Ein Generator zur Erzeugung zufälliger Identitäten

Von Oliver Theobald und Tristan Stadler

# Teil 1:

Testwerte:

|  |  |  |
| --- | --- | --- |
| male first names:  Thomas  Benno  Marvin  Erich  Timon  Keanu  Birger  Igor  Immanuel  Geoffrey  Dietmar  Joah  Gilbert  Eugen  Dario  Dirk  Grant  Hans  André  Rupert  Wieland  Djordje  Mahomet  Linus  Winand  Jo  Sören  Christopher  Philipp  Gustav  Arnold  Johannes  Markus  Loris  Silvester  Andrew  Fabius  Gregor  Dietmar  Philemon  Tristan  Clyde  Isaak  Franz  Tom  Baptist  Nikita  Mario  Helmut  Hubert | female first names:  Johanna  Cosima  Beate  Cora  Fenja  Kassandra  Aloisia  Aloisia  Hannelore  Adriana  Marianne  Christiane  Alma  Delia  Clementine  Merit  Hannelore  Sibylle  Emma  Julia  Delia  Manuela  Else  Jenny  Monika  Angelina  Adela  Mechthild  Clementine  Babette  Helga  Nadia  Gundela  Verena  Holde  Yvonne  Alrun  Ariane  Marietta  Melanie  Solveig  Cordula  Jeanette  Sibylle  Hannah  Natalie  Esther  Astrid  Wilma  Solvej | last names:  Sperner  Götz  Kaiser  Kurz  Anders  Biermann  Voigt  Herrmann  Franz  Voigtländer  Widera  Herold  Adam  Schröder  Träger  Sonntag  Pilsak  Wiegand  Schultheiß  Schnorr  Schubert  Fröhlich  Weck  Widera  Quant  Fehling  Krüger  Schmidt  Fehling  Schubert  Falk  Plagemann  Lenz  Rohlfs  Albrecht  Reich  Luhmann  Calvin  Ludwig  Querfurt  Kolmar  Widera  Kolmar  Krause  Krause  Krüger  Ossege  Widera  Krupp  Schnurre |

(Ein Spaltenumbruch erfolgte bei jeder Leerzeile)

Bei Teil 1 wurde angenommen, dass die erzeugten Namen reine Listenelemente sein sollten. Die Doppelnamen wurden in Teil 2 implementiert, als es um die vollständigen Namen ging. Daher gibt es als Testwerte „nur“ zufällig ausgewählte Namen. Es lässt sich erkennen, dass der Zufallsgenerator gut funktioniert. Die Namen sind gut durchmischt.

Die vorgegebene Datei EPR\_04.py wurde umbenannt in names.py. Außerdem wurden die Klammern so geändert, dass die Namen nun in Listen und nicht in Mengen gespeichert sind. Aus Mengen lassen sich nämlich nicht so leicht zufällige Elemente entnehmen. Zum Vorteil der leichten Veränderung, siehe Teil 6. Dazu wurde eine weitere Liste für die Adressnamen hinzugefügt.

# Teil 2:

Testwerte:

|  |  |  |
| --- | --- | --- |
| Olga Siemens  Babette-Aurora Wittgenstein  Ansgard Curtius  André Löchte  Maximilian Köhler-Lohmann  Solveig Landzettel  Kasimir Schultheiß  Eike Ruoff  Cordula Körner  Leyla Reich  Vivien Howaldt  Heron Nietzsche  Lilith Wolfram  Adele Dürrschnabel  Julia Thoma-Böll  Delia-Sandra Löchte  Karl Baumann  Beatrix Thyssen  Archibald Jakobi  Maximilian Winkler  Muriel Hintz  Franziska Schmidt  Lucia-Ilona Tenberge  Scholastika Howaldt  Aribert Frank  Lydia Wolf  Sarah Heine-Wolf  Joelle Voigtländer  Gaudenz Brühlmeier-Anders  Reinhold Frenzel  Merrit Lammers-Wiegand  Charlie Leber  Mathilde Schnurrer  Angelina Baumann | Nikolaus Freund  Magdalena Lorenz  Ivo Pügner  Tatjana Honecker  André Diederich  Samuel Wieland-Henne  Marianne Schnurre  Kira-Julia Türck  Jördis-Sibylle Teuber  Ansgard Mach  Frank Wieland  Linus Marg  Konrad Voigtländer  Alraune Selbmann  Charlie Fröhlich  Corvin Oswald  Sabrina Ingo  Ilja Becker  Kerstin Zacharias  Linda-Margarete Hoffmann  Burkhardt Friedemann  Klemens Herzog  Dr. Anastasius Meyer  Amadeus Becherer  Joelle Thyssen  Aldhelm Sygnetzki  Fabia Reich  Ada Frank  Cosimo-Calvin Amalie  Samantha Schneider  Matthias Mach  Dietrich Freud  Natascha Braun  Yvonne Krause | Hauke Preuß  Nando Ossege  Eberhard Schulthess  Vivien Henne  Nikolaus Schmidt  Claudia Dürrschnabel-Abendroth  Romy Schmid  Konstantin-Thilo Thomas  Liane Behm  Virginia Steiner-Voigt  Olivier Franz  Nadia Kosba  Evelyn Anders  Thorsten Thyssen  Arnulf Zahn-Fröhlich  Waltraud Frenzel  Gernot Marg-Schinder  Jessica Schnurre  Mona Pohl  Apollonia Heine  Zoe Schulze  Zacharias Semper  Winand-Justus Albert  Crispian Pohl  Kriemhild Ossege-Frank  Solvi Knebel  Manfred Schinder  Gilbert Ossege  Isaak-Udo Köhler-Thomas  Othmar Sauerbruch-Andres  Michaela Lehner-Herold  Jan Honecker |

(Ein Spaltenumbruch erfolgte nach jeweils 34 Zeilen)

Die Tests decken sowohl Doppelnamen, als auch Doppelnachnamen und Doktortitel ab. Auch ein Doppelvor- zusammen mit einem Doppelnachnamen (Isaak-Udo Köhler-Thomas) existiert.

Zur Doktorentscheidung:

Ein Doktortitel ist an 1% der Personen „zu vergeben“. Dabei sollen nur 45% an Frauen gehen und 55% an Männer. Das wird dadurch erreicht, dass Männer eine Wahrscheinlichkeit von 0,011 und Frauen eine von 0,009 haben, einen Doktortitel zu tragen. Das folgt nach Einsetzen in den Satz von Bayes (<https://de.wikipedia.org/wiki/Satz_von_Bayes>).

# Teil 3:

Die Tests werden so gestaltet, dass 1000 zufällige Namen erstellt werden. Die Informationen werden gesammelt (Doppelname, Dr.-Titel, etc.) und der Anteil der eingetretenen Ereignisse an allen Namen ausgerechnet und ausgegeben. Dadurch lässt sich überprüfen, ob die Werte nahe den Soll-Werten liegen. Bei 100 oder 1000 Durchläufen kann es zu Teils groben Abweichungen kommen. Wenn eine höhere Genauigkeit gewünscht ist, lässt sich die Anzahl der Testversuche leicht im Code erhöhen.

Testwerte:

0.9% are doctors.

9.3% have a double first name.

15.0% have a double last name.

48.1% are male.

48.7% are female.

3.2% have an unclear gender due to some names being suitable for both genders.

Die Statistik stimmt bis auf kleine statistische Abweichungen mit den geforderten Werten überein.

# Teil 4:

Die Hausnummernverteilung wurde folgendermaßen gewählt. In Deutschland haben die meisten Häuser eine ein- bis vierstellige Hausnummer. Aus diesem Bereich wählen wir unsere Hausnummer. Die meisten liegen in einem Bereich von 1-99. Deshalb haben wir uns für eine Wahrscheinlichkeit von 89% entschieden, dass die generiert Nummer in diesem Bereich liegt. Es gibt ja viele kleine Straßen. Mit einer Wahrscheinlichkeit von 10% ist die Nummer 3-Stellig. Sehr große Straßen mit bis zu 4-Stelligen Hausnummern gibt es selten. Deshalb haben wir uns für eine Wahrscheinlichkeit von einem Prozent entschieden, dass die Nummer zwischen 1000 und 9999 liegt.

# Teil 5: