

Handbuch Technologie Suche

1	Inhalt	
2	Lieferumfang	2
3	Start der Anwendung	2
4	Beenden der Anwendung	2
5	Protokollierung	2
6	Bedienung der der Anwendung	3
6.1	Import von einzelnen Dateien	4
6.2	Anlegen von Web-Crawljobs	4
6.3	Anlegen von Bulkimport-Jobs	5
6.4	Administration	6
6.4.1	Hintergrundverarbeitung	6
6.4.2	Dokumente	7
6.5	Dokumente	7
6.6	Suche	10

2 Lieferumfang

Die Anwendung wird als ZIP-Datei auf einen USB-Datenträger geliefert. Sie besteht aus den Dateien „techsearch_2.7z.001“ bis „techsearch_2.7z.003“. Diese bitte mit dem Programm [7-Zip](#) entpacken. Danach entsteht eine Datei „techsearch.zip“. Bitte diese auch wieder entpacken. Die doppelte Packung ist aufgrund der möglichen Dateigröße auf USB-Sticks notwendig.

Es entstehen zwei Verzeichnisse:

- techsearch
- techsearchData

In „techsearch“ befindet sich die Anwendung. In „techsearchData“ werden die Daten von Modellen gespeichert.

3 Start der Anwendung

In das Verzeichnis „techsearch“ wechseln und das Skript „run.cmd“ starten. Es wird automatisch die Datenbank und danach die Anwendung gestartet

4 Beenden der Anwendung

Im Fenster, dass beim Start der Anwendung geöffnet wurde die Tastenkombination Strg+C drücken. Bei der Frage „Batchvorgang abbrechen (J/N)?“ mit „n“ antworten. Im Skript wird noch die Datenbank und das Python-Backend beendet.

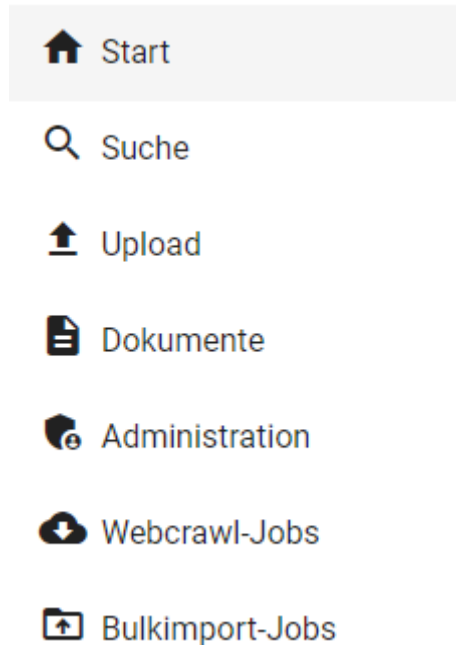
5 Protokollierung

Die Anwendung schreibt ein Protokoll in das Verzeichnis „techsearch\backend\logs\techsearch.log“. Die Datei wird aktuell maximal 10 MB groß. Ältere Dateien werden gepackt abgelegt.

6 Bedienung der der Anwendung

Einen Browser öffnen und die Adresse <http://localhost:8080/> eingeben.

Es wird die Startseite der Anwendung angezeigt. Mit einem Klick links auf das Menüsymbol wird das Menü geöffnet:



Suche: Suche nach Dokumenten.

Upload: Hochladen von einzelnen Dateien

Administration: Wartung des Systems. Vor allem die Hintergrundverwaltung starten.

Webcrawl-Job: Webseiten aus dem Internet herunterladen

Bulkimport-Jobs: Dateien von EPO-Importieren

6.1 Import von einzelnen Dateien

The screenshot shows the 'Technologie Suche' interface. At the top, there is a dark blue header with a menu icon, the text 'Technologie Suche', and links for 'Start' and 'Upload'. Below the header, a light blue box contains the text 'Hierher Dateien ziehen oder [Klicken um Dateien auszuwählen](#)'. Underneath this box is a dropdown menu labeled 'Sprache der Dokumente'. A progress bar labeled 'Fortschritt:' is shown, followed by three buttons: 'Alle Dateien hochladen', 'Alle Dateien abbrechen', and 'Alle Dateien entfernen'. Below the progress bar, it says 'Anzahl der Dateien: 1'. At the bottom, there is a table with columns: 'Name', 'Größe', 'Fortschritt', 'Status', and 'Aktionen'. The table contains one row for 'Technischer Bericht.pdf' (0,078 MB) with a progress bar and three action buttons: 'Hochladen', 'Abbrechen', and 'Entfernen'.

Dateien entweder per Drag&Drop in den grauen Bereich ziehen oder auf den dortigen Link klicken. Danach auf „Alle Dateien hochladen klicken“.

6.2 Anlegen von Web-Crawljobs

The screenshot shows the 'Webcrawl-Jobs' interface. At the top, there is a dark blue header with a menu icon, the text 'Technologie Suche', and links for 'Start' and 'Upload'. Below the header, the title 'Webcrawl-Jobs' is displayed. Underneath the title, there is a button labeled '+ Neuen Webcrawl-Job anlegen' and two search input fields labeled 'Suchen'. A modal window titled 'Webcrwal-Job erstellen' is open in the center. The modal contains four text input fields: 'Name *' (with a subtext 'Name zur einfachen Identifizierung des Jobs'), 'Beschreibung', 'Beschreibung des Jobs', and 'URLs' (with a subtext 'Initiale URLs für das Webcrawling'). At the bottom of the modal, there are two buttons: 'Speichern' and 'Abbrechen'.

Über das Symbol „+“ können neue Jobs angelegt werden.




6.3 Anlegen von Bulkimport-Jobs

The screenshot shows the 'Bulkimport-Jobs' section of the 'Technologie Suche' application. At the top, there is a dark blue header with a menu icon, the text 'Technologie Suche', and links for 'Start' and 'Upload'. Below the header, the main area is titled 'Bulkimport-Jobs'. On the left, there is a button labeled '+ Neuen Bulkimport-Job anlegen'. To the right, there are two search bars, each with the placeholder text 'Suchen'. A modal window titled 'Bulkimport-Job erstellen' is open in the center. It contains the following fields: 'Name *' with a subtext 'Name zur einfachen Identifizierung des Jobs', 'Beschreibung' with a subtext 'Beschreibung des Jobs', and 'Pfade' with a subtext 'Pfade der EPO-Dateien. Der Server muss darauf Zugriff haben!'. Each of these three fields has a small icon in the bottom right corner. At the bottom of the modal, there are two buttons: 'Speichern' and 'Abbrechen'.

Über das Symbol können neue Jobs angelegt werden. Als Dateiformat wird nur das von EPO unterstützt: <https://www.epo.org/searching-for-patents/data/bulk-data-sets/data.html>

6.4 Administration

6.4.1 Hintergrundverarbeitung

 **Technologie Suche**  **Start**  **Upload**

Administration

Hintergrundverarbeitung

Dokumente

Status aktualisieren

Jobname: Doc2Vec-Model erzeugen

Soll der Job laufen: Nein

Läuft der Job: Nein

Letztes Lebenszeichen:

Startzeit:

Joblauf einplanen

Job abbrechen

Jobname: Dokumente vorverarbeiten

Soll der Job laufen: Nein

Läuft der Job: Nein

Letztes Lebenszeichen:

Startzeit:

Joblauf einplanen

Job abbrechen

Jobname: Index neu erzeugen

Soll der Job laufen: Nein

Läuft der Job: Nein

Letztes Lebenszeichen:

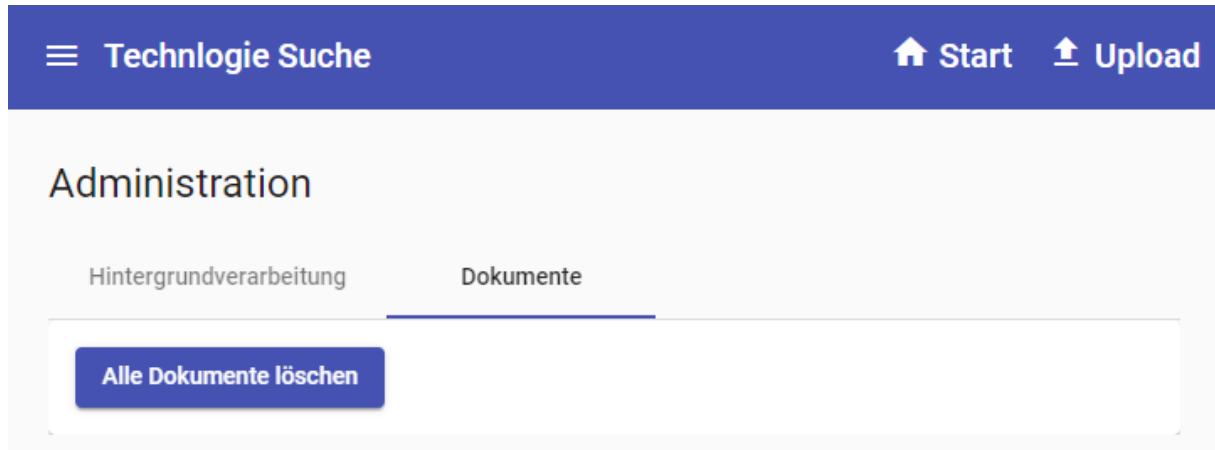
Startzeit:

Joblauf einplanen

Job abbrechen

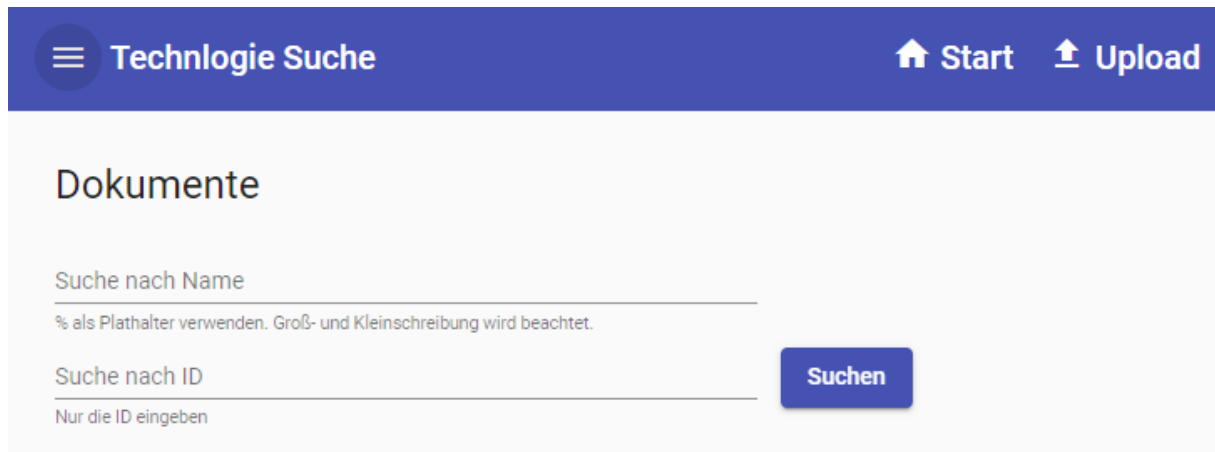
Hier können die einzelnen Jobs gestartet und abgebrochen werden. **Aktuell läuft kein Job automatisch los. Sie müssen Administrativ gestartet werden!**

6.4.2 Dokumente



Hier können alle Dokumente aus der Datenbank gelöscht werden

6.5 Dokumente



Hier kann direkt nach Dokumenten gesucht werden. Die Suche erfolgt entweder über den Namen (erzeugt aus dem Dateinamen) oder über die ID des Dokuments. Bei der Suche über den Namen kann als Platzhalter das Prozentzeichen (%) verwendet werden und es werden mehrere Dokumente geliefert. Die Suche über die ID liefert nur ein Dokument zurück.

Dokumente

%|

% als Platzhalter verwenden. Groß- und Kleinschreibung wird beachtet.

Suche nach ID

Nur die ID eingeben

Suchen

Dokument: [Patentantrag.docx](#)

Id: 102916

Name: Patentantrag.docx

Sprache: ENGLISH

Dokumentenformat: MSWORD

Dokumententyp: UPLOADED

Länge des Originaltextinhalts: 0

Länge des Originalbinärinhalts: 23190

Länge der erkannten Sätze: 0



Zu einem Dokument werden verschiedene Information ausgegeben. Über das Symbol mit dem Mülleimer kann ein Dokument gelöscht werden. Über das Symbol mit dem „i“ können die Information nur zu einem Dokument angezeigt werden:

Dokumentdetails

Id:102916

Name:Patentantrag.docx

Sprache:ENGLISH

Dokumentenformat: MSWORD

Dokumententyp: UPLOADED

Länge des Originaltextinhalts:0

Länge des Originalbinärinhalts:23190

Länge der erkannten Sätze:0

Vorverarbeitung testen

Schließen

Über die Schaltfläche „Vorverarbeitung testen“ kann die Vorverarbeitung nur für dieses Dokument gestartet werden:

Dokumentdetails

Id:102916
Name:Patentantrag.docx
Sprache:ENGLISH
Dokumentenformat: MSWORD
Dokumententyp: UPLOADED

Länge des Originaltextinhalts:0
Länge des Originalbinärinhalts:23190
Länge der erkannten Sätze:0

Vorverarbeitung testen

Sätze:

Kopfzeile Überschrift Unterüberschrift Tabelle Uhrzeit Montag Dienstag Mittwoch Donnerstag Freitag Deutsch
Englisch Mathematik Sport Geschichte Englisch Mathematik Deutsch Geschichte Sport Zu dem Text, gibt es
eine Fußnote footnoteRef Das ist die Fußnote Fußzeile

Sätze lemmatisiert:

Kopfzeile Überschrift Unterüberschrift Tabelle Uhrzeit Montag Dienstag Mittwoch Donnerstag Freitag Deutsch
Englisch Mathematik Sport Geschichte Englisch Mathematik Deutsch Geschichte Sport Zu dem text, gibt es
eine Fußnote footnoteRef Das ist die Fußnote Fußzeile

Sätze analysiert:

```
{
  "G": [
    [
      "Sport", "Zu", "S", "text", "P"
    ],
    [
      "Kopfzeile", "NNP",
      "Überschrift", "NNP",
      "Unterüberschrift", "NNP",
      "Tabelle", "NNP",
      "Uhrzeit", "NNP",
      "Montag", "NNP",
      "Dienstag", "NNP",
      "Mittwoch", "NNP",
      "Donnerstag", "NNP",
      "Freitag", "NNP",
      "Deutsch", "NNP",
      "Englisch", "NNP",
      "Mathematik", "NNP",
      "Sport", "NNP",
      "Geschichte", "NNP",
      "Englisch", "NNP",
      "Mathematik", "NNP",
      "Deutsch", "NNP",
      "Geschichte", "NNP",
      "Sport", "NNP",
      "Zu", "NNP",
      "dem", "DT",
      "text", "VBP",
      "gibt", "FW",
      "es", "FW",
      "eine", "FW",
      "Fußnote footnoteRef", "NNP",
      "Das", "NNP",
      "ist", "NNP",
      "die", "VBP",
      "Fußnote", "NNP",
      "Fußzeile", "NNP"
    ]
  ]
}
```

Schließen

6.6 Suche

Über die Suche können die verschiedenen Modelle abgefragt werden:

The screenshot shows the 'Technologie Suche' web application. The header is dark blue with a hamburger menu icon, the text 'Technologie Suche', and links for 'Start' and 'Upload'. The main content area is titled 'Suche' and features a horizontal tab bar with six options: 'Index', 'Word2Vec', 'Doc2Vec', 'Sent2Vec', 'POS-Modell', and 't-SNE'. The 'Index' tab is currently selected. Below the tabs, there is a large text input field labeled 'Anfrage', a language selection dropdown menu labeled 'Sprache', and a blue button labeled 'Index durchsuchen'.

Über die Registerkarten wird das jeweilige Modell ausgewählt. Dann kann das betreffende Modell abgefragt werden.