

Stralau ist Strafau ist Krakau

Eine Fallstudie zu real existierender Künstlicher Intelligenz

Die Halbinsel Stralau liegt im Berliner Ortsteil Friedrichshain. Kürzlich schrieb ich mit Hilfe des Programms Open Office einen Text, in dem das Wort Stralau vorkam. Ich hatte mich aber vertippt (Stradau statt Stralau). Das Rechschreibprogramm bemerkte den Fehler und schlug mir "Strafau" als Ersatzwort vor.

Ich nahm an, dass dieses Unsinn-Wort versehentlich in das Benutzer-Wörterbuch meines Computers hineingerutscht war. Aber so etwas lässt sich ja korrigieren. Ich informierte mich über Benutzerwörterbücher und wie man falsche Einträge daraus wieder entfernen kann. Allerdings kam "Strafau" in keinem der Benutzerwörterbücher auf meiner Festplatte vor. Trotzdem wurde es immer wieder als Ersatzwort vorgeschlagen. Wo war dieses Strafau nur gespeichert? Es war einfach nicht auf der Festplatte zu finden. Langsam kam mir der Verdacht, dass "Strafau" für Open Office zu den in der internen Vorschlagsliste "eingebauten" (und möglicherweise speziell codierten) korrekten Wörtern gehörte. Ich testete die Hypothese, indem ich auf einem anderen Computer denselben Fehler in eine Textdatei einbaute. Auch hier wurde "Strafau" als Korrekturbegriff vorgeschlagen. Also war "Strafau" wohl als korrektes Wort direkt in Open Office "einprogrammiert" und es lag nur an mir, dass ich dieses Wort nicht kannte.

Als nächstes suchte ich im Internet mit diversen Suchmaschinen nach "Strafau". Bing, die Suchmaschine von Microsoft Edge, lieferte viele Links zu "Stralau" aber keines zu "Strafau". Die Suche mit Google brachte überraschende Ergebnisse: Neben einigen offensichtlichen Treffern wie Strafausschließungsgrund, Strafaussetzung, Strafausmaß und ähnlichen Wörtern wurden auch die folgenden Links angezeigt:

- * Schillers Werke: Fünfter Band: Wallenstein / Demetrius
- * Geschichte der Königreiche Galizien, Lodomirien und Rothreussen

Aha, dachte ich; dass ich "Strafau" nicht kannte, war offensichtlich eine Bildungslücke von mir. Also klickte ich zunächst auf den "Schiller"- und dann auf den "Geschichts"-Link. Beide Links verwiesen auf Bücher des Google Books Projekts (deshalb wurden Sie wohl auch von Bing nicht angezeigt). Beide Texte waren in Frakturschrift gesetzt. Freundlicherweise hinterlegt Google die gesuchten Wörter farblich zur besseren Erkennung. Hinterlegt war "Krakau". Das Fraktur-Krakau war von einem (mehr oder weniger intelligenten) dazwischengeschalteten OCR-Programm (OCR: Optical Character Recognition) offenbar als "Strafau" übersetzt worden.

Offen bleibt die Frage: Wie ist "Strafau" in die Vorschlags-Liste von Open Office gekommen? Enthält diese Liste etwa alle von Google Books eingescannten und richtig wie falsch durch OCR übersetzten Wörter? Abschließend noch zur Klarstellung: Nichts gegen Open Office. Es ist ein grundsätzlich ganz ausgezeichnetes, kostenloses Textverarbeitungs-Programm.

Jochen Ziegenbalg