

## Wladimir Iossifowitsch Lewenstein

Wie schreibt man denn doch gleich "Philip Gresli" oder wars "Phillip Gresly"? Wie kann ich den Eintrag in der Datenbank finden, auch wenn ich die korrekte Schreibweise vergessen habe?

**Author:** Philipp G. Freimann  
(BBW  
(Berufsbildungsschule  
Winterthur)  
<https://www bbw.ch>)

Kein Problem für Wladimir Levenshtein. Der russische Mathematiker (\*1935) hat einen nach ihm benannten Abstand zweier Zeichenketten definiert. Die *Levenshtein-Distanz* gibt an, wie viele Ersetzungen, Einfügungen oder Löschungen von einer Zeichenkette zur anderen nötig sind. Sind zwei Zeichenketten identisch, so ist auch die Levenshtein Distanz gleich Null. In der Datenbank wird nach Eingabe von "Philip Gresli" für jeden Eintrag diese Distanz berechnet. Am Schluss werden diejenigen Namen angegeben, bei denen die Distanz minimal ist.

Doch wie funktioniert diese Levenshtein-Distanz?

Als einfaches Beispiel nehmen wir die Wörter "MEYER" und "MUELLER" und berechnen davon die Levenshtein-Distanz.

Wir beginnen mit der folgenden Tabelle:

```
##MUELLER
#01234567
M1.....
E2.....
Y3.....
E4.....
R5.....
```

In Leserichtung füllen wir nun sukzessive Werte in die leeren Zellen ein.

Das Verfahren benötigt für die Berechnung der Zahl im aktuellen Feld nur die drei Werte aus den drei angrenzenden Feldern links, oben und links/oben.

Sind die beiden Buchstaben (links und oberhalb des aktuell zu berechnenden Feldes) verschieden, so nehmen wir einfach das Minimum der drei genannten angrenzenden Werte, zählen eins (1) dazu und schreiben das Ergebnis ins aktuelle Feld.

```
Beispiel {U ≠ M → '1' eintragen = min (0, 1, 2) + 1}
##MUELLER
#01234567
M101.....
E2.....
Y3.....
E4.....
R5.....
```

Sind die beiden Buchstaben jedoch identisch, so zählen wir dem Feld links/oben in Gedanken eins ab und ermitteln wieder das Minimum der drei Zahlen. Wiederum zählen wir dem Minimum eins dazu und füllen den so errechneten Wert ins aktuelle Feld.

```
Beispiel {E = E → '1' eintragen = min (2, 1-1, 2) + 1}
##MUELLER
#01234567
M10123456
E2111....
Y3.....
E4.....
R5.....
```

Zuletzt steht im Feld unten rechts die Levenshtein-Distanz.

Beispiel: Distanz MUELLER zu MEYER = 3

##MUELLER

#01234567

M10123456

E21112345

Y32222345

E43323334

R54433443

Schreiben Sie nun eine Subroutine, welche zwei Zeichenketten entgegen nimmt und deren Levenshtein-Distanz berechnet:

Funktion:

levenshtein(a: String, b: String): int;