

Harald Bögeholz

Clipboard-Verhüterli

Rechtssicheres Kopieren von Texten

Wissenschaftliches Arbeiten im Internet ist gefährlich: Nur allzu leicht fängt man sich über die Zwischenablage eine Textpassage ein, die später zum Verlust des hart erarbeiteten Dokortitels führt. Neuronale Netze können vor solchen Missgeschicken schützen.

Getreu dem Motto „Es ist alles schon geschrieben worden, nur noch nicht von jedem“ wiederholen Online-Artikel, Schul- und Studienarbeiten oder Dissertationen fast zwangsläufig Gedanken, die andere zuvor aufgeschrieben haben. Doch wenn man anderer Leute Texte einfach wörtlich kopiert, sieht man sich schnell dem Vorwurf der Urheberrechtsverletzung oder des Plagiats ausgesetzt.

Für dieses Problem gibt es zwei Lösungen: das wörtliche Zitat mit Nennung des Urhebers oder das Umformulieren mit eigenen Worten. Letzteres macht mehr Arbeit, Ersteres kann man leicht vergessen. Bei beidem kann der Computer helfen.

Seit einiger Zeit wird unter Textschaffenden unter der Hand ein Tool namens CopyRight herumgereicht. Es ist technisch recht ausführlich dokumentiert, aber der Urheber zieht es vor, anonym zu bleiben. Über Google kann man es praktisch nicht finden, weil eine Suche nach CopyRight die Nadel mit einem Heuhaufen aus irrelevanten Treffern zuschüttet. Eine Download-Adresse finden Sie am Ende dieses Artikels.

Clever oder smart

CopyRight klinkt sich unter Windows in die Zwischenablage ein und sorgt bei allen kopierten Textpassagen dafür, dass man sie gefahrlos in die eigene Arbeit einfügen und weiterverwenden kann. Dazu kennt es zwei

Betriebsarten: Im Modus „einfach“ fügt es vor und nach jedem kopierten Text Anführungszeichen ein sowie dahinter den Text „[!!!]“. So kann man beim Zusammenstellen einer Arbeit zügig vorankommen und nicht versehentlich vergessen, Zitate zu kennzeichnen und die Quelle anzugeben.

Interessanter ist aber der Modus „intelligent“. Er gibt den kopierten Text beim Einfügen nicht wörtlich wieder aus, sondern formuliert ihn sinngemäß um. So kann man sich die lästigen Anführungszeichen und Quellenangaben sparen und muss sich auch nicht umständlich merken, welche Aussagen woher kommen. Dadurch ist CopyRight ideal für die schnelle Produktion von Online-Nachrichten: Einfach Informationen aus verschiedenen Quellen zusammenkopieren, ein paar überleitende Sätze spendieren und die Abschlüsse redigieren, fertig.

Künstliche Intelligenz

Technisch steckt hinter CopyRight ein tiefes neuronales Netz mit Long Short Term Memory (LSTM). Für eine ausführliche Erklärung neuronaler Netze siehe [1], hier nur so viel: LSTM-Netze werden unter anderem zur Übersetzung eingesetzt. Sie haben ein Kurzzeitgedächtnis und können daher nicht nur einzelne Wörter erkennen, sondern erinnern sich Wort für Wort an den Kontext. So entsteht beim Einlesen eines Satzes im neurona-

len Netz eine interne Repräsentation des Gesagten. Der zweite Schritt einer Übersetzung ist nun eine Art Gedächtnistransplantation in ein neuronales Netz der Zielsprache, das gespiegelt wird und in umgekehrter Richtung wieder einen Satz ausgibt.

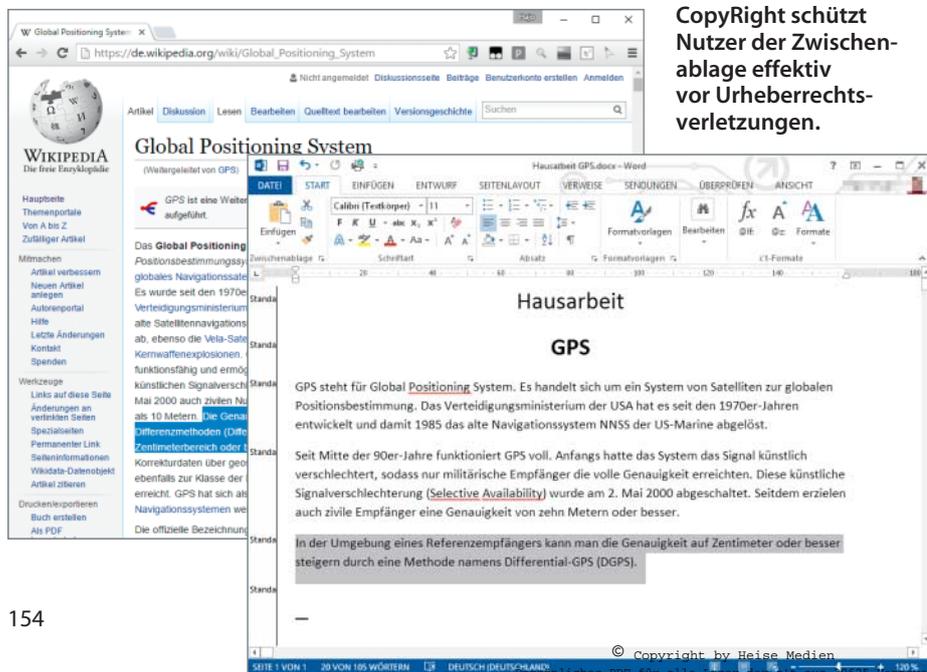
So funktioniert Google Translate, und es war wohl eine naheliegende Idee, bei CopyRight für Quell- und Zielsprache ein und dieselbe Sprache zu verwenden und das ausgebende LSTM-Netz lediglich so zu trainieren, dass es sich an die gewünschte Ausdrucksweise anpasst und eben nicht exakt dasselbe ausgibt, was das Erkennungsnetz als Eingabe bekommen hat.

Das Ganze klingt wesentlich komplizierter, als es in der Praxis ist: Um wissenschaftlich mit CopyRight zu arbeiten, müssen Sie das neuronale Netz zunächst trainieren, damit es über das Thema Ihrer Arbeit und Ihren persönlichen Schreibstil Bescheid weiß. Dazu genügt es, relevante Textmengen aus Ihrem Werk in die Zwischenablage zu kopieren. CopyRight analysiert alle Texte, die in der Zwischenablage vorbeikommen, und lernt so im Laufe der Zeit, Texte mit ähnlichem Duktus zu formulieren. Wichtig für die Lernphase ist es, dass in den gelernten Texten genügend Fachbegriffe und lange Sätze vorkommen. Andernfalls kann es passieren, dass sich CopyRight zu unwissenschaftlich ausdrückt.

Ausblick und Gefahren

Im Augenblick ist CopyRight ein ideales Werkzeug, um die Kulturtechnik des Kopierens rechtssicher anzuwenden. Vor allem bei Texten wie Online-Artikeln, die ohnehin wenig Bestand haben, bedeutet es für den Autor eine erhebliche Zeitersparnis. Wenn man einen gewissen Qualitätsanspruch hat, muss man über die erzeugten Texte noch einmal drüberlesen und hier und da nachbessern, aber das Resultat ist jedenfalls hinreichend weit von etwaigen Originalen entfernt, als dass man Urheberrechtsverletzungen befürchten müsste.

Auch die Plagiatsjäger, die in letzter Zeit immer mehr Doktorarbeiten von Prominenten ins Visier nehmen, laufen ins Leere. Doch Vorsicht: Das könnte sich mit fortschreitender technischer Entwicklung ändern. Denn ein LSTM-Netz kann nicht nur Text generieren, sondern umgekehrt auch sehr zuverlässig selbst geschriebenen Text erkennen (Autokorrelation). Im Moment ist CopyRight ein Closed-Source-Tool und das neuronale Netz nicht zugänglich. Doch sollte der Autor eines Tages beschließen, mit derselben Technik ein Werkzeug für die Suche nach Plagiaten zu programmieren, könnte es für so manche Dissertation brenzlig werden. (bo@ct.de)



Literatur

[1] Andrea Trinkwalder, Netzgespinste, Die Mathematik neuronaler Netze: einfache Mechanismen, komplexe Konstruktion, c't 6/16, S. 130

ct Download CopyRight: ct.de/you5u