

Die Flut der Texte

Notwendigkeit und Perspektiven, die Bibliothek digital zu erweitern

Lesen wandelt sich. Die Nutzung des Internets hat die tägliche Zeit für das Lesen von Büchern, Zeitschriften und Zeitungen weiter reduziert. Viele Verlage stehen wirtschaftlich unter Druck, können sich in der Medienkonkurrenz nur schwer behaupten. Dabei steigt die Zahl der digitalen Publikationen gerade jenseits der Verlage. Wie bewältigt man die Textmassen effizient und was bedeutet dieser Wandel für die Bibliotheken?

Zunächst hat es schon immer unterschiedlichste Formen des Lesens gegeben: Lesen zur Ablenkung und Entspannung, Lesen zum Nachdenken, Lesen zum Lernen, zur schnellen Orientierung, zum Nachschlagen, zum Suchen – und so fort. Manche dieser Formen werden durch die digitalen Angebote erleichtert, insbesondere das Nachschlagen (Wikipedia) und das Suchen (Google). Aber auch für die Verfügbarkeit der Texte (E-Journals, E-Books) und für ihren Kauf oder Lizenzierung (Amazon) hat sich mit dem Internet vieles vereinfacht. Dabei wird das Buch nicht überall durch seine digitale Kopie ersetzt. Kognitions- und Bildungsforschung haben nachgewiesen, dass man aus Texten auf Papier besser lernt als aus Texten am Bildschirm, unabhängig von Alter oder Vorbildung. Insbesondere für das genauere Lesen wählen viele das Buch oder den Ausdruck. Für viele Texte müsste eine Bibliothek deshalb finanziell in der Lage sein, beide Formen anzubieten. Bei seit Jahrzehnten stagnierenden Literaturretats und schwindender Kaufkraft ist dies viel zu oft nicht der Fall.

Dabei erschöpft sich das Potential digitaler Texte nicht mit der leichteren Auffindbarkeit, der besseren Verfügbarkeit, dem digitalen Annotieren, Copy and Paste oder der Möglichkeit des Ausdrucks. Digitale Texte sind mehr

als Nachbildungen oder Ersatz ihrer analogen Vorbilder. Digitale Texte lassen sich maschinell durchsuchen, aus großen Mengen digitaler Texte mächtige Such-Indices erstellen. Die Ähnlichkeiten oder wechselseitigen Zitate lassen sich viel leichter ermitteln. Man kann sie maschinell erweitern, indem man sie mit automatisch generierten Schlagwörtern anreichert oder maschinell Merkmale aus dem Leseverhalten und den persönlichen Hintergründen und Bezügen ihrer Leser (Social Graph) hinzufügt. Die großen Suchmaschinen- und Social-Media-Anbieter tun dies, ohne dass sie diese zusätzlichen Merkmale offenlegen. Auf diese Weise verbessern sie ihre Suchergebnisse, um erfolgreicher Werbung zu platzieren.

Eine Bibliothek hat in Deutschland zumeist die Aufgabe, eine ausgewogene Auswahl an Literatur anzubieten. Und eine Bibliothek wird auch nicht das Leseverhalten ihrer Nutzer protokollieren, um ein konkurrenzfähiges Angebot zu Amazon, Google und Co. aufzubauen. Die Deutsche Forschungsgemeinschaft warnte im letzten Jahr mit einem Positionspapier („Daten-tracking in der Wissenschaft“) davor, wie kommerzielle Anbieter das Leseverhalten im Bereich wissenschaftlicher E-Books und E-Journals zum Beispiel über Cookies protokollieren und auswerten, um Einfluss auf die Inhalte von Forschung zu nehmen. Bibliotheken sollten wie die Wissenschaften neutrale Orte bleiben. Dies bedeutet allerdings nicht, dass die vielen Werkzeuge, welche es für den effizienten Umgang mit maschinenlesbaren Texten gibt, in einer wissenschaftlichen Bibliothek nichts verloren hätten, im Gegenteil: Bildung und Wissenschaft benötigen die Souveränität über die Texte und können ihre Auswertung nicht kommerziellen oder politischen Interessen überlassen.

Auch die Politik hat das Potential gesehen, welches das Text Data Mining für die Forschung bedeutet, und mit der Reform des Urhebergesetzes 2021 in § 60d erste Möglichkeiten



Centi-folium Stultorum in Quarto, Oder Hundert Ausbündige Narren in Folio, Misc.qt.1, WLB Stuttgart

„Aus der Württembergischen Landesbibliothek sind insbesondere die Literatur über unsere Region, die Sammlung des Hölderlin Archivs oder die Materialgruppen (z. B. Flugblätter der Neuen Sozialen Bewegungen) unserer Bibliothek für Zeitgeschichte von höchstem Interesse.“

für die Aufbereitung und maschinelle Untersuchung urheberrechtlich geschützter Texte für die nicht kommerziellen Zwecke der Wissenschaften geschaffen. So entstehende Datenbestände (Textkorpora) dürfen abgegrenzten Gruppen zur Auswertung oder Einzelnen zur Überprüfung von Forschungsergebnissen zugänglich gemacht werden.

Grundsätzlich bildet jede Sammlung von Texten ein Korpus. Von besonderem Interesse sind natürlich vollständige, sehr umfangreiche oder repräsentative Sammlungen. Aus der

Württembergischen Landesbibliothek sind insbesondere die Literatur über unsere Region, die Sammlung des Hölderlin Archivs oder die Materialgruppen (z.B. Flugblätter der Neuen Sozialen Bewegungen) unserer Bibliothek für Zeitgeschichte von höchstem Interesse. Nun ist die Aufbereitung der Materialien mit beträchtlichen Aufwänden verbunden und kann deshalb nicht für jedes Forschungsvorhaben von Neuem bewerkstelligt werden, wie es bislang § 60d Abs. 3 UrhG vorsieht. Eine Bibliothek wird nur dann in die digitale Aufbereitung





Bücher-Narr im Original und digitalisiert, vgl. Abb.S. 5

urheberrechtlich geschützter Werke investieren, wenn die damit verbundenen Aufwände nachhaltig gesichert werden können, was oft mit den Rechteinhabern ausgehandelt werden muss.

Nun wurde bei derselben Reform des Urheberrechts mit § 61d UrhG wenigstens die Möglichkeit geschaffen, nicht mehr im Buchhandel oder bei den Verlagen verfügbare Werke nach Abschluss „kollektiver Lizenzen“ mit den einschlägigen Verwertungsgesellschaften wie z. B. der VG Wort online bereitzustellen. Diese Möglichkeit ist auch für nie kommerziell vertriebene Publikationen vorgesehen. Erforderlich ist die Registrierung beim Amt für geistiges Eigentum der Europäischen Union (EUIPO) und die Wahrung einer halbjährigen Einspruchsfrist möglicher Rechteinhaber. Für Bücher, Zeitschriften und Zeitungen aus Verlagen gilt jedoch eine zusätzliche Schutzfrist von 30 Jahren nach Erscheinen, welche aber auch durch eine freiwillige Lizenz abgegolten werden könnte.

Aufgrund des geringen Alters dieser Normen sind naturgemäß nicht alle Fragen letztlich geklärt. So ist für die wissenschaftliche Nutzung vielfach das einrichtungsübergreifende Zusammenführen von digitalen Texten erforderlich, da die Sammeltätigkeit und Überlieferung naturgemäß immer Lücken lassen. Diese Fortentwicklungen werden aber nur geschehen, indem Archive und Bibliotheken mit dem Aufbau von Korpora beginnen, sobald in Kürze die notwendigen Workflows von Nationalbibliothek

und Verwertungsgesellschaften freigegeben werden.

Die Zusammenstellung, Retrodigitalisierung und Datenaufbereitung sind gleichermaßen mit Aufwänden verbunden wie die Vermittlung der für die Anwendung erforderlichen Kompetenzen im Umgang mit den vorhandenen Werkzeugen. Hierzu braucht es zusätzlich einschlägig qualifiziertes Personal wie Computerlinguisten, Informatiker oder Data Librarians. Diese sind in der Wirtschaft gleichermaßen begehrt, weshalb die zusätzlichen Stellen ausreichend dotiert sein müssen.

Der digitale Anbau der Bibliothek ist weder billig noch kurzfristig erreichbar. Er kann nicht alleine bewerkstelligt werden, doch sind die Wissenschaftlichen Bibliotheken Deutschlands seit Jahrzehnten effizient vernetzt. Aber wenn sie mit der digitalen Erweiterung nicht zeitnah beginnen können, geraten Bildung und Wissenschaft immer stärker in kommerzielle Abhängigkeiten. Solange insbesondere jüngere Leserinnen und Leser das Text- und Data-Mining nicht lernen und ausprobieren können, bleiben auch die Interessenten für die einschlägigen Studiengänge und die entsprechenden Fachkräfte für die Wirtschaft aus. Selbst kleinere nordamerikanische Bibliotheken sind da aufgrund eines leistungsfähigeren Urheberrechts viel weiter.

Rupert Schaab