

Phonetische Suche am Beispiel von Soundex

Soundex ist ein phonetischer Algorithmus, also ein Verfahren, das sich an der Aussprache orientiert. Es dient zur Katalogisierung von englischen Namen. Das Ziel ist, Namen mit gleicher oder ähnlicher Aussprache zusammenzufassen. Eine leichte Abweichung dieses Algorithmus wurde beispielsweise von der amerikanischen Regierung in den 1930er Jahren bei einer Volkszählung verwendet. Heutzutage wird das Verfahren oder ähnliche von Suchmaschinen verwendet, um zum Beispiel trotz falscher Eingaben bei Suchbegriffen die richtigen Treffer zu liefern.

Algorithmus:

Der Soundex-Code eines Namens besteht aus einem Großbuchstaben, der von drei Ziffern gefolgt wird. Der Buchstabe ist der Anfangsbuchstabe des Namens; die Ziffern kodieren die folgenden Laute. Das folgende Beispiel erklärt den Algorithmus.

Beispiel: Die Namen „Bachmann“ und „Müller“.

1. Entferne aus dem Suchbegriff alle Zeichen, die keine Buchstaben oder Zahlen sind. Merke dir außerdem den ersten Buchstaben.
2. Weise den Buchstaben, die dem Anfangsbuchstaben folgen, die folgenden Zahlenwerte zu:

A E H I O U W Y Ä Ö Ü	0
B F P V	1
C G J K Q S X Z	2
D T	3
L	4
M N	5
R	6

3. Wenn jetzt aufeinanderfolgende Buchstaben mit dem gleichen Zahlenwert auftreten, wird einer der beiden Buchstaben, bzw. Zahlenwerte gestrichen.
4. Ersetze die erste Ziffer durch den in 1. gemerkten Buchstaben.
5. Entferne die Nullen aus der Zeichenkette. Diese entsprechen den Vokalen sowie h, w, y
6. Ein Soundex-Code besteht immer aus einem Buchstaben gefolgt von drei Zahlen.
⇒ Falls mehr als drei Zahlen vorkommen, entferne die restlichen.
⇒ Falls weniger als drei Zahlen vorkommen, werden die restlichen Stellen mit Nullen aufgefüllt.

Bachmann	Müller
10205055	504406
1020505	50406
B020505	M0406
B255	M46
B255	M460

Aufgabe 1:

Führe das Verfahren nun für die beiden Namen ‚Hohlmeier‘ und ‚Nonnenmacher‘ durch.

Aufgabe 2:

Schreibe ein C#-Programm, das für eingegebene Namen den entsprechenden Soundex Code ausgibt. Teste es für die Namen Smith, Schmitt. Welchen Soundex-Code erhält man für sie?

Aufgabe 3:

Recherchiere weitere phonetische Suchverfahren und programmiere sie in C#.