MERTON
Onlinemagazin des *Stifterverbandes*

Foto: iStock/noipornpan

Data-Science trifft Schöngeist

VERÖFFENTLICHT AM 07.12.2018

Digitale Geisteswissenschaften: Mit Computern ganze Stapel von Gedichten und Romanen auf einmal bearbeiten und sogar Musikaufführungen interpretieren. Forschungspioniere zeigen, was möglich ist.

LESEZEIT: 8 MINUTEN

TEXT:
MARION KOCH

Christof Schöch ist Romanist und sein Alltag an der Uni sah früher so aus wie der vieler seiner Kollegen. Er setzte sich an seinen Schreibtisch, schlug seine Bücher auf und vertiefte sich in französische und spanische Geschichten und Gedichte. Auf diese Weise ging er den Texten des Schriftstellers François Bon auf den Grund und promovierte über den französischen Roman im Zeitalter der Aufklärung. Ganz traditionell. Heute ist Schöch Professor für Digital Humanities an der Universität Trier. Die Texte von Denkern und Dichtern liest er jetzt auf dem Bildschirm und erforscht sie in großer Dimension: ganze Stapel von Büchern auf einmal, zusammengefasst in digitalen Dateien – indem er Computerprogramme darüber laufen lässt.

Digital Humanities, das sind computergestützte Forschungsmethoden in den Geistes- und Kulturwissenschaften: Sie zählen zu den Raritäten im Fächerkanon der deutschen Hochschulen. Fast 20.000 Studiengänge bieten die knapp 400 Hochschulen der Bundesrepublik insgesamt an, gerade einmal 15 Treffer listet die Webseite des Hochschulkompasses auf, wenn man das neue Fach in das Suchfeld eingibt. In Trier zum Beispiel, in Leipzig und Stuttgart, in Regensburg, Würzburg und München kann man digitale Geisteswissenschaften studieren. Viele Kollegen von Schöch stehen der „Data-Science“, dem

wissenschaftlichen Extrahieren von Informationen aus digitalen Daten, skeptisch gegenüber. Forscher wie Schöch hingegen, die die neuen Methoden für sich entdeckt haben, sehen darin viele neue Möglichkeiten, Literatur, historische Texte oder auch Musik zu ergründen.

> Computer sind 'Distant Reader', die aus der Vogelperspektive auf Datenmengen schauen.

Jochen Tiepmar (Foto: privat)

JOCHEN TIEPMAR

Informatiker an der Universität Leipzig

An der Universität Leipzig bringt der Informatiker Jochen Tiepmar seinen Studierenden die Digital Humanities näher: „Informations- und Geisteswissenschaften bergen zusammen ein großes Potenzial. Mit recht einfachen technischen Methoden lassen sich ganz neue Erkenntnisse generieren, die ohne computerbasierte Untersuchungen nicht möglich wären“, sagt er.

Worin Computer gut sind: „Sie sind ‚Distant Reader‘, die aus der Vogelperspektive auf Datenmengen schauen“, sagt Tiepmar. Mit „Text-Mining“ filtern Algorithmen statistische und linguistische Informationen aus Texten heraus. Sie finden Auffälligkeiten, Häufigkeiten oder Kohärenzen, können nach einzelnen Wörtern, die in einem Sinnzusammenhang stehen, oder nach Argumenten suchen, erklärt er. Eine weitere Methode ist die Schrifterkennung. Sie helfe dabei, Hieroglyphen zu deuten oder historische Handschriften zu analysieren und ihre Urheber zu identifizieren.

SHAKESPEARES WORTSCHATZ

Christof Schöch (Foto:privat)

Die Frage, was durch Digital Science möglich ist, beantwortet Christof Schöch so: „Zum Beispiel William Shakespeare.“ Der australische Wissenschaftler Hugh Craig habe Algorithmen entwickelt, mit denen er den Umfang des Wortschatzes in den Texten des weltbekannten englischen Dramatikers im Vergleich zu dessen Zeitgenossen erfasste. „Damit konnte Craig den Mythos widerlegen, dass der berühmte Schriftsteller auf ein ungewöhnlich vielfältiges Vokabular zurückgreift“, berichtet Schöch über die Untersuchung aus dem Jahr 2011. Seine literarische Stärke habe Shakespeare also offenbar nicht aus einem ungewöhnlich reichen Wortschatz geschöpft.

Seit einem Jahr koordiniert Schöch auch das internationale Projekt „Distant Reading for European Literary History“. „100 Wissenschaftler aus 30 Ländern arbeiten daran, europäische Literaturgeschichte neu zu denken“, sagt er. Dazu schaffen sie ein „digitales Korpus“ aus mindestens 1.000 europäischen Romanen, die von 1840 bis 1920 in verschiedenen europäischen Sprachen verfasst wurden. Das Material wird in einheitlichem Format gespeichert, anschließend werden computerlinguistische sowie weitere Analyseverfahren eingesetzt, um herauszufinden, wie sich das Genre des Romans im europäischen Vergleich entwickelt hat, welche Parallelen es in den Ländern gibt, wie sich die Autoren beeinflusst haben und welche Themen in welchen Ländern vorherrschend waren.

QUERFLÖTE TRIFFT COMPUTER

Auch Miriam Akkermann hat die digitalen Möglichkeiten für sich entdeckt. Die Musikwissenschaftlerin forscht seit 2015 als Medienwissenschaftlerin an der Universität Bayreuth und ist außerdem Mitglied der **Arbeitsgruppe Big Data** der Jungen Akademie. „Die Arbeitsgruppe wurde 2014 von Wissenschaftlern unterschiedlicher Disziplinen gegründet, die vor einem Berg an Daten standen und sich darüber austauschen wollten, wie man praktischen Nutzen aus solchen Datensätzen zieht“, berichtet die Wissenschaftlerin.

Für Akkermann passen Musik, Wissenschaft und Computer gut zusammen. Sie ist Flötistin. In ihrem Musikstudium hat sie begonnen, den Computer zu nutzen, um ihr Ausdrucksspektrum zu erweitern: „Mit Mikrofon und Computer konnte ich der sonst so hell und fröhlich klingenden Querflöte plötzlich auch dreckige, fiese oder tiefe Töne entlocken“, sagt sie. Nehme man es genau, gebe es heute kaum mehr Musik, bei deren Entstehung kein Computer verwendet werde. Das fange damit an, dass heute so gut wie jedes Musikstück am Bildschirm notiert oder zumindest gesetzt werde. Dass sie den Computer auch einsetzt, um Musik zu erforschen, ist für sie selbstverständlich.

Sie untersucht, wie sich die Aufführungspraxis von Musik entwickelt und welche Ausprägungen sich als „gängig“ herauskristallisieren. Im Fokus hat sie dabei vor allem, wie Computermusik, insbesondere die sogenannte Mixed Music der 80er- und 90er-Jahre, heute aufgeführt wird. „Mixed Music bezeichnet hier Stücke, die im Zusammenspiel von akustischen Instrumenten und per Computer gesteuerter Live-Elektronik aufgeführt werden“, erklärt sie. Mit analogen und statistischen Methoden analysiert die Wissenschaftlerin digitalisierte Aufführungsdaten, historische Informationen und Metadaten, die sie in Archiven sammelt, um herauszufinden: Wer hat was wie interpretiert und zu welchem Zeitpunkt? Wann hat sich etwas in der Aufführung verändert und warum, wurden Stücke zum Beispiel plötzlich schneller oder langsamer interpretiert oder wurden andere Instrumente verwendet?

> Mit Mikrofon und Computer konnte ich der sonst so hell und fröhlich klingenden Querflöte plötzlich auch dreckige, fiese oder tiefe Töne entlocken.

Miriam Akkermann (Foto:privat)

MIRIAM AKKERMANN

Medienwissenschaftlerin an der Universität Bayreuth

Ihre Erkenntnisse vergleicht Akkermann mit Aufführungskonzepten von Barockmusik, die sich Anfang des 20. Jahrhunderts herausbildeten. Die Musikstile liegen zwar sowohl zeitlich als auch inhaltlich weit auseinander. „Doch nach weniger als 30 Jahren zeigen sich in der Computermusik bereits Herausforderungen, die heute vornehmlich von der Aufführung Alter Musik oder Barockmusik bekannt sind: So fehlen Informationen zur Spielweise, zum Klangergebnis, aber auch zu den verwendeten Instrumenten, zu denen hier dann auch der Computer zählt“, erklärt sie. Weil die Quellenlage gleich sei, könne sie auf gleiche Mechanismen zurückgreifen, um die Daten zu analysieren.

INVASION AUF DIE GEISTESWISSENSCHAFTEN

Immer wieder kommt es vor, dass der Romanist Christof Schöch von skeptischen Kollegen vor der Gefahr gewarnt wird, dass plötzlich Informatiker bestimmen, welche Fragen in den Literaturwissenschaften gestellt werden. Er antwortet ihnen dann, dass es gerade deshalb so wichtig sei, dass sie sich das Zepter nicht aus der Hand nehmen ließen und diesen Bereich der Forschung selbst mitgestalten. Dazu müssten sie ja keine Vollinformatiker werden, aber die Grundlagen der Technik verstehen, um mit Informatikern auf Augenhöhe zusammenarbeiten zu können. Er ist sich aber sicher: „In gar nicht so ferner Zukunft werden Recherchestrategien im Open Access, computerlinguistische Suchverfahren oder die professionelle Verwaltung bibliografischer Referenzen zu den digitalen Schlüsselqualifikationen auch von Geisteswissenschaftlern gehören – und damit werden die etablierten Geisteswissenschaften und die Digital Humanities noch näher zusammenrücken.“

Bis dahin werden vielleicht auch die Probleme gelöst sein, die heute noch das Programmieren von Algorithmen erschweren, sagt der Informatiker Tiepmar. Die Methode „Named Entity Recognition“ (NER) etwa, mit der beispielsweise Figuren in Romanen erkannt werden sollen: „Noch ist es für den Computer nicht leicht, herauszufiltern, dass zum Beispiel mit ‚der ältere Herr‘, ‚der 70-Jährige‘ oder dem Personalpronomen ‚er‘ immer die gleiche Person gemeint ist.“ Auch ließen sich Algorithmen nur schwer in andere Sprachen, auf andere Fachbereiche oder auch nur auf andere Projekte übertragen. „Man muss bei fast jedem neuen Projekt wieder von vorn anfangen“, sagt Tiepmar.

Der Informatiker setzt darauf, dass in Zukunft Daten über nationale Grenzen hinweg vereinheitlicht werden, um sie leichter verarbeiten zu können. So wie es das europäische Projekt **CLARIN-D** vormache, in dessen Rahmen eine digitale Forschungsinfrastruktur für Geistes- und Sozialwissenschaften aufgebaut werde, die einheitlich gestaltete, einfach zu nutzende Sprachdatenbanken und Analysewerkzeuge für Wissenschaftler bereitstelle. Außerdem sollten Geisteswissenschaftler und Informatiker mehr kommunizieren, um gemeinsame Forschungsprojekte zu generieren, findet Tiepmar. Seine Uni hat dazu in diesem Jahr das **Forum für Digital Humanities Leipzig** (FDHL) gegründet.

Logo: Stifterverband/Sven Sedivy

Daten erfassen, erkunden, managen, kuratieren, analysieren, visualisieren, interpretieren, kontextualisieren, beurteilen und anwenden - diese Fähigkeiten (neudeutsch data literacy) sind in der heutigen Zeit unverzichtbar. Dennoch stehen sie noch an zu wenigen Hochschulen auf dem Lehrplan. Mit dem Förderprogramm **Data Literacy Education** möchten die Heinz Nixdorf Stiftung und der Stifterverband dies nun ändern und den Erwerb von *data literacy* für Studierende aller Fächer an deutschen

Hochschulen etablieren. Es werden insgesamt drei Hochschulen mit jeweils bis zu 250.000 Euro gefördert: Die Universitäten in Lüneburg und Göttingen sowie die Hochschule Mannheim.

Der Romanist Schöch arbeitet indes mit seinen Kollegen daran, weiter die digitalen Methoden zur Erforschung von Literatur zu verbessern. Er sagt: „Mit spannenden, verlässlichen Erkenntnissen werden wir die Kollegen von den neuen Möglichkeiten überzeugen."

QUELLE: HTTPS://MERTON-MAGAZIN.DE/DATA-SCIENCE-TRIFFT-SCHOENGEIST