

Lexikografisch sortieren

Um eine lexikographische Ordnung von Wörtern zu erreichen, kann nicht einfach die triviale aufsteigende Sortierung einer Zeichentabelle durchgeführt werden.

Zum einen müssen große und kleine Buchstaben gleich behandelt werden, und zum anderen sollen Umlaute und Ligaturen (ß) nach ihren zugrunde liegenden Buchstaben einsortiert werden:

A < a < AA < aa < AAb < Aachen < ...
... < aha < Ähre < Akne < ...
... < haschen < Häschen < hat < ...
... < Masse < Maße < Maßeinheit < ...
... < ZZ < zz

Entwickeln Sie nach obigem Beispiel eine Funktion, die zwei gegebene Wörter miteinander vergleicht.

Tipp: Ein 'ä' wird wie ein 'a' sortiert, es sei denn, es gibt auch eine ansonsten identische Wortvariante mit einem 'a' an derselben Stelle. In einem solchen Fall wird das Wort mit Umlaut nach letzterem einsortiert. Analog verhält es sich mit den Groß- und Kleinbuchstaben. Mit Vorteil werden die Wörter zunächst in ihre lexikalische Grundform gebracht: nur Kleinbuchstaben, keine Umlaute [a statt ä], keine Ligaturen [ss statt ß], keine Satzzeichen [*Leerschlag* statt '.']. Stimmen die Grundformen nicht überein, so wird nach eben dieser Grundform sortiert.

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