

Übung 1

Aufgabe 1.1

Geben Sie RegEx-Ausdrücke für folgende Anforderungen an:

1. Alle Wörter. Wörter bestehen nur aus alphabetischen Zeichen und sind durch Whitespace, Punktionszeichen, Linebreaks usw. von anderen Wörtern getrennt.
2. Alle Vorkommen der bestimmten Artikel *der*, *die*, *das* (auch solche, die am Satzanfang vorkommen)
3. Alle Zeilen, die mit einer Zahl beginnen und einem Wort aus mindestens drei Buchstaben enden
4. Alle Zeilen eines Textes, die NICHT mit einem Punkt enden
5. Mobilnummern mit abgetrennter Länder- und Mobilfunkanbieterwahl, also z.B. +49 157 1234567
6. Studentische email-Adressen der Hochschule Niederrhein
7. Alle Jahreszahlen zwischen 1900 und 2099
8. Datumsangaben der folgenden Form: Zuerst wird der Tag als ein- oder zweistellige Zahl angegeben, danach folgt ein Punkt, dann der Monat als ein- oder zweistellige Zahl, danach wieder ein Punkt, dann das Jahr als zwei- oder vierstellige Zahl.
9. Alle Teilstrings, die mit einem bestimmten HTML-Tag beginnen und dem gleichen HTML-Tag abgeschlossen werden. In dem folgenden Beispiel sollen also die jeweiligen farblich markierten Abschnitte gematcht werden:

`<h1>Lorem ipsum</h1>`

`<p>Lorem ipsum dolor sit amet. consetetur sadipscing elitr, sed diam nonumy eirmod tempor invidunt ut labore et dolore magna aliquyam erat, sed diam voluptua. </p>`

`<p>At vero eos et accusam et justo duo dolores et ea rebum.</p>`

Aufgabe 1.2

Schreiben Sie ein Programm in Python (oder einer Programmiersprache Ihrer Wahl), das alle Erwähnungen eines Herrn Josef Meier in einem Text findet und ausgibt. Sie sind sich allerdings über die Schreibweise unklar: sein Vorname könnte auch Joseph lauten oder es könnten Kurzformen wie Sepp oder Jupp oder eine Abkürzung wie J. verwendet werden. Auch bei der Orthografie seines Nachnamens gibt es Unsicherheiten (mit e oder a? Mit i oder y? Vielleicht fehlt auch das letzte e wie in Mair?). Falls die Anrede im Text zusammen mit seinem Namen auftaucht, soll diese ebenfalls gefunden werden.

Hinweise

- Gute Seite, um mit RegEx zu experimentieren: <https://regex101.com/>
- Für das Python-Skript sollten Sie die Standard-Bibliothek `re` verwenden: <https://docs.python.org/3/library/re.html>