



MASCHINELLES LERNEN
FÜR BIBLIOTHEKEN

Maßgeschneiderte Informationsversorgung

Die digitale Transformation ist in vollem Gange. Big Data – die Erzeugung, Verknüpfung und Auswertung großer Datenmengen – beeinflusst in unserer digitalisierten Gesellschaft beinahe alle Lebensbereiche und ist auch aus der Wissenschaft nicht mehr wegzudenken. Doch die Versprechen, die das digitale Zeitalter im Bereich der Literaturversorgung und Informationsinfrastrukturen mit sich bringt, verwirklichen sich nicht von allein. **Bibliotheken als zentrale Wissensanbieter** stehen vor einer gewaltigen Herausforderung: Sogenannte *Data Lakes*, die verschiedenste Arten wissenschaftlicher Daten in ihrer natürlichen Form zusammenführen, brauchen eine starke Strukturierung, ein rigoroses Metadaten-Management und maßgeschneiderte Services zur Suche und zum Datenzugriff, um nicht als nutzlose *Data Swamps* zu enden. Ein Schritt in diese Richtung sind die *Fachinformationsdienste für die Wissenschaft (FID)*, die seit 2011 die Sonder-sammelgebiete der Universitätsbibliotheken, eines der ältesten Förderprogramme der *Deutschen Forschungsgemeinschaft (DFG)*, ablösen. Die *FIDs* sollen Wissenschaftlern in Deutschland unabhängig von ihrem Standort einen möglichst direkten und komfortablen Zugriff auf Spezial-

Es geht nicht darum, dass Bibliotheken ihr gesammeltes Wissen bloß digitalisieren, sondern die Daten nutzbar machen, also in ihrer natürlichen Form zusammenführen, strukturieren und mit einem sinnvollen Metadaten-Management versehen.
→ Foto: TIB / C. Bierwagen

literatur und forschungsrelevante Informationen ermöglichen. Als bundesweites System ergänzen sie die Informationsinfrastrukturen der Hochschulen, Forschungseinrichtungen und forschenden Industrie durch überregionale Dienstleistungen für den Spitzenbedarf.

PubPharm, der *Fachinformationsdienst für die Pharmazie*, wird seit 2014 von der *Universitätsbibliothek Braunschweig* getragen – in enger Kooperation mit Wissenschaftlern des *L3S am Institut für Informationssysteme der TU Braunschweig*. Die interdisziplinäre Forschung, die das *L3S* kennzeichnet, erweist sich auch hier als Erfolgsmodell: Aus der Kombination von bibliothekarischer Expertise, fachwissenschaftlicher Kompetenz und forschender Informatik entsteht Innovation für die pharmazeutische Fachwelt.

Im Mittelpunkt von *PubPharm* steht der Nutzer mit seinen Informationsbedürfnissen.

In diese Richtung wird die erweiterbare und personalisierbare Informationsinfrastruktur weiter ausgebaut. Das Dienst-





publikationen lernt *PubPharm* mit Hilfe von Deep-Learning-Techniken, in welchem Kontext diese Entitätsinformationen auftreten, und verknüpft sie dann dauerhaft mit den entsprechenden Publikationen. Im Gegensatz zu klassischen *Linked Open Data* oder Fachdatenbanken, die vereinzelte, aus dem Kontext gerissene Informationsstücke enthalten, wird bei *PubPharm* die Verknüpfung von Konzepten sichtbar. Die Nutzer erhalten Erklärungen zu den Verknüpfungen und können sich ihrem Forschungsgegenstand aus unterschiedlichen Blickwinkeln nähern. Die wissenschaftlichen Produkte und ihre relevanten Entitäten stehen also nicht mehr für sich allein, sondern bilden ein Netzwerk, ohne dessen tieferes Verständnis kaum noch wissenschaftliche Innovation möglich ist. ¶

→ <https://www.pubpharm.de>

angebot soll damit auch flexibler und passgenauer werden. Das geschieht mithilfe von Machine-Learning und Deep-Learning-Technologien, die **eine semantisch angereicherte Suche** und den direkten Datenzugriff auf relevante Literatur ermöglichen. Der *Fachinformationsdienst* ist also noch immer auf klar umrissene wissenschaftliche Produkte wie Publikationen, Forschungsdatensätze, Patente oder Software als zentrale wissensvermittelnde Elemente ausgerichtet. Für den weiteren Erkenntnisgewinn ist es aber unverzichtbar, sie mit umfassenden und hochwertigen semantischen Metadaten zu erschließen.

In komplexen Wissensräumen konnten Nutzer bislang nur mithilfe bibliographischer Attribute wie Autoren, Publikationsjahr oder Publikationsform navigieren. Neuerdings konzentrieren sich entsprechende Dienste aber vermehrt auch auf **entitäts-**

zentrierte Informationen, also solche, die den Inhalt der einzelnen wissenschaftlichen Produkte betreffen, wie in Publikationen genannte Wirkstoffe, Moleküle oder chemische Substanzen. Aus Millionen von Fach-

KONTAKT:

Prof. Dr. Wolf-Tilo Balke

Balke@L3S.de



\\ \\ Wolf-Tilo Balke ist seit 2008 Leiter des *Instituts für Informationssysteme (IfIS)* und Ordinarius an der *Technischen Universität Braunschweig* sowie Mitglied des Direktoriums des *Forschungszentrums L3S*. In der Forschung beschäftigt er sich vor allem mit Datenbanktechnologie und Informationsdiensten, insbesondere personalisierte Anfrageverarbeitung und Deep Learning. \\ \\