

# Tabellen und Views in der Datenbank

Tabelle „Rohdaten“

Spaltenname	Datentyp	Inhalt
Id (Primary Key)	Integer (serial)	Identifikation der einzelnen Stellenanzeige
Seite	Text	Angabe der Stellen Börse
seiten_inhalt_html	Text	Inhalt der gescrapten Daten als HTML
seiten_inhalt	Text	Inhalt der gescrapten Daten als Langtext
url	Text	URL der Stellenanzeige
Datum	Zeitstempel ohne Zeitzone	Zeitstempel wann gescrapt wurde
storno	boolean	False = Stellenanzeige existiert True = Stellenanzeige existiert nicht

Es werden alle gescrapten Stellenanzeigen hinzugefügt.

Tabelle „LOG-Storno“

Spaltenname	Datentyp	Inhalt
Id (Primary Key)	Int (serial)	Identifikation der einzelnen Stellenanzeige
Seite	Text	Angabe der Stellen Börse
url	Text	URL des Stellenanzeige
grund	Text	Warum ist die Stellenanzeige nicht mehr verfügbar
Datum	Zeitstempel ohne Zeitzone	Zeitstempel wann die Änderung festgestellt wurde

Es wird dokumentiert, wann eine Stellenanzeige nicht mehr verfügbar ist.

## Tabelle „daten\_aufgearbeitet“

Spaltenname	Datentyp	Inhalt
id	BIGINT	Identifikation der einzelnen Stellenanzeige
beruf	Text	Jobtitel
firma	Text	Name des Arbeitgebers
ort	Text	Ort der Arbeitsstelle
anstellungsart	Text	Vollzeit, Teilzeit, etc.
beschreibung	Text	Beschreibung von Stelle, Tätigkeit, Firma, etc.

Abspeicherung der Extrahierten Daten aus den View „aktuell“

## View „Anzahl nach“

Spaltenname	Datentyp	Inhalt
Seite	Text	Angabe der Stellen Börse
anzahl	Integer	Anzahl der Stellenanzeigen
gesamtsumme	Integer	Anzahl aller Stellenanzeigen

Es wird die jeweilige Anzahl der einzelnen Stellenbörsen und die Gesamtanzahl aller Stellenbörsen ausgegeben.

## Tabelle „aufgeschlüsselt1“

Spaltenname	Datentyp	Inhalt
Id (Primary Key)	Int (serial)	Identifikation der einzelnen Stellenanzeige
seite	Text	Angabe der Stellen Börse
datum	Zeitstempel ohne Zeitzone	Zeitstempel des Hinzufügens zur Tabelle
storno	Boolean	Stellenanzeige noch verfügbar?
antiskills_1	Text	Skills, die gefordert werden, aber bei Data-Craft nicht gelehrt werden
antiskills_2	Text	Skills, die gefordert werden, aber bei Data-Craft nicht gelehrt werden

antiskills_3	Text	Skills, die gefordert werden, aber bei Data-Craft nicht gelehrt werden
antiskills_4	Text	Skills, die gefordert werden, aber bei Data-Craft nicht gelehrt werden
antiskills_5	Text	Skills, die gefordert werden, aber bei Data-Craft nicht gelehrt werden
antiskills_6	Text	Skills, die gefordert werden, aber bei Data-Craft nicht gelehrt werden
antiskills_7	Text	Skills, die gefordert werden, aber bei Data-Craft nicht gelehrt werden
antiskills_8	Text	Skills, die gefordert werden, aber bei Data-Craft nicht gelehrt werden
antiskills_9	Text	Skills, die gefordert werden, aber bei Data-Craft nicht gelehrt werden
antiskills_10	Text	Skills, die gefordert werden, aber bei Data-Craft nicht gelehrt werden
antiskills_11	Text	Skills, die gefordert werden, aber bei Data-Craft nicht gelehrt werden
antiskills_12	Text	Skills, die gefordert werden, aber bei Data-Craft nicht gelehrt werden
antiskills_13	Text	Skills, die gefordert werden, aber bei Data-Craft nicht gelehrt werden
antiskills_14	Text	Skills, die gefordert werden, aber bei Data-Craft nicht gelehrt werden
antiskills_15	Text	Skills, die gefordert werden, aber bei Data-Craft nicht gelehrt werden
antiskills_16	Text	Skills, die gefordert werden, aber bei Data-Craft nicht gelehrt werden
skills_1	Text	Skills, die gefordert sind und bei Data-Craft gelehrt werden
skills_2	Text	Skills, die gefordert sind und bei Data-Craft gelehrt werden

skills_3	Text	Skills, die gefordert sind und bei Data-Craft gelehrt werden
skills_4	Text	Skills, die gefordert sind und bei Data-Craft gelehrt werden
skills_5	Text	Skills, die gefordert sind und bei Data-Craft gelehrt werden
skills_6	Text	Skills, die gefordert sind und bei Data-Craft gelehrt werden
skills_7	Text	Skills, die gefordert sind und bei Data-Craft gelehrt werden
skills_8	Text	Skills, die gefordert sind und bei Data-Craft gelehrt werden
skills_9	Text	Skills, die gefordert sind und bei Data-Craft gelehrt werden
skills_10	Text	Skills, die gefordert sind und bei Data-Craft gelehrt werden
benefits_1	Text	Was bietet die Firma ihren Mitarbeitern?
benefits_2	Text	Was bietet die Firma ihren Mitarbeitern?
benefits_3	Text	Was bietet die Firma ihren Mitarbeitern?
benefits_4	Text	Was bietet die Firma ihren Mitarbeitern?
benefits_5	Text	Was bietet die Firma ihren Mitarbeitern?
benefits_6	Text	Was bietet die Firma ihren Mitarbeitern?
benefits_7	Text	Was bietet die Firma ihren Mitarbeitern?
benefits_8	Text	Was bietet die Firma ihren Mitarbeitern?
benefits_9	Text	Was bietet die Firma ihren Mitarbeitern?
benefits_10	Text	Was bietet die Firma ihren Mitarbeitern?

Es werden aus dem Langtext der Spalte „seiten\_inhalt“ der Rohdatentabelle drei unterschiedliche Kategorien herausgefiltert. Diese werden in einzelne Spalten aufgeteilt. Die Spaltenanzahl kann sich bei mehr Übereinstimmungen noch erhöhen. Hier wird dann jeweils die letzte Zahl um eins erhöht.

Tabelle “aufgeschlüsselt2”

Spaltenname	Datentyp	Inhalt
Id (Primary Key)	Int (serial)	Identifikation der einzelnen Stellenanzeige
seite	Text	Angabe der Stellen Börse
datum	Zeitstempel ohne Zeitzone	Zeitstempel des Hinzufügens zur Tabelle
storno	Boolean	Stellenanzeige noch verfügbar?
jobtitel_1	Text	Jobtitel falls angegeben aus der Recherche
jobtitel_2	Text	Jobtitel falls angegeben aus der Recherche
jobtitel_3	Text	Jobtitel falls angegeben aus der Recherche
jobtitel_4	Text	Jobtitel falls angegeben aus der Recherche
jobtitel_5	Text	Jobtitel falls angegeben aus der Recherche
jobtitel_6	Text	Jobtitel falls angegeben aus der Recherche
jobtitel_7	Text	Jobtitel falls angegeben aus der Recherche
jobtitel_8	Text	Jobtitel falls angegeben aus der Recherche
jobtitel_9	Text	Jobtitel falls angegeben aus der Recherche
jobtitel_10	Text	Jobtitel falls angegeben aus der Recherche
jobtitel_11	Text	Jobtitel falls angegeben aus der Recherche
jobtitel_12	Text	Jobtitel falls angegeben aus der Recherche
jobtitel_13	Text	Jobtitel falls angegeben aus der Recherche
anstellungsart_1	Text	die Zeitparameter des Jobs sofern angegeben
anstellungsart_2	Text	die Zeitparameter des Jobs sofern angegeben
anstellungsart_3	Text	die Zeitparameter des Jobs sofern angegeben
arbeitsort_1	Text	Präsenz oder remote
arbeitsort_2	Text	Präsenz oder remote

arbeitsort_3	Text	Präsenz oder remote
--------------	------	---------------------

Es werden aus dem Langtext der Spalte „seiten\_inhalt“ der Rohdatentabelle drei unterschiedliche Kategorien herausgefiltert. Diese werden in einzelne Spalten aufgeteilt. Die Spaltenanzahl kann sich bei mehr Übereinstimmungen noch erhöhen. Hier wird dann jeweils die letzte Zahl um eins erhöht.

## View „abgelaufen“

Spaltenname	Datentyp	Inhalt
id	Int (serial)	Identifikation der einzelnen Stellenanzeige
seite	Text	Angabe der Stellen Börse
seiten-inhalt	Text	Inhalt der gescrapten Daten als Langtext
url	Text	Link zur Stellenanzeige
datum	Zeitstempel ohne Zeitzone	Zeitstempel wann gescrappt wurde
storno	Boolean	False = Stellenanzeige existiert True = Stellenanzeige existiert nicht

Es werden die Stellenanzeigen angezeigt, deren Link abgelaufen ist.

## View „alt“

Spaltenname	Datentyp	Inhalt
id	Int (serial)	Identifikation der einzelnen Stellenanzeige
seite	Text	Angabe der Stellen Börse
seiten-inhalt	Text	Inhalt der gescrapten Daten als Langtext
url	Text	Link zur Stellenanzeige
datum	Zeitstempel ohne Zeitzone	Zeitstempel wann gescrappt wurde
storno	Boolean	False = Stellenanzeige existiert True = Stellenanzeige existiert nicht

Es werden Stellenanzeigen angezeigt, die älter als 60 Tage alt sind.

## View „gestern“

Spaltenname	Datentyp	Inhalt
id	Int (serial)	Identifikation der einzelnen Stellenanzeige
seite	Text	Angabe der Stellen Börse
seiten-inhalt	Text	Inhalt der gescrapten Daten als Langtext
url	Text	Link zur Stellenanzeige
datum	Zeitstempel ohne Zeitzone	Zeitstempel wann gescrapt wurde
storno	Boolean	False = Stellenanzeige existiert True = Stellenanzeige existiert nicht

Es werden die aktuellen Stellenanzeigen von gestern und heute angezeigt, deren Link nicht abgelaufen ist.

### View „aktuell“

id	Int (serial)	Identifikation der einzelnen Stellenanzeige
seite	Text	Angabe der Stellen Börse
seiten_inhalt_html	Text	Inhalt der gescrapten Daten als HTML
seiten-inhalt	Text	Inhalt der gescrapten Daten als Langtext
url	Text	Link zur Stellenanzeige
datum	Zeitstempel ohne Zeitzone	Zeitstempel wann gescrapt wurde
storno	Boolean	False = Stellenanzeige existiert True = Stellenanzeige existiert nicht

Es werden die aktuellen Stellenanzeigen der letzten 2 Tage angezeigt, deren Link nicht abgelaufen ist.

### View „aufgearbeitete\_daten“

Spaltenname	Datentyp	Inhalt
id	BIGINT	Identifikation der einzelnen Stellenanzeige
beruf	Text	Jobtitel
firma	Text	Name des Arbeitgebers

ort	Text	Ort des Arbeitsstelle
anstellungsart	Text	Vollzeit, Teilzeit, etc.
beschreibung	Text	Beschreibung von Stelle, Tätigkeit, Firma, etc.
seite	Text	Angabe der Stellen Börse
datum	Zeitstempel ohne Zeitzone	Zeitstempel wann gecrapt wurde
url	Text	Link zur Stellenanzeige

Es wird eine Zusammenführung aus den Tabellen "Rohdaten" und "daten\_aufgearbeitet" angezeigt.

## View „antiskill\_negativ“, View „antiskill\_positiv“, View „skill\_negativ“, View „skill\_positiv“, View „benefits\_negativ“, View „benefits\_positiv“

Diese Views sind auf Grundlage des Tabelle "aufgeschlüsselt1" erstellt, somit sind die Columns identisch mit der Tabelle "aufgeschlüsselt1".

Die Views antiskill\_negativ, skill\_negativ und benefits\_negativ enthalten die Datensätze, die die jeweiligen skills, antiskills oder benefits auf übereinstimmungen überprüft und nur die zeilen auswirft, in denen keine Übereinstimmungen gefunden wurden.

Die Views antiskill\_positiv, skill\_positiv und benefits\_positiv hingegen werfen nur die Ergebnisse aus, wo eine Übereinstimmung gefunden wurde.

Hierzu reicht es aus, nur die jeweilige Spalte mit der Endung "\_1" zu prüfen.