

## Testat 1: Reguläre Ausdrücke v1.3\*

Abgabe: 07. Juli 2021, 23:59 Uhr

**Aufgabenstellung** Die Studierendendaten der HS Fulda sollen in ein neues Datenverwaltungssystem "Best-DBMS" umgezogen werden. Best-DBMS ist restriktiver als das alte System und akzeptiert nur Werte die bestimmten Formaten folgen. Falls nur ein Wert eines Eintrages nicht korrekt formatiert ist, bricht Best-DBMS den Import der gesamten Studierendendaten ab ohne eine Rückmeldung zum Fehler zu geben. Leider wurde das Format der alten Studierendendaten nicht konsequent eingehalten, was das Umziehen der Daten derzeit blockiert.

**Aufgabe** Entwickeln Sie zur Unterstützung der Systemumstellung ein Tool, das das Format der zu importierenden Studierendendaten auf Konsistenz mit Best-DBMS prüft. Die einzelnen Datenfelder der Studierendendaten, deren Formateinschränkungen sowie Positiv- und Negativbeispiele finden Sie in Tabelle 1 auf der nächsten Seite. Für das Tool gelten die folgenden Randbedingungen:

- Es dürfen ausschließlich reguläre Ausdrücke zur Überprüfung des Formates genutzt werden
- Jedes Datenfeld darf nur durch genau einen regulären Ausdruck überprüft werden
- Das Tool soll die falsch formatierten Werte auf der Kommandozeile getrennt durch Zeilenumbrüche ausgeben

Die Programmiersprache, in der das Tool entwickelt wird, ist frei wählbar.

**Studierendendaten** Die umzuziehenden Studierendendaten finden Sie in der Datei "Studierendendaten.txt" ([https://elearning.hs-fulda.de/ai/pluginfile.php/84541/mod\\_assign/introattachment/0/Studierendendaten.txt?forcedownload=1](https://elearning.hs-fulda.de/ai/pluginfile.php/84541/mod_assign/introattachment/0/Studierendendaten.txt?forcedownload=1)). Insgesamt gibt es 500 Einträge zu Studierenden wobei jeder Eintrag einen Wert zu jedem der 7 Datenfelder besitzt. Die einzelnen Werte jedes Eintrags sind durch einen Tabulator (\t) voneinander getrennt. Die verschiedenen Einträge sind durch ein Zeilenende (Windows) voneinander getrennt:

```
S1_Abschluss\t S1_FD-Nr \t S1_Name \t S1_Geb-Datum \t S1_Tel-Nr \t S1_eMail \t S1_Passwort
S2_Abschluss\t S2_FD-Nr \t S2_Name \t S2_Geb-Datum \t S2_Tel-Nr \t S2_eMail \t S2_Passwort
....
```

**Abgabe** Laden Sie den Quellcode Ihres Tools unter dem Namen *Vorname\_Nachname* in Moodle hoch: <https://elearning.hs-fulda.de/ai/mod/assign/view.php?id=49846>. Falls Ihr Tool nicht out-of-the-box auf einem Windows/Linux System mit standardmäßiger Entwicklungsumgebung kompiliert werden kann, geben Sie bitte die benötigten Abhängigkeiten sowie eine Anleitung zum Kompilieren in einer beiliegenden Datei "Readme.txt" an. Tools die nicht ohne Änderung im Quellcode kompiliert werden können werden als nicht abgegebener Versuch gezählt. Die Abgabe schließt automatisch am 07 Juli 2021 um 23:59 Uhr. Spätere Einreichungen werden nicht berücksichtigt.

**Bestehenskriterien** Die Bewertung des Testates erfolgt anhand Ihres Tool unter Eingabe der bereitgestellten Studierendendaten (500 Einträge mit je 7 Werten). Insgesamt gibt es 3500 Werte in den Studierendendaten unter denen sich 100 falsch formatierte Werte befinden. Das Testat gilt als bestanden, wenn in der Ausgabe Ihres Tools mindestens 66 der falsch formatierten Werte und gleichzeitig nicht mehr als 34 der korrekt formatierten Werte enthalten sind. Gruppenabgaben sowie Mehrfacheinreichungen des gleichen Quellcodes zählen als Betrugsversuch und führen automatisch zum Nichtbestehen des Testates. Sollten Sie um ein Datenfeld zu prüfen andere Methoden als reguläre Ausdrücke verwenden bzw. mehrere reguläre Ausdrücke zur Prüfung nutzen, werden die Ausgaben Ihres Tools für das entsprechende Datenfeld nicht gewertet.

**Disclaimer** Die Aufgabenstellung ist fiktional, die Studierendendaten wurden zufällig generiert und jegliche Ähnlichkeit zu existierenden Personen ist rein zufällig. Die beschriebenen Datenformate wurden speziell auf einen sinnvollen Lerneffekt für reguläre Ausdrücke angepasst und entsprechen nicht den tatsächlichen Formaten von realen Anwendungen.

**\*Versionshistorie** V1.1: Einschränkungen im Datenfeld Namen upgedated. V1.2: Einschränkungen in der eMail Adresse präzisiert. V1.3: Bei Datenfeld "Nummer", Bedingung 3 muss die Nummer statt der Durchwahl aus mind. einer Ziffer bestehen.

Datenfeld	Einschränkung von Best-DBMS	Korrekte Beispiele	Fehlerhafte Beispiele
Abschluss	Darf nur die Werte "Bachelor" oder "Master" enthalten	- Bachelor - Master	- bachelor - Diplom
FD-Nummer	Beginnt mit "fd", gefolgt von zwei Kleinbuchstaben mit dem Studiengang (ai, ei, lt, pg, wi) und zum Schluss 4 Ziffern	- fdei2032 - fdlt2000	- fDai2000 - fdai000 - fdaa2000
Name	Besteht aus ein oder mehreren Vornamen gefolgt von ein bis zwei Nachnamen mit den folgenden Einschränkungen: 1. Es dürfen nur Buchstaben des lateinischen Alphabetes verwendet werden (keine deutschen Umlaute, gilt für alle Datenfelder!) 2. Der erste Buchstabe der Vor- und Nachnamen muss groß geschrieben sein. Alle anderen Buchstaben müssen klein geschrieben sein 3. Namen dürfen nur durch den Bindestrich "-" oder ein Leerzeichen getrennt sein 4. Der letzte Vorname und der erste Nachname müssen durch ein Leerzeichen getrennt sein	- Angela Apfel - Adam–August Aue - Beate Barbara Braun-Becker	- zacharias Zoellner - Yves - Xena XiaoXu - Vladimir Voß
Geburtsdatum	Das Datum soll im Format tt.mm.jjjj angegeben sein, wobei tt, mm und jjjj die einzelnen Ziffern des Tages, Monats und Jahres sind mit den folgenden Einschränkungen: 1. Der Tag muss zwischen 1 und 31 liegen 2. Der Monat muss zwischen 1 und 12 liegen Hinweis: Das Jahr sowie die korrekte Anzahl der Tage für einen konkreten Monat (29, 30 oder 31) sollen nicht geprüft werden.	- 04.03.2000 - 31.02.1900	- 3.03.2000 - 04.März.2000 - 32.03.2000 - 04.13.2000
Telefon-Nr.	Besteht aus Ziffern und unterteilt sich der Reihenfolge nach in: Internationale Vorwahl, Vorwahl, Nummer und optionale Durchwahl. Dabei gelten die folgenden Einschränkungen: 1. Die internationale Vorwahl wird entweder angegeben als "0" für nationale Rufnummern oder "+" gefolgt von bis zu zwei Ziffern bei denen die erste Ziffer ungleich 0 ist für internationale Nummern 2. Bei nationalen Nummern folgt die Vorwahl direkt nach der 0, bei internationalen Nummern folgt ein Leerzeichen nach der internationalen Vorwahl 3. Die Länge der Vorwahl, Nummer und Durchwahl ist beliebig, wobei Vorwahl und Nummer mindestens aus einer Ziffer bestehen müssen 4. Vorwahl und Nummer sind durch ein Leerzeichen getrennt 5. Falls eine Durchwahl existiert, ist sie durch Bindestrich "-" von der Nummer getrennt.	- 0661 4212-0001 - +49 661 4212-01 - 0661 4212	- 49 661 4212 - +49661 4212 - +494 661 4212 - 0661 4212 0001 - +04 661 4212
eMail Adresse	Unterteilt sich der Reihenfolge nach in: Den lokalen Teil, das @-Zeichen, den Domänenteil und die Top-Level Domäne. Dabei gelten die folgenden Einschränkungen: 1. Der lokale- und Domänenteil dürfen aus ein bis beliebig vielen Ziffern, Klein- und Großbuchstaben sowie den Sonderzeichen ".", ",", "-", "_" und "\" bestehen 2. Das @-Zeichen trennt den lokalen- und Domänenteil 3. Der Domänenteil muss mit einem Klein- oder Großbuchstaben oder einer Ziffer starten. 4. Die Top-Level Domäne folgt am Ende und besteht aus zwei bis drei Kleinbuchstaben die durch einen Punkt vom Domänenteil getrennt sind.	- foo@bar.com - FOO-@bar.com - foo@bar.co.uk - _,-@bar.com	- @bar.com - foo@?bar.com - foobar.com - foo@,bar.com - foo@bar.come - foo@bar.Com - foo@bar-com
Passwort	Besteht aus Klein- und Großbuchstaben, Ziffern sowie den Sonderzeichen: "!", " ", "\$", "%", "&", "/", "(", ")", "?", ".", "-", "_", "\" und \". Das Passwort muss mindestens einen Kleinbuchstaben, einen Großbuchstaben, eine Ziffer und eine der oben genannten Sonderzeichen enthalten. Die Länge des Passworts ist beliebig. Hinweis: Nutzen Sie Lookaround Operationen von regulären Ausdrücke zur Vereinfachung des Ausdrucks.	- aA4\$	- aA4 - aA\$ - a4\$ - A4\$ - aA4\${}

Table 1: Datenfelder von Best-BMS und deren Einschränkungen. Hinweis: Der Begriff "Buchstaben" bezieht sich immer nur auf die lateinischen Buchstaben, nicht auf deutsche Umlaute.