

Phonetik

Wie schreibt sich eigentlich Herr Meier?



Holger Darjus

(Jg. 1962) ist Technischer Leiter des Medienvorstufen-Unternehmens EINSATZ Creative Production GmbH & Co. KG in Hamburg mit rund 45 Mitarbeitern. Er entwickelt seit 1996 FileMaker-Datenbanken speziell für die Medienvorstufe.
h.darjus@einsatz.de

Datenbanken sind doch was Tolles! In sie packen wir alles hinein, was wir uns nicht merken können, um es jederzeit abrufbereit zu haben. Das, was unser Gehirn nicht schafft, meistert eine Datenbank bravourös – oder? Will ich z. B. Herrn Mayr anrufen, habe ich seine Rufnummer im Kopf und greife gleich zum Hörer, oder ich schlage mal eben in meiner FileMaker-Datenbank nach: Suchmodus aktivieren, im Feld „Namen“ Mayer eingeben, suchen ... und ups, nix gefunden.

Warum? Weil ich vergessen hatte, dass Herr Mayr zwischen dem „y“ und dem „r“ eben kein „e“ hat. Kann auch sein, dass er sich „Meier“ oder „Meyer“ oder „Mair“ oder ... schreibt. Um jetzt ans Ziel zu kommen, wäre doch eine phonetische Suche sehr hilfreich. Phonetik, auf eine kurze Formel gebracht, bedeutet in etwa „klingt wie“. Das heißt nicht die Schreibweise, sondern der Klang ist das Entscheidende.

Der Algorithmus für eine Umsetzung ist vergleichsweise simpel. So werden bestimmte Laute und Buchstaben einem von sechs Klangbildern zugeordnet, Vokale einfach ausgeblendet (genauso wie „J“ und „Y“, die ja wie „I“ klingen) und doppelte, direkt aufeinanderfolgende Zuordnungen zusammengefasst – fertig. Die Umsetzung in FileMaker ist ähnlich einfach. Hier muss jedoch berücksichtigt werden, dass FileMaker, warum auch immer, ein „ß“ in Großbuchstaben nicht durch „SS“ ersetzt. Hier die passende Formel:

```
Austauschen (
  (Austauschen (Text; "ß"; "SS"));
  ["R"; "7"]; ["L"; "6"]; ["N"; "5"];
  ["M"; "5"]; ["Z"; "4"]; ["X"; "4"];
  ["SCH"; "4"]; ["Q"; "4"]; ["K"; "4"];
  ["G"; "4"]; ["C"; "4"]; ["S"; "4"];
  ["W"; "3"]; ["V"; "3"]; ["P"; "3"];
  ["F"; "3"]; ["B"; "3"]; ["T"; "2"];
  ["D"; "2"]; ["Y"; "" ]; ["J"; "" ];
  ["H"; "" ]; ["U"; "" ]; ["O"; "" ];
  ["I"; "" ]; ["E"; "" ]; ["A"; "" ];
);
["77"; "7"]; ["66"; "6"]; ["55"; "5"];
["444"; "4"]; ["44"; "4"]; ["333"; "3"];
["33"; "3"]; ["22"; "2"]; ["11"; "1"] )
```

Das Ergebnis ist verblüffend. Angewendet auf alle denkbaren Schreibweisen von Herrn Meir lautet es nämlich immer 57. Und ob sich Herr Schmid nun mit „d“ oder „dt“ oder „tt“ schreibt, ist mir bei Suchabfragen jetzt auch egal, weil ein Script dafür sorgt, dass bei einer phonetischen Suche der Wert 452 übergeben wird.

Der hier angewandte Algorithmus ist übrigens nur einer von vielen. Die beiden am weitesten verbreiteten Reduzierungsverfahren heißen „Soundex“ und „Metaphone“ und unterscheiden sich vor allem in der „Schärfe“ (Fuzzy-Logic). So gibt es Verfahren, die bestimmte Buchstaben, sofern sie am Wortanfang stehen, nicht ersetzen. Das wirkt sich je nach angewandeter Sprache sehr unterschiedlich aus.

Auch die Wortlänge wird in dem hier gezeigten Beispiel nicht berücksichtigt. So fallen Worte wie „Uhu“ durch das Raster. Meine Erfahrung hat jedoch gezeigt, dass jede Verfeinerung der Berechnung auch immer ungeahnte Seiteneffekte mit sich bringt. Im hier beschriebenen Anwendungsfall tut die simple Formel aber schon sehr gute Dienste. Die Formel ist auch als Eigene Funktion auf den Internetseiten des K&K-Verlags abzurufen.

Viel Spaß damit!