

Christian Herzog, Ulrich Wirth

## Lebenswissenschaftler suchen Antworten, keine Dokumente

**Waren Redundanzen bisher nur lästig, so sind sie im Internetzeitalter riskant. Zwar helfen Informationen, Ungewissheit zu reduzieren. Überschüssige Information provoziert jedoch einen Zuwachs an Unsicherheit, wodurch Recherchen zunehmend zeitraubender und ineffizienter sind. Ressourcen wie Zeit und Geld werden vergeudet, was so inakzeptabel wie überflüssig ist. In wissensintensiven Arbeitsgebieten wie den Life Sciences, in denen sowohl Grundlagenforschung betrieben als auch gezielt marktwirtschaftlich orientiert gearbeitet wird, kann der Zeitverlust zudem zu Wettbewerbsnachteilen führen. Die Lösung bieten neuartige Recherche-Tools auf der Basis intelligenter semantischer Web-Technologien. So ist es z.B. möglich, Text-Mining-Verfahren an bereits vorhandene Fachdatenbanken anzudocken. Dadurch wird Wissenstransfer nicht nur effizienter, es eröffnen sich auch zusätzliche Wissensmöglichkeiten.**

### Wissensmanagement in den Life Sciences

In den interdisziplinär ausgerichteten Life Sciences wie Biologie, Chemie und Medizin und der noch jungen Disziplin Life Science Engineering, worunter z.B. Umwelt- und Medizintechnik fallen, ist Wissensmanagement geschäftskritisch, denn Wissen ist essenzieller Bestandteil entsprechender Marktleistungen. Das gilt sowohl für Grundlagen- als auch für Anwendungsforschung.

Biowissenschaftler z.B. fokussieren primär auf Grundlagenforschung. Um Ideen und Trends, die sich in ihrem spezifischen Forschungsfeld entwickeln, überhaupt identifizieren zu können, müssen sie sich regelmäßig einen mög-

lichst vollständigen Überblick über ihr Fachgebiet verschaffen. Das ist aufwändig und zeitintensiv. Aber auch stärker anwendungsbezogene Forscher, etwa in der Pharmaindustrie, betreiben ständig Internet- und Literaturrecherchen, wobei sie in oft mühevoller Kleinarbeit aus der Informationsflut die Spreu vom Weizen trennen. Der klinisch tätige Arzt wiederum sucht in erster Linie Antworten auf Fragen, die sich unmittelbar bei der Ausübung seines Berufes stellen, etwa welche Therapie für den Patienten die beste ist. Allen gemein ist die Arbeitsgrundlage, nämlich aktuelle und qualitätsgesicherte Informationen, die sie zeitnah benötigen. Dazu verfügen Lebenswissenschaftler über spezifische Recherchezugänge, wie kommerzielle Datenbanken. Die Datenbanken, die z.B. in PubMed verfügbar sind, stellen einen Datenpool dar, der sowohl abgeschlossene als auch aktuell laufende Publikationen aus Medizin und Life Sciences nahezu vollständig dokumentiert.

Über diese Suchzugänge gelangen die Wissenschaftler zwar zu durchaus passablen Näherungslösungen oder den gewünschten Suchergebnissen selbst. Doch Zeit ist eine kostbare Ressource; wird sie vergeudet, produziert das unnötige Kosten. Und Näherungslösungen sind genau genommen auch Zielverfehlungen. Das ist nicht nur berufsethisch bedenklich, sondern verträgt sich vor allem mit Gütemaßen wie Qualität, Precision und Recall nur schlecht. Hier kommen semantische Suchlösungen ins Spiel, die den Wissenstransfer erleichtern sowie neue Wissensmöglichkeiten für die Forschung eröffnen.

### Was ist semantisch am Semantischen Web?

Semantische Suchlösungen verfolgen das Konzept, verschiedene Ressourcen

wie Dokumente, Bilder und Personen semantisch zu verknüpfen. Das ermöglicht den Übergang vom gegenwärtigen Web mit seinen einfachen Hyperlinks zu einem semantisch angereicherten Web, in dem Beziehungen zwischen den verteilt gelagerten Ressourcen geknüpft werden. Intelligente Suchmaschinen können diese Beziehungen mit Hilfe von semantischen und statistischen Prozeduren lesbar machen, wodurch sie den Textinhalt quasi automatisch verstehen. Dadurch schließen Suchergebnisse auch solche Dokumente mit ein, die den Suchterm nicht explizit enthalten, aber relevant sind. Eine Suche nach Begleitsymptomen von Migräne verwies sowohl auf Lichtempfindlichkeit als auch auf verwandte Konzepte wie Geräusch- oder Geruchsempfindlichkeit. Dadurch lassen sich Suchanfragen auf wenige, relevante Dokumente verdichten, die bei herkömmlichen Volltextsuchen umständlich in umfangreichen Ergebnislisten hätten identifiziert werden müssen.

Für die Life Sciences bedeutet das: Damit die Suche in PubMed eine semantische wird, muss dessen Datenbasis semantisch aufbereitet werden. Dafür sorgen semantische Technologien, wie das Knowledge Dashboard 2.0. Sie kombinieren die automatisierte Analyse großer Textmengen mit schnellen und vielfältigen Zugriffsmöglichkeiten. Die technische Basis bildet eine ontologiebasierte Indexierungs- und Retrievalösung, die eigens für Medizin und Life Sciences entwickelt wurde, mit alternativen Thesauri aber auch außerhalb der Life Sciences Anwendung findet. Diese Technik verwandelt wissenschaftliche Publikationen wie PubMed-Abstracts in so genannte Fingerprints. Diese enthalten die ontologiebasiert identifizierten Fachbegriffe und Bewertungsrelationen für jeden einzelnen Begriff. Bei der Indexierung und in den Anwendungen kommen Dokumentations Sprachen zum Einsatz, etwa die Medical Subject Headings (MeSH), ein Thesaurus zur Sacherschließung von Büchern und Zeitschriftenartikeln in Medizin und Biowissenschaften, und der weltweit größte Metathesaurus Unified Medical Language System (UMLS).

Der klassische Rechercheeinstieg ist der themenbezogene, bei Migräne zum Bei-

spiel nach „Gesichtsfeldausfällen“. Anschließend kann der Nutzer wie gewohnt die dazugehörigen Abstracts aus PubMed aufrufen. Ein Mehrwert besteht in der übersichtlichen Anzeige aller biomedizinischen Konzepte, die in den 37 Treffern vorkommen. Die Technologie lässt konzeptionelles Browsen auch innerhalb größter Textmengen zu – und zwar nach einem genormten Vokabular (zum Beispiel Disorders, Chemicals & Drugs, Procedures, Anatomy, Physiology) – ganz ohne Eingabe von Suchtermen, nur per Klick. Wird unter „Disorders“ zum Beispiel der Suchbegriff „Aura“ aktiviert, öffnet sich ein neues Fenster, welches die Suche feiner spezifiziert. Bei einer herkömmlichen Benutzeroberfläche hätte der Nutzer seine Suchanfrage selbstständig umformulieren und weitere Suchbegriffe händisch in die Suchmaske eingeben müssen, um in die Tiefe zu gehen. Das strukturierte Browsen spart also wertvolle Recherchezeit und damit Geld.

### Expertenprofile, Trend-Mining und Knowledge Discovery

Wer hat sich mit wem und zu welcher Zeit weltweit zu einem Aspekt der aktuellen Suchanfrage auseinandergesetzt? Wer gilt als ausgewiesener Fachmann? Wer taugt zum Key Note Speaker? Bei der Umwandlung der PubMed-Abstracts in Fingerprints werden Expertenprofile generiert, aus denen sich relevante Veröffentlichungen und deren Autoren identifizieren lassen.

Welche Fachbegriffe werden zunehmend häufig im Kontext des Themas genannt? Der Bereich „Trends“ lässt das rasch erkennen. Die statistische Auswertung ergibt z.B., dass die Botox-Therapie zunehmend in den Fokus der Migräne-Forschung gerückt ist und 2007 so viele Publikationen erschienen sind wie in den drei Jahren zuvor. Über die einzelnen Reiter können anschließend die einschlägigen Publikationen, Experten etc. betrachtet werden. Die automatisierte Inhaltsanalyse nimmt dem Wissenschaftler also enorme Recherchezeit ab.

Semantische Suche kann auch die Arbeitsgebiete von morgen und über-

morgen identifizieren. Wie ist das möglich? Ganz einfach: Dank technischer Unterstützung können Recherchertools analysieren, welche (Such-)Begriffe erstmals in einer Publikation genannt werden und welche Fachbegriffe voraussichtlich mit dem Thema in Zukunft gemeinsame Relevanz entwickeln werden. Wer weiß, für die Migränetherapie könnte dies zum Beispiel das einschläffördernde und angstlösende „Wunderhormon“ Progesteron sein. Dieses Wissen ist für Kliniker, vor allem aber für die Pharmaforschung von besonderer Bedeutung.

### Fazit:

Mit den Möglichkeiten thematischer Verknüpfungen und der schlussfolgernden Suche nach impliziten Informationen helfen semantische Technologien,

die Limitationen herkömmlicher Volltext- und Datenbanksuchen zu überwinden.

Das Knowledge Dashboard 2.0 bietet Zugang nicht nur zu qualitätsgesichertem und professionellem Know-how, sondern beschleunigt den Wissenstransfer durch vertikales Suchen. Es lässt sich auch in anderen Bereichen als den Life Sciences einsetzen und kann zudem in Intranet-Lösungen integriert oder auf das „Web of Science“ aufgesetzt werden. Der sinkende Rechercheaufwand führt zu einem Zeitgewinn und zu Einsparpotenzialen.

Die Möglichkeiten der Trendanalyse, das Erkennen von Zukunftsthemen, die Identifikation von Experten – das ist Wissensvorsprung per Knopfdruck, der solche Informationsprofis überzeugen dürfte, die an mehr als nur an Dokumenten interessiert sind.

### Die Autoren:



Christian Herzog, MD, Executive Vice President and General Manager for Life Sciences ist Mitgründer und CEO der SsynX Solutions GmbH, die im Oktober 2007 von dem langjährigen Partner Collexis Holdings Inc., USA übernommen wurde. Er ist heute für das weltweite Geschäft von Collexis im Bereich Life Sciences als EVP/GM verantwortlich. Christian Herzog ist Arzt und hat parallel zu seinem Medizinstudium Wirtschaftswissenschaften studiert. Vor seiner Tätigkeit bei SsynX/Collexis war er in verantwortlicher Position bei verschiedenen Beratungsunternehmen mit Fokus auf den Gesundheits- und Life-Science-Bereich tätig.

herzog@wissensmanagement.net



Ulrich Wirth, M.A., leitet die Euro-Schulen Trier, Höhere Berufsfachschule für Medizinische Dokumentationsassistenten. Er verantwortet die Bereiche Bildungsmanagement, Marketing und Demografieberatung. Als Sachverständiger in der Rahmenplankommission „Geprüfter Fachwirt für Informationsdienste“ des Deutschen Industrie- und Handelskammertages (DIHK), Berlin, vertritt er die Interessen des Deutschen Verbands Medizinischer Dokumentare (DVMD) e.V. Auf diese Tätigkeiten ist er im Studium, in der Unternehmenskommunikation der Volkswagen AG, durch ein anschließendes Volontariat in der Pressedokumentation der Frankfurter Allgemeine Zeitung GmbH und als Information Specialist am Potsdamer Institut für Information und Dokumentation (IID) bestens vorbereitet worden.

wirth@wissensmanagement.net