

Gerichtsentscheidungen in Datenbank schreiben

Der Unique Constraint des Primary Key schlägt fehl, wenn 'nur' Aktenzeichen und Datum als Schlüssel verwendet werden und es mehrere Dokumente mit demselben Aktenzeichen und Datum gibt, daher werden die Anhängsel '_a', '_b' etc. mit in die Datenbank genommen.

Getestet mit `sqlite 3.13.0`.

In []:

```
path_to_db = '../data/bgh.db'
```

In []:

```
path_to_files = # Pfad zu den einzulesenden Dateien hier (als String)
```

In []:

```
import sqlite3, re, os
```

In []:

```
con = sqlite3.connect(path_to_db)
cur = con.cursor()
sql = '''CREATE TABLE entscheidungen (
            aktenzeichen TEXT,
            entschdatum TEXT,
            volltext TEXT,
            PRIMARY KEY (aktenzeichen, entschdatum)
        )'''
cur.execute(sql)
con.close()
```

In []:

```
con = sqlite3.connect(path_to_db)
for file in [f for f in os.listdir(path_to_files) if f.endswith('.txt')]:
    cur = con.cursor()
    aktenzeichen = re.sub('-', '/', re.search(r'XI.*(?=\.txt)', file).group(0))
    entschdatum = file[:10]
    with open(f'{path_to_files}/{file}', 'r') as f:
        volltext = f.read()
    sql = 'INSERT INTO entscheidungen VALUES (?, ?, ?)'
    row = (aktenzeichen, entschdatum, volltext)
    cur.execute(sql, row)
con.commit()
con.close()
```

Ende.