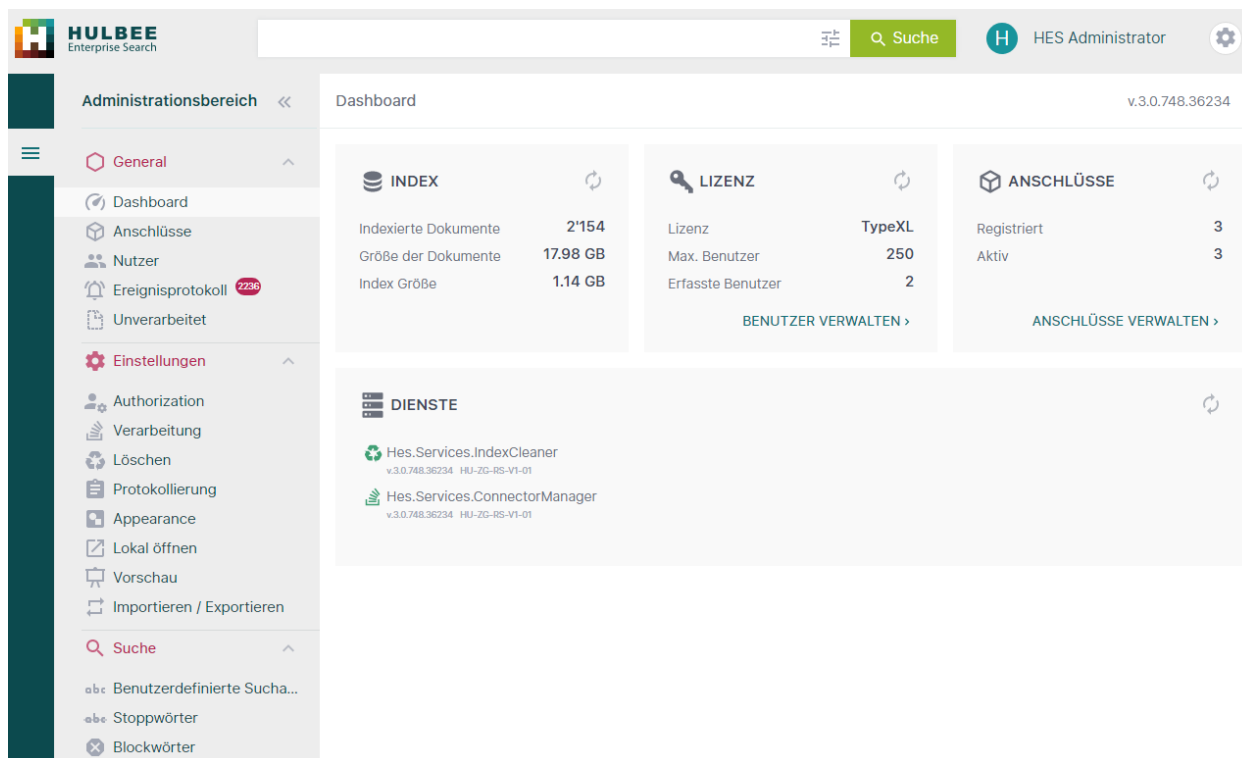


Abb. 14 Administratorsite. Allgemeine Ansicht.

Bei weiteren Screenshots wird nur ein Arbeitsbereich des Admin-Panels angezeigt.

#### 4.5.1 Dashboard

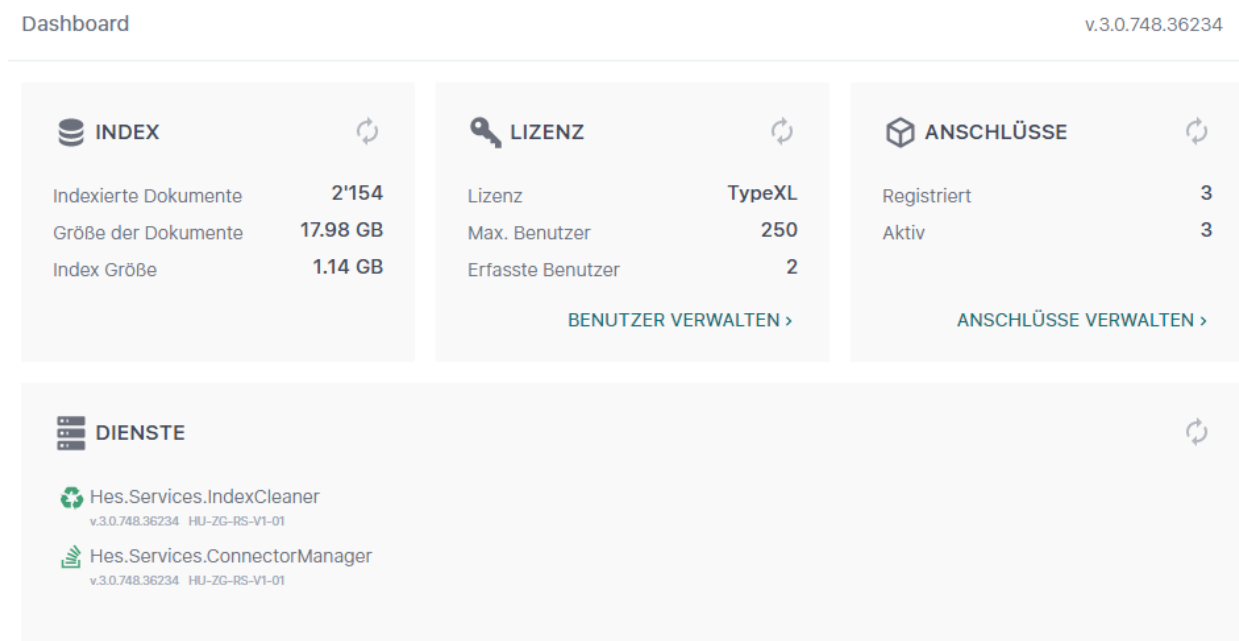


Abb. 15 "Dashboard".

Die Seite stellt vier Widgets dar, auf denen Sie verschiedene Arten von Statistiken sehen können.

"INDEX" Widget zeigt Informationen über den Index, nämlich:



- **Indexed documents.** Anzahl der indexierten Dokumente. Möglicherweise entspricht diese Zahl nicht der Gesamtzahl der Dateien auf allen verfügbaren Ressourcen, 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 kann bei dem von der Datei benötigten Speicherplatz aus den vorher genannten Gründen abweichen.
- **Index size.** Indexgrösse. Elasticsearch auf der Festplatte.

„LIZENZ“ {id} widget ist der Lizenztyp angegeben. Anzahl der Nutzer, die den Server bereits genutzt haben und Anzahl der potenziellen Nutzer.

„ANSCHLÜSSE“. Informationen zu Connectoren die im System registriert sind.

„DIENSTE“. 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

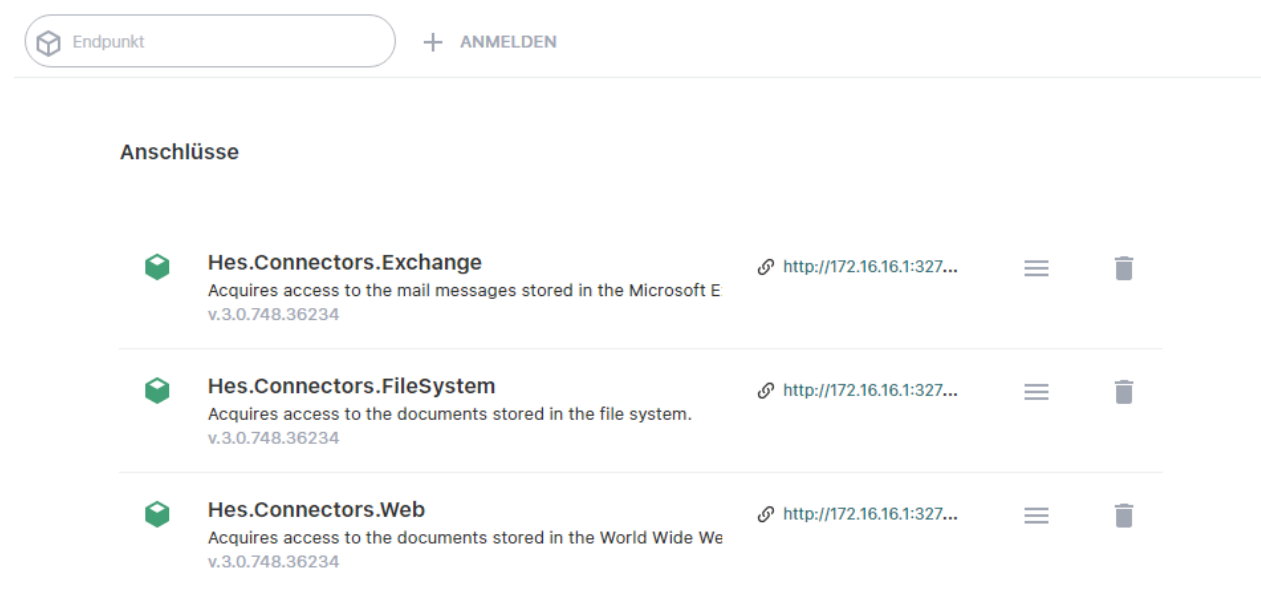


Abb. 16 „Anschlüsse“ Seite.

„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 gleichen Einstellungen stehen nach einem Klick auf den Connector zur Verfügung.

Die Arbeit mit Connectoren erfolgt über http-Protokoll. Das Feld „**Endpunkt**“ dient der Eingabe der Server-url eines bestimmten Connectors, die der Konfigurationsdatei entnommen werden kann im Falle von Connectoren die im HES zur Verfügung gestellt werden. Die url eines benutzerdefinierten Connectors muss bei dem Entwickler des Connectors beantragt werden.

Vorhandene Connectoren können entfernt (Abschnitt 5.1.4) oder neue registriert werden (Abschnitt 4.3).

### 4.5.3 Nutzer

< > Page 1 of 1

#### Users

<input type="checkbox"/>	USER NAME ▲	STATUS	
<input type="checkbox"/>	sces\la	Active ^	
		Active	
		LockedOut	

Abb. 17 „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 („Locked Out“ status). 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 Ereignisprotokoll

< > Seite 1 von 140



Kritisch, , Warnung, Infor... ▼



Hes.Services.ConnectorM... ▼

#### Ereignisprotokoll

●

**Heute, 10:29:00** **Start**

Started processing of the feed Demo@Hes.Connectors.FileSystem .

Hes.Services.ConnectorManager

●

**Heute, 10:29:00** **Warning**

Could not find any active feed on connector Hes.Connectors.Exchange .

Hes.Services.ConnectorManager

Abb. 18 „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 5

### Unverarbeitete Dokumente

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.





<input type="checkbox"/>	STATUS	URI		
<input type="checkbox"/>	Aborted	<a href="\\hu-wt-rs-v1-09\\Common\\Daten\\Books\\boehlau_ges_werke_5.pdf">\\hu-wt-rs-v1-09\\Common\\Daten\\Books\\boehlau_ges_werke_5.pdf</a> The document is too large to be processed		 Wiederholen
<input type="checkbox"/>	Aborted	<a href="\\hu-wt-rs-v1-09\\Common\\Daten\\Books\\mayreder_rosa_der_corregidor.pdf">\\hu-wt-rs-v1-09\\Common\\Daten\\Books\\mayreder_rosa_der_corregidor.pdf</a> The document is too large to be processed		 Wiederholen

Abb. 19 „Unverarbeitet“ Seite.

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

Verarbeitung von Dateien ist mit dem Wert “Processing timeout” (Abschnitt 4.6.2) 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

#### 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.

hes.hulbee.local

##### NetBIOS Domainname

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

HES

##### 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.

Administrator



##### 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.

<input type="checkbox"/>	IDENTITÄT	ZUGANG	
+	Benutzer oder Gruppe SAMAccountName, z.B. DOMAIN\	Zulassen 	

## API Token

Gewähren Sie Zugriff auf HES Search API mit OAuth2 Zugriffstoken

<input type="checkbox"/>	CLIENT ID	CLIENT SECRET	
+	Geben Sie neue Client-ID ein		

Einstellungen speichern 

Abb. 20 „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, da es in den Einstellungen in symbolischer Form dargestellt wird. Um dieses Problem zu lösen, gehen Sie bitte wie folgt vor:

1. Gehen Sie zur Systemsteuerung.
2. Klicken Sie auf Network and Internet > Network and Sharing Center > Change adapter settings.
3. Wählen Sie die Verbindung aus, für die Sie DNS konfigurieren möchten.

4. Wählen Sie die Registerkarte Network. Unter dieser Verbindung werden die folgenden Elemente verwendet: Wählen Sie Internetprotokoll Version 4 (TCP/IPv4) aus, und klicken Sie dann auf Properties.
5. Klicken Sie auf Advanced und wählen Sie die Registerkarte DNS.
6. 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“:** Der Pfad zu dem Container auf dem Speicher mit dem Benutzer von Active Directory, dessen Daten in den Feldern „Username“ und „Password“ angezeigt werden. Dieses Feld ist optional. Dieses Feld muss in dem Fall ausgefüllt 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.
- **„API Token“:** Gewähren Sie Zugang zu HES Search API mit OAuth 2 Zugangstoken. Diese API erlaubt dem Benutzer, eine alternative Schnittstelle zur HES-Suche (zum Beispiel zum Einfügen in verschiedene CMS) bereitzustellen. Einstellung enthält die Client-ID - String-Wert zur Angabe eines bestimmten API, und Client Secret - der Wert, der automatisch für jede spezifische Client-ID generiert wird. Das erhaltene Paar "Client Secret: Client-ID" wird für die Generierung von API-Token verwendet und vom Administrator an Entwickler weitergegeben, die es benötigen, um mit den Anwendungen zu arbeiten. Eine Anwendung, die ein bestimmtes API-Token verwendet, hat Zugriff auf die Daten, auf die der bestimmte Benutzer (Benutzername und Kennwort) zugreifen kann. Es muss ein vorhandener Benutzer der Domäne sein. Der HES-Administrator kann nicht als solcher Benutzer handeln. Die Dokumentation der HES-Search-API finden Sie unter: {URL\_to\_HES}/api/docs.

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

### Verarbeitung



#### Anzahl der Parallelarbeiten

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

#### Extraction Tool

The type of content extraction tool which should be used to process documents.

Configure the type of Extraction Tool depending on the document MIME type.

<input type="checkbox"/>	MIME TYPE	EXTRACTION TOOL	
+	Type MIME type, extension or Document Format	Keine 	

#### 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.

Abb. 21 „Verarbeitung“ Seite.

Wählen Sie „Verarbeitung“ und füllen Sie folgende Felder aus:

- **“Concurrent workers count”:** Die maximale Anzahl von gleichzeitigen Operationen, die von der Verarbeitungsmaschine ausgeführt werden können. 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 HESbox 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.
- **“Extraction Tool”:** Die Art des Inhalts-Extraktionswerkzeugs, das zur Bearbeitung von Dokumenten verwendet werden soll. Es ist möglich Ixonik Konverter, Apache Tika Konverter (siehe 6.1.1) oder keinen Konverter zu verwenden. Im letzteren Fall wird der

Inhalt einiger Dateien, meist mit Text: Archiv, Mails, etc. nicht verarbeitet. Hier sind Meta-Angaben zugänglich, die aus dem Connector entnommen werden. Mit der Einstellung können Sie bestimmte Konverter für Dokumente mit bestimmten MIME TYPE (z.B. Video / \*) angeben.

- **“Processing 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“ (Abschnitt 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). Einstellung ermöglicht die Eingabe von Grenzwerten für die Dokumentengröße abhängig von Dokumenten MIME TYPE (Video/\*, Video/x-msvideo).
- **“Maximum size of the extracted content”**: 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.





### Der maximal zu verarbeiteten Grenzwert

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

^  
v

Configure the maximum document size depending on the document MIME type.

	MIME TYPE	GRENZWERT	
+	Type MIME type, extension or Document Format	107374182400	

### Die maximale extrahierte Größe

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

^  
v

### Die maximale Größe des Anhangs

Die maximale Grösse der Dokumentanhänge in Bytes.

^  
v

Einstellungen speichern ✓

Abb. 22 „Verarbeitung“ Seite, Abschluss.

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 eine neue Indexierung des Speichers benötigt wird.

## 4.6.3 Löschen

### Löschen

#### Anzahl gleichzeitig Beschäftigter

Die maximale Anzahl gleichzeitiger Vorgänge, die vom Reiniger ausgeführt werden können. Es wird empfohlen, dass dieser Wert nicht kleiner als die Anzahl der registrierten Connectoren ist.

^  
v

#### Intervall zum Ermitteln der Connectoren

Das Intervall zwischen den Connector-Ermittlungs-Sessions. Der Wert muss das Format "HH:MM:SS", z.B. 00:05:00 – fünf Minuten haben.

Einstellungen speichern ✓

Abb. 23 „Löschen“ Seite.

Bitte wählen Sie „Löschen“ und passen Sie die folgenden Felder an:

- **“Concurrent workers count”**: Die maximale Anzahl von gleichzeitigen Operationen die durch den Cleaner ausgeführt werden können. Es wird empfohlen, dass dieser Wert nicht kleiner als die Anzahl der registrierten Connectoren ist.
- **“Connectors discovery interval”**: Das Intervall zwischen den Sessions, die Zeit in welcher der Cleaner überprüft, welche Connectoren vorhanden sind. Es wird empfohlen, diese Überprüfung nicht zu häufig (nicht öfter als einmal in 5 min) durchzuführen, da während der Ausführung der Index abgefragt wird. Dieses Verfahren kann das Suchen verlangsamen, das von den Nutzern in dieser Zeit durchgeführt wird. Der Wert muss im Format "HH:MM:SS", z.B. 00:05:00 - fünf Minuten, angegeben werden.

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

#### 4.6.4 Protokollierung

##### Protokollierung

Aufzeichnungstiefe

Information

Entfernt automatisch Protokolle älter als

Woche

Einstellungen speichern



Event types

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

Vor

12.04.2017



Löschen



Abb. 24 „Protokollierung“ Seite.

Die Seite Protokollierung ermöglicht folgende Einstellungen:

- **“Trace level”**: Hier öffnet sich eine Liste, in der eingestellt werden kann, welche Ereignisse im Systemjournal aufgezeichnet werden sollen. Hes.Services.ConnectorManager und IndexCleaner werden nach dem Neustart neue Einstellungen verwenden.

- **“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:\ProgramData\Hulbee AG\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 Appearance

Mit diesen Einstellungen ist es möglich, das Aussehen einiger Elemente der HES-Startseite wie Titel, Untertitel und Bannerbild zu ändern.

##### Appearance

###### Application Title

The value will be used as the HES web-application title, shown on the homepage, in a browser's title bar or on the page's tab.

###### Application Subtitle

The value will be used as the HES web-application subtitle, shown on the homepage, in a browser's title bar or on the page's tab.

###### Homepage Banner Image

The background image which is displayed on the HES homepage.









Abb. 25 “Appearance” Seite.

#### 4.6.6 Local öffnen

Diese Einstellung bietet die Möglichkeit, die Standardeinstellung für Dokumente zu konfigurieren, je nach Dateierweiterung (.exe, .zip). Für Dateien, die über das Protokoll übertragen werden und sich im lokalen Netzwerk befinden, ist es möglich, die Originaldatei oder den Ordner mit der Datei zu

öffnen. Verwenden Sie dazu das DesktopManager-Dienstprogramm. Der Wert von None geht davon aus, dass eine weitere Arbeit mit dem Dokument nicht auf der Ergebnisanzeige verfügbar ist.

### Lokal öffnen

Sie können die Standardmethode für das Öffnen von Dokumenten konfigurieren, abhängig von der Dateierweiterung.

<input type="checkbox"/>	EXTENSION	ACTION	
<input type="checkbox"/>	(Standard)	Ausführen oder Ordne...	
<input type="checkbox"/>	.pdf	Ausführen oder Ordne...	
+	Geben Sie eine Dateierweiterung ein	Ausführen oder Ordne...	

Einstellungen speichern

Abb. 26 "Lokal öffnen" Seite.

## 4.6.7 Vorschau

### Vorschau

#### Maximale Dokumentengröße

Dokumente, deren Größe in Byte den angegebenen Wert überschreitet, stehen für die Vorschau nicht zur Verfügung.

#### Per-domain Vorschauereinstellungen

Bestimmen Sie die per-domain Vorschauereinstellungen, wie z.B. Proxy für Web-Dokumente verwenden, JavaScript erlauben oder verweigern.

<input type="checkbox"/>	DOMAIN	PROXY BEN...	SCRIPT ER...	
<input type="checkbox"/>	default	<input checked="" type="checkbox"/>	<input type="checkbox"/>	
+	URL-Adresse eingeben...	<input type="checkbox"/>	<input checked="" type="checkbox"/>	

Einstellungen speichern

Abb. 27. „Vorschau“ Seite.

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.

Gehen Sie folgendermassen vor, um Proxy-Einstellungen für Ihre Website hinzuzufügen:

1. Dafür fügen Sie die URL Ihrer Website in das Feld „type url address...“.
2. Nehmen das Häkchen in der Spalte USE PROXY raus.
3. Einstellungen speichern.

Wenn vor dem Anschauen einer Seite in der Vorschau der Website eine Authentifizierung aufgefodert wird, so geben Sie Ihren Benutzernamen und Ihr Passwort ein.

## 4.6.8 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.

Choose File No file chosen



Abb. 28. „Importieren/Exportieren“ Seite.

Die Importieren/Exportieren 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.

## 4.7 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

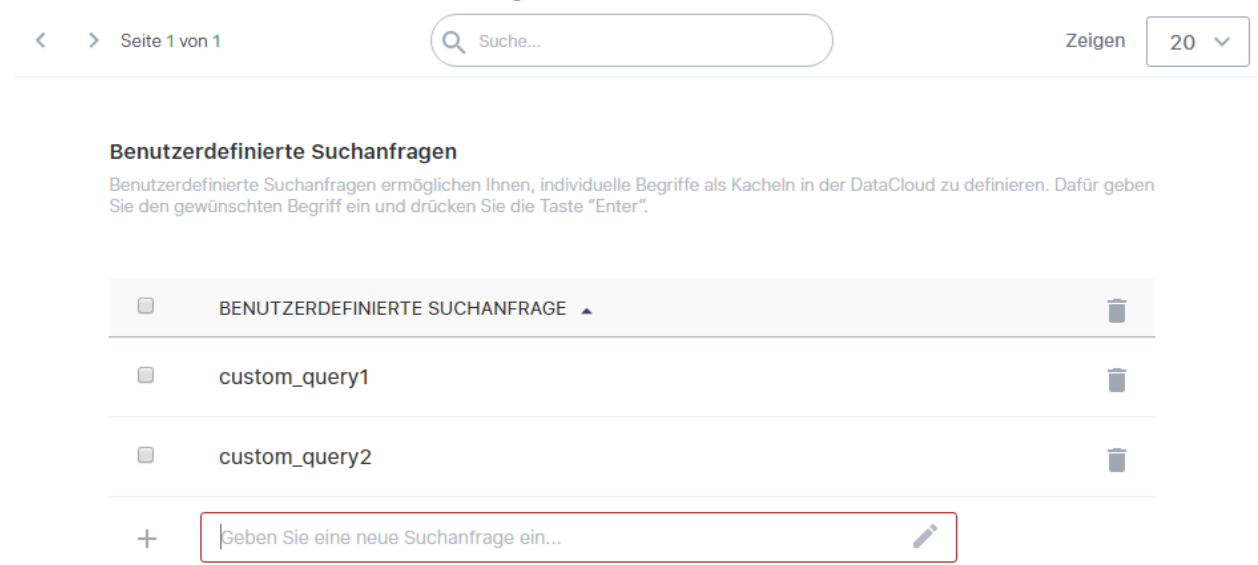


Abb. 29. „Benutzerdefinierte Suchanfragen“ Seite.

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

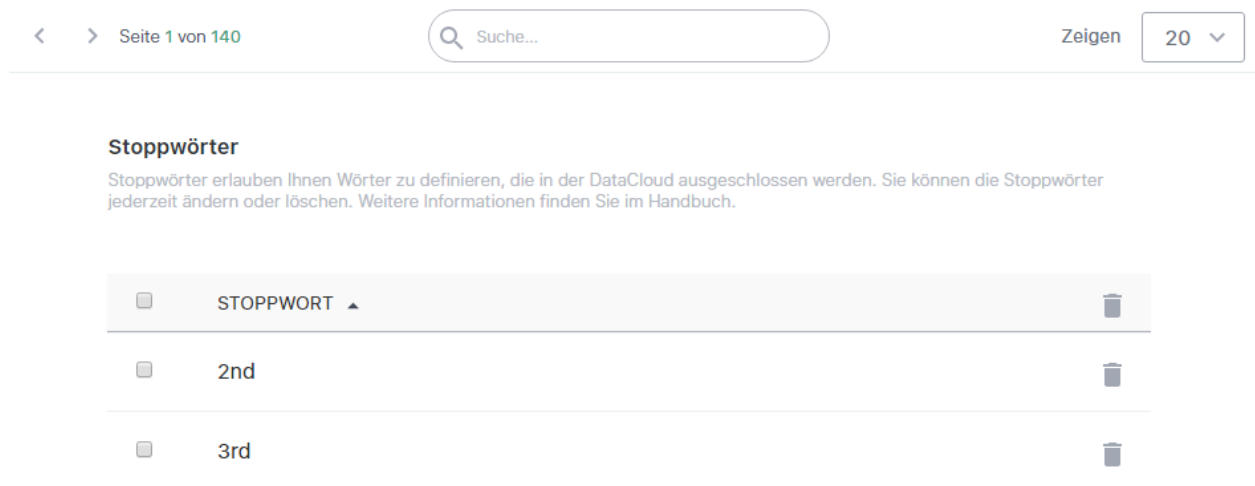


Abb. 30. „Stoppwörter“ Seite.

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 „Stoppwörter“ Registerkarte können Sie Wörter angeben, die nicht in der Data Cloud erscheinen werden. 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.

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

## 4.7.3 Blockwörter

Die Liste der Schlüsselwörter, die die Anforderung blockieren, wenn die Suchanfrage einen oder mehrere Einträge aus der Liste enthält.

# 5 Connectoren

## 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 Connectoren 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
```

HES Installer führt diese Operation automatisch für http-Protokoll aus. 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 „ALLGEMEIN“ 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“ (Abb. 10), über die Sie zu den Registerkarten (die Einstellungen für Connectoren) gehen können.

Auf der Registerkarte „ALLGEMEIN“ sind folgende Einstellungen:

☒ **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.

Exchange Connector

**Authentifizierungsverfahren**

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

None

Einstellungen speichern






Abb. 31. Anschlüsse. „ALLGEMEIN“.

- **“Enable data processing”**: Definiert ob die Daten, die vom Connector kommen, verarbeitet werden sollten. 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“ (Abb. 32) 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.
- **“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 (Abschnitt 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.

Liste der Einstellungen auf der Registerkarte „FEEDS“ wird erst angezeigt sobald sie in einer der Speicherlisten geöffnet sind.

 Demo  http://172.16.16.1:32770/feed/Demo 



☒ **Enable processing**  
Check to enable indexing of the feed.

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

**DURCHLAUF**

☐ **Traverse with Modified-Since condition (experimental)**  
If this feature is enabled, the traversing requests will contain the date of the last successive traversing session to filter out the documents which haven't changed and have been already indexed. This can improve the performance of the feed traversing and speed up the synchronization of the document changes with the search index.  
*Note: the feature is experimental, its behavior could be changed in future.*

**Losgrösse abfragen**  
Die Anzahl der Dokumente pro durchquerende Abfrage abrufen.

**Limit throughput**  
Limit the number of documents which can be processed per certain amount of time.

None ^

**Abfrageintervall**  
Minimale Auszeit in Millisekunden für das Durchqueren der Anfragen.

1000 ^  
v

**REINIGUNG**

**Reinigung der Chargengröße**  
Anzahl der zu analysierenden Dokumente pro Anfrage.

1000 ^  
v

**Abfrageintervall**  
Minimaler Timeout zwischen Reinigungsaufträgen in Millisekunden.

10000 ^  
v

**Zeitüberschreitung der Reinigungsanforderung**  
Die Anzahl der Sekunden, die gewartet wird, bis die Reinigungsanfrage abläuft.

100 ^  
v

Einstellungen speichern ✓

Abb. 32. Anschlüsse. „FEEDS“.

- **“Enable processing”**: Anklicken, um die Indizierung der Dokumente aus dem Feed zu aktivieren.
- **“Display Name”**: Name, der für den Feed in Filtern, Metadaten usw. angezeigt wird.

„DURCHLAUF“ auf der Registerkarte „FEEDS“:

- **“Traverse with Modified-Since condition”**: Ob die Durchlaufanfrage das Datum der letzten aufeinanderfolgenden Durchlaufsitzen enthält, um die Dokumente zu filtern, die sich nicht geändert haben und bereits indiziert wurden. Dies kann den Durchlauf des Feeds verbessern und die Synchronisation der Dokumentänderungen mit dem Suchindex beschleunigen. **Achtung**: Diese Einstellung ist experimentell, also achten Sie auf die korrekte Arbeit des Systems bei der Verwendung dieser Einstellung.
- **“Polling batch size”**: Die Anzahl der Dokumente, die pro Durchlaufanfrage zu holen sind.
- **“Limit throughput”**: Begrenzen Sie die Anzahl der Dokumente, die pro Zeitdauer verarbeitet werden können. Kann verwendet werden, wenn nötig, um die Belastung bei der Verarbeitung von Storage zu reduzieren. Wählen Sie eine der passenden Varianten aus der

Dropdown-Liste aus, verwenden Sie None - keine Limits aufstellen oder Custom - Limit wird manuell eingestellt mit Hilfe der Einstellung "Request interval".

- **"Request interval"** (milliseconds): Minimales Timeout zwischen Anfragen, um Inhalte zu erhalten. Wird manuell eingestellt, nur wenn vorherige Einstellung "Limit throughput" als Custom eingestellt war.
- **"Polling interval"** (milliseconds): Minimaler Timeout zwischen Durchlaufanfragen.

„REINIGUNG“ auf der Registerkarte „FEEDS“:

- **"Cleaning batch size"**: Eine ungefähre Anzahl von zu analysierenden Dokumenten pro Anfrage zur Entfernung der veralteten Daten aus dem Index (Reinigungsanforderung). Der tatsächliche Wert kann je nach Anzahl der Dokumente im Feed variieren. Empfohlener Wert für diese Einstellung ist 1000.
- **"Polling interval"** (milliseconds): Minimales Timeout zwischen Reinigungsaufträgen.
- **"Cleaning request timeout"** (seconds): Die maximale Zeit für eine Antwort von dem Connector nach Reinigungsauftrag. Für Connectoren, die eine lange Zeit brauchen, um auf Reinigungsanforderung zu reagieren, kann die Zeit erhöht werden. Empfohlener Wert für diese Einstellung ist nicht mehr als 100 Sek.

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 „Abmelden“ neben dem jeweiligen Header des Connectors (siehe „Connectors“ Administrations-Bereich).

#### Anschlüsse

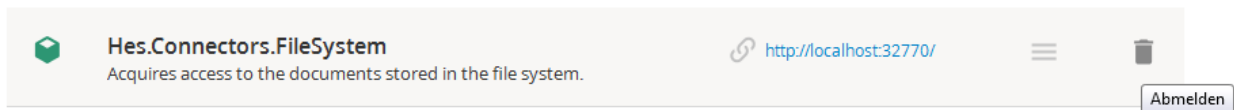


Abb. 33. Connector Bereich, „Abmelden“.

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 „EINSTELLUNGEN“ 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-Connectors

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.
- **“Default checkpoint”**: Der Pfad zu dem Verzeichnis, das im "Scan-Ort" angegeben ist, ist der Anfang für die Indexierung (ist der Startpunkt für den Durchlauf des Speichers). Wenn der Wert nicht angegeben wurde, wird der Speicher von Anfang an durchforstet. Die Einstellung „Default checkpoint“ macht Sinn, wenn beim durchforsten des Speichers aus irgendeinem Grund ein Fehler auftritt und das Durchforsten vom neuem beginnen muss.

- **“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.
- **“Username”, “Password”**: Daten des Benutzers, der in “Access credentials” definiert ist und Zugang zum repository hat, werden in “Scan location” angezeigt. Wenn die Netzwerkressource gefunden wurde, stellt eine Active Directory-Domäne eine Domäne innerhalb des Benutzernamens zur Verfügung (Domäne\Benutzername, Benutzername@Domain).
- **“Feed traversing timeout (milliseconds)”**: Die maximale Zeit, die der Connector beim Sammeln von Dokumenten zur Verarbeitung verbrauchen kann. Wenn das Timeout überschritten wurde, wird die Anzahl der Dokumente angezeigt, die bis dahin gesammelt wurden. 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. Der Wert 0 bedeutet, dass die Antwort sofort gebildet wird, unabhängig von der Anzahl der Dateien die im ersten Verzeichnis gefunden werden, aber der Wert ist nicht höher als “Polling batch size”.
- **“Maximum count of files in directory”**: Maximale Anzahl von Dateien im Verzeichnis (ohne Unterverzeichnisse) für die Verarbeitung. Wenn die Anzahl der Dateien im Verzeichnis grösser ist, wird nur das indiziert was angegeben wurde und in der Registerkarte „DIAGNOSEN“ wird eine entsprechende Warnung angezeigt. Wenn einige grössere Verzeichnisse noch durchlaufen werden müssen, sollte dieser Wert erhöht werden sowie der Wert im "Polling timeout" in der Registerkarte „ALLGEMEIN“. Bei null hat diese Einstellung keine Auswirkung.

### 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 PowerShell url”**: Url von Exchange Server PowerShell. Beispiel: <https://my-exchange.com/PowerShell>.
- **“Use internal link for OWA”**: Ob die interne url von OWA (Outlook Web App) unter Beibehaltung der Verweise auf das Dokument verwendet werden soll. Diese Links werden im Index gespeichert und zum Öffnen der Dokumente in der MS-Exchange Web-Oberfläche verwendet. Im ausgeschalteten Zustand, wird eine externe Referenz verwendet. Externe und interne Links können unterschiedlich sein. Die Konfiguration der internen und externen Links erfolgt über das Exchange Admin Center.
- **“Use internal link for EWS”**: Ob die interne URL für EWS (Exchange Web Services) verwendet werden soll um den Connector für die Kommunikation mit dem Exchange-Server zu aktivieren.
- **“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”**: Active Directory-Container in dem Speicher, der für das Durchsuchen von Benutzerpostfächern verwendet wird. 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 indexiert.
- **“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.
- **“Crawl ‘Drafts’ folder”**: Wenn aktiviert, werden Mail- Nachrichten aus dem Ordner „Entwürfe“ indiziert.
- **“Crawl ‘Junk Email’ folder”**: Wenn aktiviert, werden Mail-Nachrichten aus dem „Junk Email“ (Spam) Ordner indiziert.
- **“Crawl ‘Deleted Items’ folder”**: Wenn aktiviert, werden Mail-Nachrichten aus dem Ordner „Gelöschte Objekte“ indiziert.

## 5.4 Web-Connector

Bietet Zugriff auf die gespeicherten Dokumente im World Wide Web. Enthält Crawler, der die Seiten im lokalen und globalen Netzwerk durchsucht. Crawler hat seinen eigene Zeit für den Durchlauf und



durchläuft kontinuierlich. Durch das Auschecken des Connectors wird der Durchlauf des Crawlers nicht beendet (siehe 5.1.4).

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 Liste der Einstellungen enthält Einstellungen für den Connector sowie für den Crawler. Zu den Einstellungen für Web-Connector gehören folgenden Punkte:

- **“Storage name”**: Name des Speicherorts. Name des Speichers, der verarbeitet werden soll. Wenn er geändert wird, wird der Speicher erneut durchlaufen.
- **“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 “Crawl external pages” und “Crawl external pages links”, 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”**: Die Möglichkeit die Seite zu crawlen, die über den kanonischen Link verfügbar ist (siehe 5.4.4), anstelle der Seite, die über den ursprünglichen Link geöffnet wurde. Das eine oder andere Vorgehen hängt von der Verfügbarkeit von rel=“canonical“ (Canonical URL Tag) ab oder „og:url“ (Open-Graph-meta-tag) im Abschnitt <head> der Seite. Wenn auf der Seite keine Unterstützung der Daten-tags vorhanden ist, dann sollten die Originalseiten durchlaufen werden.
- **“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 Anfragen, bevor ein Dokument aus den gecrawlten Daten gelöscht wird.
- **“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. Dabei können Sie den Zugriff auf Dokumente für eine Gruppe von Benutzern gewähren, aber den Zugriff für bestimmte Benutzer aus dieser Gruppe verbieten (benutzen Sie dafür “Access list denied”).
- **“Access list denied”**: Eine Liste der Benutzer und Gruppen, denen Zugriff auf die Ressource verweigert wird, selbst wenn diese in der Erlaubten-Liste von „Access list allowed“ vorhanden sind.
- **“Active Directory credentials”**: Die Identitätsanmeldeinformationen, die für den Zugriff auf das Active Directory verwendet werden, um die symbolischen Namen von Benutzern und Gruppen in SID zu konvertieren. In den Feldern (“Domain”, “Container”, “User name”, “Password”) werden die Daten eines Benutzers angegeben, der Leserechte aus der Liste der Benutzer aus Active Directory hat. Der Benutzername im entsprechenden Feld sollte dem Domänennamen vorangestellt sein. Dieser Abschnitt sollte ausgefüllt werden, wenn „X-HES-User“ – Funktion verwendet wird und eine Website die X-HES-User-Header-Daten enthält Benutzer/Benutzergruppen in symbolischen Namen zurückgibt (Beispiel: X-HES-Users-Allowed: HES\Domänen –Benutzer;HES\j.doe).
- **“Taboo rules”**: Eine Liste der regulären Ausdrücke zum Ausschluss einiger Webseiten aus gecrawlten Ergebnissen. Jeder einzelne reguläre Ausdruck wird in eine separate Zeile platziert. Auf diese Weise durchläuft der Crawler keine unwichtigen Teile der Website (oder jene, wo er in die Schleife gehen kann, wodurch eine unendliche Anzahl von Seiten vermieden wird). Ein weiteres Szenario der Funktion ist die Aufteilung einer Website auf mehrere Repositories. Beispiele für reguläre Ausdrücke finden Sie unter: [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: Abschnitt der Website, die den GET-Parameter actpos=2 enthält, muss in einem anderen Speicher mit anderen Zugriffseinstellungen abgelegt werden. Geben Sie hierzu „. \*actpos=2.\*“ in die „Taboo rules“ des Primärspeichers ein, sodass alle Seiten, die diesen Parameter enthalten, beim Crawlen übersprungen werden. Und für einen anderen Speicher muss die gleiche „Url“ und ihre eigene Regel hinzugefügt werden „. \*actpos=(?! 2).\*“. Daher werden Seiten mit dem Parameter actpos=2 gecrawlt. Bitte beachten Sie, wenn der reguläre Ausdruck falsch zusammengestellt ist, führt dies zur Überlastung des Systems.

**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. Dazu kann man einen neuen Storage erstellen, aber das führt zu einem Ausfall

von Voreinstellungen. Um dies zu verhindern, ändern Sie einfach seinen Namen im Feld "Storage name" und speichern Sie die Einstellungen.

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:

- **userAgentString** – Die User-Agent-Zeichenfolge, die für HTTP-Anfragen verwendet werden soll.
- **isRespectUrlNamedAnchorOrHashbangEnabled** – Ob URL-Namen Anchors oder Hashbangs als Teil der URL betrachtet werden. Wenn falsch, werden sie ignoriert. Falls zutreffend, werden sie als Teil der URL betrachtet.
- **minAvailableMemoryRequiredInMb** – Verwendet das nächste Vielfache von 16 zu dem eingestellten Wert. Wenn vor dem Starten eines Crawlings nicht mindestens so viel Speicher verfügbar ist, wirft es unzureichende Speicherausnahmen aus. Wenn null, hat diese Einstellung keine Auswirkung.
- **maxMemoryUsageInMb** – Die maximale Menge an Speicher, damit der Prozess verwendet wird. Wenn diese Grenze überschritten wird, stoppt der Crawler vorzeitig. Wenn null, hat die Einstellung keine Wirkung.
- **maxMemoryUsageCacheTimeInSeconds** – Maximale Zeitdauer vor dem Aktualisieren des Werts, der verwendet wird, um die Menge an Speicher zu bestimmen, die von dem Prozess verwendet wird, der die Crawlerinstanz hostet. Dieser Wert hat keine Auswirkung, wenn maxMemoryUsageInMb Null ist.
- **isForcedLinkParsingEnabled** – Legt einen Wert fest, der angibt, ob der Crawler die Seitenverweise analysieren soll, auch wenn einige Einstellungen festlegen, dass diese Links nicht gecrawlt werden.

#### 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-->
```

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 HTTP 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:

- Falsch geschriebene Namen von Benutzern/Benutzergruppen im symbolischen Format (SID-Format). Die Namen der Benutzer/Gruppe von Benutzern, die im SID-Format geschrieben wurden, werden nicht überprüft, so dass es keine Meldung in „DIAGNOSTIK“ gibt.
- 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](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 2.4.\*
- <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.