

DIAGNOSTIK & FÖRDERUNG

Wenn Computerprogramme Tests korrigieren

Dr. Fabian Zehner erklärt wie eine Software Textantworten bewertet und warum das fairer sein kann als die menschliche Auswertung

12. APRIL 2023

LENA STERZ

LESEZEIT: 4 MINUTEN

Forschende arbeiten seit vielen Jahren daran, dass PISA und andere Tests in Zukunft von einer Software ausgewertet werden können. Dr. Fabian Zehner erklärt, wie seine Software ReCo funktioniert und wie sie schon bald den Unterricht verändern und bereichern könnte.

Redaktion: Was genau kann die Software ReCo?

Dr. Fabian Zehner: ReCo steht für Automatic Text Response Coder, das heißt, es dreht sich eigentlich alles darum, Textantworten automatisch auszuwerten, beziehungsweise zu kodieren. Die Textantworten werden dafür jeweils einer Benotungs-Kategorie zugeordnet. Entstanden ist das Ganze im Kontext von PISA.

Redaktion: Warum ist so eine Software gerade für die PISA-Tests interessant?

Zehner: Speziell bei PISA gibt es sehr viele Aufgaben, bei denen man mit einem kurzen Text antworten muss. Und da steckt natürlich sehr viel menschliche Arbeit dahinter, diese Hunderttausende von Textantworten von Schüler:innen weltweit auszuwerten. Aber vor allem ist so ein automatisches Auswerten im Endeffekt konsistenter - eine Software macht keine Flüchtigkeitsfehler und es gibt auch keine kulturellen oder personenabhängigen Unterschiede, was die Auswertung angeht; manche Auswertenden sind nachlässiger als andere, manche bewerten die gleiche Antwort besser als andere. Die Software bewertet alle gleich.

Redaktion: Wie funktioniert die Software genau?

Zehner: Wenn zu einem Text zum Beispiel die Frage gestellt wird: Was hat die Protagonistin am Ende der Geschichte wütend gemacht? Und die richtige Antwort wäre: Weil der Vater eine ihr wichtige Zeitschrift weggeschmissen hat, dann

erkennt die Software nicht nur, ob in der Antwort die richtigen Worte vorkommen, sondern sie berücksichtigt die Semantik. Dafür werden die Antworten quasi in Zahlen übersetzt und anhand dieser Zahlen lernt die Software dann, welche Zahlen, also Antwortsemantik, bedeuten, ob eine Antwort richtig oder falsch ist. Das muss man ReCo beibringen, bevor es einen Text auswerten kann.

Und das Tolle daran ist, wenn ReCo weiß, dass „Zeitschrift“ ein richtiges Wort ist, dann erkennt sie auch „Magazin“ als richtig an, weil die Software weiß, dass es ein Synonym ist.

Redaktion: Das klingt so, als wäre es im Moment noch sehr aufwändig ReCo beizubringen, welche Antworten richtig und falsch sind. Es ist deshalb wahrscheinlich nicht so realistisch, dass die Software schon bald Lehrkräfte bei der Klausuren-Auswertung entlasten kann, oder?

Zehner: Richtig, in den nächsten Jahren jedenfalls nicht. ReCo ist eher für groß angelegte Studien geeignet, bei denen wir Bildungsforschung betreiben und etwas über große Gruppen herausfinden.

Für die Individualdiagnostik, gerade bei wichtigen Prüfungen wie dem Abitur, von denen das Leben und die Bildungskarriere einzelner Schülerinnen und Schüler abhängen, würde ich von automatischer Auswertung aktuell noch absehen, vor allem rechtlich. Mit der Zeit, glaube ich, wird sich das ändern, aber dafür müssen auch die rechtlichen Rahmenbedingungen noch entwickelt werden. Ich denke, das wird mindestens noch ein, zwei Jahrzehnte dauern.

Redaktion: Sie haben außerhalb von Prüfungsauswertungen noch eine weitere Anwendungsmöglichkeit der ReCo-Software für den Unterricht entwickelt, wie funktioniert die?

Zehner: Wenn eine Lehrkraft im Unterricht eine Frage stellt und alle Schüler:innen ein Mobilgerät haben, auf dem sie eine Antwort eingeben können, dann können über die Software alle antworten – es wird also nicht nur der oder die eine Schüler:in drangenommen, der oder die sich gerade gemeldet hat. Die Lehrkraft bekommt dann die Antworten von allen Schüler:innen angezeigt, wobei inhaltlich ähnliche Antworten dabei von ReCo automatisch zu Antworttypen

gruppiert.

Darstellung der Antworttypen in Anlehnung an die Automatic Text Response Coder Software ReCo

Dann sieht man direkt auf den ersten Blick, dass sich zum Beispiel 50 Prozent der Antworten um die Zeitschrift der Tochter drehen, um beim Beispiel von eben zu bleiben. Andere Antworten gehen vielleicht auf irgendeinem anderen Aspekt ein, der in der Geschichte passiert; und das wird dann der Lehrkraft angezeigt, so dass sie nicht mit 30 Antworten klarkommen muss, sondern nur mit einigen wenigen Antworttypen. So kann sie einschätzen, wie viel Prozent der Lerngruppe einen bestimmten Unterrichtsinhalt verstanden haben. Die App lässt sich auch bei Diskussionen einsetzen, etwa wenn in der Klasse die Vor- und Nachteile sozialer Medien diskutiert werden.

Redaktion: Kann die Lehrkraft auch direkt sehen, welche Antworten richtig oder falsch sind?

Zehner: Nein, die ReCo-Software kann in diesem Fall nicht anzeigen, ob die Antwort richtig oder falsch ist. Dafür ist dann die Lehrkraft da. In diesem Anwendungsszenario soll es gar nicht so sehr um Fragen gehen, die richtig beantwortet werden müssen. Vielmehr kann es um offene Diskussionen gehen oder was die Schüler:innen aus der letzten Stunde erinnern. Aufgrund des, ich nenne es mal, „semantischen Wörterbuchs“, werden aber ähnliche Antworten wie „Der Vater ist wütend“ und „Der Papa ist sauer“ übersichtlich auf dem Smartboard oder dem Endgerät der Lehrkraft zusammengefasst. Die Software wird in diesem Jahr in Klassenzimmern erprobt.

Redaktion: Welche Vorteile sehen Sie beim Einsatz von ReCo gegenüber dem normalen Unterrichtsgespräch?

Zehner: Durch die Software kann man mehr Schülerinnen und Schüler mitnehmen; im Idealfall sind dann wirklich alle mit am Ball. Niemand kann sich

hinter anderen verstecken, die sich sonst immer melden. So kann die Unterrichtsqualität gesteigert werden, indem durch die richtigen Fragen alle mental mitmachen; etwas, das kognitive Aktivierung genannt wird. Und die Schülerinnen und Schüler können sehen, wo sie im Vergleich zu den anderen stehen.

Redaktion: In den nächsten Jahren werden neben ChatGPT auch andere Software-Programme die Schule und das Lernen verändern. Von welchen wissen Sie?

Zehner: Es gibt zum Beispiel das große Feld der sogenannten intelligenten tutoriellen Systeme. Das sind Lernprogramme, deren Inhalte oder Formate sich gezielt an die Lernenden anpassen. Ein solches System ist etwa AutoTutor, in dem Lernende per Chat mit dem Programm interagieren. AutoTutor erkennt zum Beispiel auch den aktuellen Gefühlszustand der Lernenden, sodass wenn von Frustration erkannt wird, langsamer vorangegangen werden kann oder wenn Langeweile erkannt wird, schneller im Lernmaterial vorangegangen werden kann. Ein anderes Beispiel ist FeedBook, das im Fremdsprachenlernen eingesetzt wird und von Computerlinguist:innen der Uni Tübingen entwickelt wird. FeedBook liefert beispielsweise beim Ausfüllen von Lückentexten differenzierte Rückmeldung, was die Lernenden falsch gemacht haben und liefert eine entsprechende Erklärung.

Redaktion: Herr Doktor Zehner, wir danken Ihnen für dieses Gespräch.

Zur Person

Dr. Fabian Zehner ist Habilitand und Projektleiter am DIPF (Leibniz-Institut für Bildungsforschung und Bildungsinformation), unter anderem zur Beforschung automatischer Auswertung von Textantworten. Nach einem Frühstudium der Informatik schloss er ein Psychologie-Studium ab und forscht heute an der Schnittstelle beider Gebiete.

Weiterführende Literatur

- [Website Automatic Text Response Coder ReCo](#)