

Am Zitat überführt

Der Informatiker Prof. Dr. Bela Gipp entwickelt neue Methoden der Plagiatserkennung – sogar über Sprachgrenzen hinweg

Die Plagiatsoftware CitePlag überprüft Texte auf inhaltlich-semantic Ähnlichkeiten. Farbige Hervorhebungen und Verbindungslinien markieren Ähnlichkeiten in den Vergleichstexten – sogar dann, wenn Texte umformuliert oder in eine Fremdsprache übersetzt wurden.

„Wort für Wort“ ist das Credo der verbreitetsten Plagiatsoftware: Verdachtstexte werden maschinell mit Textdatenbanken verglichen; die Software überprüft, wo Textteile eins zu eins übereinstimmen. Solche textbasierte Plagiatserkennung kommt jedoch schnell an ihre Grenzen, wenn eben nicht Wort für Wort abgeschrieben wird. Plagiatoren sind bekanntlich erfindungsreich, wenn es darum geht, ihr Plagiat zu vertuschen: Texte werden paraphrasiert und im Wortlaut verfremdet, umstrukturiert oder durch Übersetzung verschleiert. Viele Computerprogramme zur Plagiatserkennung scheitern daran, solche umformulierten Texte als Plagiat zu erkennen.

„Plagiatsoftware sollte nicht nur nach übereinstimmendem Text schauen“, empfiehlt der Konstanzer Informatiker Prof. Dr. Bela Gipp. Mit seinem Projekt CitePlag entwickelt er eine Plagiatsoftware, die inhaltlich-semantic Ähnlichkeiten zur Plagiatserkennung heranzieht.

Zitationen – ein zuverlässiges Unterscheidungsmerkmal

Eines der ergiebigsten Erkennungsmerkmale von CitePlag bilden dabei ausgerechnet Zitate. Damit sind nicht etwa die „vergessenen“ Fußnoten und Anführungszeichen eines Plagiats gemeint, sondern abbeschriebene Zitationen. Kein wissenschaftlicher Text, der ohne Zitate und Fußnoten auskäme, der ganz auf Querverweise auf andere wissenschaftliche Arbeiten verzichten könnte. Plagiatoren kopieren diese Zitate und ihre Quellenangaben in den Fußnoten einfach mit. Immerhin wäre es extrem aufwändig, Zitate in einem abbeschriebenen Text durch andere passende Zitate auszutauschen – dafür müsste ein Plagiator so tief in die Materie eingearbeitet sein, dass er sich durch das Plagiat kaum Aufwand ersparen würde. „Zitationen sind ein sehr zuverlässiges Unterscheidungsmerkmal“, erläutert Bela Gipp: „Es ist extrem unwahrscheinlich, dass in zwei wissenschaftlichen Texten dieselben Zitationen in derselben Reihenfolge vorkommen.“

Zitationen sind nur eines von mehreren Verfahren, die CitePlag kombiniert, um Texte auf inhaltlich-semantic Ähnlichkeiten zu überprüfen. Die Software stellt zwei Texte visuell gegenüber und verdeutlicht Ähnlichkeiten mittels farbiger Hervorhebungen und Verbindungslinien – selbst dann, wenn abbeschriebene Textabschnitte umformuliert oder im Textverlauf weit auseinandergezogen wurden. Sogar Plagiate durch Übersetzungen in Fremdsprachen kann Bela Gipp's Software erkennen.

CitePlag ist in der Testphase

„Ich möchte betonen: Ich bin kein Plagiatjäger“, bekräftigt Bela Gipp. „Mein Ziel ist, den Aufwand des Plagierens so zu erhöhen, dass es sich nicht mehr lohnt.“ Bela Gipp's CitePlag befindet sich augenblicklich in der Testphase und soll künftig kostenlos als universitärer Dienst angeboten werden. Dem Informatiker ist aber wichtig, dass Nutzer nicht ausschließlich auf die maschinelle Plagiatserkennung vertrauen: „Es sollte sich immer noch ein Experte den Text anschauen und gegenprüfen, ob es sich wirklich um ein Plagiat handelt.“ Plagiatsoftware ist eine Methode, das Plagieren zu erschweren, viel entscheidender ist für Gipp aber der öffentliche Diskurs um qualitatives wissenschaftliches Arbeiten: „Das Problem des Plagierens wird nicht mit technischen Mitteln allein gelöst werden können. Ideenplagiate stellen nach wie vor das größte Problem dar. Entscheidend wird der Umgang der Gesellschaft mit Plagiaten sein.“

– informatik.uni-konstanz.de/information-science/

„Mein Ziel ist, den Aufwand des Plagierens so zu erhöhen, dass es sich nicht mehr lohnt.“

Prof. Dr. Bela Gipp



Prof. Dr. Bela Gipp ist Juniorprofessor für Informationswissenschaft an der Universität Konstanz. Zuvor forschte er an der University of California in Berkeley (USA) und am National Institute of Informatics in Tokio (Japan). Zu seinen Forschungsschwerpunkten zählen Methoden der Gewinnung und Visualisierung von Informationen, Wissensmanagementsysteme sowie die Entwicklung von semanticen Methoden für Plagiatsoftware und Empfehlungssysteme.